2019 Research Compendium

December 30, 2019

About Newfound Research

While other asset managers focus on alpha, our first focus is on managing risk. We know investors care deeply about protecting the capital they have worked hard to accumulate. And as investors approach and enter retirement, managing "sequence risk" becomes even more important.

How? We believe in systematic, disciplined, and repeatable decision-making powered by the evidence-based insights of consistent, thoughtful research. Specifically, we focus on the high conviction application of the major quantitative investment "styles" – i.e. value, momentum, carry, defensive, and trend – within tactical asset allocation.

Separately Managed Accounts	• Available on most major TAMP and UMA platforms • Model delivery is available to qualifying institutions
Mutual Funds	• Select Newfound strategies are made available in open-end fund format • See www.thinknewfoundfunds.com
ETFs	• Select Newfound strategies are made available in exchange-traded fund format
Custom Index Solutions	• We offer custom tailored investment solutions that are built with and powered by our research and insights.

For more information, visit www.thinknewfound.com

Table of Contents

Fixed Income

Portable Beta

Rebalance Timing Luck

Sector Rotation

Sequence Risk

Trend

IS MULTI-MANAGER DIVERSIFICATION WORTH IT?

January 7, 2019

SUMMARY

- Portfolio risk is traditionally quantified by volatility. The benefits of diversification are measured in how portfolio volatility is changed with the addition or subtraction of different investments.
- Another measure of portfolio risk is the dispersion in terminal wealth: a measure that attempts to capture the potential difference in realized returns. For example, two equity managers that each hold 30 stock portfolios may exhibit similar volatility levels but will likely have very different realized results.
- In this commentary we explore existing literature covering the potential diversification benefits that can arise from combining multiple managers together.
- Empirical evidence suggests that in heterogeneous categories (e.g. many hedge fund styles), combining managers can reduce portfolio volatility. Yet even in homogenous categories (e.g. equity style boxes), combining managers can have a pronounced effect on reducing the dispersion in terminal wealth.
- Finally, we address the question as to whether manager diversification leads to dilution, arguing that a combination of managers will reduce idiosyncratic process risks but maintain overall style exposure.

1. Introduction

In their 2014 paper *The Free Lunch Effect: The Value of Decoupling Diversification and Risk*, Croce, Guinn, and Robinson draw a distinction between the risk reduction effects that occur due to de-risking and those that occur due to diversification benefits.

To illustrate the distinction, the authors compare the volatility of an all equity portfolio versus a balanced stock/bond mix. In the 1984-2014 sample period, they find that the all equity portfolio has an annualized volatility of 15.25% while the balanced portfolio has an annualized volatility of just 9.56%.

Over 75% of this reduction in volatility, however, is due simply to the fact that bonds were much less volatile than stocks over the period. In fact, of the 568-basis-point reduction, only 124 basis points was due to actual diversification benefits.

Why does this matter?

Because de-risking carries none of the benefits of diversification. If there is a commensurate trade-off between expected return and risk, then all we have done is reduced the expected return of our portfolio.[1]

It is only by combining assets of like volatility – and, it is assumed, like expected return – that should allow us to enjoy the free lunch of diversification.

Unfortunately, unless you are willing to apply leverage (e.g. risky parity), the reality of finding such free lunch opportunities across assets is limited. The classic example of inter-asset diversification, though, is taught in Finance 101: as we add more stocks to a portfolio, we drive the contribution of idiosyncratic volatility towards zero.

Yet volatility is only one way to measure risk. If we build a portfolio of 30 stocks and you build a portfolio of 30 stocks, the portfolios may have nearly identical levels of volatility, but we almost assuredly will end up with different realized results. This difference between the *expected* and the *realized* is captured by a measure known as *terminal wealth dispersion*, first introduced by Robert Radcliffe in his book *Investment: Concepts, Analysis, Strategy*.

This form of risk naturally arises when we select between investment managers. Two managers may both select securities from the same universe using the same investment thesis, but the realized results of their portfolios can be starkly different. In rare cases, the specific choice of one manager over another can even lead to catastrophic results.

The selection of a manager reflects not only an allocation to an asset class, but also reflects an allocation to a process. In this commentary, we ask: how much diversification benefit exists in process diversification?

2. The Theory Behind Manager Diversification

In *Factors from Scratch*, the research team at O'Shaughnessy Asset Management (OSAM), in partnership with anonymous blogger Jesse Livermore, digs into the driving elements behind value and momentum equity strategies.

They find that value stocks do tend to exhibit negative EPS growth, but this decay in fundamentals is offset by multiple expansion. In other words, markets do appear to correctly identify companies with contracting fundamentals, but they also exaggerate and over-extrapolate that weakness. The historical edge for the strategy has been that the re-rating – measured via multiple expansion – tends to overcompensate for the contraction in fundamentals.

For momentum, OSAM finds a somewhat opposite effect. The strategy correctly identifies companies with strengthening fundamentals, but during the holding period a valuation contraction occurs as the market recognizes that its outlook might have been too optimistic. Historically, however, the growth outweighed the contraction to create a net positive effect.

These are the true, underlying economic and behavioral effects that managers are trying to capture when they implement value and momentum strategies.

[1] For the sake of brevity, we are just going to charge forward and completely ignore the empirical evidence of semi-flat security market lines and the question of, "how do you measure risk?"

These are not, however, effects we can observe directly in the market; they are effects that we have to forecast. To do so, we have to utilize semi-noisy signals that we believe are correlated. Therefore, every manager's strategy will be somewhat inefficient at capturing these effects.

For example, there are a number of quantitative measures we may apply in our attempt to identify value opportunities; e.g. price-to-book, price-to-earnings, and EBITDA-to-enterprise-value to name a few. Two different noisy signals might end up with different performance just due to randomness.

This noise between signals is further compounded when we consider all the other decisions that must be made in the portfolio construction process. Two managers may use the same signals and still end up with very different portfolios based upon how the signals are translated into allocations.

Consider this: Morningstar currently[2] lists 1,217 large-cap value funds in its mutual fund universe and trailing 1-year returns ranged from 1.91% to -22.90%. This is not just a case of extreme outliers, either: the spread between the 10th and 90th percentile returning funds was 871 basis points.

It bears repeating that these are funds that, in theory, are all trying to achieve the same goal: large-cap value exposure.

Yet this result is not wholly surprising to us. In *Separating Ingredients and Recipe in Factor Investing* we demonstrated that the performance dispersion between different momentum strategy definitions (e.g. momentum measure, look-back length, rebalance frequency, weighting scheme, et cetera) was larger than the performance dispersion between the traditional Fama-French factors themselves in 90% of rolling 1-year periods. As it turns out, *intra*-factor differences can cause greater dispersion than *inter*-factor differences.

Without an ex-ante view as to the superiority of one signal, one process, or one fund versus another, it seems prudent for a portfolio to have diversified exposure to a broad range of signals that seem plausibly related to the underlying phenomenon.

3. Literature Review

While foundational literature on modern portfolio diversification extends back to the 1950s, little has been written in the field of manager diversification. While it is a well-established teaching that a portfolio of 25-40 stocks is typically sufficient to reduce idiosyncratic risk, there is no matching rule for how many managers to combine together.

One of the earliest articles on the topic was written by Edward O'Neal in 1997, titled *How Many Mutual Funds Constitute a Diversified Mutual Fund Portfolio?*

Published in the Financial Analysts Journal, this article explores risk across two different dimensions: the volatility of returns over time and the dispersion in terminal period wealth. Again, the idea behind the latter measure is that two investors with

[2] Data as of 1/3/2019

identical horizons and different investments will achieve different terminal wealth levels, even if those investments have the same volatility.

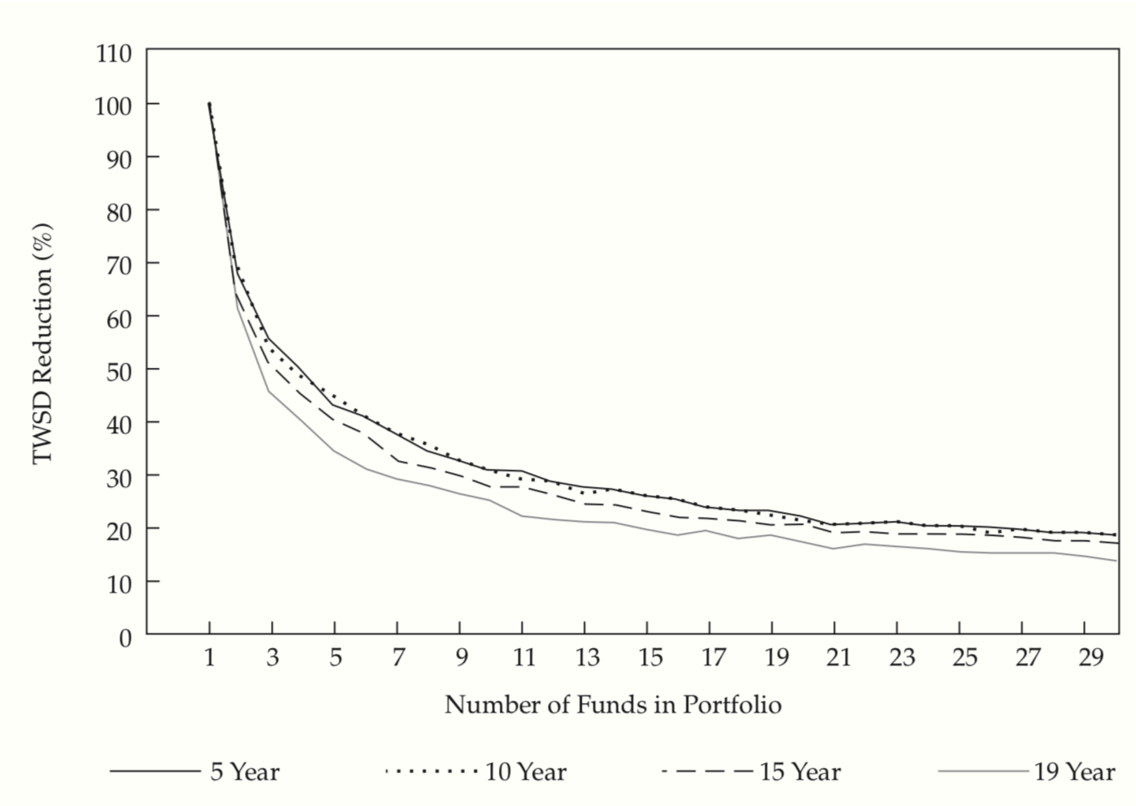
Exploring equity mutual fund returns from 1986 to 1997, the study adopts a simulation-based approach to constructing portfolios and tracking returns. Multi-manager portfolios of varying sizes are randomly constructed and compared against other multi-manager portfolios of the same size.

O'Neal finds that while combining managers has little-to-no effect on volatility (manager returns were too homogenous), it had a significant effect upon the dispersion of terminal wealth. To quote the article,

> *"Holding more than a single mutual fund in a portfolio appears to have substantial diversification benefits. The traditional measure of volatility, the time-series standard deviation, is not greatly influenced by holding multiple funds. Measures of the dispersion in terminal-wealth levels, however, which are arguably more important to long-term investors than time-series risk measures, can be reduced significantly. The greatest portion of the reduction occurs with the addition of small numbers of funds. This reduction in terminal-period wealth dispersion is evident for all holding periods studied. Two out of three downside risk measures are also substantially reduced by including multiple funds in a portfolio. These findings are especially important for investors who use mutual funds to fund fixed-horizon investment goals, such as retirement and college savings."*

Allocating to three managers instead of just one could reduce the dispersion in terminal wealth by nearly 50%, an effect found to be quite consistent across the different time horizons measured.

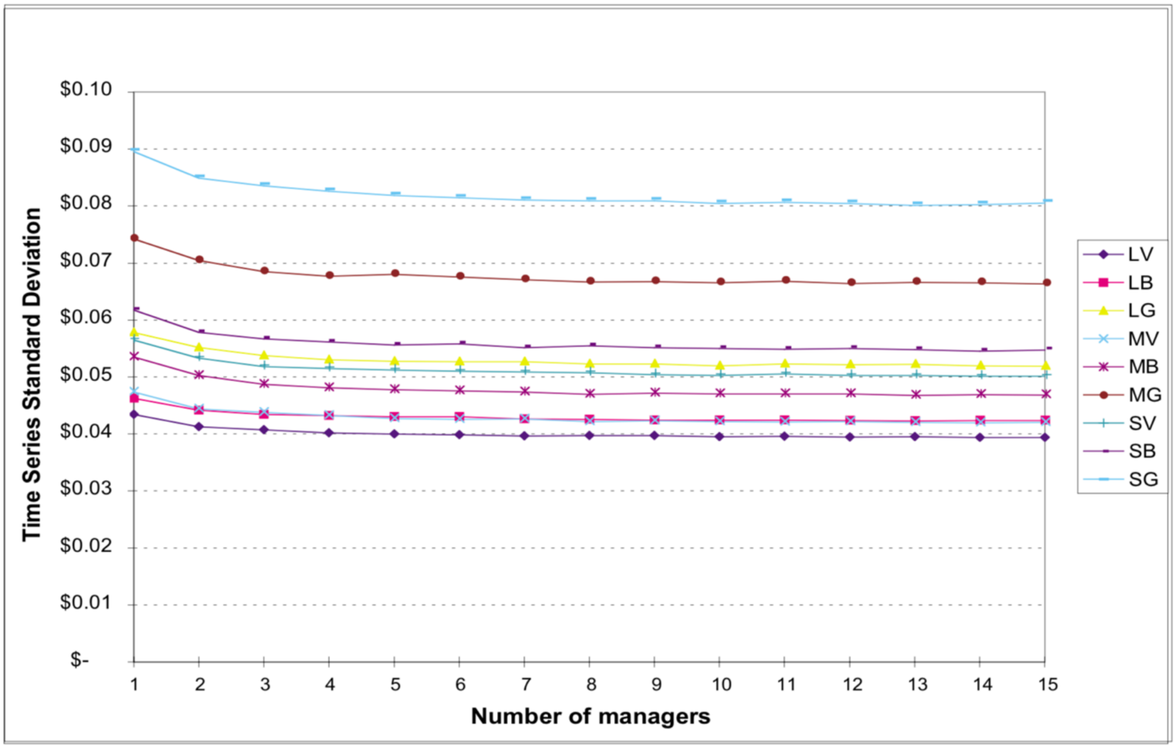
Figure 1. Reduction in Terminal-Wealth Standard Deviation for Growth Fund Portfolios over Different Holding Periods



In 1999, O'Neal teamed up with L. Franklin Fant to publish *Do You Need More than One Manager for a Given Equity Style?* Adopting a similar simulation-based approach, Fant and O'Neal explored multi-manager equity portfolios in the context of the style-box framework.

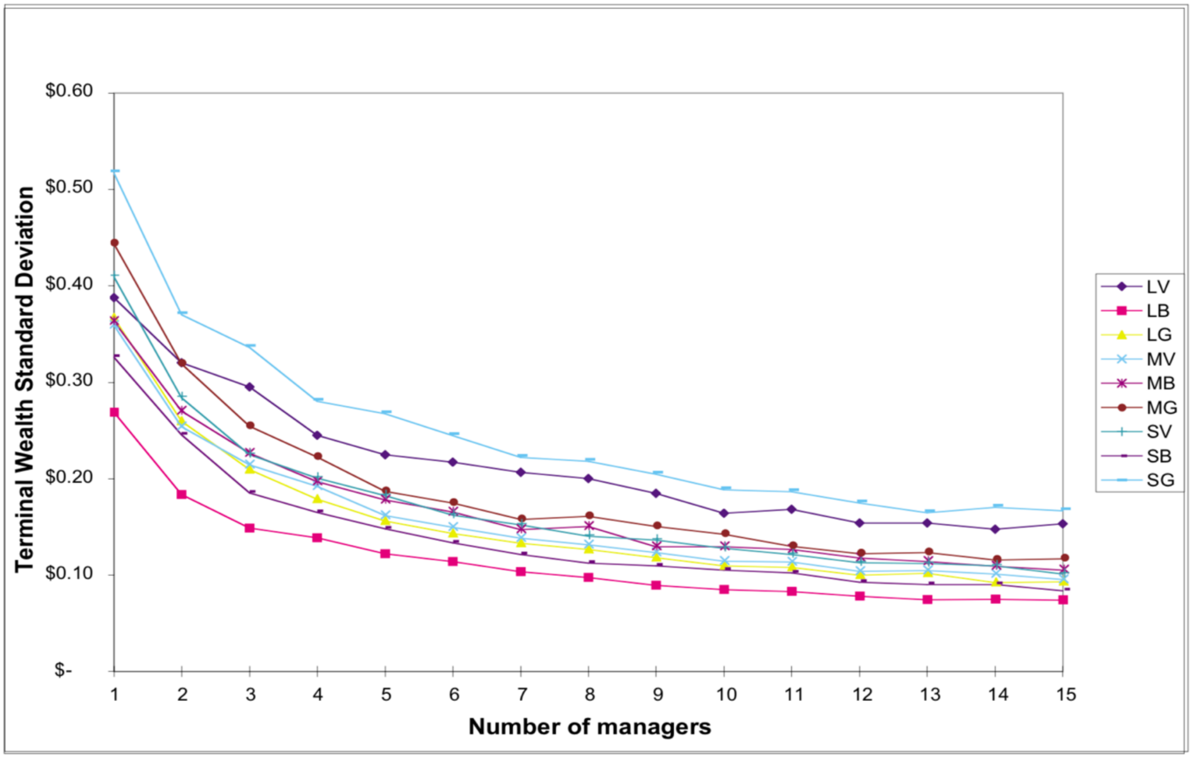
And, as before, they find that taking a multi-manager approach has little effect upon portfolio volatility.

EXHIBIT 3
TIME SERIES STANDARD DEVIATION OF QUARTERLY RETURNS FOR MULTIMANAGER PORTFOLIOS



It does, however, again prove to have a significant effect on the deviation in terminal wealth.

EXHIBIT 4
TERMINAL WEALTH STANDARD DEVIATION FOR MULTIMANAGER PORTFOLIOS



To quote the paper,

> *"Regardless of the style category considered, the variability in terminal wealth levels is significantly reduced by using more managers. The first few additional managers make the most difference, as terminal wealth standard deviation declines at a decreasing rate with the number of managers. Concentrating on the variability of periodic portfolio returns fails to document the advantage of using multiple managers within style categories.*
>
> *Second, some categories benefit more from additional managers than others. Plan sponsors would do well to allocate relatively more managers to the styles that display the greatest diversification benefits. Growth styles and small-cap styles appear to offer the greatest potential for diversification."*

In 2002, François-Serge Lhabitant and Michelle Learned pursued a similar vein of research in the realm of hedge funds in their article *Hedge Fund Diversification: How Much is Enough?* They employ the same simulation-based approach but evaluate diversification effects within the different hedge fund styles.

They find that while diversification does little to affect the expected return for a given style, it does appear to help reduce portfolio volatility: sometimes quite significantly so. This somewhat contradictory result to the prior research is likely due to the fact that hedge funds within a given category exhibit far more heterogeneity in process and returns than do equity managers in the same style box.

(Note that while the graphs below only show the period 1990-1993, the paper explores three time periods: 1990-1993, 1994-1997, and 1998-2001 and finds a similar conclusion in all three).

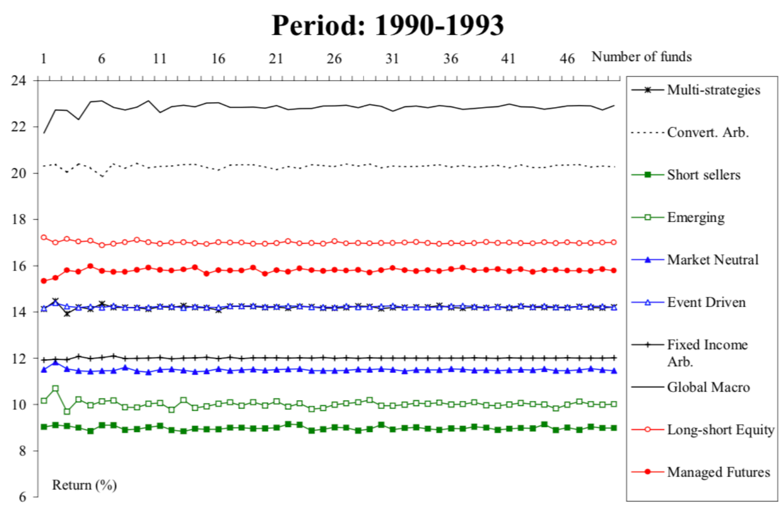


Figure 1: Evolution of the mean return of a hedge fund portfolio as a function of the number of underlying hedge funds.

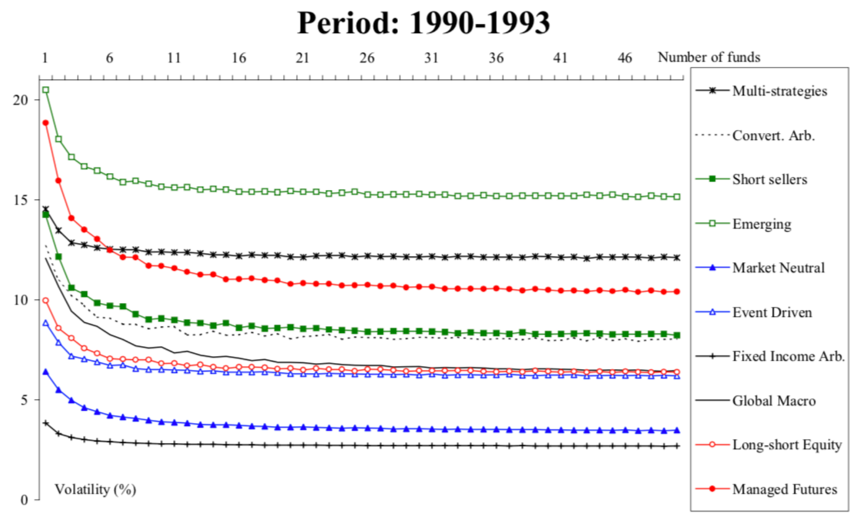


Figure 2: Impact of diversification on volatility

Perhaps most importantly, however, they find a rather significant reduction in risk characteristics like a portfolio's realized maximum drawdown.

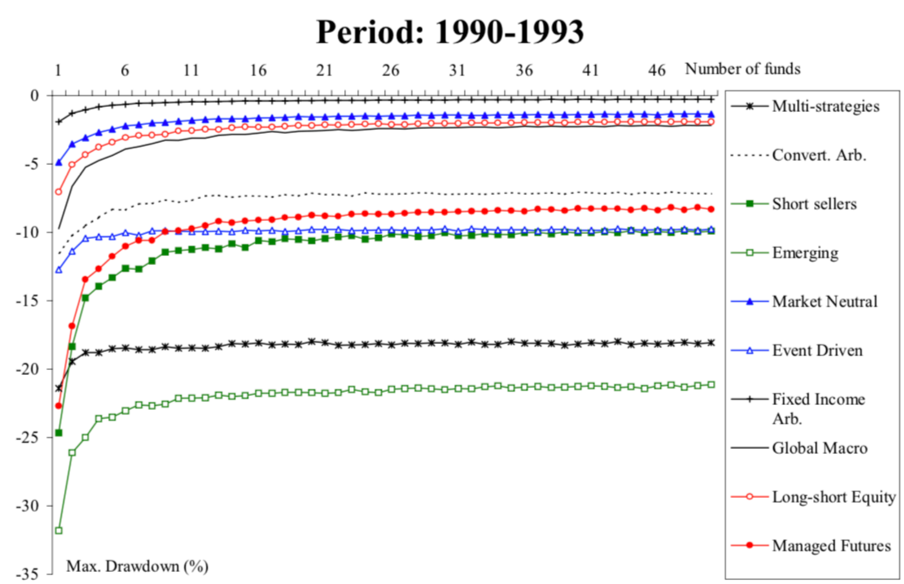


Figure 7: Impact of diversification on maximum draw-down

To quote the article,

> *"We find that naively adding more funds to a portfolio tends to leave returns stable, decrease the standard deviation, and reduce downside risk. Thus, diversification should be increased as long as the marginal benefits of adding a new asset to a portfolio exceeds the marginal cost.*
>
> *…*
>
> *If a sample of managers is relatively style pure, then a fewer number of managers will minimize the unsystematic risk of that style. On the contrary, if the sample is really heterogeneous, increasing the number of managers may still provide important diversification benefits."*

Taken together, this literature paints an important picture:

- Diversifying across managers in the same category will likely do little to reduce portfolio volatility, except in the cases where categories are broad enough to capture many heterogeneous managers.
- Diversifying across managers appears to significantly reduce the potential dispersion in terminal wealth.

But why is minimizing "the dispersion of terminal wealth" important? The answer is the same reason why we diversify in the first place: risk management.

The potential for high dispersion in terminal wealth means that we can have dramatically different outcomes based upon the choices we are making, placing significant emphasis on our skill in manager selection. Choosing just one manager is *more right* style thinking rather than our preferred *less wrong*.

4. But What About Dilution?

The number one response we hear when we talk about manager diversification is: "when we combine managers, won't we just dilute our exposure back to the market?"

The answer, as with all things, is: "it depends." For the sake of brevity, we're just going to leave it with, "no."

No?

No.

If we identify three managers as providing exposure to value, then it makes little logical sense that somehow a combination of them would suddenly remove that exposure. Subtraction through addition only works if there is a negative involved; i.e. one of the managers would have to provide *anti-value* exposure to offset the others.

Remember that an active manager's portfolio can always be decomposed into two pieces: the benchmark and a dollar-neutral long/short portfolio that isolates the active over/under-weights that manager has made.

To "dilute back to the benchmark," we'd have to identify managers and then weight them such that all of their over/under-weights net out to equal zero.

Candidly, we'd be impressed if you managed to do that. Especially if you combine managers within the same style who should all be, at least directionally, taking similar bets. The dilution that occurs is only across those bets which they *disagree* on and therefore reflect the idiosyncrasies of their specific process.

What a multi-manager implementation allows us to diversify is our *selection risk*, leading to a return profile more "in-line" with a given style or category. In fact, Lhabitant and Learned (2002) demonstrated this exact notion with a graph that plots the correlation of multi-manager portfolios with their broad category. While somewhat tautological, an increase in manager diversification leads to a return profile closer to the given style than to the idiosyncrasies of those managers.

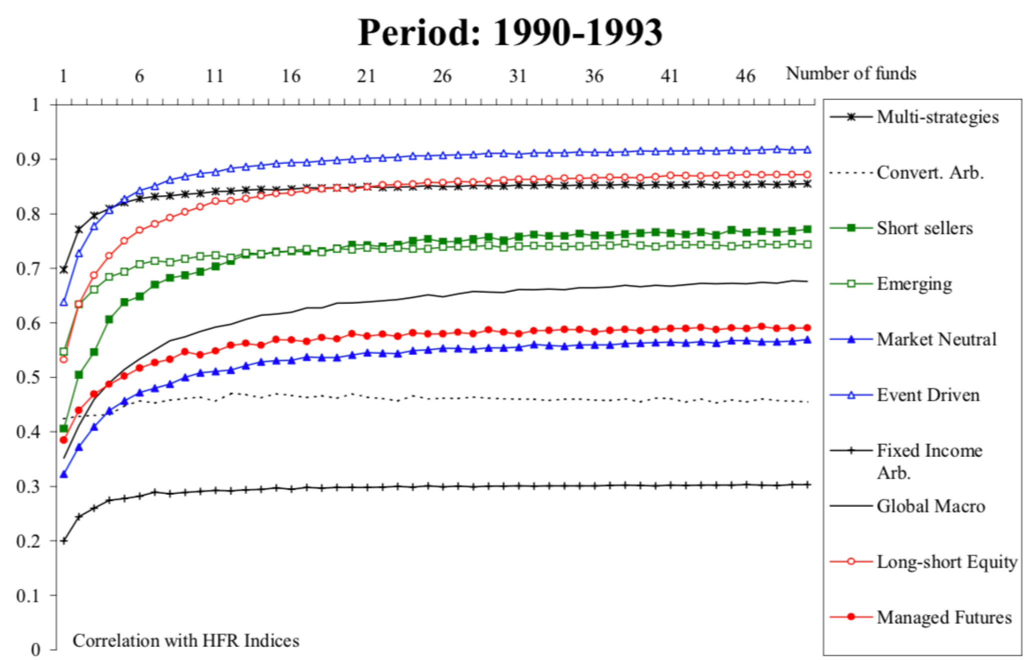
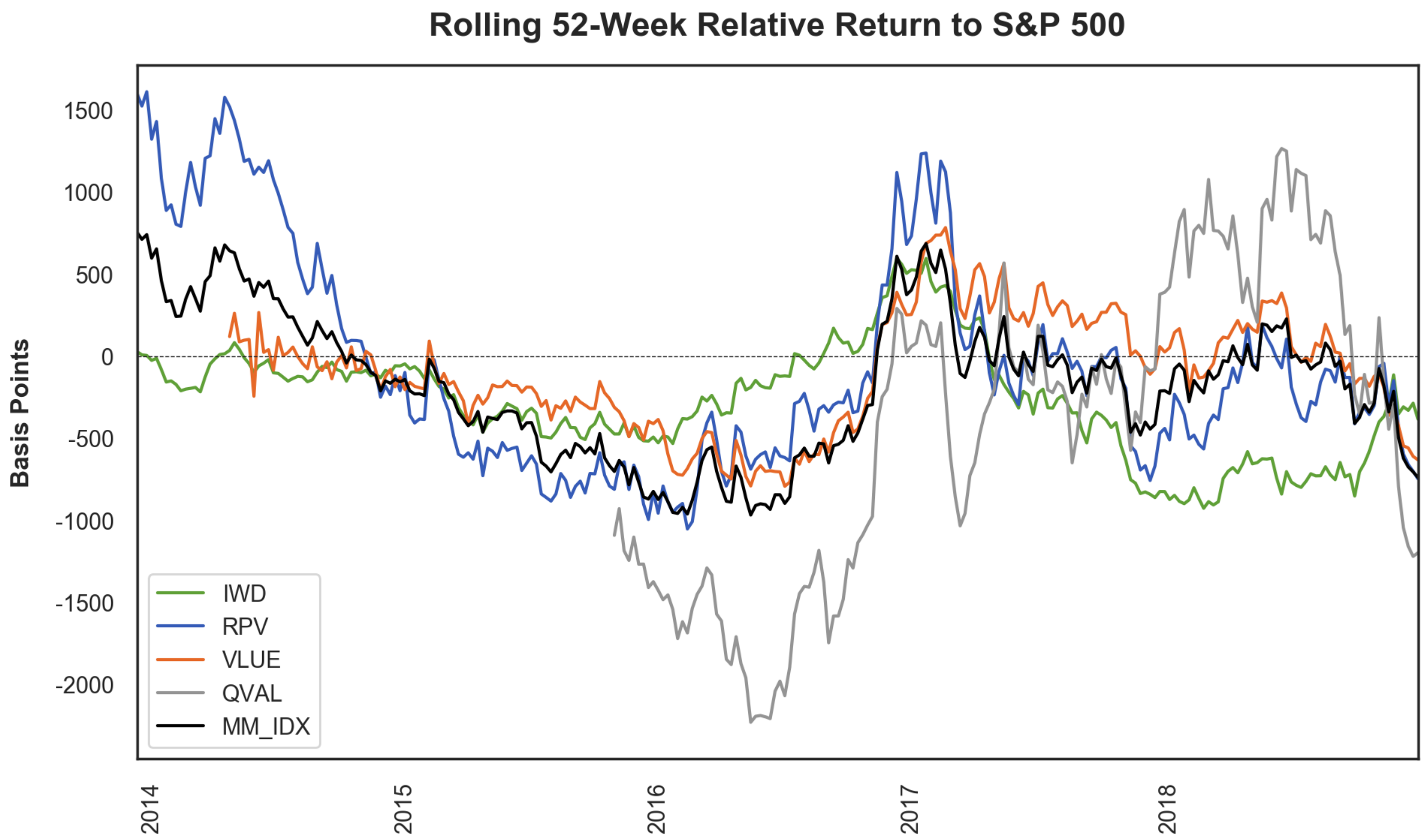


Figure 9: Impact of diversification on correlation with hedge fund style indices

We can also see this with a practical example. Below we take several available ETFs that implement quantitative value strategies and plot their rolling 52-week return relative to the S&P 500. We also construct a multi-manager index

("MM_IDX") that is a naïve, equal-weight portfolio. The only wrinkle to this portfolio is that ETFs are not introduced immediately, but rather slowly over a 12-month period.[3]



Source: CSI Analytics. Calculations by Newfound Research. It is not possible to invest in an index. Returns are total returns (i.e. assume the reinvestment of all distributions) and are gross of all fees except for underlying expense ratios of ETFs. Past performance does not guarantee future results.

We can see that while the multi-manager blend is never the best performing strategy, it is also never the worst. Never the hero; never a zero.

It should be noted that while manager diversification may be able to reduce the idiosyncratic returns that result from process differences, it will not prevent losses (or relative underperformance) of the underlying style itself. In other words,

[3] This is done to make the graph's interpretation clearer; otherwise you would have an ETF like QVAL influencing the return of MM_IDX before the line for QVAL is plotted, which can lead to strange-looking results. Blending does not perfectly fix this issue, but it helps.

we might avoid the full brunt of losses specific to the Sequoia Fund, but no amount of diversification would prevent the relative drag seen by the quantitative value style in general over the last decade.

We can see this in the graph above by the fact that all the lines generally tend to move together. 2015 was bad for value managers. 2016 was much better. But we can also see that every once in a while, a specific implementation will hit a rough patch that is idiosyncratic to that approach; e.g. IWD in 2017 and most of 2018.

Multi-manager diversification is the tool that allows us to avoid the full brunt of this risk.

5. Conclusion

Taken together, the research behind manager diversification suggests:

- In heterogeneous categories (e.g. many hedge fund styles), manager diversification may reduce portfolio volatility.
- In more homogenous categories (e.g. equity style boxes), manager diversification may reduce the dispersion in terminal wealth.
- Multi-manager implementations appear to reduce realized portfolio risk metrics such as maximum drawdown. This is likely partially due to the reduction in portfolio volatility, but also due to a reduction in exposure to funds that exhibit catastrophic losses.
- Multi-manager implementations do not necessarily "dilute" the portfolio back to market exposure, but rather "dilute" the portfolio back to the style exposure, reducing exposure idiosyncratic process risk.

For advisors and investors, this evidence may cause a sigh of relief. Instead of having to spend time trying to identify the best manager or the best process, there may be significant advantages to simply "avoiding the brain damage"[4] and allocating equally among a few. In other words, if you don't know which low-volatility ETF to pick, just buy a couple and move on with your life.

But what are the cons?

- A multi-manager approach may be tax inefficient, as we will need to rebalance allocations back to parity between the exposures.
- A multi-manager approach may lead to fund bloat within a portfolio, doubling or tripling the number of holdings we have. While this is merely optical, except possibly in small portfolios, we recognize there exists an aversion to it.

[4] Language courtesy of friend-of-the-firm Wesley Gray.

- By definition, performance will be middling: the cost of avoiding the full brunt of losers is that we also give up the full benefit of winners. We're reluctant to label this as a con, as it is arguably the whole point of diversification, but it is worth pointing out that the same behavioral biases that emerge in portfolio reviews of asset allocation will likely re-emerge in reviews of manager selection, especially over short time horizons.

For investment managers, a natural interpretation of this research is that approaches blending different signals and portfolio construction methods together should lead to more consistent outcomes. It should be no surprise, then, that asset managers adopting machine learning are finding significant advantages with ensemble techniques. After all, they invoke the low-hanging fruit of manager diversification.

Perhaps most interesting is that this research suggests that fund-of-funds really are not such bad ideas so long as costs can be kept under control. As the asset management business continues to be more competitive, perhaps there is an edge – and a better client result – found in cooperation.

FRAGILITY CASE STUDY: DUAL MOMENTUM GEM

January 14, 2019

SUMMARY

- Recent market volatility has caused many tactical models to make sudden and significant changes in their allocation profiles.
- Periods such as Q4 2018 highlight model specification risk: the sensitivity of a strategy's performance to specific implementation decisions.
- We explore this idea with a case study, using the popular Dual Momentum GEM strategy and a variety of lookback horizons for portfolio formation.
- We demonstrate that the year-to-year performance difference can span hundreds, if not thousands, of basis points between the implementations.
- By simply diversifying across multiple implementations, we can dramatically reduce model specification risk and even potentially see improvements in realized metrics such as Sharpe ratio and maximum drawdown.

1. Introduction

Among do-it-yourself tactical investors, Gary Antonacci's Dual Momentum is the strategy we tend to see implemented the most. The Dual Momentum approach is simple: by combining both relative momentum and absolute momentum (i.e. trend following), Dual Momentum seeks to rotate into areas of relative strength while preserving the flexibility to shift entirely to safety assets (e.g. short-term U.S. Treasury bills) during periods of pervasive, negative trends.

In our experience, the precise implementation of Dual Momentum tends to vary (with various bells-and-whistles applied) from practitioner to practitioner. The most popular benchmark model, however, is the Global Equities Momentum ("GEM"), with some variation of Dual Momentum Sector Rotation ("DMSR") a close second.

Recently, we've spoken to several members in our extended community who have bemoaned the fact that Dual Momentum kept them mostly aggressively positioned in Q4 2018 and signaled a defensive shift at the beginning of January 2019, at which point the S&P 500 was already in a -14% drawdown (having peaked at over -19% on December 24th). Several DIYers even decided to override their signal in some capacity, either ignoring it entirely, waiting a few days for "confirmation," or implementing some sort of "half-and-half" rule where they are taking a partially defensive stance.

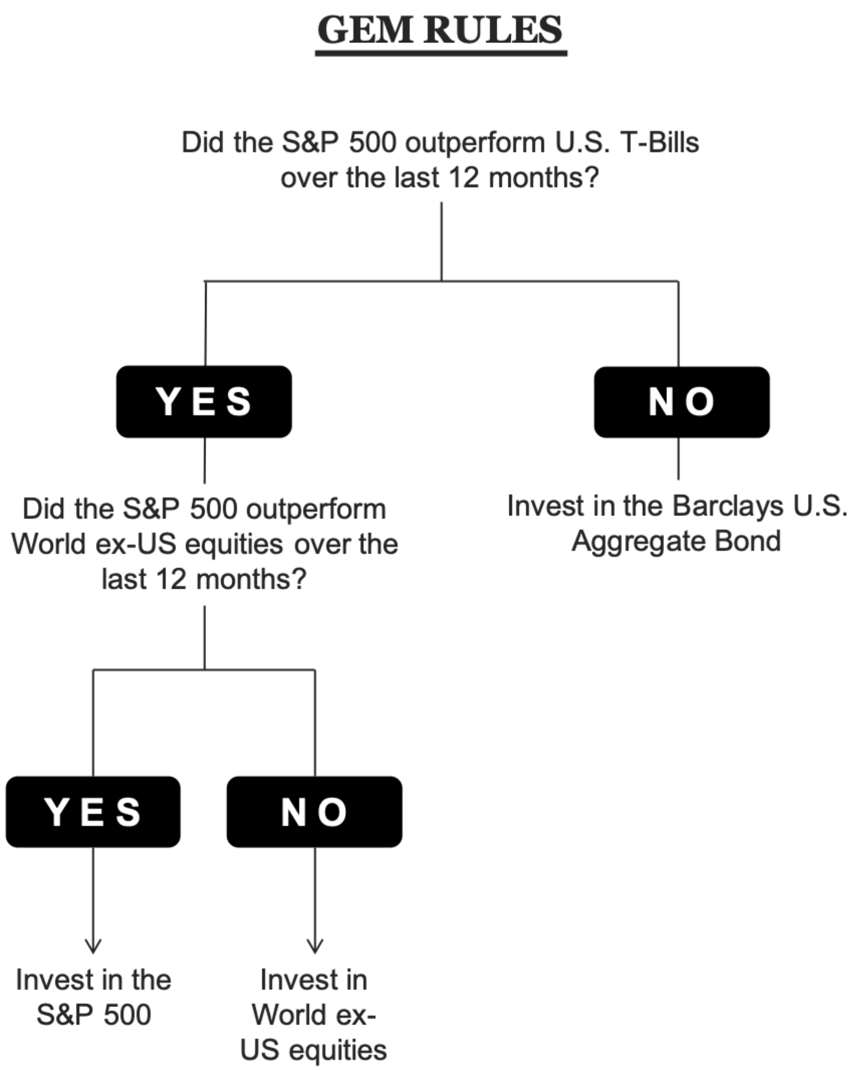
Ignoring the fact that a decision to override a systematic model somewhat defeats the whole point of being systematic in the first place, this sort of behavior highlights another very important truth: there is a significant gap of risk that exists between the long-term supporting evidence of an investment style (e.g. momentum and trend) and the precise strategy we attempt to implement with (e.g. Dual Momentum GEM).

At Newfound, we call that gap *model specification risk*. There is significant evidence supporting both momentum and trend as quantitative styles, but the precise means by which we measure these concepts can lead to dramatically different portfolios and outcomes. When a portfolio's returns are highly sensitive to its specification – i.e. slight variation in returns or model parameters lead to dramatically different return profiles – we label the strategy as *fragile.*

In this brief commentary, we will use the Global Equities Momentum ("GEM") strategy as a case study in fragility.

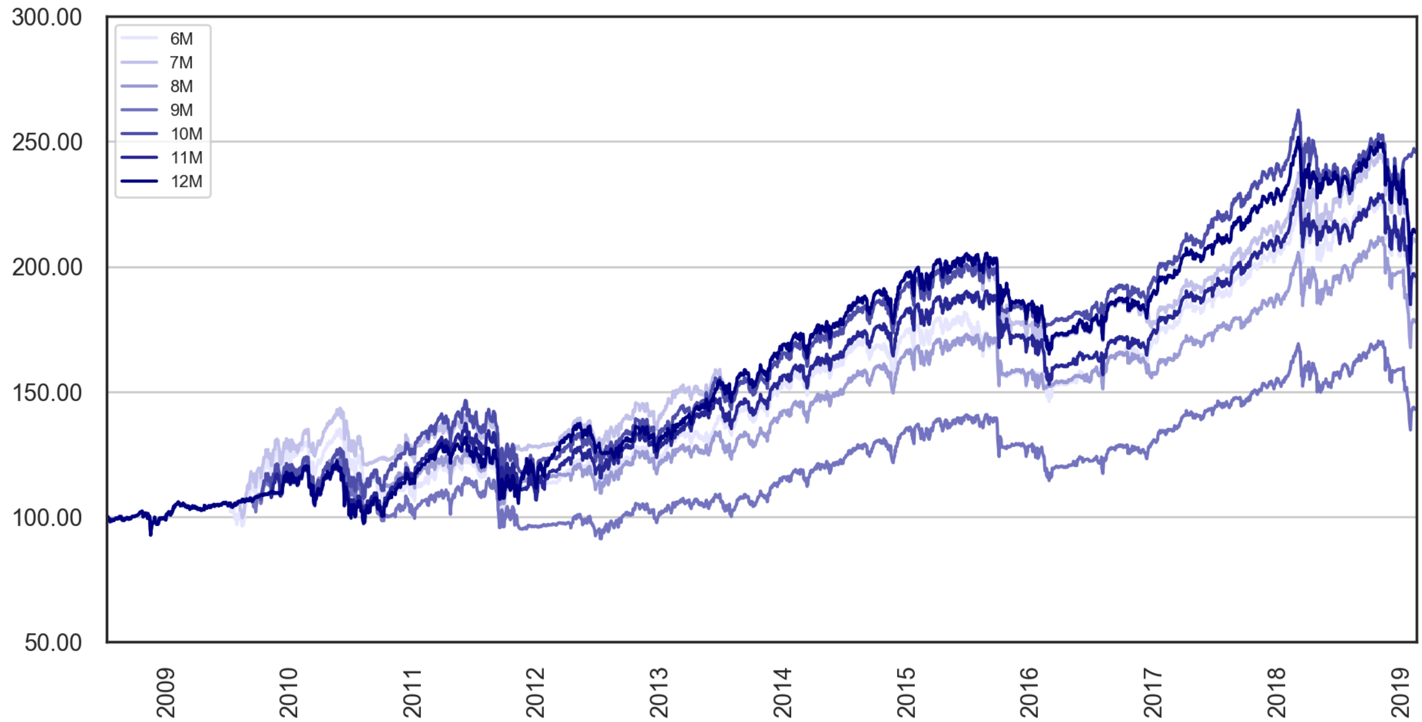
2. Global Equities Momentum ("GEM")

To implement the GEM strategy, an investor merely needs to follow the decision tree below at the end of each month.



From a practitioner stand-point, there are several attractive features about this model. First, it is based upon the long-run evidence of both trend-following and momentum. Second, it is very easy to model and generate signals for. Finally, it is fairly light-weight from an implementation perspective: only twelve potential rebalances a year (and often much less), with the portfolio only holding one ETF at a time.

Despite the evidence that "simple beats complex," the simplicity of GEM belies its inherent fragility. Below we plot the equity curves for GEM implementations that employ different lookback horizons for measuring trend and momentum, ranging from 6- to 12-months.



Source: CSI Analytics. Calculations by Newfound Research. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees except for underlying ETF expense ratios. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

We can see a significant dispersion in potential terminal wealth. That dispersion, however, is not necessarily consistent with the notion that one formation period is inherently better than another. While we would argue, ex-ante, that there should be little performance difference between a 9-month and 10-month lookback – they both, after all, capture the notion of "intermediate-term trends" – the former returned just 43.1% over the period while the latter returned 146.1%.

These total return figures further hide the year-to-year disparity that exists. The 9-month model, for example, was not a consistent loser. Below we plot these results, highlighting both the best (blue) and worst (orange) performing specifications. We see that the yearly spread between these strategies can be hundreds-to-thousands of basis points; consider that in 2010, the strategy formed using a 10-month lookback returned 12.2% while the strategy formed using a 9-month lookback returned -9.31%.

Same thesis. Same strategy. Slightly different specification. Dramatically different outcomes. That single year is likely the difference between hired and fired for most advisors and asset managers.

Formation Period

	6M	7M	8M	9M	10M	11M	12M
2009	19.57%	21.52%	10.80%	10.80%	12.37%	8.68%	9.13%
2010	-9.24%	2.37%	1.72%	-9.31%	12.21%	7.07%	4.65%
2011	-1.00%	-1.96%	-2.57%	-8.88%	-7.50%	-6.05%	0.06%
2012	14.42%	14.42%	6.43%	6.43%	10.38%	12.88%	10.25%
2013	20.94%	17.23%	18.06%	15.24%	24.51%	22.91%	29.91%
2014	10.57%	13.46%	13.46%	13.46%	13.46%	13.46%	13.46%
2015	-9.87%	-12.00%	-6.86%	-6.86%	-6.86%	-7.88%	-7.88%
2016	10.57%	4.93%	4.09%	5.67%	10.76%	5.67%	6.71%
2017	20.33%	21.50%	18.89%	18.34%	22.76%	20.94%	20.94%
2018	-6.73%	-4.34%	-6.46%	-8.61%	0.72%	-8.15%	-8.15%

Source: CSI Analytics. Calculations by Newfound Research. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees except for underlying ETF expense ratios. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

For those bemoaning their 2018 return, note that the 10-month specification would have netted a *positive* result! That specification turned defensive at the end of October.

Now, some may cry "foul" here. The evidence for trend and momentum is, after all, centuries in length and the efficacy of all these horizons is supported. Surely the noise we see over this ten-year period would average out over the long run, right?

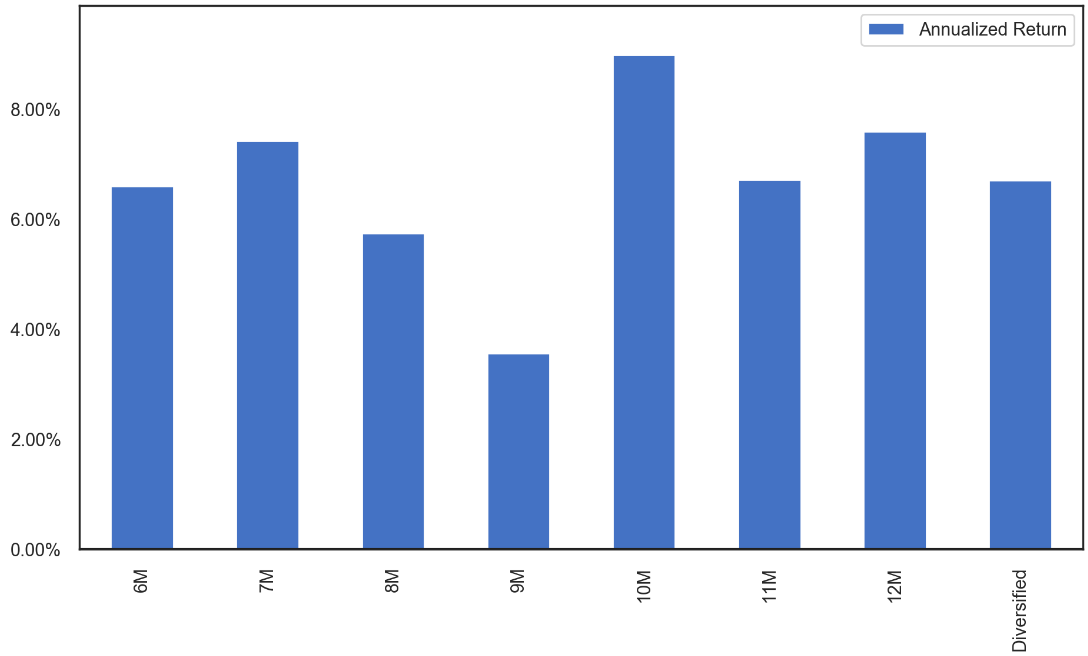
The unfortunate reality is that these performance differences are not expected to mean-revert. The gambler's fallacy would have us believe that bad luck in one year should be offset by good luck in another and vice versa. Unfortunately, this is not the case. While we would expect, at any given point in time, that each strategy has equal likelihood of experiencing good or bad luck going forward, that luck is expected to occur completely independently from what has happened in the past.

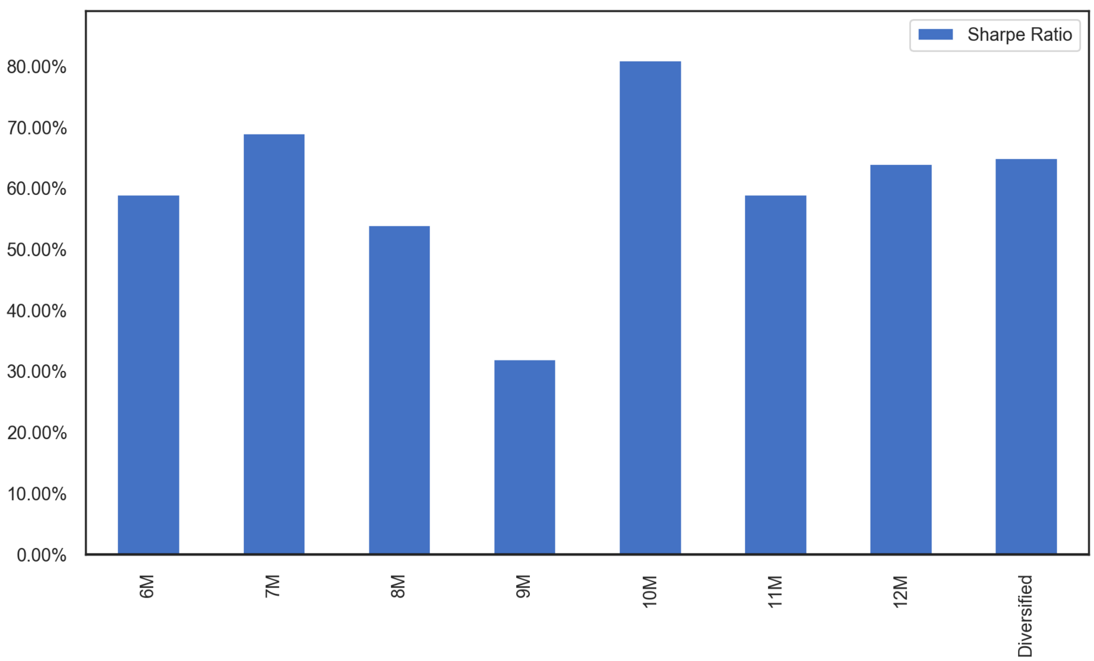
The implication is that performance differences due to model specification are not expected to mean-revert and are therefore expected to be random, but very permanent, return artifacts.[5]

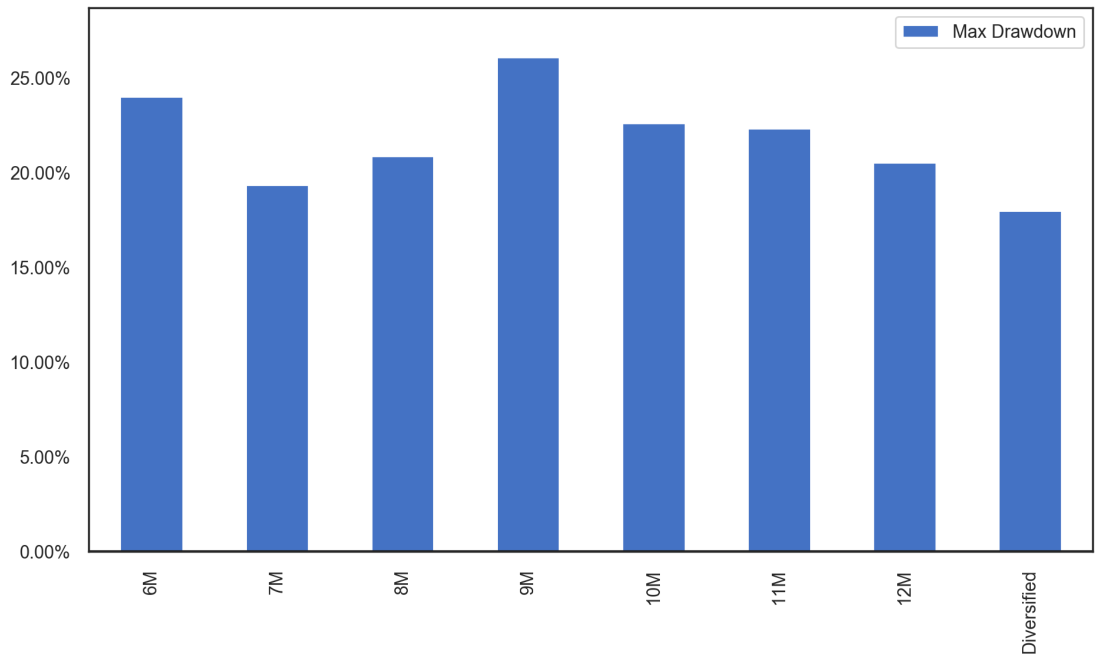
The larger problem at hand is that none of us have a hundred years to invest. In reality, most investors have a few decades. And we act with the temperament of having just a few years. Therefore, bad luck can have very permanent and very scarring effects not only upon our psyche, but upon our realized wealth.

[5] This can be more precisely analyzed using Augmented Dick-Fuller test to establish that the various pairs to not exhibit mean-reversionary behavior. A one-way ANOVA test can also be employed to establish that the various implementations have the same population mean. For the sake of brevity, we leave this as an exercise to the reader.

But consider what happens if we try to neutralize the role of model specification risk and luck by diversifying across the seven different models equally (rebalanced annually). We see that returns closer in line with the median result, a boost to realized Sharpe ratio, and a *reduction* in the maximum realized drawdown.







Source: CSI Analytics. Calculations by Newfound Research. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees except for underlying ETF expense ratios. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

These are impressive results given that all we employed was naïve diversification.

5. Conclusion

The odd thing about strategy diversification is that it *guarantees* we will be wrong. Each and every year, we will, by definition, allocate at least part of our capital to the worst performing strategy. The potential edge, however, is in being vaguely wrong rather than precisely wrong. The former is annoying. The latter can be catastrophic.

In this commentary we use the popular Dual Momentum GEM strategy as a case study to demonstrate how model specification choices can lead to performance differences that span hundreds, if not thousands, of basis points a year. Unfortunately, we should not expect these performance differences to mean revert. The realizations of good and bad luck are permanent, and potentially very significant, artifacts within our track records.

By simply diversifying across the different models, however, we can dramatically reduce specification risk and thereby reduce strategy fragility.

To be clear, no amount of diversification will protect you from the risk of the *style.* As we like to say, "risk cannot be destroyed, only transformed." In that vein, trend following strategies will always incur some sort of whipsaw risk. The question is whether it is whipsaw related to the style as a whole or to the specific implementation.

For example, in the graphs above we can see that Dual Momentum GEM implemented with a 10-month formation period experienced whipsaw in 2011 when few of the other implementations did. This is more specification whipsaw than style whipsaw. On the other hand, we can see that almost all the specifications exhibited whipsaw in late 2015 and early 2016, an indication of style whipsaw, not specification whipsaw.

Specification risk we can attempt to control for; style risk is just something we have to bear.

At Newfound, evidence such as this to informs our own trend-following mandates. We seek to diversify ourselves across the axes of *what* ("what are we investing in?"), *how* ("how are we making the decisions?"), and *when* ("when are we making those decisions?") in an effort to reduce specification risk and provide the greatest style consistency possible.

DRAWDOWNS AND PORTFOLIO LONGEVITY

January 22, 2019

SUMMARY

- While retirement planning is often performed with Monte Carlo simulations, investors only experience a single path.
- Large or prolonged drawdowns early in retirement can have a significant impact upon the probability of success.
- We explore this idea by simulation returns of a 60/40 portfolio and measuring the probability of portfolio failure based upon a quantitative measure of risk called the Ulcer Index.
- We find that a high Ulcer Index reading early in an investor's retirement can dramatically increase the probability of failure as well as the expected longevity of a portfolio.

1. Introduction

At Newfound we often say, "while other asset managers focus on alpha, our first focus is on risk."

Not that there is anything wrong with the pursuit of alpha. We'd argue that the pursuit of alpha is actually a necessary component for well-functioning financial markets.

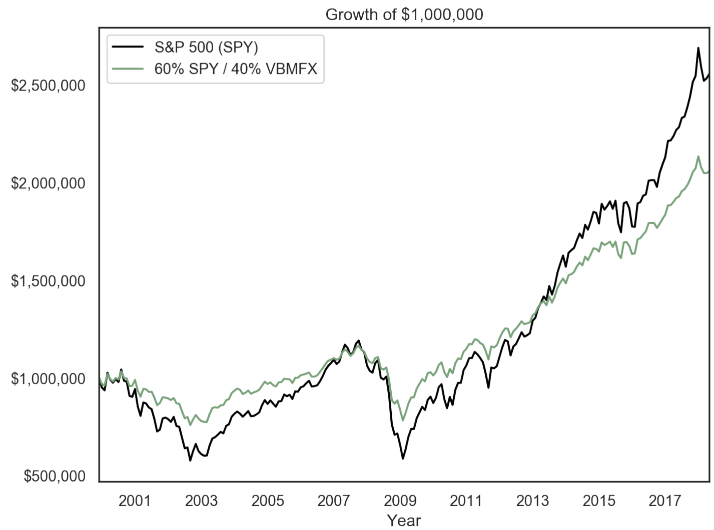
It's simply that we have never met a financial advisor who has built a financial plan that assumed any sort of alpha. Alpha is great if we can harvest it, but the empirical evidence suggesting how difficult that can be (both for the manager net-of-fees as well as the investor behaviorally) would make the *presumption* of achieving alpha rather bold.

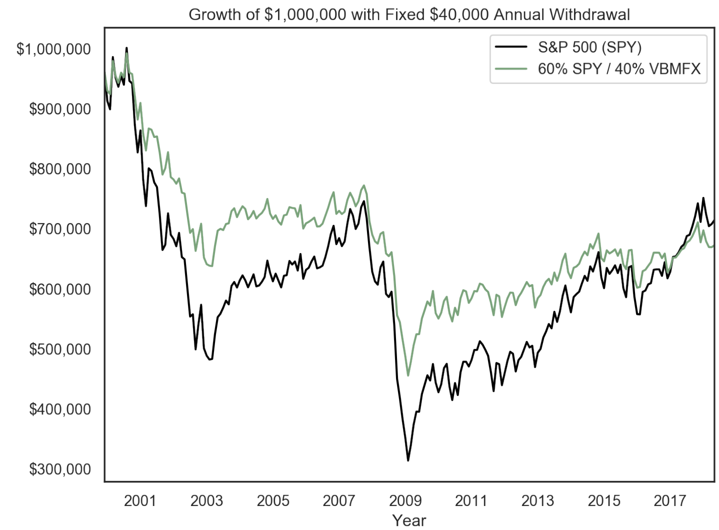
Furthermore, alpha is a zero-sum game: we can't all plan for it.

Risk, however, is a crucial element of every investor's plan. Bearing too little risk can lead to a portfolio that "fails slowly," falling short of achieving the escape velocity required to outpace inflation. Bearing too much risk, however, can lead to sudden and catastrophic ruin: a case of "failing fast."

When investors hit retirement, the usual portfolio math changes. While we're taught in Finance 101 that the order of returns does not matter, the introduction of portfolio withdrawals makes the order of returns a large determinant of plan success. This phenomenon is known as "sequence risk" and it peaks in the years just before and after retirement.

Typically, we look at returns through the lens of the investment. In retirement, however, what really matters is the returns of the investor.





We're often told that our primitive brain, trained on the African veldt, is unsuited for investing. Yet our brain seems to understand quite well that we do not get to live our lives as the average of a Monte Carlo simulation.

If we lose our arm to a lion because we did not flee when we heard a rustle in the bushes, we do not end up with half of an arm because of all the other parallel universes where we did flee. On the timeline we live, the situation is binary.

As investors, the same is true. We live but a single path and there are very real, very permanent knock-out conditions we need to be aware of. Prolonged and significant drawdowns during the first years of retirement rank among the most dangerous.

2. Drawdowns and the Risk of Ruin

A retirement plan typically establishes a safe withdrawal rate. This is the amount of inflation-adjusted money an investor can withdraw from their portfolio every year and still retain a sufficiently high probability that they will not run out of money before they die.

A well-established (albeit controversial) rule is that 4% of an investor's portfolio level at retirement is usually an appropriate withdrawal amount. For example, if an investor retires with a $1,000,000 portfolio, they can theoretically safely withdraw $40,000 a year. Another way to think of this is that the portfolio reflects 25 years of spending assuming growth matches inflation.

The problem with portfolio drawdowns is that the withdrawal rate now reflects a larger proportion of capital unless it is commensurately adjusted downward. For example, if the portfolio falls to $700,000, a $40,000 withdrawal is now 5.7% of capital and the portfolio reflects just 17.5 years of spending units.

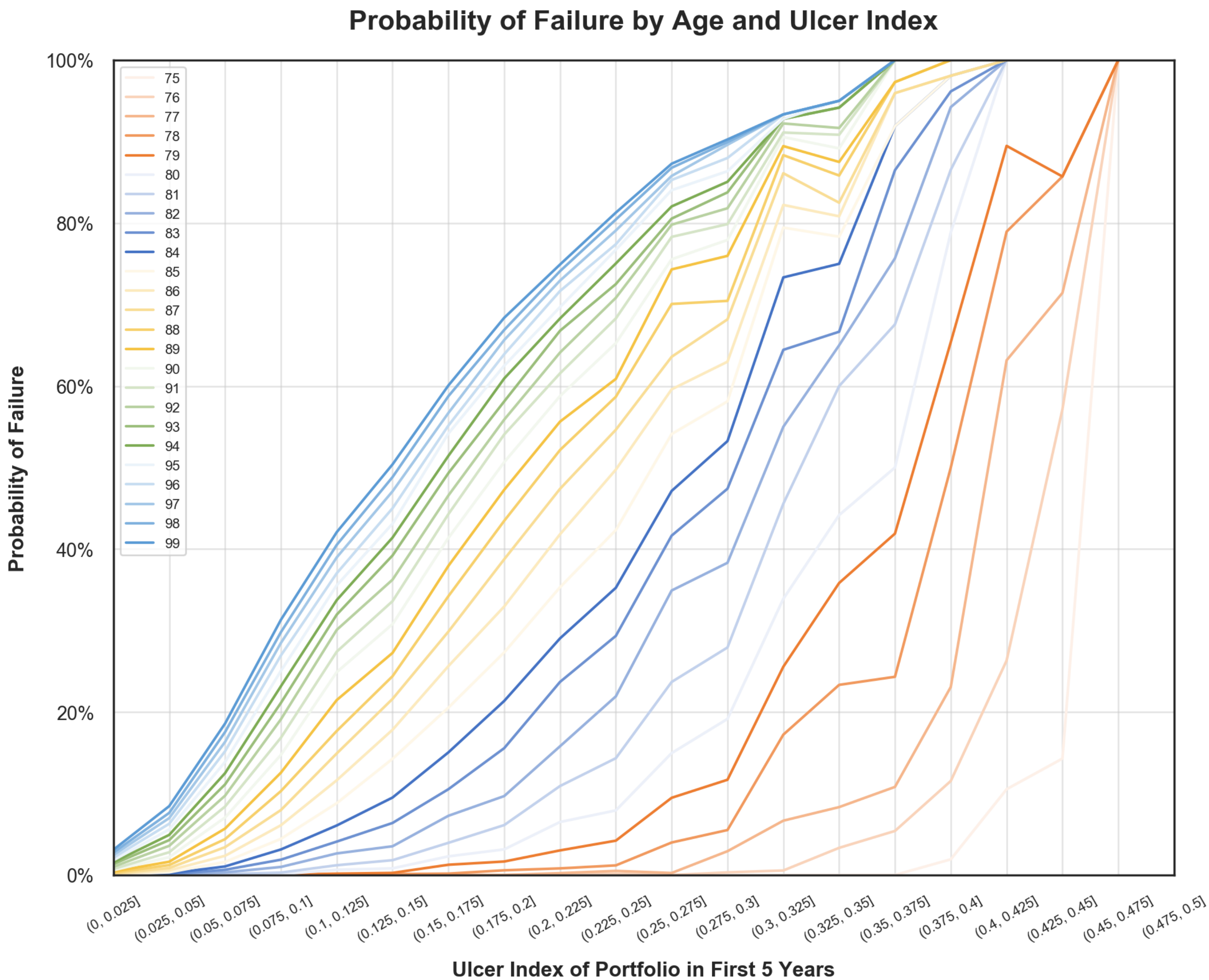
Even shallow, prolonged drawdowns can have a damaging effect. If the portfolio falls to $900,000 and stays stagnant for the next five years, the $40,000 withdrawals grow from representing 4% of the portfolio to nearly 5.5% of the portfolio. If we do not adjust the withdrawal, at five years into retirement we have gone from 25 spending units to 18.5, losing a year and a half of portfolio longevity.

As sudden and steep drawdowns can be just as damaging as shallow and prolonged ones, we prefer to use a quantitative measure known as the Ulcer Index to measure this risk. Specifically, the Ulcer Index is calculated as the root mean square of monthly drawdowns, capturing both severity and duration simultaneously.

In an effort to demonstrate the damaging impact of drawdowns early in retirement, we will run the following experiment:

- Generate 250,000 simulations, each block-bootstrapped from monthly real U.S. equity and real U.S. 5-year Treasury bond returns from 1918 – 2018.
- Assume a 65 year old investor with a $1,000,000 starting portfolio and a fixed real $3,333 withdrawal monthly ($40,000 annual).
- Assume the investor holds a 60/40 portfolio at all times.
- For each simulation:
  - Calculate the *Ulcer Index* of the first five years of portfolio returns (ignoring withdrawals).
  - Determine how many years until the portfolio runs out of money.

Based upon this data, below we plot the probability of failure – i.e. the probability we run out of money before we die – given an assumed age of death as well as the Ulcer Index realized by the portfolio in the first five years of retirement.



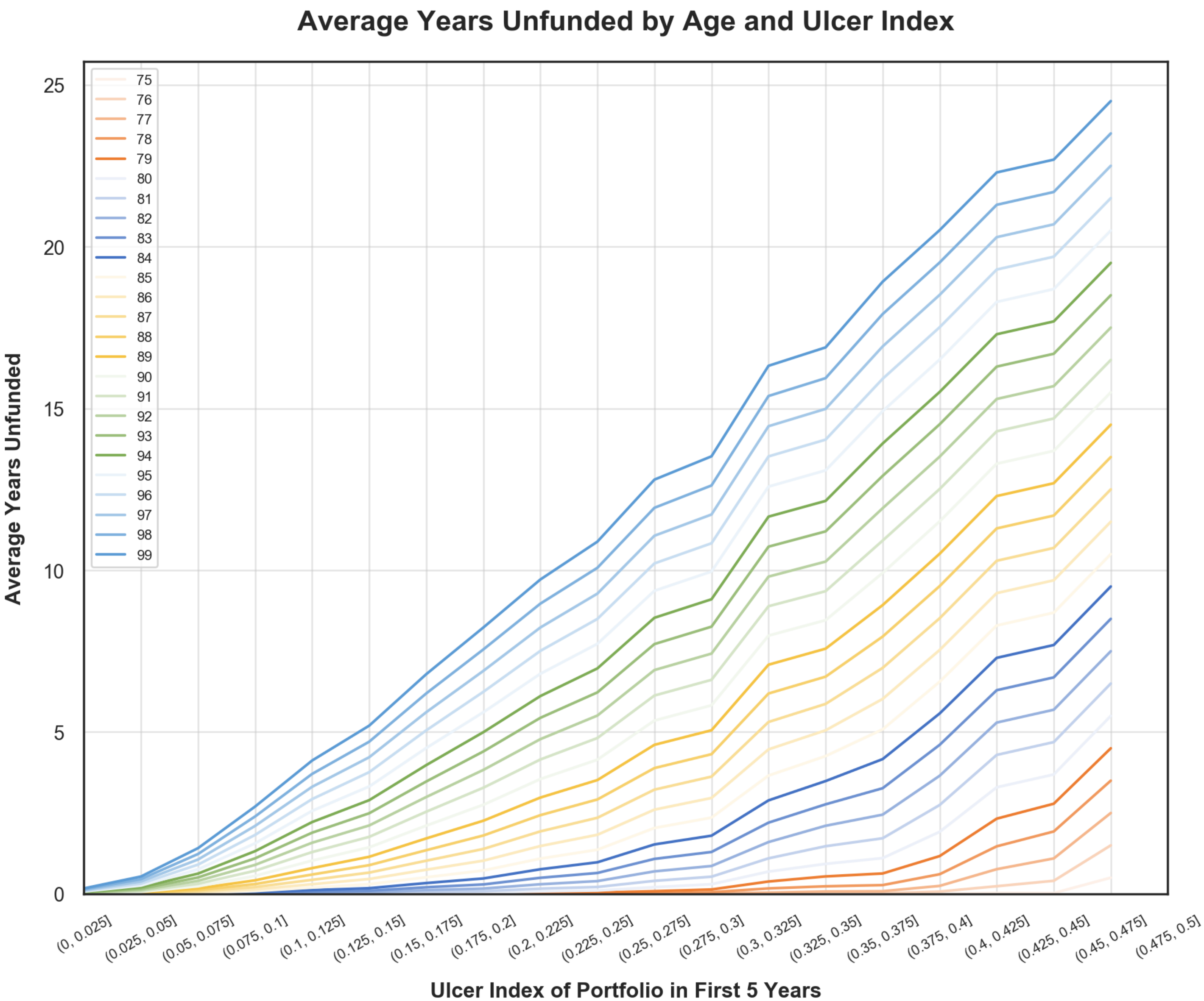
Source: Global Financial Data. Calculations by Newfound Research.

As an example of how to read this graph, consider the darkest blue line in the middle of the graph, which reflects an assumed age of death of 84. Along the x-axis are different bins of Ulcer Index levels, with lower numbers reflecting fewer and less severe drawdowns, while higher numbers reflect steeper and more frequent ones.

As we trace the line, we can see that the probability of failure – i.e. running out of money before death – increases dramatically as the Ulcer Index increases. While for shallow and infrequent drawdowns the probability of failure is <5%, we can see that the probability approaches 50% for more severe, frequent losses.

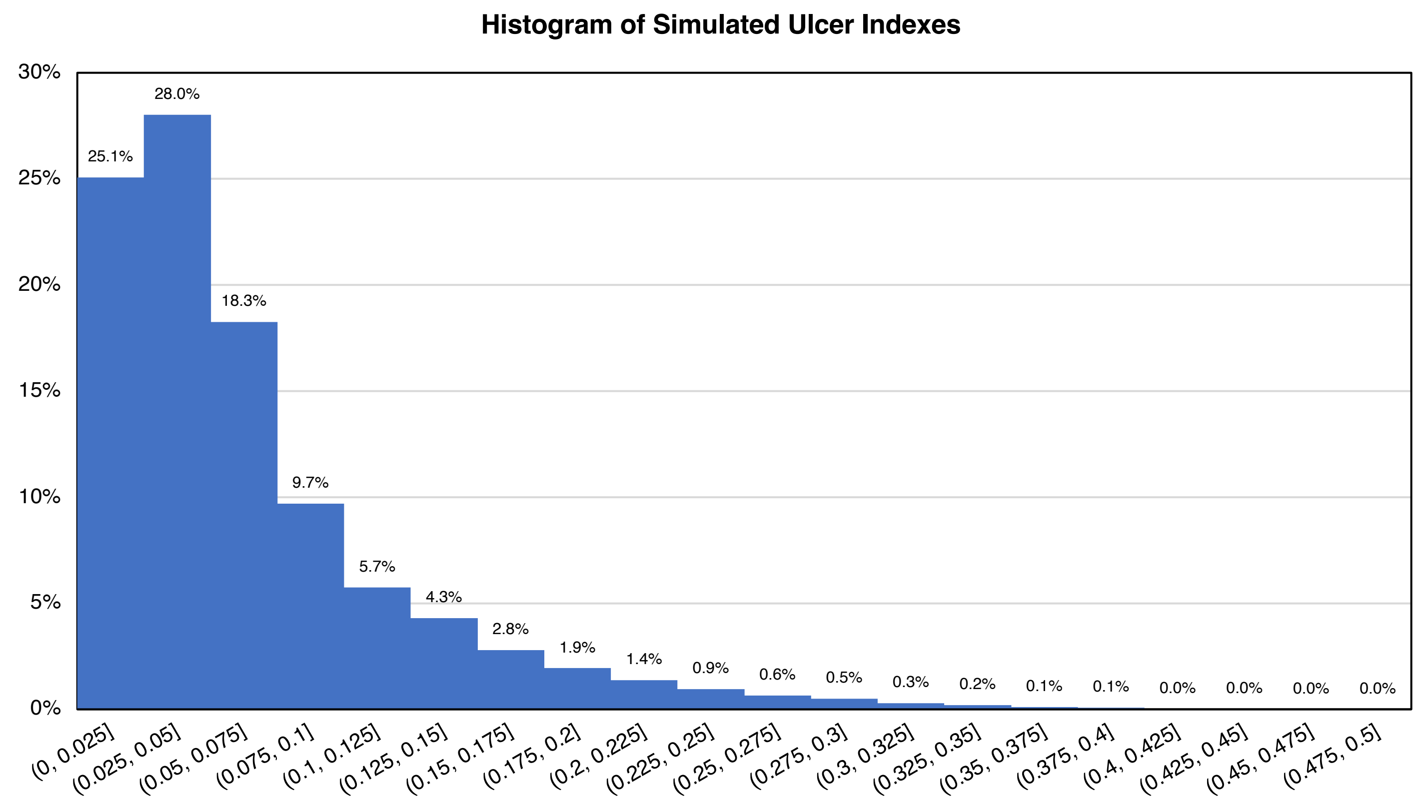
Beyond the binary question of failure, it is also important to consider when a portfolio runs out of money relative to when we die. Below we plot how many years prior to death a portfolio runs out of money, on average, based upon the Ulcer Index.

Once again using the darkest blue line as an example, we can see that for most minor-to-moderate Ulcer Index levels, the portfolio would only run out of money a year or two before we die in the case of failure. For more extreme losses, however, the portfolio can run out of money a full decade before we kick the bucket.



Source: Global Financial Data. Calculations by Newfound Research.

It is worth stressing here that these Ulcer Index readings are derived using simulations based upon prior realized U.S. equity and fixed income returns. In other words, while improbable (see the histogram below), extreme readings are not impossible.



It is worth further acknowledging that U.S. assets have experienced some of the highest realized risk premia in the world, and more conservative estimates may put a higher probability mass on more extreme Ulcer Index readings.

3. Conclusion

For early retirees, large or prolonged drawdowns early in retirement can have a significant impact on the probability of success.

In this commentary, we capture both the depth and duration of drawdowns using a single metric known as the Ulcer Index. We simulate 250,000 possible return paths for a 60/40 portfolio and calculate the Ulcer Index in the first five years of returns. We then plot the probability of failure as well as expected portfolio longevity conditional upon the Ulcer Index level realized.

We clearly see a positive relationship between failure and Ulcer Index, with larger and more prolonged drawdowns earlier in retirement leading to a higher probability of failure. This phenomenon is precisely why investors tend to de-risk their portfolios over time.

While the right risk profile and a well-diversified portfolio make for a strong foundation, we believe that investors should also consider expanding their investment palette to include alternative assets and style premia that may be more defensive oriented in nature. For example, defensive equities (e.g. low-volatility and quality approaches) have historically demonstrated an ability to reduce drawdown risk. Diversified, multi-asset style premia also tend to exhibit low correlation to traditional risk factors and a low intrinsic style premia.

Here at Newfound, we focus on trend equity strategies, which seek to overlay trend-following approaches on top of equity exposures in an effort to reduce left-tail risk and create a higher quality of return profile.

However, an investor chooses to build their portfolio, however, it should be risk that is on the forefront of their mind.

TIGHTENING THE UNCERTAIN PAYOUT OF TREND-FOLLOWING

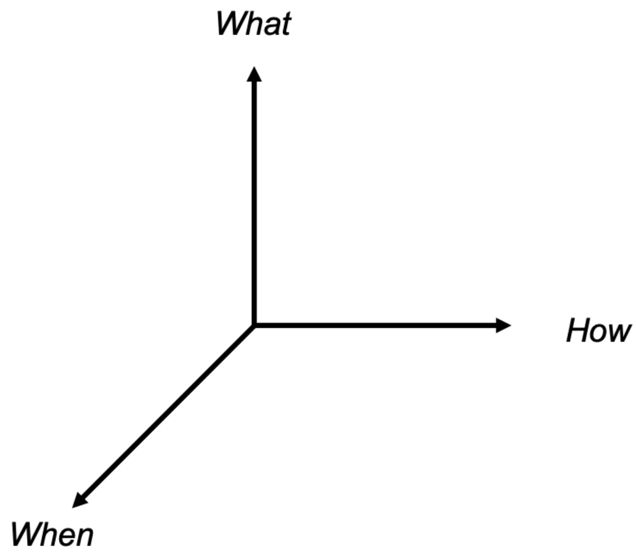
January 28, 2019

SUMMARY

- Long/flat trend-following strategies have historically delivered payout profiles similar to those of call options, with positive payouts for larger positive underlying asset returns and slightly negative payouts for near-zero or negative underlying returns.
- However, this functional relationship contains a fair amount of uncertainty for any given trend-following model and lookback period.
- In portfolio construction, we tend to favor assets that have a combination of high expected returns or diversifying return profiles.
- Since broad investor behavior provides a basis for systematic trend-following models to have positive expected returns, taking a multi-model approach to trend-following can be used to reduce the variance around the expected payout profile.

1. Introduction

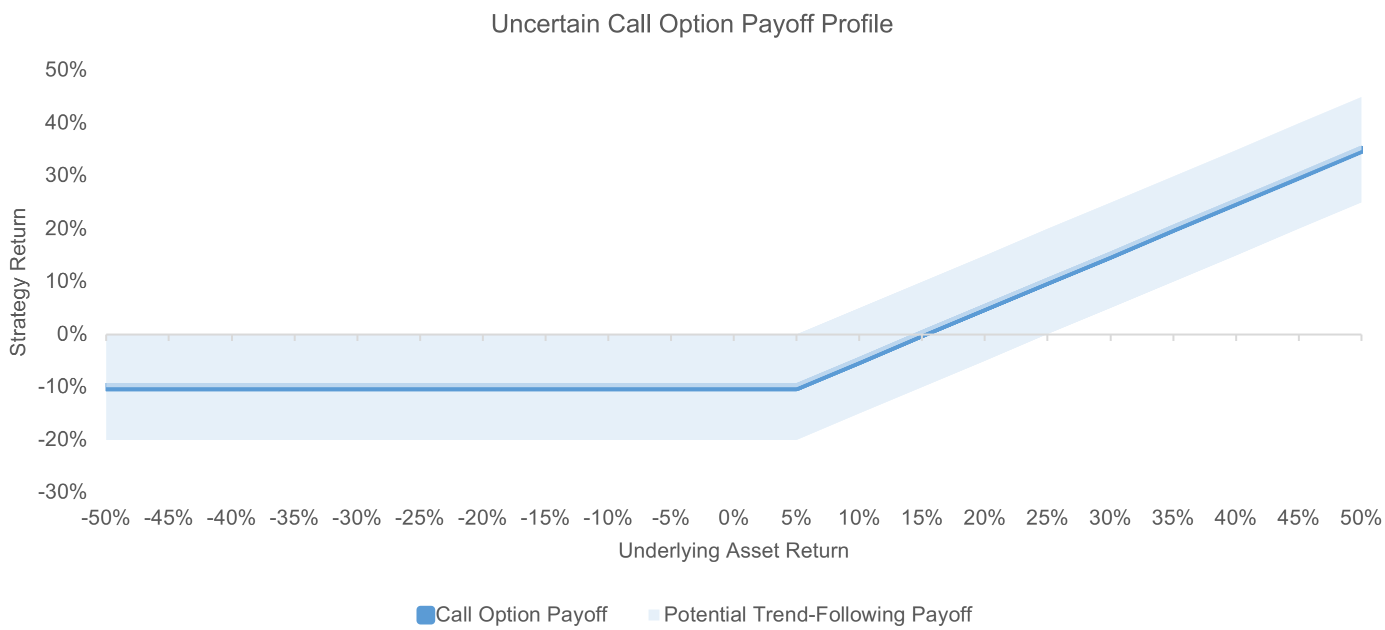
Over the past few months, we have written much about model diversification as a tactic for managing specification risk, even with specific case studies. When we consider the three axes of diversification, model diversification pertains to the "how" axis, which focuses on strategies that have the same overarching objective but go about achieving it in different ways.



Long/flat trend-following, especially with equity investments, aims to protect capital on the downside while maintaining participation in positive markets. This leads to a payout profile that looks similar to that of a call option.[6]

However, while a call option offers a defined payout based on the price of an underlying asset and a specific maturity date, a trend-following strategy does not provide such a guarantee. There is a degree of uncertainty.

[6] We have written about decomposing trend-equity strategies into option overlays before. See https://blog.thinknewfound.com/2018/09/decomposing-trend-equity/



The good news is that uncertainty can potentially be diversified given the right combinations of assets or strategies.

In this commentary, we will dive into a number of trend-following strategies to see what has historically led to this benefit and the extent that diversification would reduce the uncertainty around the expected payoff.

2. Diversification in Trend-Following

The justification for a multi-model approach boils down to a simple diversification argument.

Say you would like to include trend-following in a portfolio as a way to manage risk (e.g. sequence risk for a retiree). There is academic and empirical evidence that trend-following works over a variety of time horizons, generally ranging from 3 to 12 months. And there are many ways to measure trends, such as moving average crossovers, trailing returns, deviations from moving averages, risk adjusted returns, etc.

The basis for deciding ex-ante which variant will be the best over our own investment horizon is tenuous at best. Backtests can show one iteration outperforming over a given time horizon, but most of the differences between strategies are either noise from a statistical point of view or realized over a longer time period than any investor has the lifespan (or mettle) to endure.

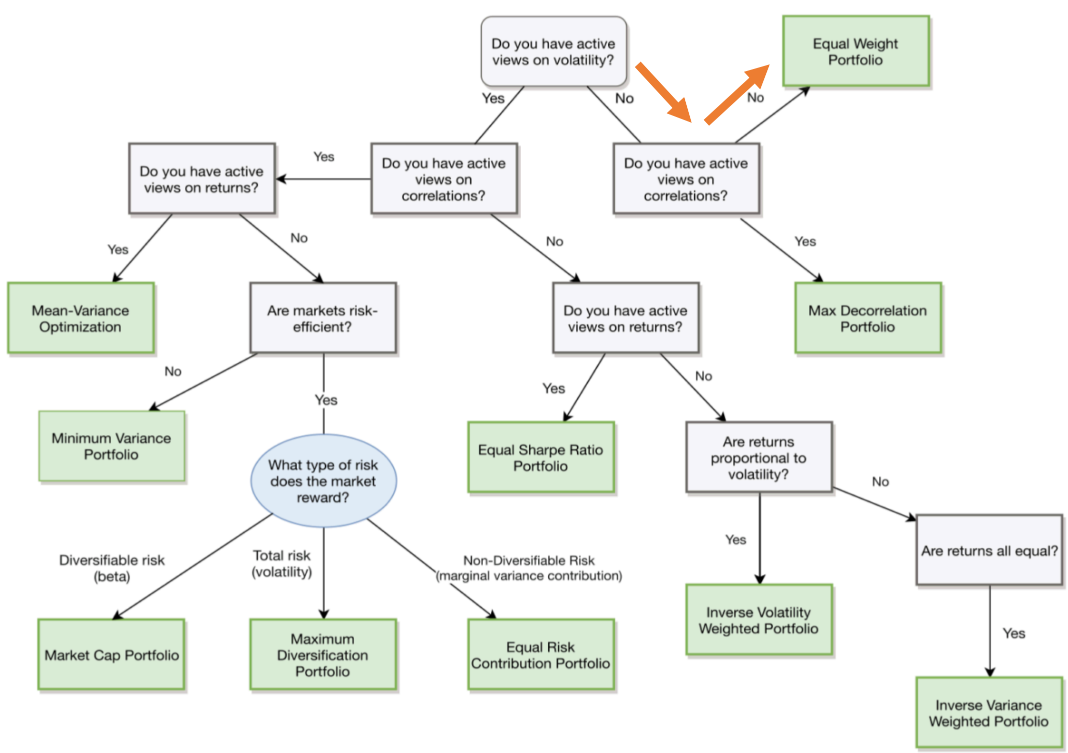
However, we expect each one to generate positive returns over a sufficiently long time horizon. Whether this is one year, three years, five years, 10 years, 50 years… we don't know. What we do know is that out of the multitude the variations of trend-following, we are very likely to pick one that is *not* the best or even in the top segment of the pack in the short-term.

From a volatility standpoint, when the strategies are fully invested, they will have volatility equal to the underlying asset. Determining exactly when the diversification benefits will come in to play – that is, when some strategies are invested and others are not – is a fool's errand.

Modern portfolio theory has done a disservice in making correlation seem like an inherent trait of an investment. It is not.

Looking at multiple trend-following strategies that can coincide precisely for stretches of time before behaving completely differently from each other, makes many portfolio construction techniques useless. We do not expect correlation benefits to always be present. These are nonlinear strategies, and fitting them into a linear world does not make sense.

If you have pinned up ReSolve Asset Management's flow chart of portfolio choice above your desk (from Portfolio Optimization: A General Framework for Portfolio Choice), then the decision on this is easy.



Source: ReSolve Asset Management. Reprinted with permission

From this simple framework, we can break the different performance regimes down as follows:

Feeling	Scenario	Description
👍	High correlation and on the right side of the trend	This is trend-following at its finest.
👎	High correlation and on the wrong side of the trend	This is whipsaw and is from style risk, which is tough to diversify *within* trend-following strategies.
💪	Low correlation (therefore mixed sides of the trend)	This can be disappointing because the underperforming trend-following strategies can feel like a drag on overall performance, but this is better than picking the worst performer. This is the robust behavior that leads to long-term outperformance.

3. The Math Behind the Diversification

The expected value of a trend-following strategy can be thought of as a function of the underlying security return:

$$\mathbb{E}[r_{i,trend}] = f_i(r_{underlying})$$

Where the subscript *i* is used to indicate that the function is dependent on the specific trend-following strategy.

If we combine multiple trend-following strategies into a portfolio, then the expectation is the average of these functions (assuming an equal weight portfolio per the ReSolve chart above):

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} r_{i,trend}\right] = \frac{1}{N}\sum_{i=1}^{N} f_i(r_{underlying})$$

What's left to determine is the functional form of *f*.

Continuing in the vein of the call option payoff profile, we can use the Black-Scholes equation as the functional form (with the risk-free rate set to 0). This leaves three parameters with which to fit the formula to the data: the volatility (with the time to expiration term lumped in, i.e. $\sigma\sqrt{T-t}$), the strike, and the initial cost of the option.

$$f_i(r_{underlying}) = \left(\frac{1 + r_{underlying}}{1 + r_K}\right) N(d_1) - N(d_2)$$

where d_1 and d_2 are defined in the standard fashion and N is the cumulative normal distribution function.

r_K is the strike price in the option formula expressed as a percent relative to the current value of the underlying security.

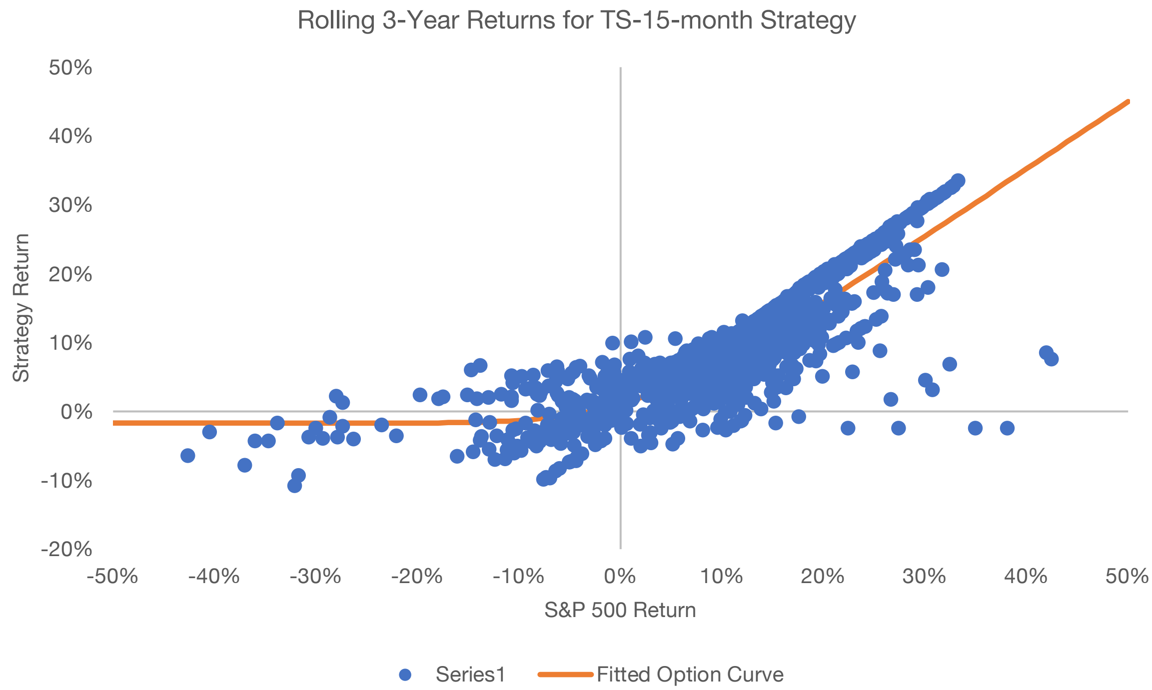
In the following example, we will attempt to provide some meaning to the fitted parameters. However, keep in mind that any mapping is not necessarily one-to-one with the option parameters. The functional form may apply, but the parameters are not ones that were set in stone ex-ante.[7]

4. An Example: Trend-Following on the S&P 500

As an example, we will consider a trend-following model on the S&P 500 using monthly time-series momentum with lookback windows ranging from 4 to 16 months. The risk-free rate was used when the trends were negative.

The graph below shows an example of the option price fit to the data using a least-squares regression for the 15-month time series momentum strategy using rolling 3-year returns from 1927 to 2018.

[7] Using options to describe trend-following strategies has been done before, e.g. by Dao, et. al.(2016) in their paper entitled Tail protection for long investors: Trend Convexity at Work. In that paper, the authors replicate a long/short trend-following strategy using a portfolio of strangle option strategies. In this context, the parameters of the options would have a definite meaning.



Source: Global Financial Data and Kenneth French Data Library. Calculation by Newfound. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

The volatility parameter was 9.5%, the strike was 2.3%, and the cost was 1.7%.

What do these parameters mean?

As we said before this can be a bit tricky. Painting in broad strokes:

- The volatility parameter describes how "elbowed" payoff profile is. Small values are akin to an option close to expiry where the payoff profile changes abruptly around the strike price. Larger values yield a more gentle change in slope.

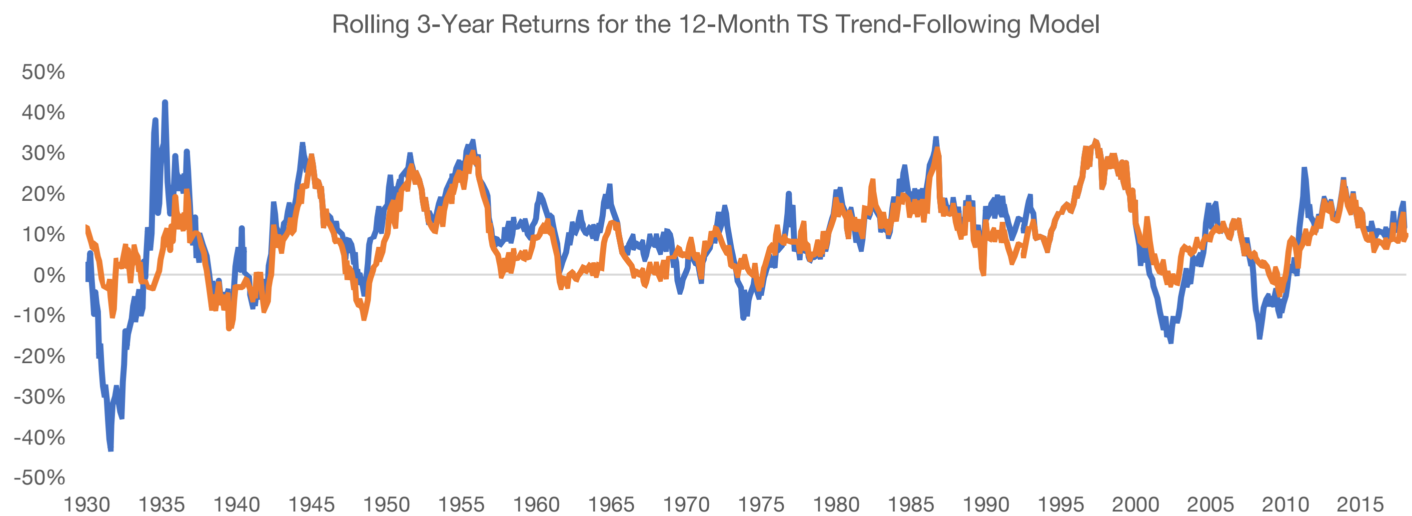
- The strike represents the point at which the payoff profile changes from participation to protection using trend-following lingo. In the example where the strike is 2.3%, this means that the strategy would be expected to start protecting capital when the S&P 500 return is less than 2.3%. There is some cost associated with this value being high.

- The cost is the vertical shift of the payoff profile, but it is not good to think of it as the insurance premium of the trend-following strategy. It is only one piece. To see why this is the case, consider that the fitted volatility may be large and that the option price curve may be significantly above the final payout curve (i.e. if the time-scaled volatility went to zero).

So what is the actual “cost” of the strategy?

With trend-following, since whipsaw is generally the largest potential detractor, we will look at the expected return on the strategy when the S&P 500 is flat, that is, an absence of an average trend. It is possible for the cost to be negative, indicating a positive expected trend-following return when the market was flat.

Looking at the actual fit of the data from a statistical perspective, the largest deviations from the expected value (the residuals from the regression) are seen during large positive returns for the S&P 500, mainly coming out of the Great Depression. This characteristic of individual trend-following models is generally attributable to the delay in getting back into the market after a prolonged, severe drawdown due to the time it takes for a new positive trend to be established.

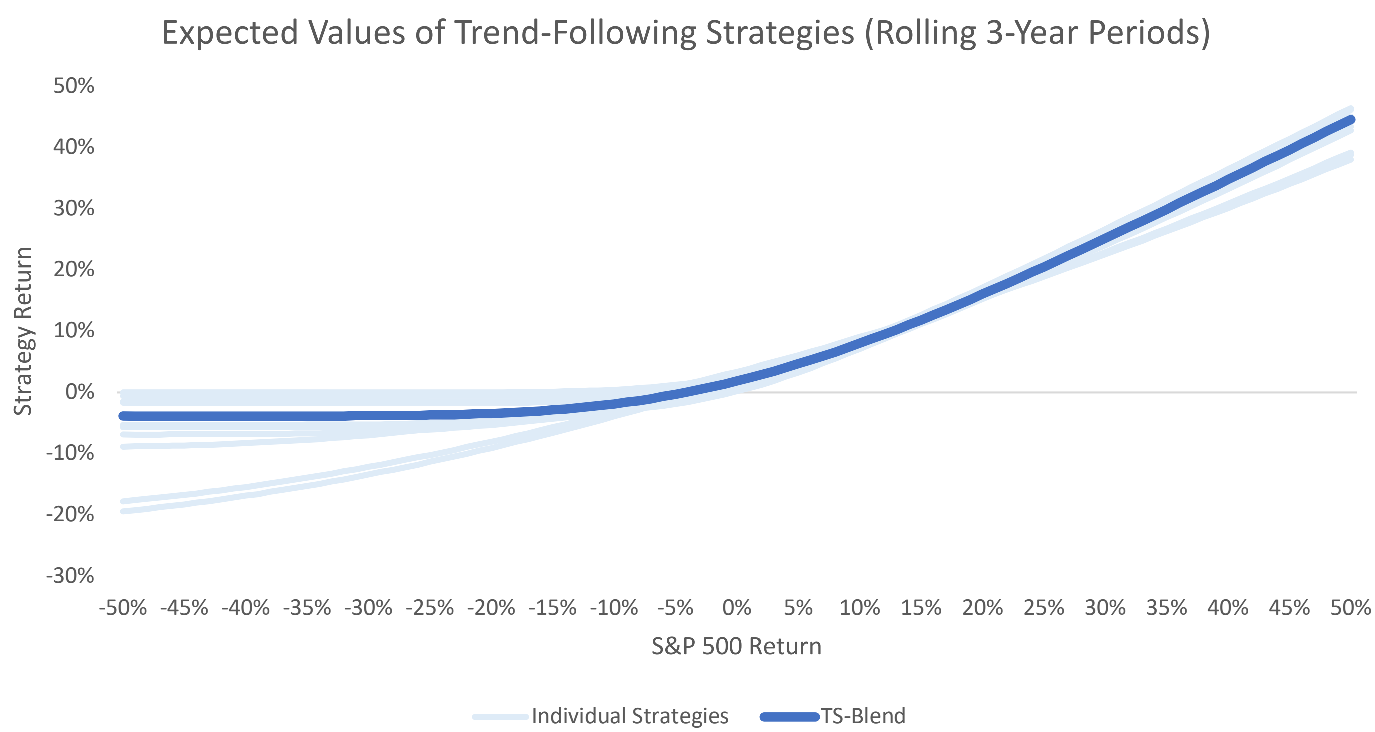


Source: Global Financial Data and Kenneth French Data Library. Calculation by Newfound. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

Part of the seemingly large number of outliers is simply due to the fact that these returns exhibit autocorrelation since the periods are rolling, which means that the data points have some overlap. If we filtered the data down into non-overlapping periods, some of these outliers would be removed.

The outliers that remain are a fact of trend-following strategies. While this fact of trend-following cannot be totally removed, some of the outliers may be managed using multiple lookback periods.

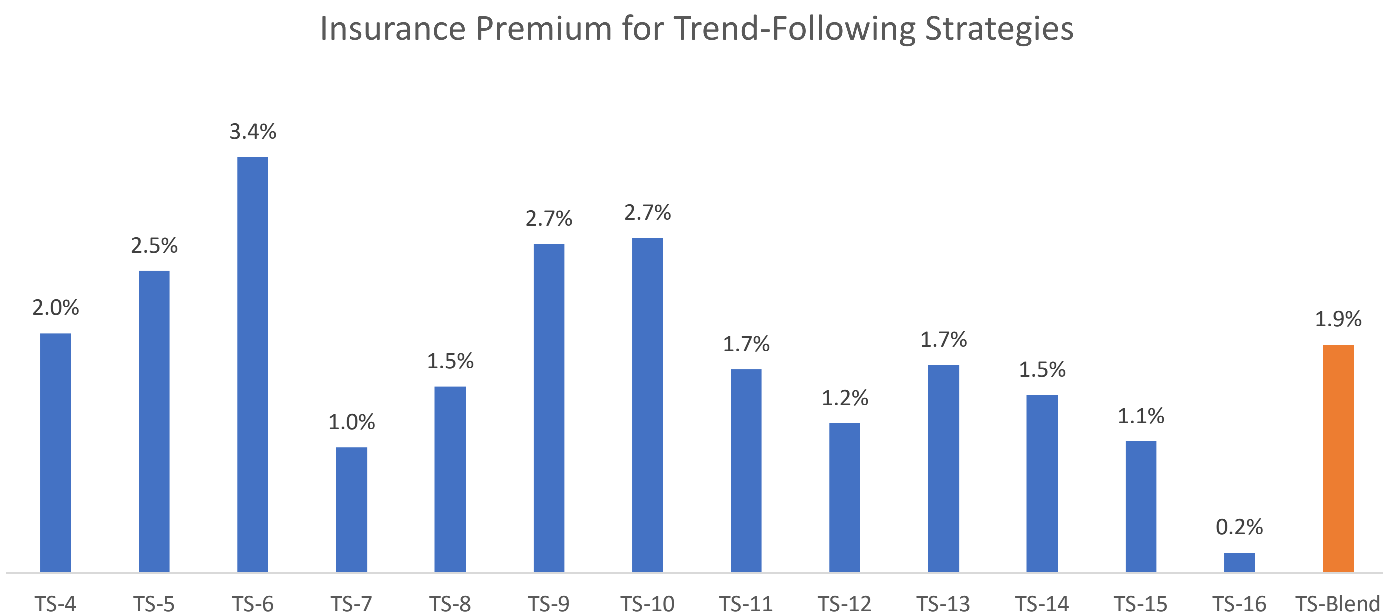
The following chart illustrates the expected values for the trend-following strategies over all the lookback periods.



Source: Global Financial Data and Kenneth French Data Library. Calculation by Newfound. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

The shorter-term lookback windows have the expected value curves that are less horizontal on the left side of the chart (higher volatility parameter).

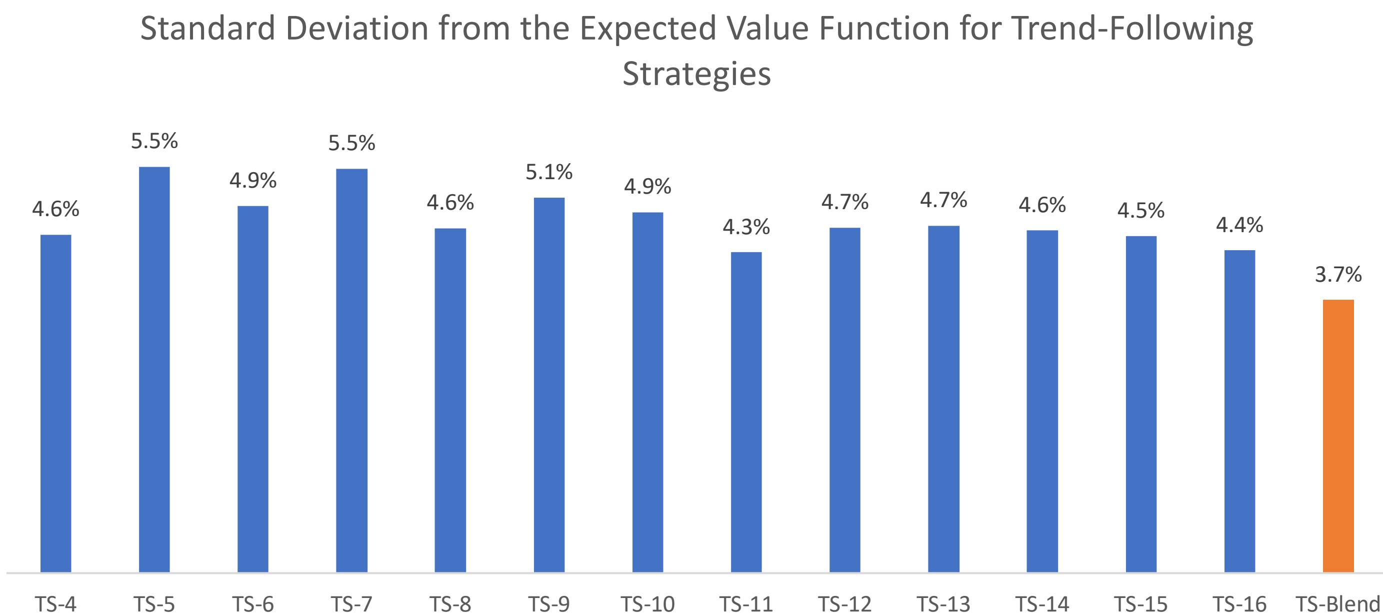
As we said before the cost of the trend-following strategy can be represented by the strategy's expected return when the S&P 500 is flat. This can be thought of as the premium for the insurance policy of the trend-following strategies.



Source: Global Financial Data and Kenneth French Data Library. Calculation by Newfound. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

The blend does not have the lowest cost, but this cost is only one part of the picture. The parameters for the expected value functions do nothing to capture the distribution of the data *around* – either above or below – these curves.

The diversification benefits are best seen in the distribution of the rolling returns around the expected value functions.



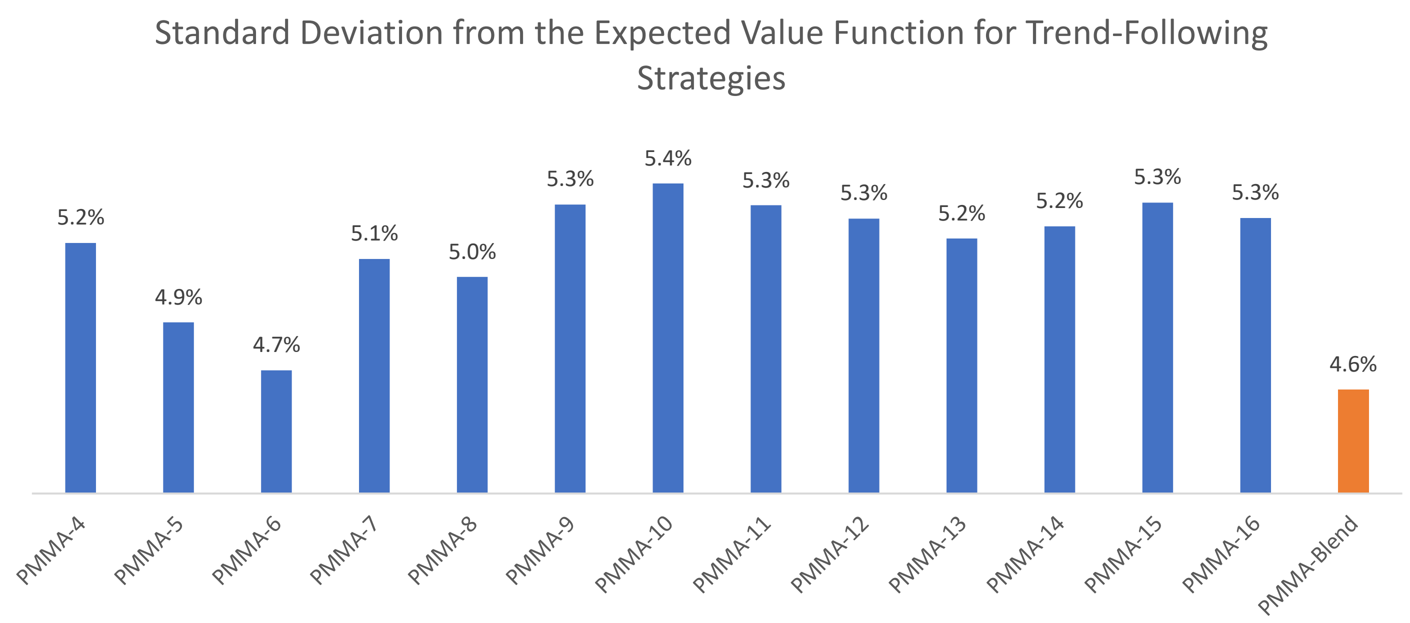
Source: Global Financial Data and Kenneth French Data Library. Calculation by Newfound. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

Now with a more comprehensive picture of the potential outcomes, a cost difference of even 3% is less than one standard deviation, making the blended strategy much more robust to whipsaw for the potential range of S&P 500 returns.

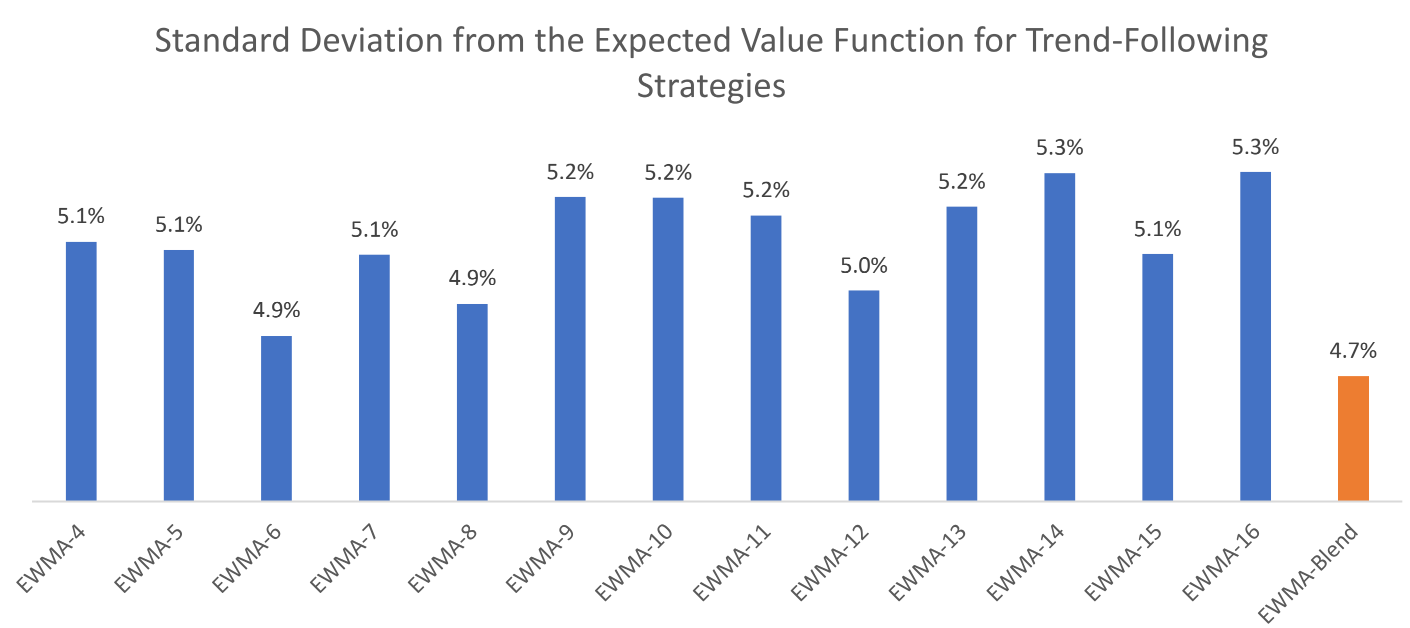
As a side note, the cost of the short window (4 and 5 month) strategies is relatively high. However, since there are many rolling periods when these models are the best performing of the group, there can still be a benefit to including them. With them in the blend, we still see a reduction in the dispersion around the expected value function.

5. Expanding the Multitude of Models

To take the example even further down the multi-model path, we can look at the same analysis for varying lookback windows for a price-minus-moving-average model and an exponentially weighted moving average model.

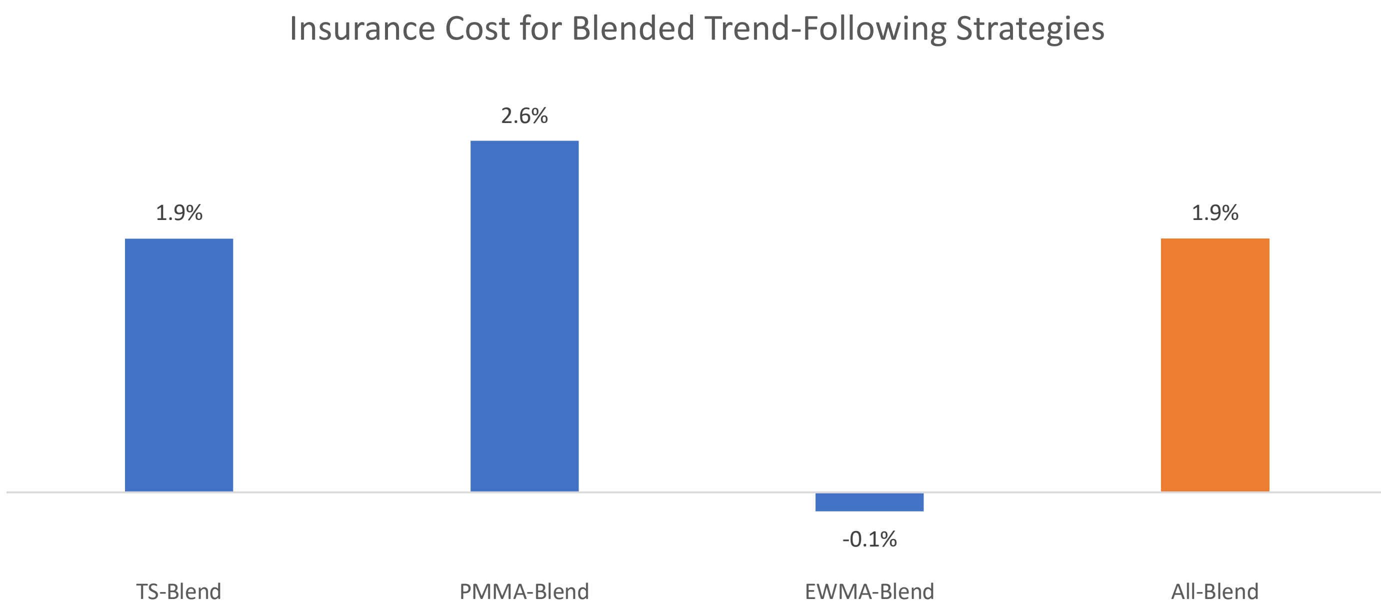


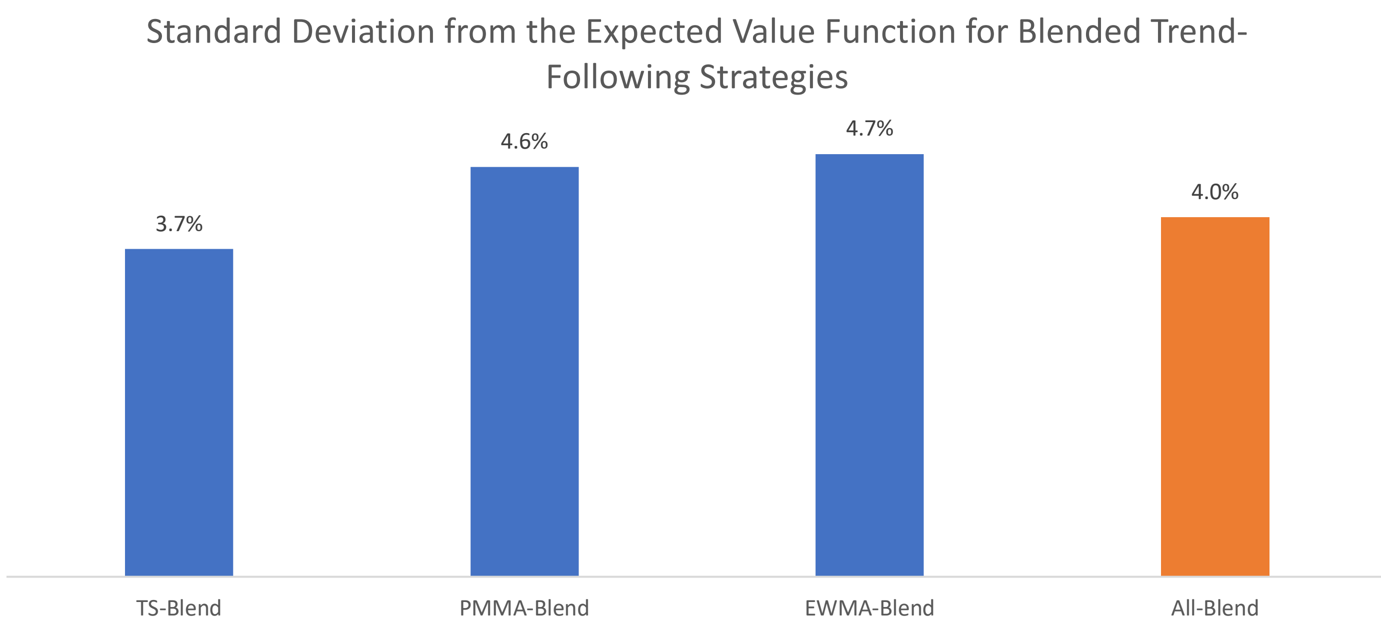
Source: Global Financial Data and Kenneth French Data Library. Calculation by Newfound. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.



Source: Global Financial Data and Kenneth French Data Library. Calculation by Newfound. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

And finally, we can combine all three trend-following measurement style blends into a final composite blend.





Source: Global Financial Data and Kenneth French Data Library. Calculation by Newfound. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

As with nearly every study on diversification, the overall blend is not the best by all metrics. In this case, its cost is higher than the EWMA blended model and its dispersion is higher than the TS blended model. But it exhibits the type of middle-of-the-road characteristics that lead to results that are robust to an uncertain future.

6. Conclusion

Long/flat trend-following strategies have payoff profiles similar to call options, with larger upsides and limited downsides. Unlike call options (and all derivative securities) that pay a deterministic amount based on the underlying securities prices, the payoff of a trend-following strategy is uncertain,

Using historical data, we can calculate the expected payoff profile and the dispersion around it. We find that by blending a variety of trend-following models, both in how they measure trend and the length of the lookback window, we can often reduce the implied cost of the call option and the dispersion of outcomes.

A backtest of an individual trend-following model can look the best over a given time period, but there are many factors that play into whether that performance will be valid going forward. The assets have to behave similarly, potentially both on an absolute and relative basis, and an investor has to hold the investment for a long enough time to weather short-term underperformance.

A multi-model approach can address both of these.

It will reduce the model specification risk that is present ex-ante. It will not pick the best model, but then again, it will not pick the worst.

From an investor perspective, this diversification reduces the spread of outcomes which can lead to an easier product to hold as a long-term investment. Diversification among the models may not always be present (i.e. when style risk dominates and *all* trend-following strategies do poorly), but when it is, it reduces the chance of taking on uncompensated risks.

Taking on compensated risks is a necessary part of investing, and in the case of trend-following, the style risk is something we desire. Removing as many uncompensated risks as possible leads to more pure forms of this style risk and strategies that are robust to unfavorable specifications.

TREND: CONVEXITY & PREMIUM

February 11, 2019

SUMMARY

- Trend following is unique among style premia in that it has historically exhibited a convex payoff profile with positive skew.
- While the historical premium is anomalous, the convexity makes sense when we use options to replicate trend following strategies.
- We explore reasons why frequent rebalancing in trend following strategies is necessary and decompose the return contributions from different portions of the option replication model.
- Most of the historical premium associated with trend following comes from the trading impact that is tied to the asset's Sharpe ratio rather than the convexity.
- By separating the impacts of convexity and trading impact, we can gain a deeper understanding of the types of risk exchanges that come with investing using trend following strategies.

1. Introduction

Unlike many of the other style premia, trend following has historically exhibited a convex payoff profile with positive skew. In less mathematical terms, that means it tends to harvest many small losses (due to reversals) and just a few large gains (when trends take off).

This is unique, as most risk and style premia exhibit the opposite: concave payoffs with negative skew.

To simplify, we can think of concave, negative-skew trades as akin to *selling* insurance, while convex, positive-skew trades are akin to buying it. With respect to traditional financial literature, the mental model of selling insurance makes quite a bit of sense, as we can think of expected excess returns as being the reward earned for being willing to bear the risk others wish to transfer away.

Trend following, then, is a bit of an anomaly. Not because it exhibits a convex, positive skew profile, but because it does so *and* has historically exhibited a positive premium. That is not supposed to happen – you don't expect to profit when you buy insurance – and it leaves many scratching their head asking, "could this be an *actual* market anomaly?"

Unfortunately, a discussion of *why* trend following works often conflates the convexity of the strategy with the oddness of the historically positive premium. The latter is, for sure, anomalous. But what we hope to show in this commentary is that the former is just a byproduct of the trading strategy itself and does not require any investor misbehavior.

How will we do this? Whenever we talk about buying or selling insurance, a very natural language to use is that of put and call options. Thus, our goal in this commentary is to approach trend following through the lens of options and demonstrate that simple trend-following strategies can be thought of as naively replicating the pay-off of a straddle.

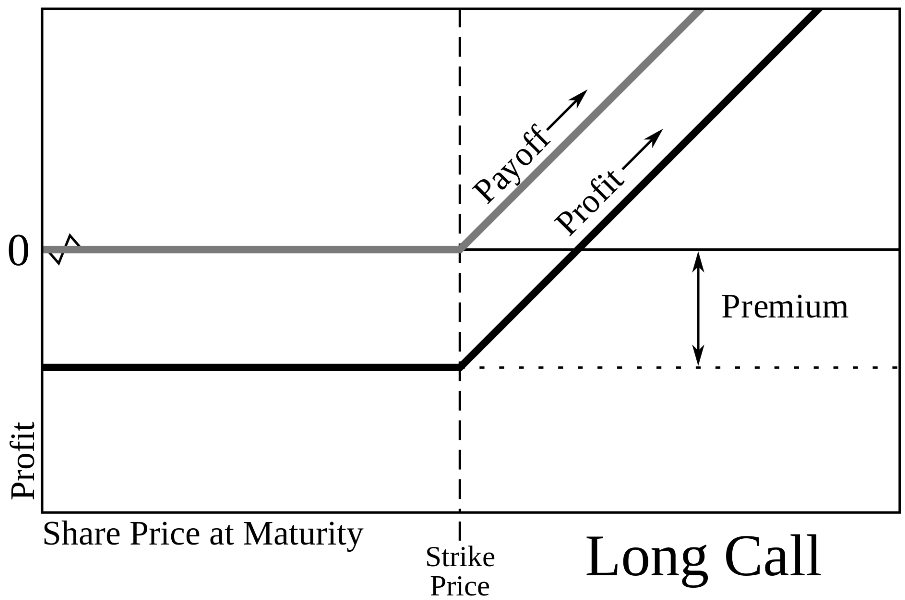
In doing so, our goal is to differentiate between two key elements of trend following: the convexity of returns it exhibits and the historically positive premium it has generated.

Please note, for all the option geeks out there, that this will be a highly simplified interpretation with lots of hand-waving. For example, we will generally assume that interest rates are zero, dividends do not exist, that price is continuous (i.e. no jumps) and there are no trading frictions. None of this is true, of course, but we do not think it meaningfully takes away from the intuition established.

2. Pricing is Replication

A foundational principle in financial engineering is the Law of One Price, which states that any two securities with identical future payouts, no matter how the future turns out, must have an identical price. Otherwise, we could construct an arbitrage.

Thus, to price an option, we only need to replicate its payoff. This is, of course, easier said than done, as options have non-linear payoffs. A call option, for example, pays nothing when price is below the strike at maturity and pays the difference between price and the strike otherwise.



Source: Wikipedia

Given access to a risk-free bond and the underlying stock, we cannot easily replicate this payoff with some sort of static portfolio. We can, however, attempt to replicate it using a *dynamic* trading strategy that adjusts our mixture of the stock and bond over time.

Consider the following example: a stock is priced at $80 and will be worth either $100 or $60 in one year. We have a call option with a strike of $90. This means if the stock ends up at $100, the payoff will be $10 while if the stock ends up at $60, the payoff will be zero.

How can we replicate this?

We need to solve the simultaneous equations:

$$\$100\Delta - B(1+r) = 10$$

$$\$60\Delta - B(1+r) = 0$$

Here, Δ is the number of shares of stock to buy, B is how much to borrow, and r is the risk-free rate. Given r, we can solve the equations for the replicating portfolio. If we assume r=0%, we find that Δ=0.25 and therefore B=$15.

Thus, to replicate the call option, we need to borrow $15 and buy 0.25 shares at $80, for a total cost of $5. Since this portfolio replicates the option payout exactly, this must also be the price of the option!

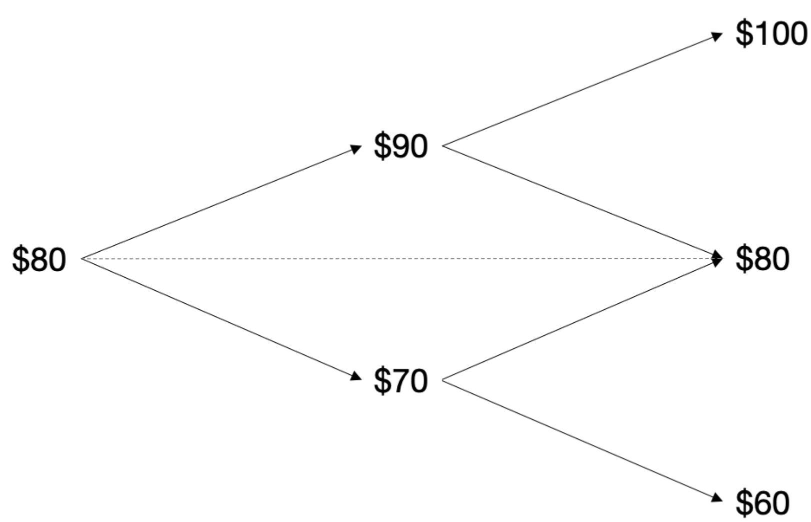
Of course, these are highly simplified assumptions. But if we collapse the time period down from 1-year to an infinitesimally small unit of time, we can repeat this exercise over and over such that we have a dynamic trading strategy that will replicate the option's final payoff, and therefore the option's value over time.

(For a more thorough – but still highly accessible – introduction to this concept, we recommend Emanuel Derman's *The Boy's Guide to Pricing & Hedging*.)

3. Trend Following is "Long Gamma"

Let us now connect trend following to options.

Consider the following case where the underlying stock price follows a binomial tree and we want to replicate the payoff of a call option with a strike of $80. Again, to simplify things, we will just assume that our risk-free rate is 0%.



Let's start when price is at $90. At this point, to replicate the final payoff, we have the following two equations:

$$\$100\Delta - B = 20$$

$$\$80\Delta - B = 0$$

Again, solving simultaneously, we find that we need to buy 1 share (at $90) and therefore B=$80. The option, therefore, is equal to $10.

Now let's consider the bottom case when the stock price is $70,

$$\$80\Delta - B = 0$$

$$\$60\Delta - B = 0$$

In this case, we find that the number of shares is equal to zero and therefore B=$0.

Finally, using these prices for the option in those two states, we can step back to the starting case, where we now know:

$$\$90\Delta - B = 10$$

$$\$70\Delta - B = 0$$

Here we find that the number of shares is equal to 0.5 and B is equal to $35, making the option therefore worth $5.

This highly simplified model tells us that:

- The number of shares held in the replicating portfolio informs how sensitive the option price is to movement in the underlying stock. Note that at the initial step the option value was $5 and the number of shares held was 0.5. As price changed by +/- $10, the value of the option changed by +/- $5.

  The more shares held, the more sensitive the option price is to the stock price change. In option's parlance, this is known as the option's "delta."

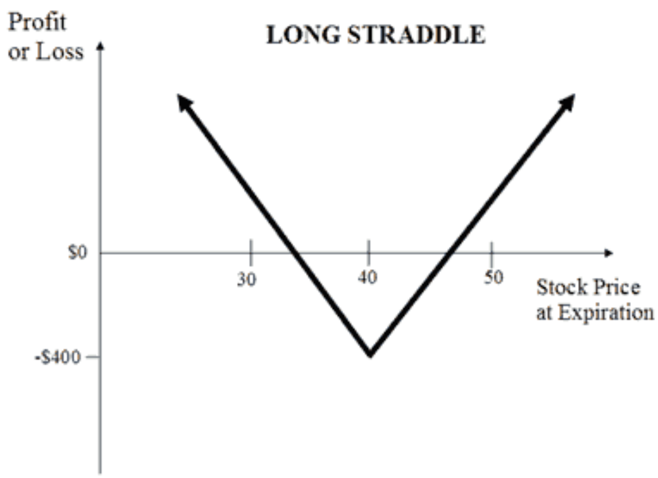
- The number of shares required by the replicating portfolio also changed based upon changes in the underlying stock price. In option's parlance, this is known as an option's "gamma." This measures how the options delta changes with changes in the stock price.

Note that as the stock price increased, the number of shares required to replicate the option increased. This implies that a call option has positive gamma.

If we repeated this whole exercise but used a put option instead, we would similarly find that a put option has positive gamma: the more price depreciates below the strike, the more shares we need to short to replicate.

Which means that these options can be roughly replicated using a very naïve trend-following strategy.

If we buy a put and a call at the same strike and same maturity, we have constructed a trade known as a "straddle". We plot an example straddle payoff profile below.



Source: theoptionsguide.com

Since a long straddle is simply the combination of a put and a call, we can replicate its payoff by just replicating both positions independently and summing up our total exposure.

It should come as no surprise that the replication of this straddle is, in essence, a trend following strategy. As the underlying stock price increases, we buy shares, and as underlying stock price decreases, we sell shares.

With this naïve model, we can already see a few interesting trade-offs:

- The convexity of trend following may have nothing to do with any sort of market "anomaly," but rather is a function of the trading strategy employed.
- Purchasers of a straddle will realize the payoff minus the up-front cost of the options, which will be a function of implied volatility. The replicating trend-following strategy will realize the same payoff minus the trading costs, which will be a path-dependent function of returns (and, therefore, realized volatility).

For a more nuanced dive into deriving this relationship, we recommend the paper "Tail protection for long investors: Trend convexity at work" by Dao, Nguyen, Deremble, Lempérière, Bouchaud, and Potters (2016).

4. Straddle-*ish*

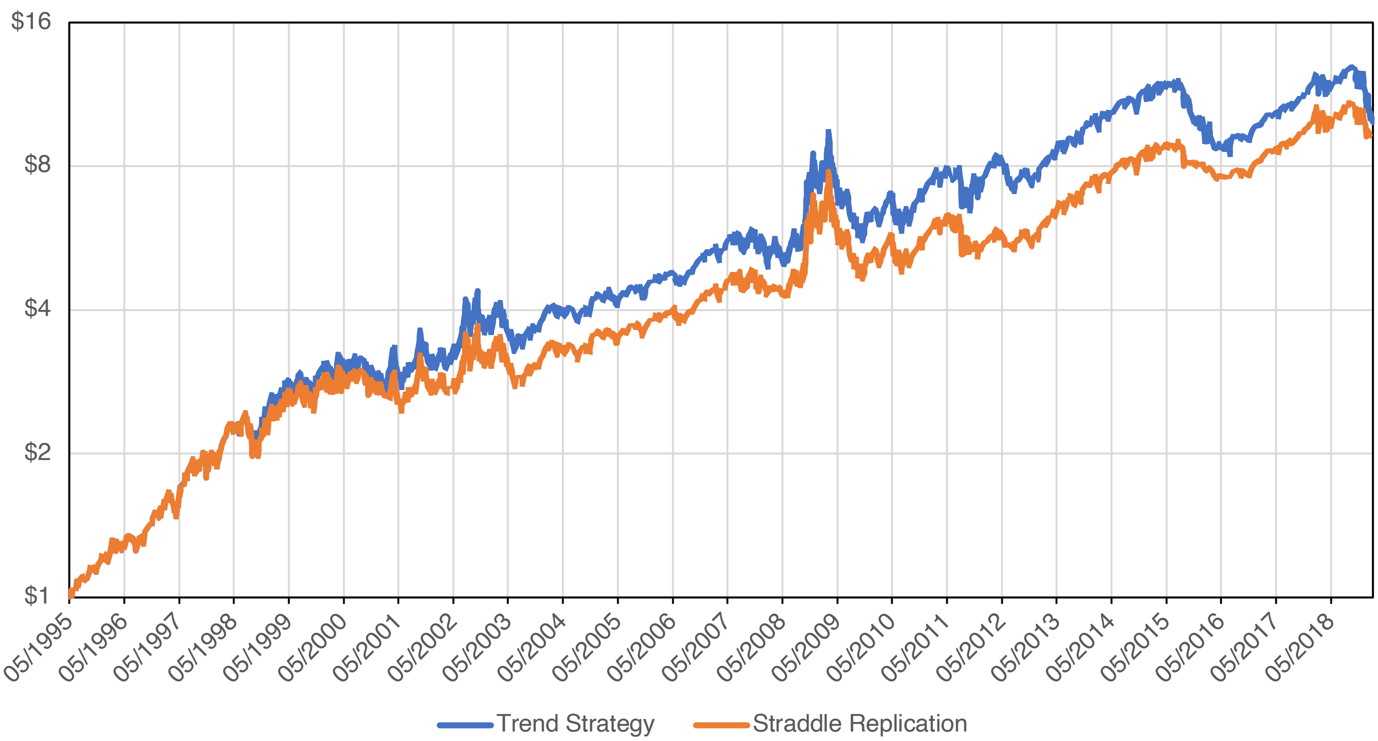
While we have demonstrated that a straddle can replicated with a (continuous) trend-following strategy, it is not the traditional trend-following archetype by any means. Furthermore, trend-following strategies are continuous in nature, while straddles have a defined expiration date.

So, let us consider a more realistic (albeit, still a toy) trend-following implementation. We will go long the S&P 500 when its prior 12-month return is positive and short when it is negative. We will rebalance the strategy at the end of each month. For simplicity, we will assume any available capital is invested in risk-free bonds that return 0%.

How might we translate this strategy into a semi-equivalent straddle replication? One interpretation may be that at the end of each month, we use the price from 12 months ago to set the strike of our straddle. To compute the delta, we will lean on the Black-Scholes equation, where we will assume that the time until expiration is one month, we'll assume no dividend payment, and we'll use prior short-term realized volatility as our input for implied volatility.[8] Since the delta will vary between -1 and 1, we will use it as our allocation to the S&P 500, investing remaining capital in risk-free bonds that return 0%.

It is important to note that in the prior section, the simplified trend following strategy replicated the straddle payoff because we were able to delta hedge over infinitesimally small time horizons at zero cost. Here, we are rebalancing monthly, applying a much more static mode of replication.

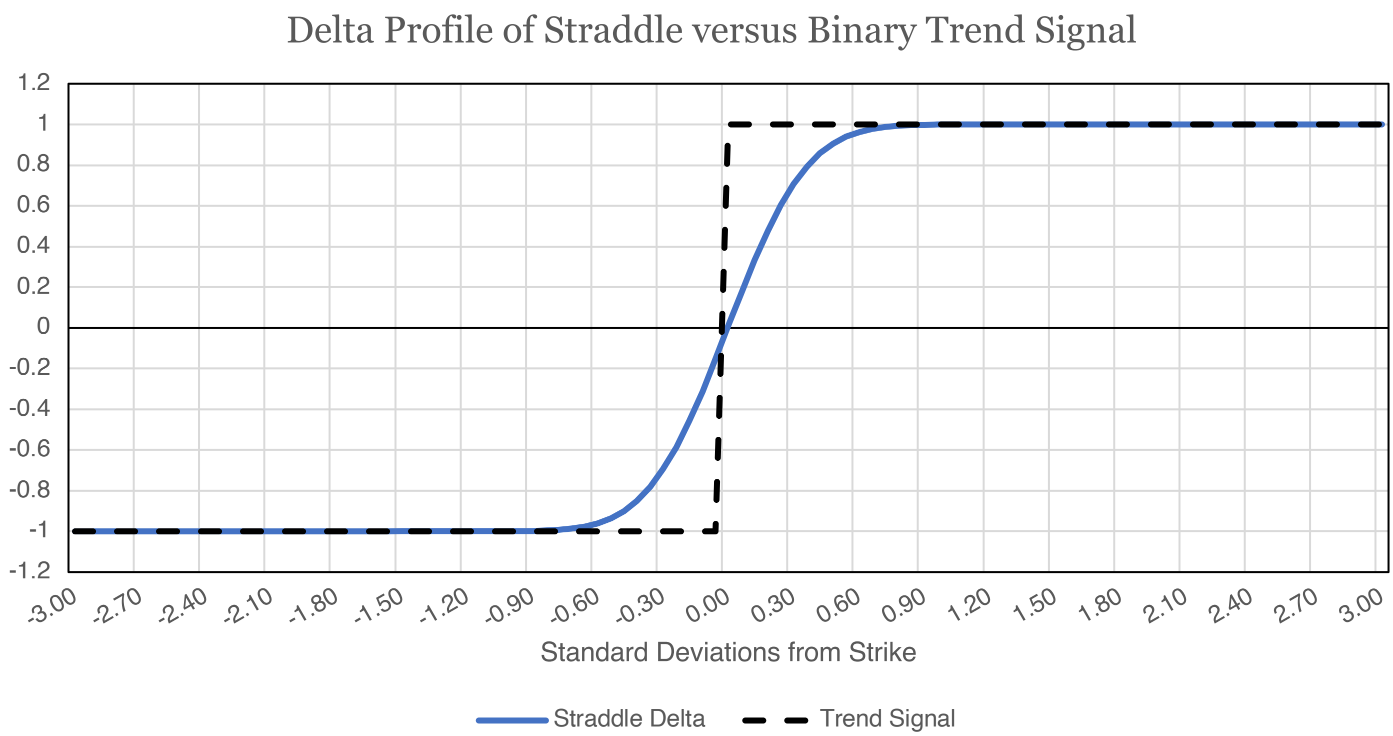
Below we plot the growth of $1 in each strategy. The correlation in monthly log-returns between the two strategies is 95.8%.



[8] This could be made more accurate by incorporating a more accurate estimate of the risk-free rate, dividend yield, and implied volatility. However, for our purposes, a toy approximation remains useful.

Source: CSI Data. Calculations by Newfound Research. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees except for underlying ETF expense ratios. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

We can see that a trend following strategy looks incredibly similar to a strategy that replicates a straddle. Why? Let's look at the trend response function versus delta.

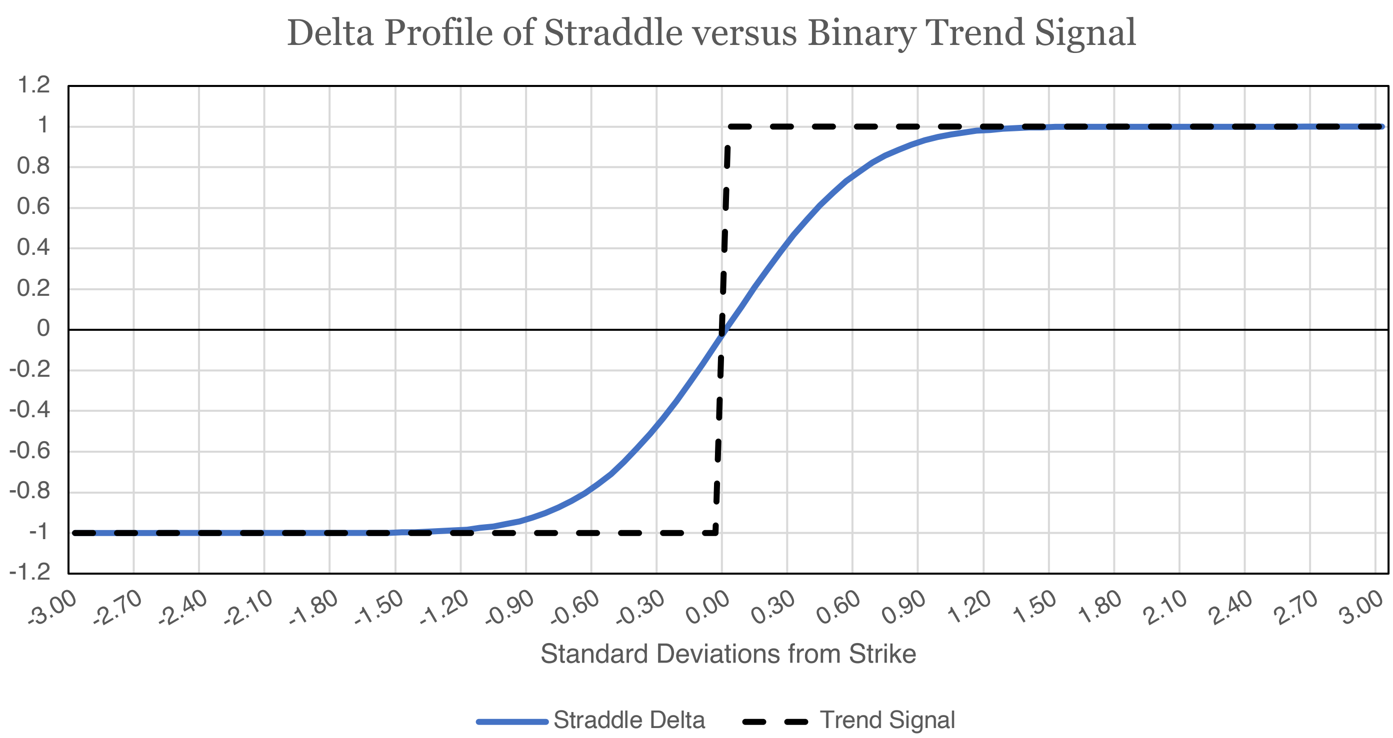


Note that the x-axis is measured in *standard deviations* from the strike. This is important. A 10% move in one asset class may be highly significant while a 10% move in another may not. Furthermore, a 10% move in one market environment may be significant while a 10% move in another may not. The delta function captures this as implied volatility is an input to the measure, while the trend signal does not.

We can see that the binary trend signal and the straddle delta are nearly identical when price is greater than 0.6 standard deviations from the strike in either direction. In this model, this can be interpreted as a strong trend where the trend strategy allocation and delta straddle allocation will coincide.

Within that range, however, we can see that the trend signal over-estimates the delta. Therefore, in cases where price continues away from the strike, the binary signal will out-perform and in cases where price reverts back towards the strike, the binary signal will under-perform.

We can use this insight to explore a few questions. For example, why do trend following strategies have to be traded frequently? Let's consider the case where we only rebalance quarterly. Note what happens to the delta function of the straddle:



We can see that the trend signal and the straddle delta only meet when price is 1.2 standard deviations from strike. The delta is taking into account how long there is until expiration and therefore adjusting itself downward in magnitude in acknowledgement that price might revert back towards the strike. The binary signal does not.

So, is the answer just to rebalance as frequently as possible? After all, as the ratio of the rebalance period to the lookback period goes to zero, the shape of the delta function approaches the binary trend signal. Conversely, as the ratio goes to infinity (i.e. the holding period length far exceeds the lookback), the shape of the delta function approaches y=0 at all points.

The answer in the real world, where prices are not continuous, we are not trading infinitesimally small horizons, and there are trading costs, is "no."

But if we just statically replicate the remaining time until option expiration, doesn't that remove the entire long gamma aspect? Have we not lost the convexity created by the trading strategy?

In an attempt to answer this question, we can ask a slightly different one: "how different is the change in delta from rolling into the new straddle versus replicating the original straddle?" We derive the math in the appendix, but under some general assumptions we can say that the deltas will converge when the ratio of the rebalance time-step and the time until option expiration goes to zero.

$$\frac{dt}{(T-t)} \to 0$$

As this ratio becomes smaller, therefore, the delta change from rolling into the new straddles will approximate the delta change of replicating the prior straddle, and thus conserve the replicating strategy's natural convexity.

There are two important drivers for this limit here: we want *dt* to be as small as possible and *T-t* to be as large as possible.

Unfortunately, this implies that we are delta hedging over an infinitesimally small time period at the beginning of the life of the straddle, a time at which the delta is approximately zero because it is struck at-the-money! On the other hand, if we go towards expiration, *T-t* goes to zero!

The dilemma at hand, then, is for a simple binary long/short trend strategy to reflect the delta of a straddle, it needs to be close to expiration when the delta function looks more like a step change. However, for the changes in delta from rolling the straddle position to reflect changes in delta from replication, the roll must occur near inception!

Using this information, we can attempt to get a rough approximation of how much a binary trend strategy's return comes from: (1) the replication of a straddle, (2) excess delta exposure from rolling straddle exposure, and (3) excess delta exposure from using a binary signal.

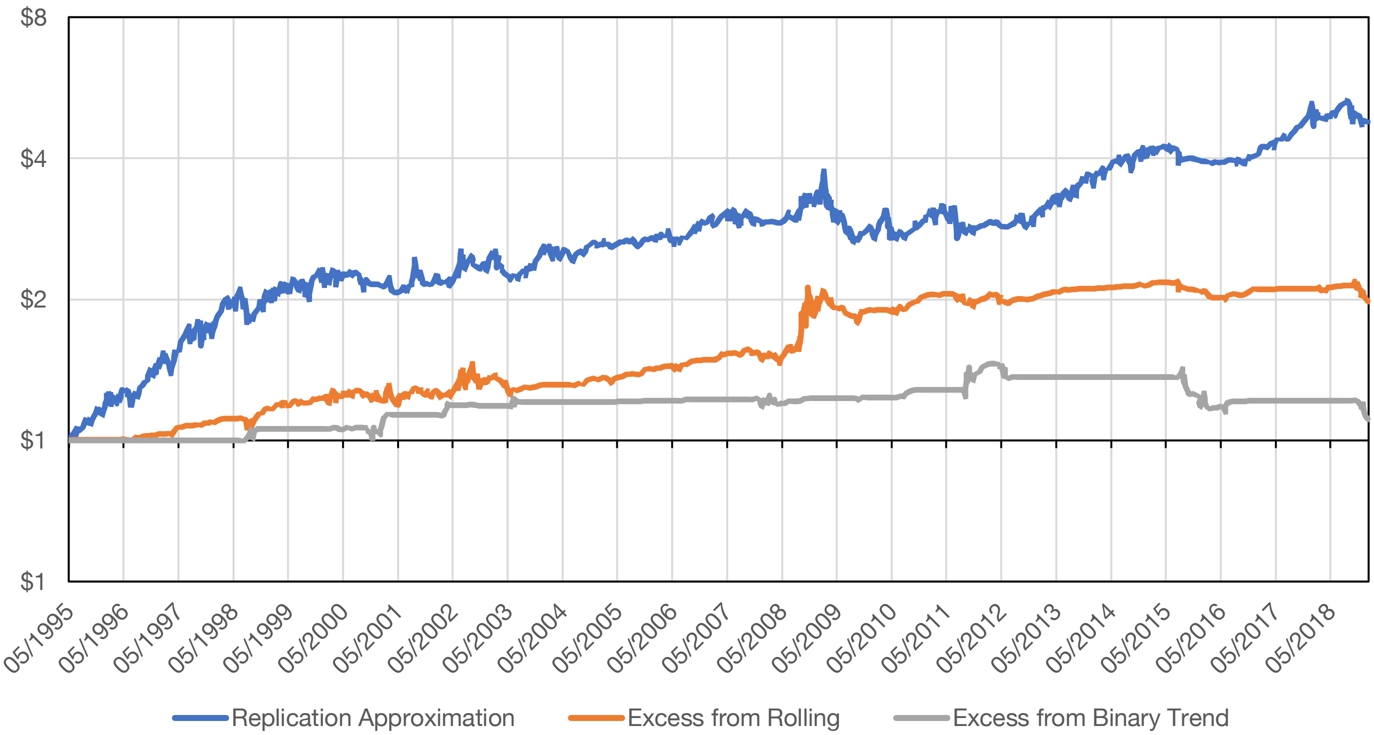
Specifically, rebalancing our portfolios once a month we will:

- Calculate the delta of a 24-month straddle with 12-months left until expiration. This allows the straddle to reflect the same strike price as the rolling and binary trend signals, but with enough life until expiration that *dt/(T-t)* is small enough that the delta may closely reflect the delta from replication.

- Calculate the delta of a 13-month straddle with 1-month until expiration. Calculate the difference in delta exposure from this step to the last and label this the excess delta from taking a rolling approach.

- Finally, calculate the binary trend signal and calculate the difference in exposure versus the rolling approach. This will reflect the excess exposure from our binary approximation.

Below we plot the total return from each of these three series.



Source: CSI Data. Calculations by Newfound Research. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees except for underlying ETF expense ratios. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

While we can see that replication has had a meaningful positive return, we can also see that the rolling model has a non-insignificant positive impact on returns. The trend model, despite being a net contributor through 2012, was more-or-less insignificant over the full period.

However, there is a bit of a head-scratcher here worth discussing: why would we expect the replication of a straddle to have a positive return? Positive *convexity*, sure. But a positive return?

5. Convexity versus Premium

Bruder and Gaussel (2011)[9] suggest that any single-asset trading strategy can be broken down into two component pieces: an option profile and trading impact. A simple constant stop-loss level, for example, can be thought of as a perpetual call

[9] Bruder, Benjamin and Gaussel, Nicolas, Risk-Return Analysis of Dynamic Investment Strategies (June 2011). Available at SSRN: https://ssrn.com/abstract=2465623 or http://dx.doi.org/10.2139/ssrn.2465623

option payoff with trading costs that alter exposure between long and flat, creating frictions due to non-continuous prices and delays in rebalancing.

For trend-following strategies, they suggest a framework whereby the trend is measured as an exponential moving average of daily returns and exposure is proportional to this measured trend, the asset's variance, and the investor's risk tolerance (a suggestion that is theoretically consistent with an optimal Markowitz/Merton strategy).

With these assumptions in hand, they introduce the following proposition:

Proposition 4 *The logarithmic return of a trend-following strategy that follows the exposure function (9) can be broken down into an option profile and some trading impact:*

$$\ln \frac{X_T}{X_0} = m\left(\underbrace{\frac{\tau}{2\sigma^2}\left(\hat{\mu}_T^2 - \hat{\mu}_0^2\right)}_{option\ profile} + \underbrace{\int_0^T \left[\frac{\hat{\mu}_t^2}{\sigma^2}\left(1 - \frac{1}{2}m\right) - \frac{1}{2\tau}\right] \mathrm{dt}}_{trading\ impact}\right) \tag{11}$$

It is worth remembering that the trend $\hat{\mu}_t$ is not a model assumption but the actual measured trend.

Using this proposition, we can decompose long/short trend equity strategies into these two component pieces. Using this model, Jusselin, Malongo, Roncalli, Lezmi, Masselin, and Dao (2017)[10] demonstrate that theoretical model returns can be decomposed as:

$$\tilde{R}_k^V = \tilde{R}_k^G + \tilde{R}_k^g$$

where:

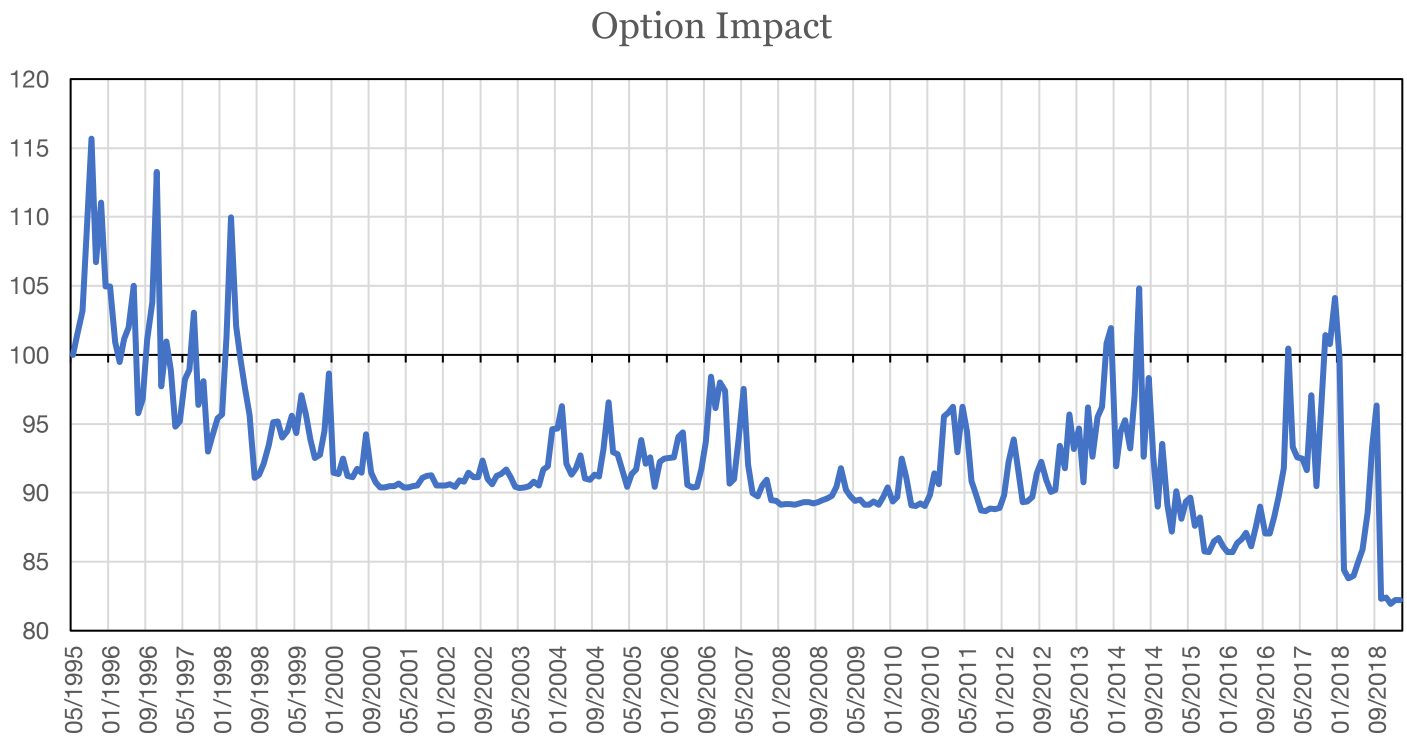
$$\tilde{R}_k^G = \frac{\alpha}{2\lambda}\left(\hat{\mu}_k^2 - \hat{\mu}_{k-1}^2\right)$$

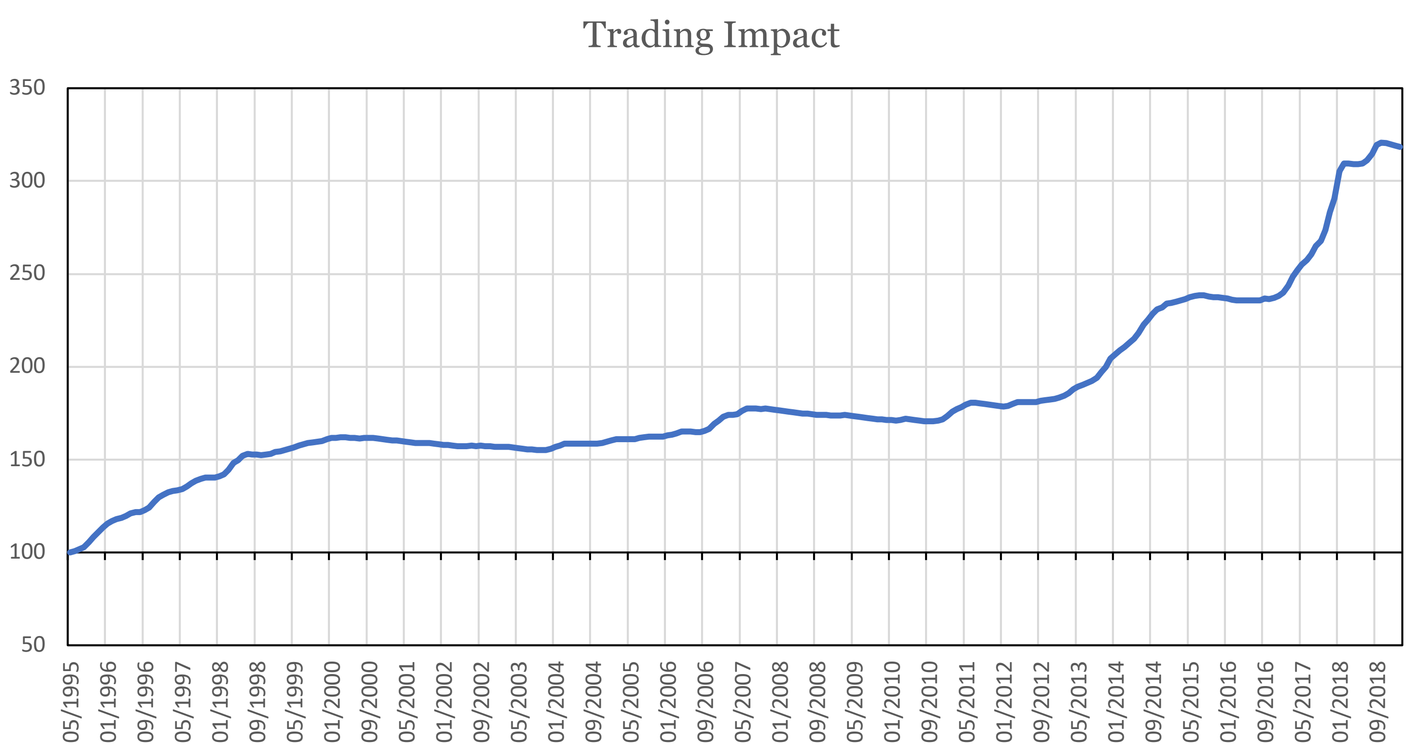
and[32]:

$$\tilde{R}_k^g = \alpha\sigma^2\left(\frac{\hat{\mu}_{k-1}^2}{\sigma^2}\left(1 - \frac{\alpha\sigma^2}{2}\right) - \frac{\lambda}{2}\right)(t_k - t_{k-1})$$

Following this model, we can extract the option payoff (G) and trading impact (g) of a trend following S&P 500 strategy.

[10] Jusselin, Paul and Lezmi, Edmond and Malongo, Hassan and Masselin, Côme and Roncalli, Thierry and Dao, Tung-Lam, Understanding the Momentum Risk Premium: An In-Depth Journey Through Trend-Following Strategies (September 24, 2017). Available at SSRN: https://ssrn.com/abstract=3042173 or http://dx.doi.org/10.2139/ssrn.3042173


Option Impact
120
115
110
105
100
95
90
85
80
05/1995
01/1996
09/1996
05/1997
01/1998
09/1998
05/1999
01/2000
09/2000
05/2001
01/2002
09/2002
05/2003
01/2004
09/2004
05/2005
01/2006
09/2006
05/2007
01/2008
09/2008
05/2009
01/2010
09/2010
05/2011
01/2012
09/2012
05/2013
01/2014
09/2014
05/2015
01/2016
09/2016
05/2017
01/2018
09/2018


Trading Impact
350
300
250
200
150
100
50
05/1995
01/1996
09/1996
05/1997
01/1998
09/1998
05/1999
01/2000
09/2000
05/2001
01/2002
09/2002
05/2003
01/2004
09/2004
05/2005
01/2006
09/2006
05/2007
01/2008
09/2008
05/2009
01/2010
09/2010
05/2011
01/2012
09/2012
05/2013
01/2014
09/2014
05/2015
01/2016
09/2016
05/2017
01/2018
09/2018

What is clear from this analysis is that trend following strategies have two different components driving returns.

The first is the underlying option payoff. In the Bruder and Gaussell (2011) model, the option is similar to a straddle struck on the *trend* of the underlying asset. We can see in the graph above that this component has a short memory and provides much of the convexity often associated with trend-following returns.

The second component is the trading impact. We can see that this component is low frequency and the driving factor is the realized Sharpe ratio. If we squint our eyes a bit at the return formula, it even resembles the gamma gain minus gamma loss of a traditional straddle delta-hedging strategy.[11]

We can see that the long-term driver of returns in trend-following strategies, then, is *not* the convexity, but rather the trading impact. Given that trend following has positive gamma, we would expect the trading impact to be positive for returns that exhibit autocorrelation (i.e. trending returns). Interestingly, Jusselin, Malongo, Roncalli, Lezmi, Masselin, and Dao (2017) demonstrate that autocorrelation may not even be a necessary component for positive returns; rather, for this particular trend following model, the trading impact will have positive profit or loss based upon the underlying asset's Sharpe ratio.

It is worth acknowledging that this analysis is based upon a particular exposure model that is driven by the underlying asset's realized trend, volatility, and the investor's risk tolerance. Thus, the delta-based model and binary signal model explored in prior sections will not match as neatly. Nevertheless, it serves as further evidence for trend following's inherent return convexity with respect to the underlying asset.

6. Conclusion

Trend following is unique among style premia in that it has historically exhibited a convex payoff profile with positive skew. By replicating example trend following strategies using straddle options, we demonstrated how convexity is *inherent* to trend following strategies outside of any historical premium.

While the historical premium is anomalous, the convexity makes sense.

By replicating the payoff of a rolling straddle strategy, we saw potential reasons why frequent rebalancing in trend following strategies is necessary and were able to decompose the return contributions of the replication, the rolling model, and a binary trend following approach under more realistic assumptions.

In the simplified model we found that most of the historical premium associated with trend following comes from the trading impact that is tied to the asset's Sharpe ratio.

[11] The approximation is closer when volatility is constant for the underlying asset; see Jusselin, Malongo, Roncalli, Lezmi, Masselin, and Dao (2017)

A key step in sticking with any trading strategy is an understanding of why it may work in the future. A good historical backtest is nice to see, but there has to be a reason – be it behavioral, economic, or structural – for the backtest to have any reliability.

Using options to isolate the return sources in trend following strategies is a way to separate the impacts of convexity and trading impact while gaining a deeper understanding the types of risk exchanges that come with investing using trend following strategies.

Appendix

For simplicity, we will assume volatility is constant. Using the Black-Scholes equation, the formula for the delta of a straddle is 2N(d1)-1, where N is the CDF of a standard normal distribution. Written more explicitly, at time *t*, the d1 of a straddle struck at time 0 and expiring at time *T* equals (where *t* and *T* are measured in annual units):

$$d1_{0.t} = \frac{\ln\left(\frac{S_t}{S_0}\right) + \frac{1}{2}\sqrt{(T-t)}\sigma^2}{\sqrt{(T-t)}\sigma}$$

When we roll into the next straddle, the new d1 becomes:

$$d1_{1,t+dt} = \frac{\ln\left(\frac{S_{t+dt}}{S_1}\right) + \frac{1}{2}\sqrt{(T-t)}\sigma^2}{\sqrt{(T-t)}\sigma}$$

Therefore, the difference in the delta exposure when we roll is:

$$\left(2N\left(d1_{0,t}\right) - 1\right) - \left(2N\left(d1_{1,t+dt}\right) - 1\right) = 2\left(N\left(d1_{0,t}\right) - N\left(d1_{1,t+dt}\right)\right)$$

We can approximate the change using the first derivative:

$$2\left(N\left(d1_{0,t}\right) - N\left(d1_{1,t+dt}\right)\right) \approx 2\left(N\left(d1_{0,t}\right) + \phi\left(d1_{1,t+dt} - d1_{0,t}\right)\left(d1_{1,t+dt} - d1_{0,t}\right)\right)$$

The difference in d1's is:

$$d1_{1,t+dt} - d1_{0,t} = \frac{\ln\left(\frac{S_{t+dt}}{S_0}\right)}{\sqrt{(T-t)}\sigma}$$

The question we now ask is, "how different is the change in delta from rolling the straddle versus replicating the original straddle?" Using the same logic, the change in delta for replication is:

$$2\left(N\left(d1_{0,t}\right) - N\left(d1_{0,t+dt}\right)\right) \approx 2\left(N\left(d1_{0,t}\right) + \phi\left(d1_{0,t+dt} - d1_{0,t}\right)\left(d1_{0,t+dt} - d1_{0,t}\right)\right)$$

Where,

$$d1_{0,t+dt} - d1_{0,t} = \frac{\ln\left(\frac{S_{t+dt}}{S_0}\right) + \frac{1}{2}\sqrt{(T-(t+dt)}\sigma^2}{\sqrt{(T-(t+dt))}\sigma} - \frac{\ln\left(\frac{S_t}{S_0}\right) + \frac{1}{2}\sqrt{(T-t)}\sigma^2}{\sqrt{(T-t)}\sigma}$$

$$d1_{0,t+dt} - d1_{0,t} = \frac{\ln\left(\frac{S_{t+dt}}{S_0}\right) - \frac{\sqrt{(T-(t+dt))}}{\sqrt{(T-t)}}\ln\left(\frac{S_t}{S_0}\right)}{\sqrt{(T-(t+dt)}\sigma}$$

And thus, as:

$$\frac{\sqrt{(T-(t+dt))}}{\sqrt{(T-t)}} \to 1$$

$$\frac{(T-t)-dt}{(T-t)} \to 1$$

$$\frac{dt}{(T-t)} \to 0$$

The delta change from rolling into the new straddles will approximate the delta change of replicating the prior straddle.

GLITCH

February 19, 2019

SUMMARY

- Trend following's simple, systematic, and transparent approach does not make it any less frustrating to allocate to during periods of rapid market reversals.
- With most trend equity strategies exhibiting whipsaws in 2010, 2011, 2015-2016, and early 2018, it is tempting to ask, "is this something we can fix?"
- We argue that there are three historically-salient features that make trend following attractive: (1) positive skew, (2) convexity, and (3) a positive premium.
- We demonstrate that the convexity exhibited by trend equity strategies is both a function of the strategy itself (i.e. a fast- or slow-paced trend model) as well as the horizon we measure returns over.
- We suggest that it may be more consistent to think of convexity as an element than can provide *crisis beta,* where the nature of the crisis is defined by the speed of the trend following system.
- The failure of a long-term trend strategy to de-allocate in Q4 2018 or meaningfully re-allocate in Q1 2019 is not a glitch; it is encoded in the DNA of the strategy itself.

"There's an old saying in Tennessee – I know it's in Texas, probably in Tennessee – that says, fool me once, shame on – shame on you. Fool me – you can't get fooled again!" - George W. Bush

It feels like we've seen this play before. It happened in 2010. Then again in 2011. More recently in 2015-2016. And who can forget early 2018? To quote Yogi Berra, "It's déjà vu all over again." We're starting to think it is a glitch in the matrix.

Markets begin to deteriorate, losses begin to more rapidly accelerate, and then suddenly everything turns on a dime and market's go on to recover almost all their losses within a few short weeks.

Trend following – like the trend equity mandates we manage here at Newfound – requires trends. If the market completely reverses course and regains almost all of its prior quarter's losses within a few short weeks, it's hard to argue that trend following should be successful. Indeed, it is the prototypical environment that we explicitly warn trend following will do quite poorly in.

That does not mean, however, that changing our approach in these environments would be a warranted course of action. We embrace a systematic approach to explicitly avoid contamination via emotion, particularly during these scenarios. Plus, as we like to say, "risk cannot be destroyed, only transformed." Trying to eliminate the risk of whipsaw not only risks style pollution, but it likely introduces risk in unforeseen scenarios.

So, we have to scratch our heads a bit when clients ask us for an explanation as to our current positioning. After all, trend following is fairly transparent. You can probably pull up a chart, stand a few feet back, squint, and guess with a reasonable degree of accuracy as to how most trend models would be positioned.

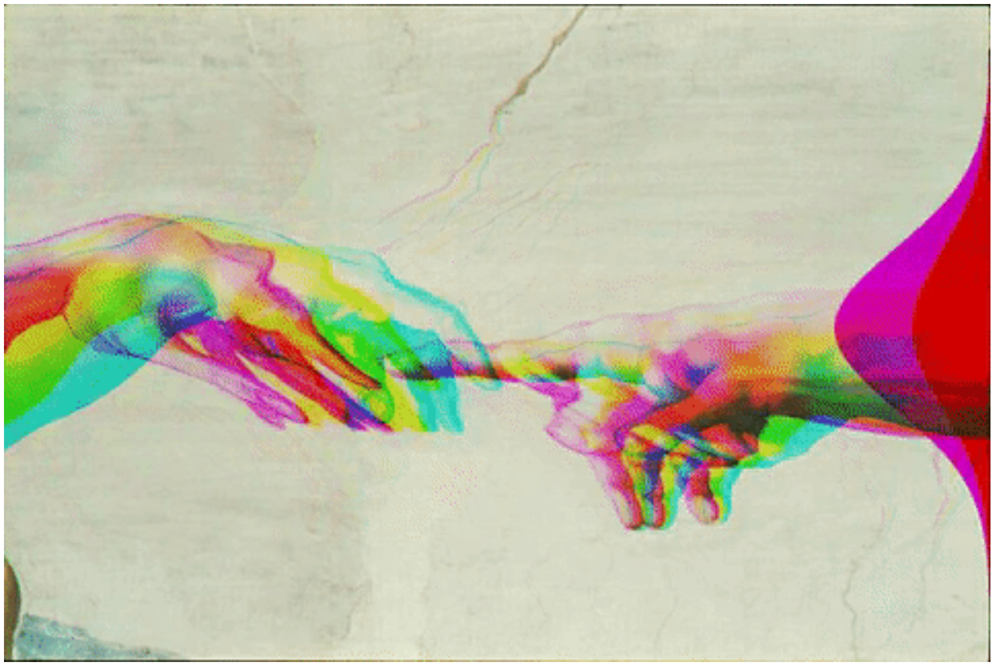
When 12-month, 6-month, and 3-month returns for the S&P 500 were all negative at the end of December, it is a safe guess that we're probably fairly defensively positioned in our domestic trend equity mandates. Despite January's record-breaking returns, not a whole lot changed. 12-, 6-, and 3-month returns were negative, negative, and just slightly positive, respectively, entering February.

To be anything but defensively positioned would be a complete abandonment of trend following.

It is worth acknowledging that this may all just be Act I. Back when this show was screening in 2011 and 2015-2016, markets posted violent reversals – with the percent of stocks above their 50-day moving average climbing from less than 5% to more than 90% – only to roll over again and retest the lows.

Or this will be February 2018 part deux. We won't know until well after the fact. And that can be frustrating depending upon your perspective of markets.

If you take a deterministic view, incorrect positioning implies an error in judgement. *You should have known to abandon trend following and buy the low on December 24th.* If you take a probabilistic view, then it is possible to be correctly positioned for the higher probability event and still be wrong. *The odds were tilted strongly towards continued negative market pressure and a defensive stance was warranted at the time.*



We would argue that there is a third model as well: *sustainability* (or, more morbidly, *survivability*). It does not matter if you have a 99% chance of success while playing Russian Roulette: play long enough and you're eventually going to lose. Permanently. Sustainability argues that the low-probability bet may be the one worth taking if the payoff is sufficient enough or it protects us from ruin.

Thus, for investors for whom *failing fast* is a priority risk, a partially defensive allocation in January and February may be well warranted, even if the intrinsic probabilities have reversed course (which, based on trends, they largely had not).

But sustainability also needs to be a discussion about being able to stick with a strategy. It does not matter if the strategy survives over the long run if the investor does not participate.

That is why we believe transparency and continued education are so critical. If we do not know what we are invested in, we cannot set correct expectations. Without correct expectations, *everything* feels unexpected. And when everything feels unexpected, we have no way to determine if a strategy is behaving correctly or not.

Which brings us back to trend equity strategies in Q4 2018 and January 2019. Did trend equity behave as expected?

Trend following has empirically exhibited three attractive characteristics:

- **Positive Skew:** The return distribution is asymmetric, with a larger right tail than left tail (i.e. greater frequency of larger, positive returns than large, negative returns).

- **Convex Payoff Profile:** As a function of the underlying asset the trend following strategy is applied on, upside potential tends to be greater than downside risk.

- **Positive Premium:** The strategy has a positive expected excess return.

While the first two features can be achieved by other means (e.g. option strategies), the third feature is downright anomalous, as we discussed in our recent commentary *Trend: Convexity & Premium*. Positive skew and convexity create and insurance-like payoff profile and therefore together tend to imply a *negative* premium.

The first two characteristics make trend following a potentially interesting portfolio diversifier. The last element, if it persists, makes it *very* interesting.

Yet while we may talk about these features as historically intrinsic properties of trend following, the nature of the trend-following strategy will significantly impact the horizon over which these features are observed. What is most important to acknowledge here is that skew and convexity are more akin to *beta* than they are *alpha*; they are byproducts of the trading strategy itself. While it can be hard to say things about alpha, we often can say quite a bit more about beta.

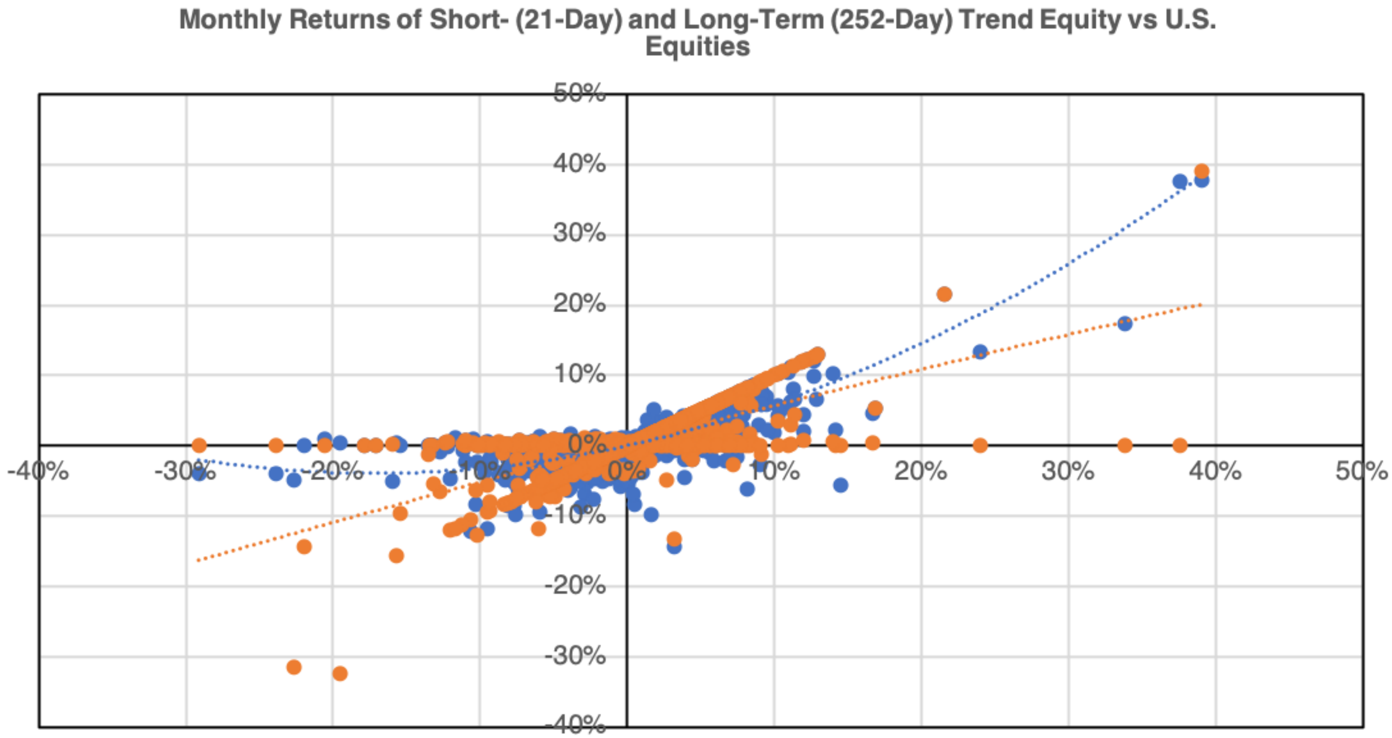
For example, a fast trend following system (typically characterized by a short lookback horizon) would be expected to rapidly adapt to changing market dynamics. This allows the system to quickly position itself for emerging trends, but also potentially makes the strategy more susceptible to losses from short-term reversals.

A slow trend following system (characterized by a long lookback period), on the other hand, would be less likely to change positioning due to short-term market noise, but is also therefore likely to adapt to changing trend dynamics more slowly.

Thus, we might suspect that a fast-paced trend system might be able to exhibit convexity over a shorter measurement period, whereas a slow-paced system will not be able to adapt rapidly. On the other hand, a fast trend following system may have less average exposure to the underlying asset over time and may compound trading losses due to whipsaw more frequently.

To get a better sense of these tradeoffs, we will construct prototype trend equity strategies which will invest either in broad U.S. equities or risk-free bonds. The strategies will be re-evaluated on a daily basis and are assumed to be traded at the close of the day following a signal change. Trend signals will be based upon prior total returns; e.g. a 252-day system will have a positive (negative) signal if prior 252-day total returns in U.S. equity markets are positive (negative).

Below we plot the monthly returns of a -short-term trend equity system (21 day)- and a -long-term trend equity system (252 day)- versus U.S. equity returns.

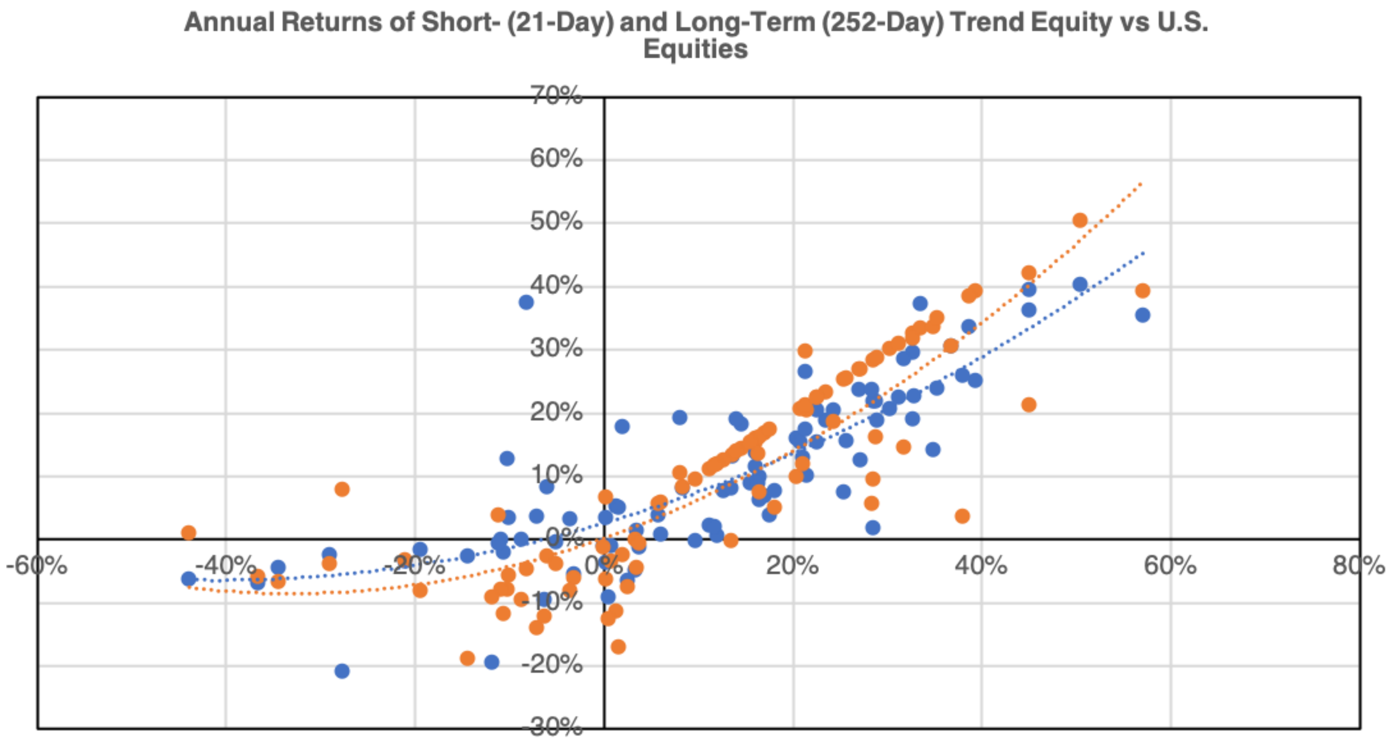


Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. For the avoidance of doubt, neither the Short-Term nor Long-Term Trend Equity strategy reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

We can see that the fast-paced system exhibits convexity over the monthly measurement horizon, while the slower system exhibits a more linear return profile.

As mentioned above, however, the more rapid adaptation in the short-term system might cause more frequent realization of whipsaw due to price reversals and therefore an erosion in long-term convexity. Furthermore, more frequent changes might also reduce long-term participation.

We now plot annual returns versus U.S. equities below.



Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. For the avoidance of doubt, neither the Short-Term nor Long-Term Trend Equity strategy reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

We can see that while the convexity of the short-term system remains intact, the long-term system exhibits greater upside participation.

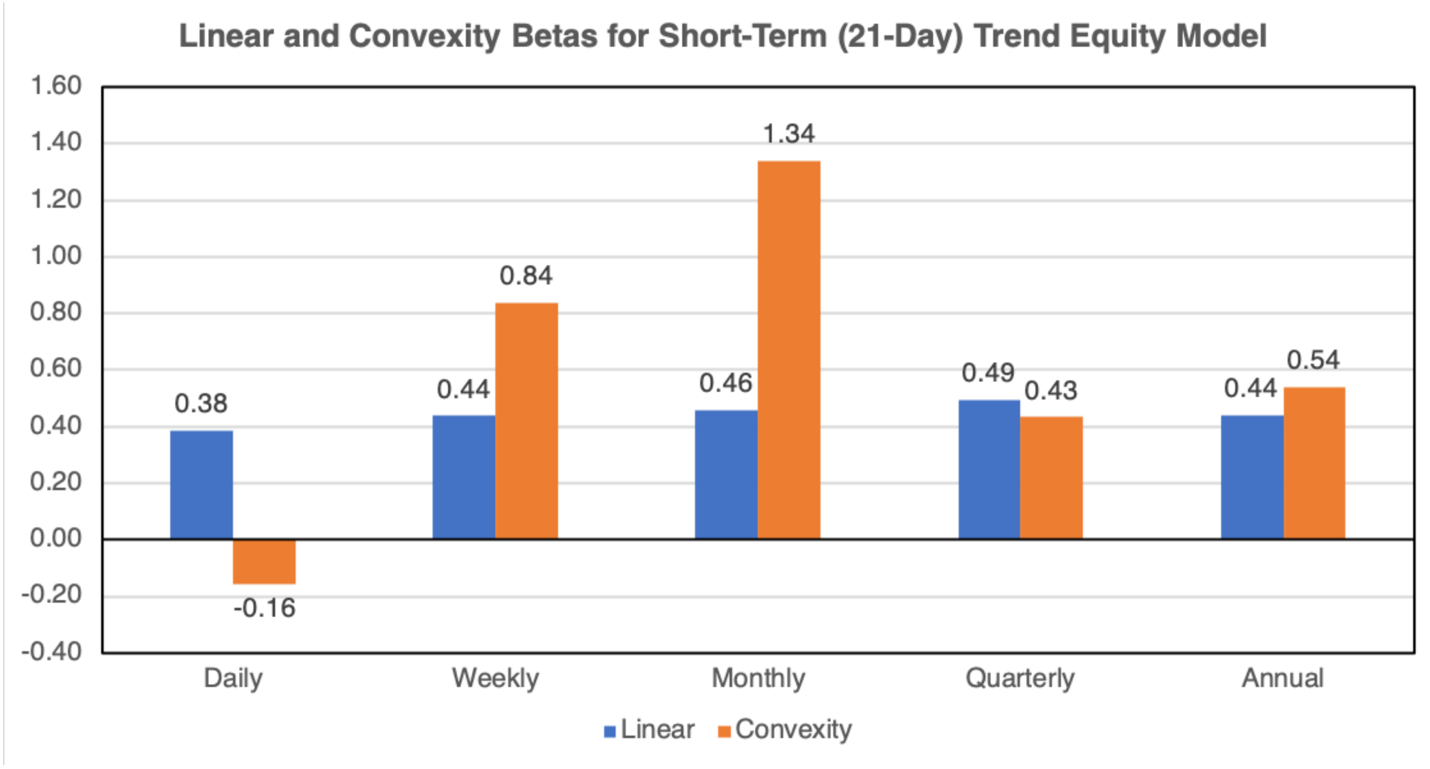
To get a better sense of these trade-offs, we will follow Sepp (2018)[12] and use the following model to deconstruct our prototype long/flat trend equity strategies:

$$r_{strategy} = \propto + \beta_{linear} r_{U.S.\ Equities} + \beta_{convexity} r^2_{U.S.\ Equities}$$

By comparing daily, weekly, monthly, quarterly, and annual returns, we can extract the linear and convexity exposure fast- and slow-paced systems have historically exhibited over a given horizon.

[12] Sepp, Artur. *Trend-following strategies for tail-risk hedging and alpha generation* by (23 April 2018). Available at https://artursepp.com/wp-content/uploads/2018/04/Trend-following-strategies-for-tail-risk-hedging-and-alpha-generation.pdf.

Below we plot the regression coefficients ("betas") for a fast-paced system.

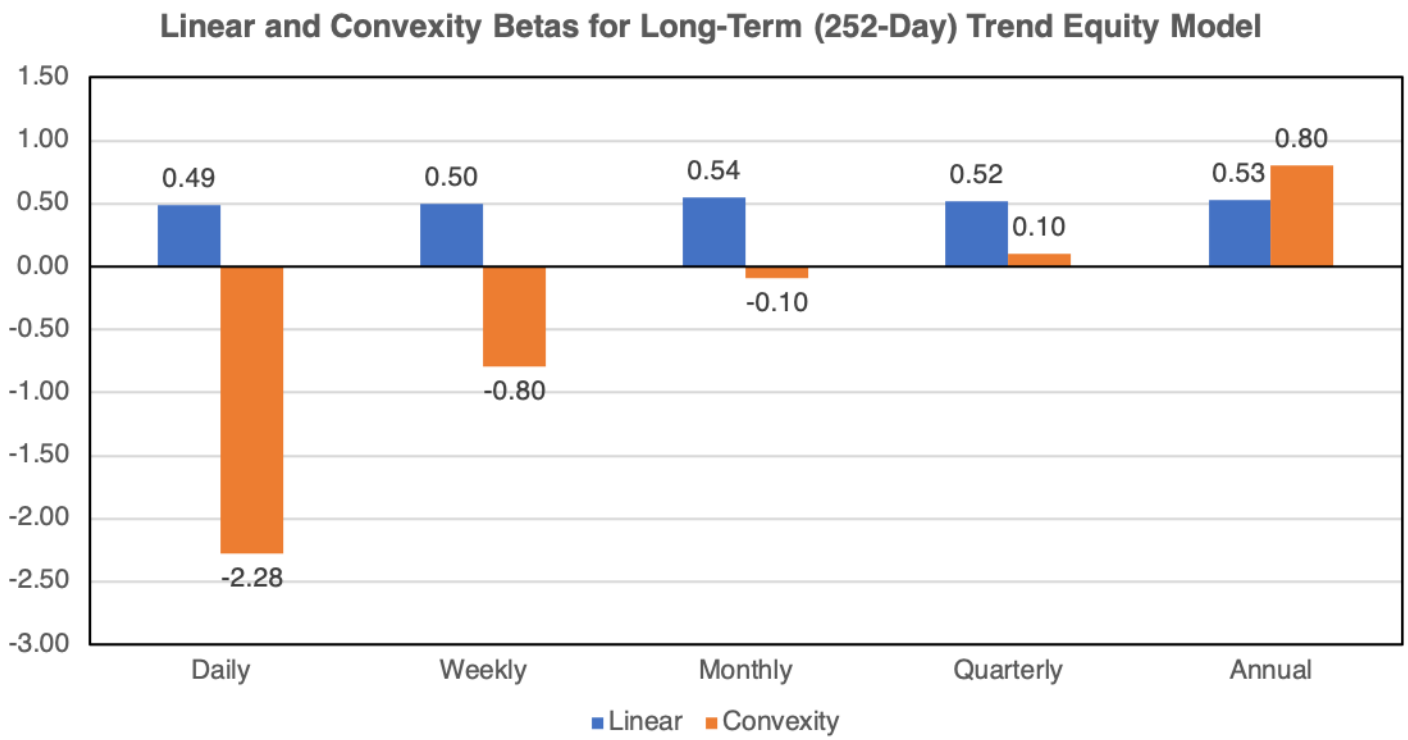


Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. For the avoidance of doubt, the Short-Term Trend Equity strategy does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

We can see that the linear exposure remains fairly constant (and in line with decompositions we've performed in the past which demonstrate that long/flat trend equity can be thought of as a 50/50 stock/cash strategic portfolio plus a long/short overlay[13]). The convexity profile, however, is most significant when measured over weekly or monthly horizons.

Long-term trend following systems, on the other hand, exhibit negative or insignificant convexity profiles over these horizons. Even over a quarterly horizon we see insignificant convexity. It is not until we evaluate returns on an annual horizon that a meaningful convexity profile is established.

[13] See our September 2018 commentary *Decomposing Trend Equity.*



Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. For the avoidance of doubt, the Long-Term Trend Equity strategy does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

These results have very important implications for investors in trend following strategies.

We can see that long-term trend following, for example, is unlikely to be successful as a tail risk hedge for short-term events. Short-term trend following may have a higher probability of success in such a scenario, but only so long as the crisis occurs over a weekly or monthly horizon.[14]

Short-term trend following, however, appears to exhibit *less* convexity with annual returns and has lower linear exposure. This implies less upside capture to the underlying asset.

Neither approach is likely to be particularly successful at hedging against daily crises (e.g. a 1987-type event), as the period is meaningfully shorter than the adaptation speed of either of the strategies.

[14] Sepp (2018) draws a connection between the half-life of the trend signal and the convexity profile, arguing that the evaluation horizon should be longer than the half-life before significant positive convexity is observed.

These results are neither feature nor glitch. They are simply the characteristics we select when we choose either a fast or slow trend-following strategy. While trend-following strategies are often pitched as *crisis alpha*, we believe that skew and convexity components are more akin to *crisis beta.* And this is a good thing. While alpha is often ephemeral and unpredictable, we can more consistently plan around beta.

Thus, when we look back on Q4 2018 and January 2019, we need to acknowledge that we are evaluating results over a monthly / quarterly horizon. This is fine if we are evaluating the results of fast-paced trend-following strategies, but we certainly should not expect any convexity benefits from slower trend models. Quite simply, it all happened too fast.

Conclusion

When markets rapidly reverse course, trend following can be a frustrating style to allocate to. With trend equity styles exhibiting whipsaws in 2010, 2011, 2015-2016, and early 2018, the most recent bout of volatility may have investors rolling their eyes and thinking, "again?"

"Where's the crisis alpha?" investors cry. "Where's the crisis?" managers respond back.

Yet as we demonstrated in our last commentary, two of the three salient features of trend following – namely positive skew and positive convexity – may be byproducts of the trading strategy and not an anomaly. Rather, the historically positive premium that trend following has generated has been the anomaly.

While the potential to harvest alpha is all well and good, we should probably think more in the context of *crisis beta* than *crisis alpha* when setting expectations. And that beta will be largely defined by the speed of the trend following strategy.

But it will also be defined by the period we are measuring the crisis over.

For example, we found that fast-paced trend equity strategies exhibit positive convexity when measured over weekly and monthly time horizons, but that the convexity decays when measured over annual horizons.

Strategies that employ longer-term trend models, on the other hand, fail to exhibit positive convexity over shorter time horizons, but exhibit meaningful convexity over longer-horizons. The failure of long-term trend strategies to meaningfully de-allocate in Q4 2018 or rapidly re-allocate in Q1 2019 is not a glitch: it is encoded into the DNA of the strategy.

Put more simply: if we expect long-term trend models to protect against short-term sell-offs, we should prepare to be disappointed. On the other hand, the rapid adaptation of short-term models comes at a cost, which can materialize as lower up-capture over longer horizons.

Thus, when it comes to these types of models, we have to ask ourselves about the risks we are trying to manage and the trade-offs we are willing to make. After all, "risk cannot be destroyed, only transformed."

HOW MUCH ACCURACY IS ENOUGH?

March 4, 2019

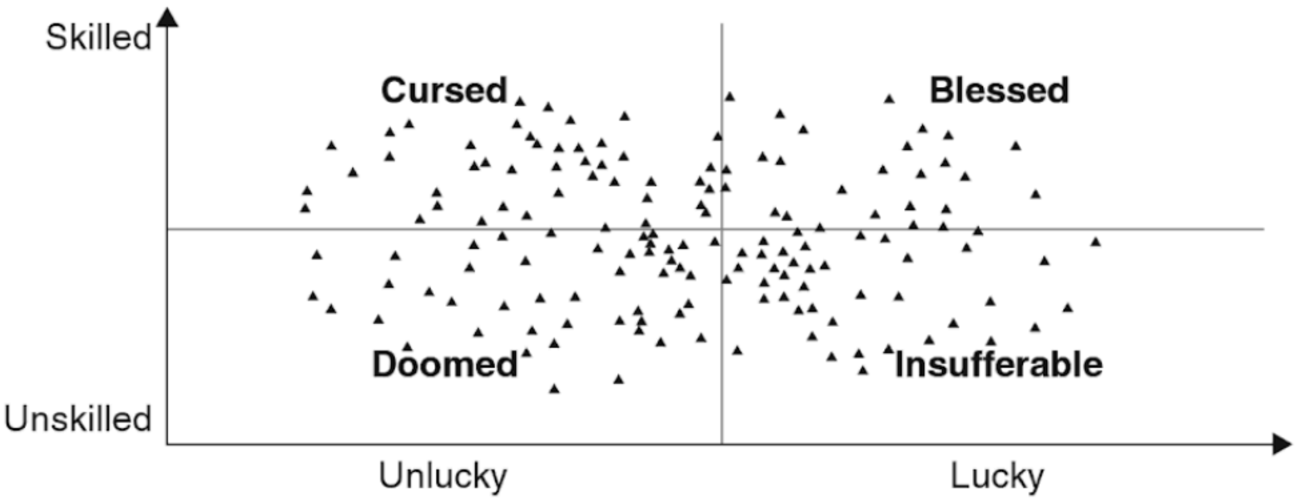
SUMMARY

- It can be difficult to disentangle the difference between luck and skill by examining performance on its own.
- We simulate the returns of investors with different prediction accuracy levels and find that an investor with the skill of a fair coin (i.e. 50%) would likely under-perform a simple buy-and-hold investor, even before costs are considered.
- It is not until an investor exhibits accuracy in excess of 60% that a buy-and-hold investor is meaningfully "beaten" over rolling 5-year evaluation periods.
- In the short-term, however, a strategy with a known accuracy rate can still masquerade as one far more accurate or far less accurate due to luck.
- Further confounding the analysis is the role of skewness of the return distribution. Positively skewed strategies, like trend following, can actually exhibit accuracy rates lower than 50% and still be successful over the long run.
- Relying on perceptions of accuracy alone may lead to highly misguided conclusions.

"The only thing sure about luck is that it will change." – Bret Harte[15]

The distinction between luck and skill in investing can be extremely difficult to measure. Seemingly good or bad strategies can be attributable to either luck or skill, and the truth has important implications for the future prospects of the strategy.

[15] The American short-story writer and poet, not the professional wrestler (spelled without the 'e').



Source: Grinold and Kahn, Active Portfolio Management. (New York: McGraw-Hill, 1999).

Time is one of the surest ways to weed out lucky strategies, but the amount of time needed to make this decision with a high degree of confidence can be longer than we are willing to wait. Or, sometimes, even longer than the data we have.

For example, in order to be 95% confident that a strategy with a 7% historical return and a volatility of 15% has a true expected return that is greater than a 2% risk-free rate, we would need 27 years of data. While this is possible for equity and bond strategies, we would have a long time to wait in order to be confident in a Bitcoin strategy with these specifications.

Even after passing that test, however, that same strategy could easily return less than the risk-free rate over the *next* 5 years *(the probability is 25%)*.

Regardless of the skill, would you continue to hold a strategy that underperformed for that long?

In this commentary, we will use a sample U.S. sector strategy that isolates luck and skill to explore the impacts of varying accuracy and how even increased accuracy may only be an idealized goal.

The (In)Accurate Investor

To investigate the historical impact of luck and skill in the arena of U.S. equity investing, we will consider a strategy that invests in the 30 industries from the Kenneth French Data Library.

Each month, the strategy independently evaluates each sector and either holds it or invests the capital at the risk-free rate. The term “evaluates” is used loosely here; the evaluation can be as simple as flipping a (potentially biased) coin.

The allocation allotted to each sector is 1/30th of the portfolio (3.33%). We are purposely not reallocating capital among the sectors chosen so that the sector calls based on the accuracy straightforwardly determine the performance.

To get an idea for the bounds of how well – or poorly – this strategy would have performed over time, we can consider three investors:

1. **The Plain Investor –** This investor simply holds all 30 sectors, equally weighted, all the time.

2. **The Perfect Investor –** This investor allocates with 100% accuracy. Using a crystal ball to look into the future, if a sector will go up in the subsequent month, this investor will allocate to it. If the sector will go down, this investor will invest the capital in cash.

3. **The Anti-Perfect Investor –** This investor not merely imperfect, they are the complete opposite of the Perfect Investor. They make the wrong calls to invest or not without fail. Their accuracy is 0%. They are so reliably bad that if you could short their strategy, you would be the Perfect Investor.

The Perfect and Anti-Perfect investors set the bounds for what performance is possible within this framework, and the Plain Investor denotes the performance of not making any decisions.

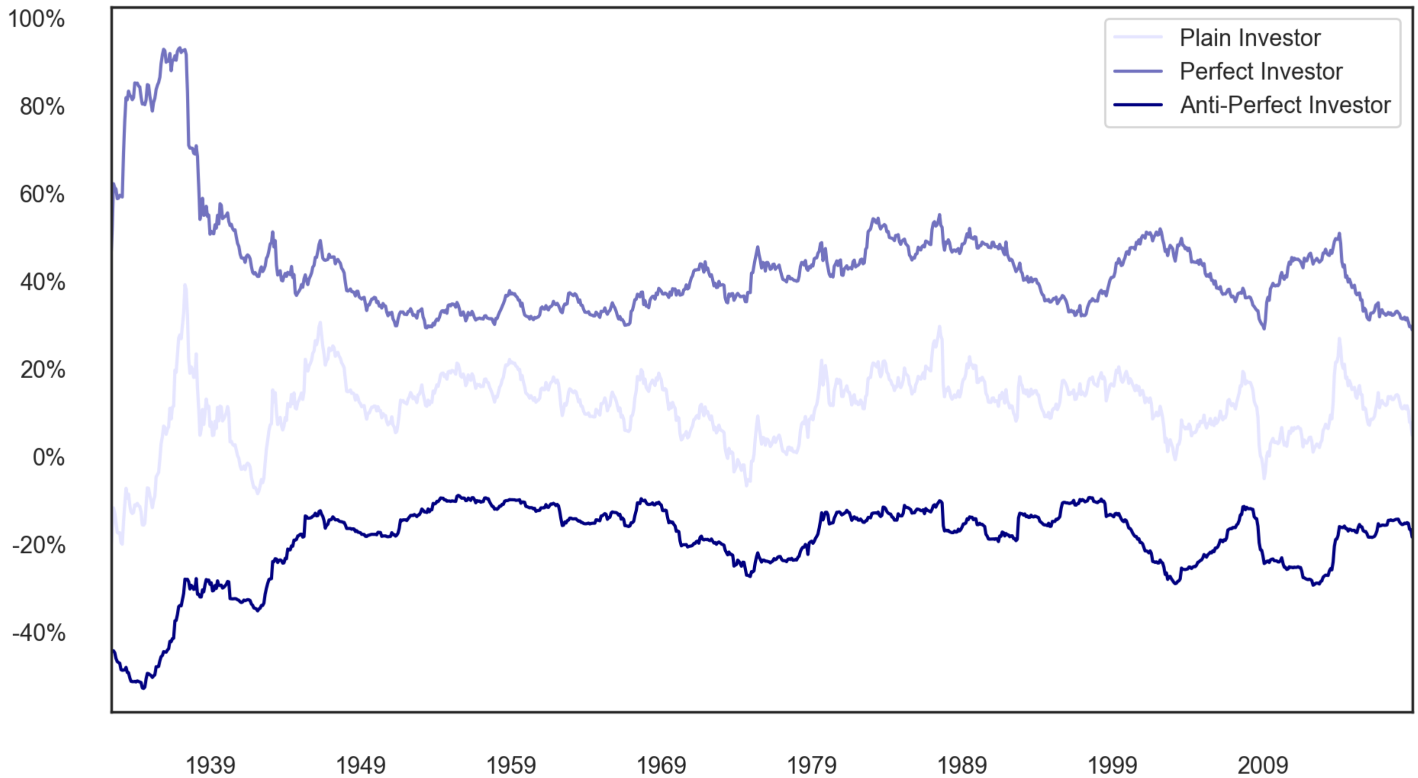
The growth of each boundary strategy over the entire time period is a little outrageous.

	Annualized Return	Annualized Volatility	Maximum Drawdown
Plain Investor	10.5%	19.3%	83.9%
Perfect Investor	42.6%	11.0%	0.0%
Anti-Perfect Investor	-20.0%	12.1%	100.0%

Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

A more informative illustration is the rolling annualized 5-year return for each strategy.

Rolling 5-Year Return of US Sector Strategies



Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

While the spread between the Perfect and Anti-Perfect investors ebbs and flows, its median value Is 59,000 basis points ("bps"). Between the Perfect and Plain investors, there is still 29,000 bps of annualized outperformance to be had. A natural wish is to make calls that harvest some of this spread.

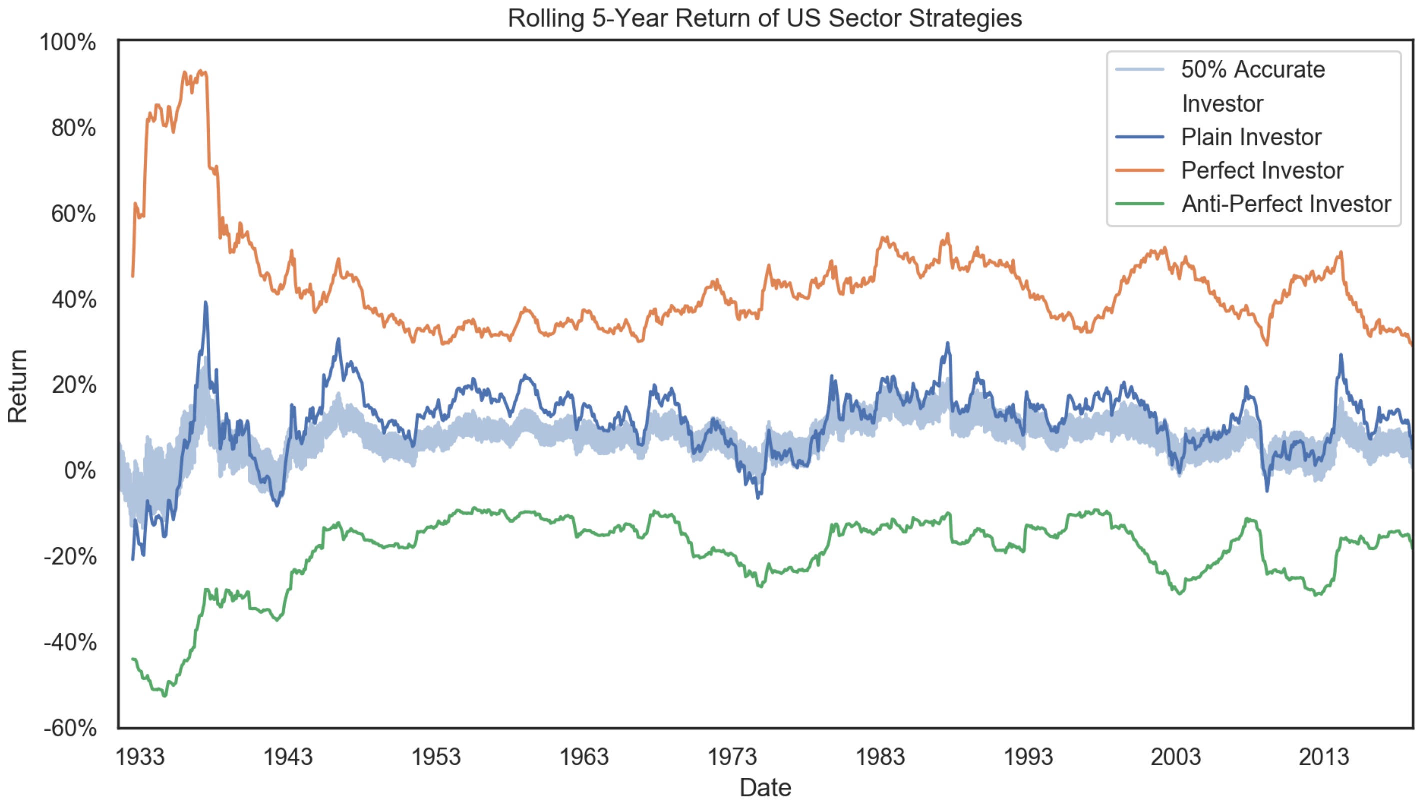
Accounting for Accuracy

Now we will look at a set of investors who are able to evaluate each sector with some known degree of accuracy.

For each accuracy level between 0% and 100% (i.e. our Anti-Perfect and Perfect investors, respectively), we simulate 1,000 trials and look at how the historical results have played out.

A natural starting point is the investor who merely flips a fair coin for each sector. Their accuracy is 50%.

The chart below shows the rolling 5-year performance range of the simulated trials for the 50% Accurate Investor.



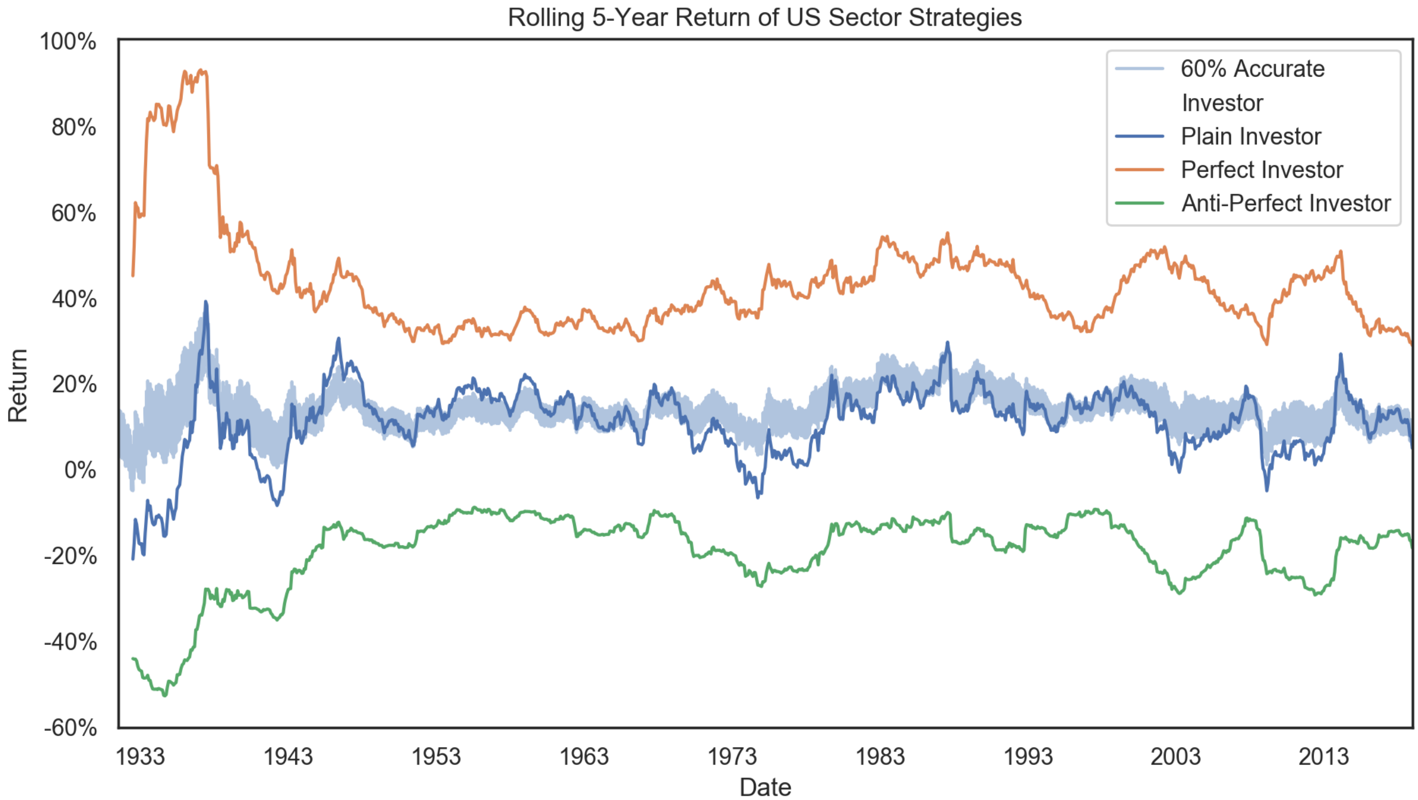
Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

In 59% of the rolling periods, the buy-and-hold Plain Investor beat even the best 50% Accurate Investor. The Plain Investor was only worse than the worst performing coin flip strategy in 6% of rolling periods.

Beating buy-and-hold is hard to do reliably if you rely only on luck.

In this case, having a neutral hit rate with the negative skew of the sector equity returns leads to negative information coefficients. Taking more bets over time and across sectors did not help offset this distributional disadvantage.

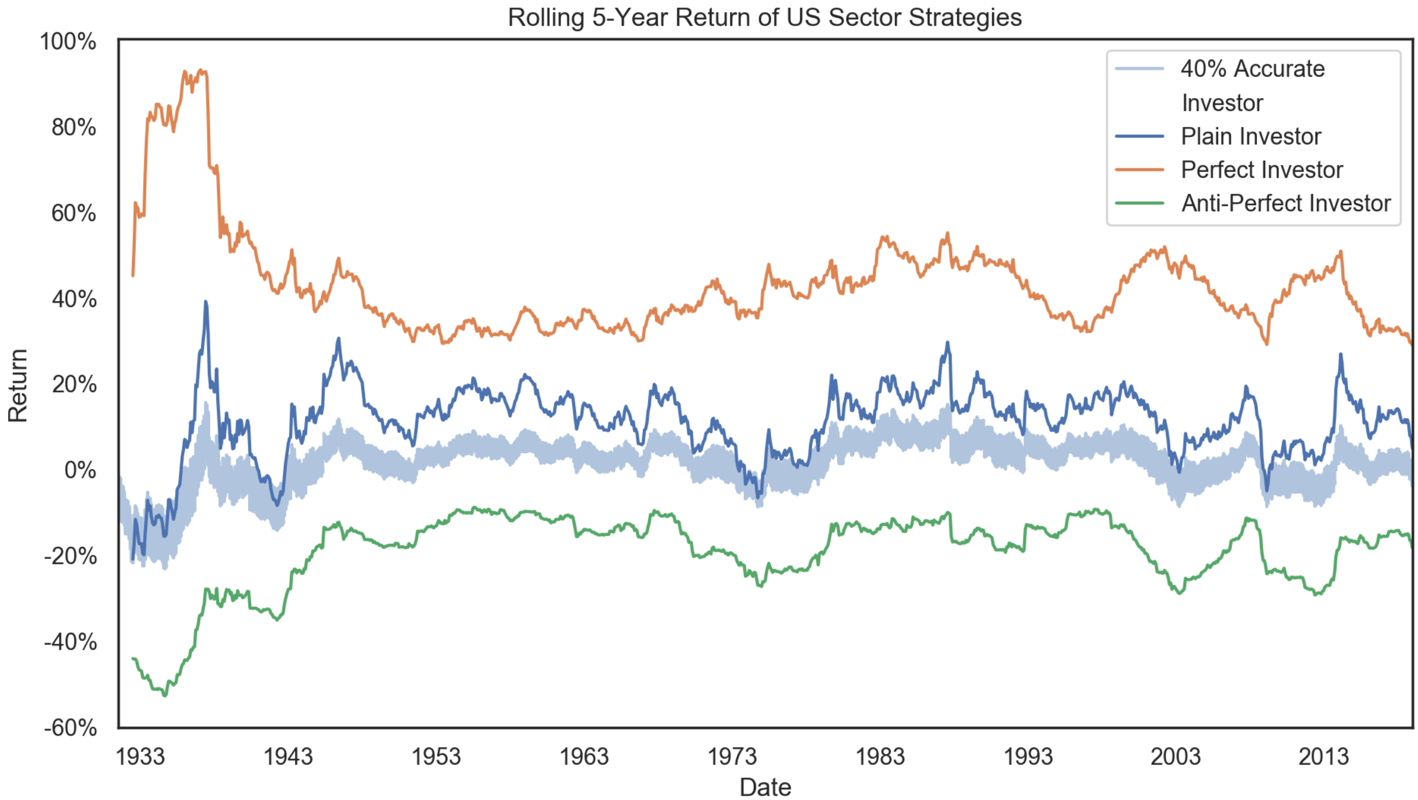
So, let's improve the accuracy slightly to see if the rolling results improve. Even with negative skew (-0.42 median value for the 30 sectors), an improvement in the accuracy to 60% is enough to bring the theoretical information coefficient back into the positive realm.



Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

The worst of these more skilled investors is now beating the Plain Investor in 41% of the rolling periods, and the best is losing to the buy-and-hold investor in 13% of the periods.

Going the other way, to a 40% accurate investor, we find that the best one was beaten by the Plain investor 93% of the time, and the worst one never beats the buy-and-hold investor.



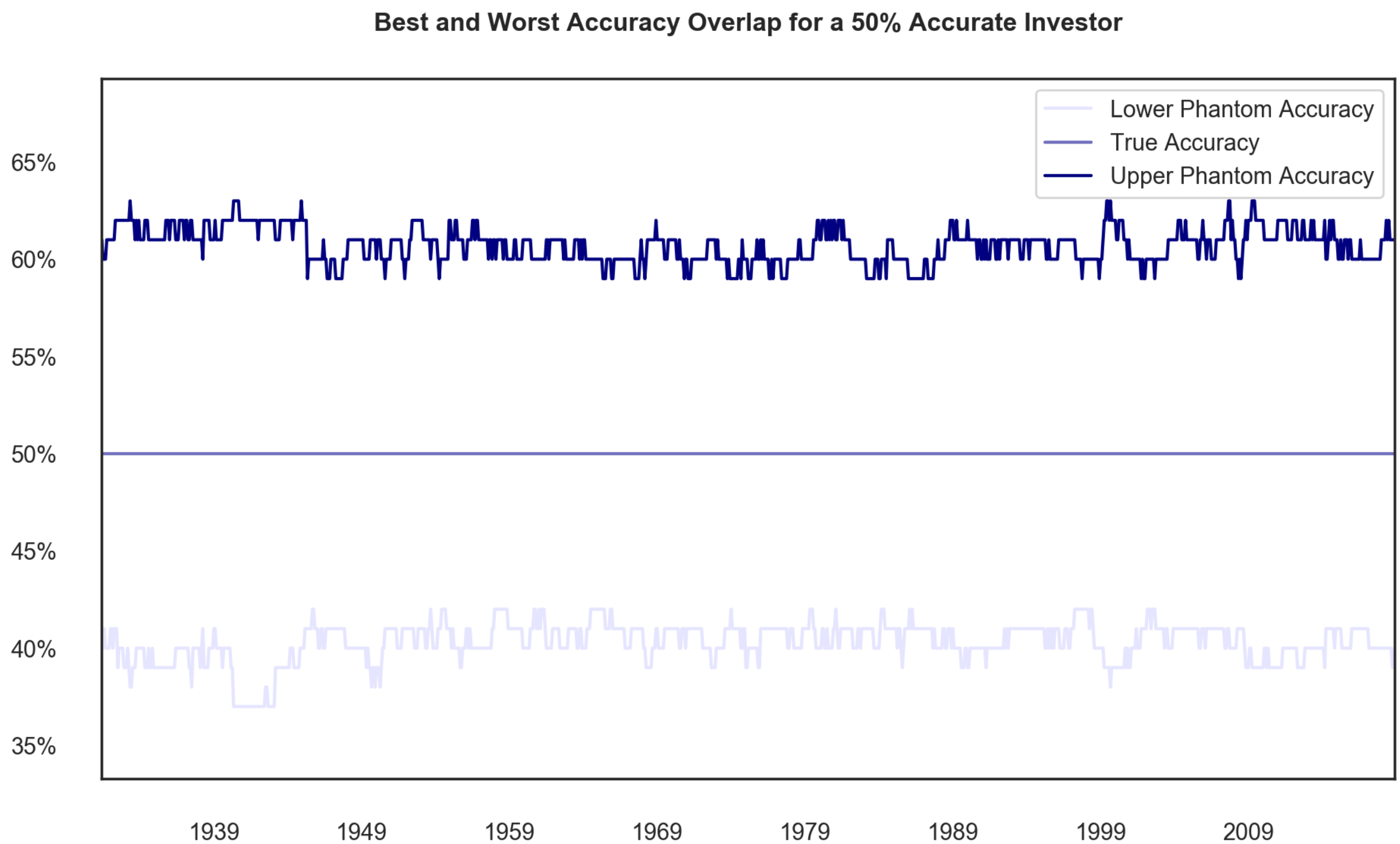
Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

If we only require a modest increase in our accuracy to beat buy-and-hold strategies over shorter time horizons, why isn't diligently focusing on increasing our accuracy an easy approach to success?

In order to increase our accuracy, we must first find a reliable way to do so: a task easier said than done due to the inherent nature of probability. Something having a 60% probability of being right does not preclude it from being wrong for a long time. The Law of Large Numbers can require larger numbers than our portfolios can stand.

Thus, even if we have found a way that will reliably lead to a 60% accuracy, we may not be able to establish confidence in that accuracy rate. This uncertainty in the accuracy can be unnerving. And it can cut both ways.

A strategy with a hit rate of less than 50% can masquerade as a more accurate strategy simply for lack of sufficient data to sniff out the true probability.



Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

You may think you have an edge when you do not. And if you do not have an edge, repeatedly applying it will lead to worse and worse outcomes.[16]

Accuracy Schmaccuracy

Our preference is to rely on systematic bets, which generally fall under the umbrella of factor investing. Even slight improvements to the accuracy can lead to better results when applied over a sufficient breadth of investments. Some of these factors also alter the distribution of returns (i.e. the skew) so that accuracy improvements have a larger impact.

[16] That is, of course, unless you are dealing with a strategy that has a *positive* skew.

Consider two popular measures of trend, used as the signals to determine the allocations in our 30 sector US equity strategy from the previous sections:

- **12-1 Momentum:** We calculate the return over an 11-month period, starting one month ago to account for mean reversionary effects. If this number is positive, we hold the sector; if it is negative, we invest that capital at the risk-free rate.

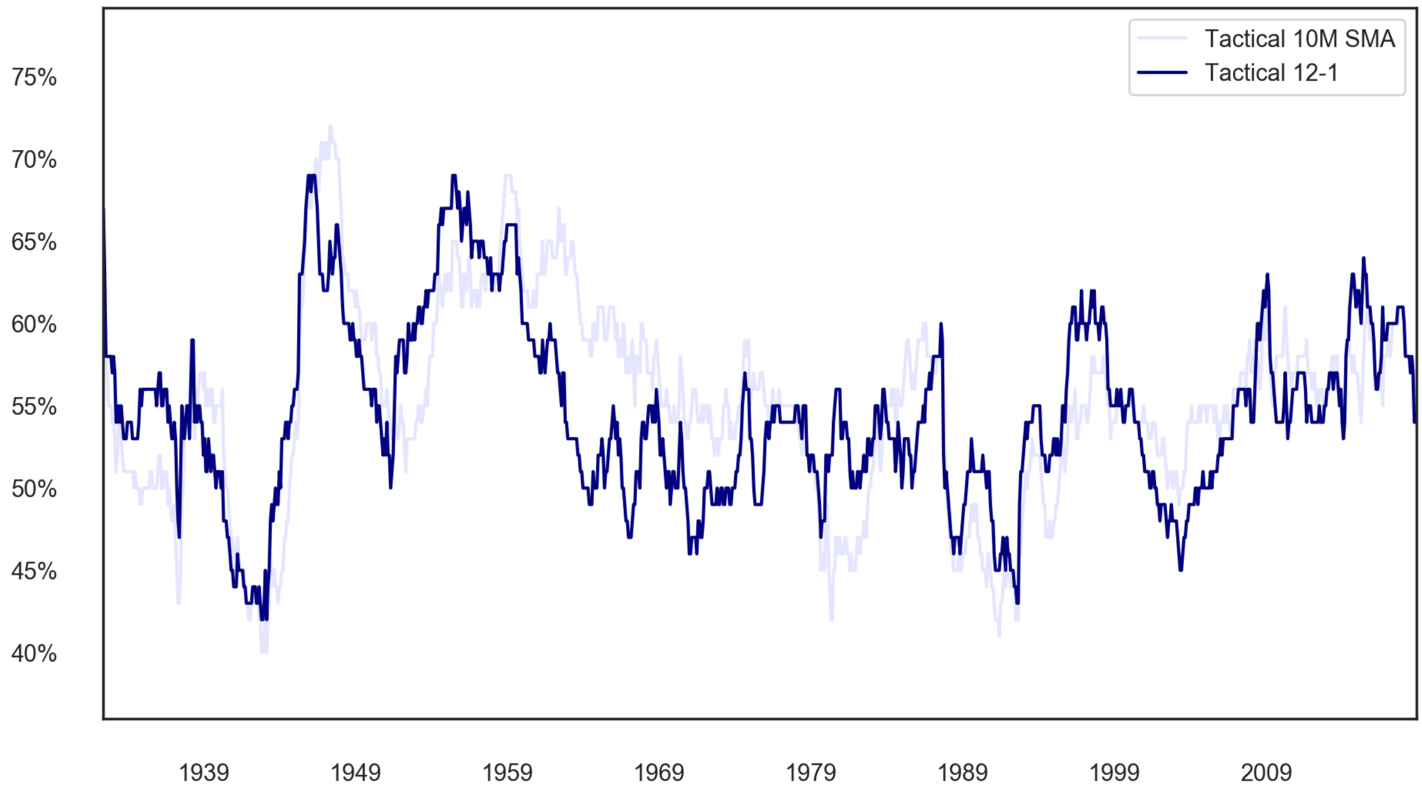
- **10-month Simple Moving Average (SMA):** We average the prices over the prior 10 months and compare that value to the current price. If the current price is greater than or equal to the average, we hold the sector; if it is less than, we invest that capital at the risk-free rate.

These strategies have volatilities in line with the Perfect and Anti-Perfect Investors and returns similar to the Plain Investor.

Using our measure of accuracy as correctly calling the direction of the sector returns over the subsequent month, it might come as a surprise that the accuracies for the 12-1 Momentum and 10-month SMA signals are only 42% and 41%, respectively.

Even with this low accuracy, the following chart shows that over the entire time period, the returns of these strategies more closely resemble those of the 55% Accurate Investor and have even looked like those of the 70% Accurate Investor over some time periods. What gives?

Nearest Average Accuracy for Trend-Following Signals



Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

This is an example of how addressing the negative skew in the underlying asset returns can offset a sacrifice in accuracy. These trend following strategies may have overall accuracy of less than 50%, but they have been historically right when it counts.

Consistently removing large negative returns – at the expense of giving up some large positive returns – is enough to generate a return profile that looks much like a strategy that picks sectors with above average accuracy.

Whether investors can stick with a strategy that exhibits below 50% accuracy, however, is another question entirely.

Conclusion

While most investors expect the proof to be in the eating of the pudding, in this commentary we demonstrate how luck can have a meaningful impact in the determination of whether skill exists. While skill should eventually differentiate itself from luck, the horizon over which it will do so may be far, far longer than most investors suspect.

To explore this idea, we construct portfolios comprised of all thirty industry groups. We then simulate the results of investors with known accuracy rates, comparing their outcomes to 100% Accuracy, 100% Inaccurate, and Buy-and-Hold benchmarks.

Perhaps somewhat counter-intuitively, we find that an investor exhibiting 50% accuracy would have fairly reliably underperformed a Buy-and-Hold Investor. This seems somewhat counter-intuitive until we acknowledge that equity returns have historically exhibit negative skew, with the left tail of their return distribution ("losses") being longer and fatter than the right ("gains"). Combining a neutral hit rate with negative skew creates negative information coefficients.

To offset this negative skew, we require increased accuracy. Unfortunately, even in the case where an investor exhibits 60% accuracy, there are a significant number of 5-year periods where it might masquerade as a strategy with a much higher or lower hit-rate, inviting false conclusions.

This is all made somewhat more confusing when we consider that a strategy can have an accuracy rate *below* 50% and still be successful. Trend following strategies are a perfect example of this phenomenon. The positive skew that has been historically exhibited by these strategies means that frequently inaccurate trades of small magnitude are offset by infrequent, by very large accurate trades.

Yet if we measure success by short-term accuracy rates, we will almost certainly dismiss this type of strategy as one with no skill.

When taken together, this evidence suggests that not only might it be difficult for investors to meaningfully determine the difference between skill and luck over seemingly meaningful time horizons (e.g. 5 years), but also that short-term perceptions of accuracy can be woefully misleading for long-term success. Highly accurate strategies can still lead to catastrophe if there is significant negative skew lurking in the shadows (e.g. an ETF like XIV), while inaccurate strategies can be successful with enough positive skew (e.g. trend following).

The Monsters of Investing: Fast and Slow Failure

March 11, 2019

Summary

- Successful investing requires that investors navigate around a large number of risks throughout their lifecycle. We believe that the two most daunting risks investors face are the risk of failing fast and the risk of failing slow.

- Slow failure occurs when an investor does not grow their investment capital sufficiently over time to meet future real liabilities. This often occurs because they fail to save enough or because they invest too conservatively.

- Fast failure occurs when an investor – often those who are living off of portfolio withdrawals and for whom time is no longer an ally – suffers a significant drawdown that permanently impairs their portfolio.

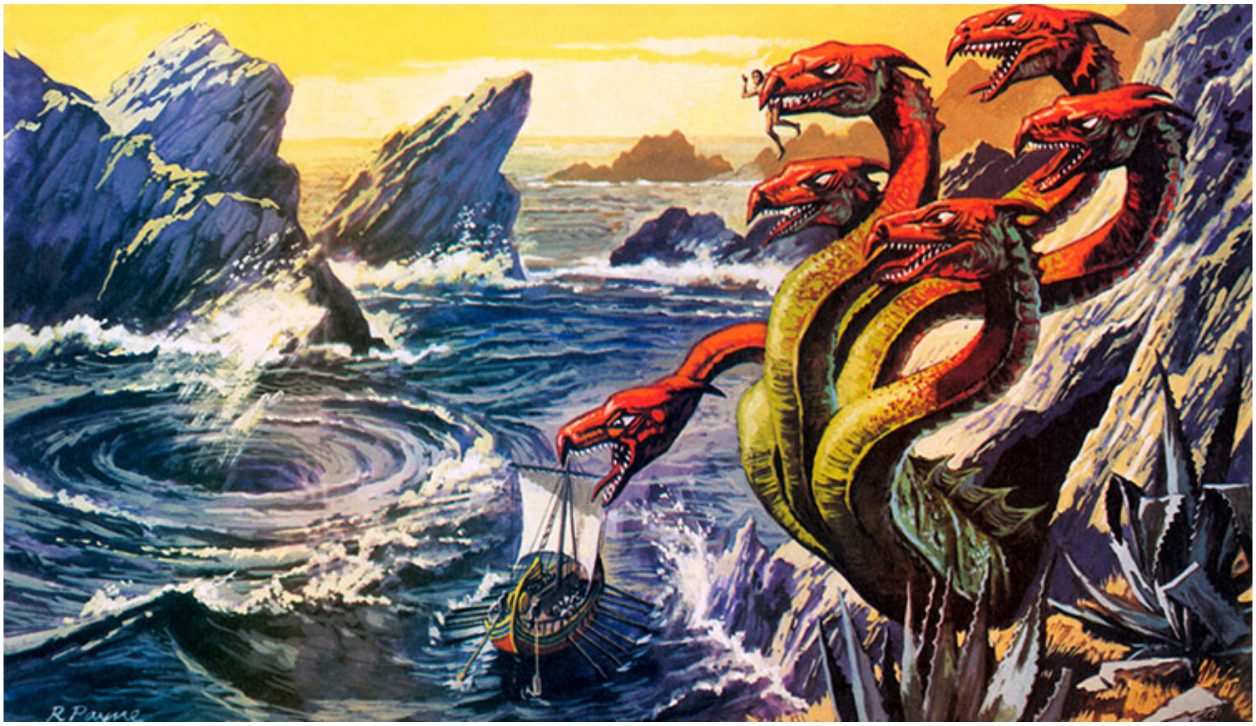
- We believe that sensitivity to these risks should dictate an investor's allocation profile. Investors sensitive to slow failure should invest more aggressively and bear more risk in certain bad states of the world for the potential to earn excess returns in good states. On the other hand, investors sensitive to fast failure should invest more conservatively, sacrificing returns in order to avoid catastrophe.

- We believe this framework can also be used to inform how investors can fund an allocation from their strategic policy to trend equity strategies.

Homer's Odyssey follows the epic ten-year journey of Odysseus and his men as they try to make their way home after the fall of Troy. Along the way, the soldiers faced a seemingly endless string of challenges, including a cyclops who ate them alive, a sorceress who turned them into pigs, and sirens that would have lured them to their deaths with a song had they not plugged their ears with beeswax.

In one trial, the men had to navigate the Strait of Messina between the sea monsters Scylla and Charybdis. With her six serpentine heads, each with a triple row of sharp teeth, Scylla haunted the cliffs that lined one edge of the strait. Ships that came too close would immediately lose six sailors to the ravenous monster. Living under a rock on the other side of the strait was Charybdis. A few times a day, this monster would swallow up large amounts of water and belch it out, creating whirlpools that could sink an entire ship.

December 30th, 2019



The strait was so narrow that the monsters lived within an arrow's range of one another. To safely avoid one creature meant almost necessarily venturing too close to the other. On the one hand was almost certain, but limited, loss; on the other, the low probability of complete catastrophe.

Investors, similarly, must navigate between two risks: what we have called in the past the risks of failing slow and failing fast.

Slow failure results from taking too little risk, often from investors allocating too conservatively or holding excessive cash. In doing so, they fail to grow their capital at a sufficient rate to meet future real liabilities. Failure in this arena does not show up as a large portfolio drawdown: it creeps into the portfolio over time through opportunity cost or the slow erosion of purchasing power.

Fast failure results from the opposite scenario: taking too much risk. By allocating too aggressively (either to highly skewed or highly volatile investments), investors might incur material losses in their portfolios at a time when they cannot afford to do so.

We would argue that much of portfolio design is centered around figuring out which risk an investor is most sensitive to at a given point in their lifecycle and adjusting the portfolio accordingly.

Younger investors, for example, often have significant human capital (i.e. future earning potential) but very little investment capital. Sudden and large losses in their portfolios, therefore, are often immaterial in the long run, as both time and savings are on their side. Investing too conservatively at this stage in life can rely too heavily on savings and fail to exploit the compounding potential of time.

Therefore, younger, growth-oriented investors should be willing to bear the risk of failing fast to avoid the risk of failing slow. In fact, we would argue that it is the willingness to bear the risk of failing fast that allows these investors to potentially earn a premium in the first place. No pain, no premium.

Over time, investors turn their human capital into investment capital through savings and investment. At retirement, investors believe that their future liabilities are sufficiently funded, and so give-up gainful employment to live off of their savings and investments. In other words, the sensitivity to slow failure has significantly declined.

However, with less time for the potential benefits of compounding and no plan on replenishing investments through further savings, the sensitivity to the risk of fast failure is dramatically heightened, especially in the years just prior to and just after retirement. This is further complicated by the fact that withdrawals from the portfolio can heighten the impact of sustained and large drawdowns.

Thus, older investors tend shift from riskier stocks to safer bonds, offloading their fast failure risk to those willing to bear it. Yet we should be hesitant to de-risk entirely; we must also acknowledge longevity risk. Too conservative a profile may also lead to disaster if an investor outlives their nest-egg.

As we balance the scales of failing fast and slow, we can see why trying to invest a perpetual endowment is so difficult. Consistent withdrawals invite the risk of failing fast while the perpetual nature invites the risk of failing slow. A narrow strait to navigate between Scylla and Charybdis, indeed!

We would be remiss if we did not acknowledge that short-term, high quality bonds are not a panacea for fail fast risk. Inflation complicates the calculus and unexpected bouts of inflation (e.g. the U.S. in the 1970s) or hyper-inflation (e.g. Brazil in the 1980s, Peru from 1988-1991, or present-day Venezuela) can cause significant, if not catastrophic, declines in real purchasing power if enough investment risk is not borne.

Purchasing seemingly more volatile assets may actually be a hedge here. For example, real estate, when marked-to-market, may exhibit significant relative swings in value over time. However, as housing frequently represents one the largest real liabilities an investor faces, purchase of a primary residence can lock in the real cost of the asset and provide significant physical utility. Investors can further reduce inflation risk by financing the purchase with a modest amount of debt, a liability which will decline in real value with unexpected positive inflation shocks.

The aforementioned nuances notwithstanding, this broad line of thinking invites some interesting guidance regarding portfolio construction.

Investors sensitive to fast failure should seek to immunize their real future liabilities (e.g. via insurance, real asset purchases, cash-flow matching, structured products, et cetera). As they survey the infinite potential of future market states, they should be willing to give up returns in all states to avoid significant failure in any given one of them.

Investors sensitive to slow failure should seek to bear a diversified set of risk premia (e.g. equity risk premium, bond risk premium, credit premium, value, momentum, carry, et cetera) that allows their portfolios to grow sufficiently to meet future real liabilities. These investors, then, are willing to pursue higher returns in the vast majority of future market states, even if it means increased losses in a few states.

I personally imagine this as if the investor sensitive to failing slow has piled up all their risk – like a big mound of dough – in the bad outcome states of the world. For their willingness to bear this risk, they earn more return in the good outcome states. The investor sensitive to failing fast, on the other hand, smears that mound of risk across all the potential outcomes. In their unwillingness to bear risk in a particular state, they reduce return potential across all states, but also avoid the risk of catastrophe.

Source: BuzzFeed

Quantitatively, we saw exactly this trade-off play out in our piece *The New Glide Path*, where we attempted to identify the appropriate asset allocation for investors in retirement based upon their wealth level. We found that:

- Investors who were dramatically under-funded – i.e. those at risk of failing slow – relative to real liabilities were allocated heavily to equities.

- Investors who were near a safe funding level – i.e. those at risk of failing fast – were tilted dramatically towards assets like Treasury bonds in order to immunize their portfolio against fast failure.

- The fortunate few investors who were dramatically over-funded could, pretty much, allocate however they pleased.

We believe this same failing slow and failing fast framework can also inform how trend equity strategies – like those we manage here at Newfound Research – can be implemented by allocators.

In our recent commentary *Three Applications of Trend Equity* we explored three implementation ideas for trend equity strategies: (1) as a defensive equity sleeve; (2) as a tactical pivot; or (3) as an alternative. While these are the most common approaches we see to implementing trend equity, we would argue that a more philosophically consistent route might be one that incorporates the notions of failing fast and failing slow.

In *Risk Ignition with Trend Following* we examined the realized efficient frontier of U.S. stocks and bonds from 1962-2017 and found that an investor who wanted to hold a portfolio targeting an annualized volatility of 10% would need to hold between 40-50% of their portfolio in bonds. If we were able to magically eliminate the three worst years of equity returns, at the cost of giving up the three best, that number dropped to 20-30%. And if we were able to eliminate the worst five at the cost of giving up the best five? Just 10%.

One interpretation of this data is that, with the benefit of hindsight, a moderate-risk investor would have had to carry a hefty allocation to bonds for the 55 years just to hedge against the low-probability risk of failing fast. If we believe the historical evidence supporting trend equity strategies, however, we may have an interesting solution at hand:

- A strategy that has historically captured a significant proportion of the equity risk premium.
- A strategy that has historically avoided a significant proportion of prolonged equity market declines.

Used appropriately, this strategy may help investors who are sensitive to failing slowly tactically increase their equity exposure when trends are favorable. Conversely, trend equity may help investors who are sensitive to failing fast de-risk their portfolio during negative trend environments.

To explore this opportunity, we will look at three strategic profiles: an 80% U.S. equity / 20% U.S. bond mix, a 50/50 mix, and a 20/80 mix. The first portfolio represents the profile of a growth investor who is sensitive to failing slow; the second portfolio represents a balanced investor, sensitive to both risks; the third represents a conservative investor who is sensitive to failing fast.

We will allocate a 10% slice of each portfolio to a naïve trend equity strategy in reverse proportion to the stock/bond mix. For example, for the 80/20 portfolio, 2% of the equity position and 8% of the bond position will be used to fund the trend equity position, creating a 78/12/10 portfolio. Similarly, the 20/80 will become an 12/78/10 and the 50/50 will become a 45/45/10.

We will use the S&P 500 index for U.S. equities, Dow Jones Corporate Bond index for U.S. bonds, and a 1-Year U.S. Government Note index for our cash proxy. The trend equity strategy will blend signals generated from trailing 6-through-12-month total returns, investing in the S&P 500 over the subsequent month in proportion to the number of positive signals. Remaining capital will be invested in the cash proxy. All portfolios are rebalanced monthly from 12/31/1940 through 12/31/2018.

Below we report the annualized returns, volatility, maximum drawdown, and Ulcer index (which seeks to simultaneously measure the duration and depth of drawdowns and can serve as a measure to a portfolio's sensitivity to failing fast) for each profile.

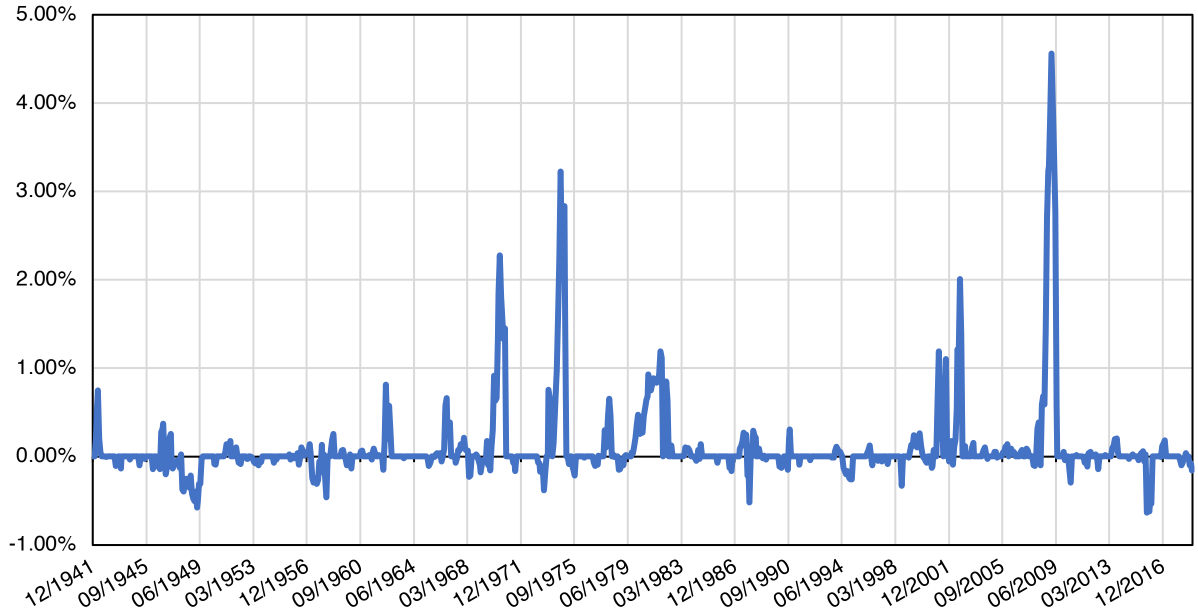
	Fail Fast		*Blend*		*Fail Slow*	
	20/80	**12/78/10**	**50/50**	**45/45/10**	**80/20**	**78/12/10**
Annualized Return	7.9%	8.0%	9.4%	9.6%	10.7%	11.0%
Annualized Volatility	5.8%	5.6%	8.4%	8.4%	11.9%	12.4%
Maximum Drawdown	16.9%	16.6%	28.8%	26.6%	42.9%	42.5%

Ulcer Index	0.025	0.025	0.045	0.044	0.083	0.087

Source: Global Financial Data. Calculations by Newfound Research. Returns are backtested and hypothetical. Past performance is not a guarantee of future results. Returns are gross of all fees. Returns assume the reinvestment of all distributions. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

For conservative investors sensitive to the risk of failing fast, we can see that the introduction of trend equity not only slightly increased returns, but it reduced the maximum drawdown and Ulcer index profile of the portfolio. Below we plot the actual difference in portfolio drawdowns between a 12/78/10 mix and a 20/80 mix over the backtested period.

Difference in Drawdowns: 12/78/10 versus 20/80

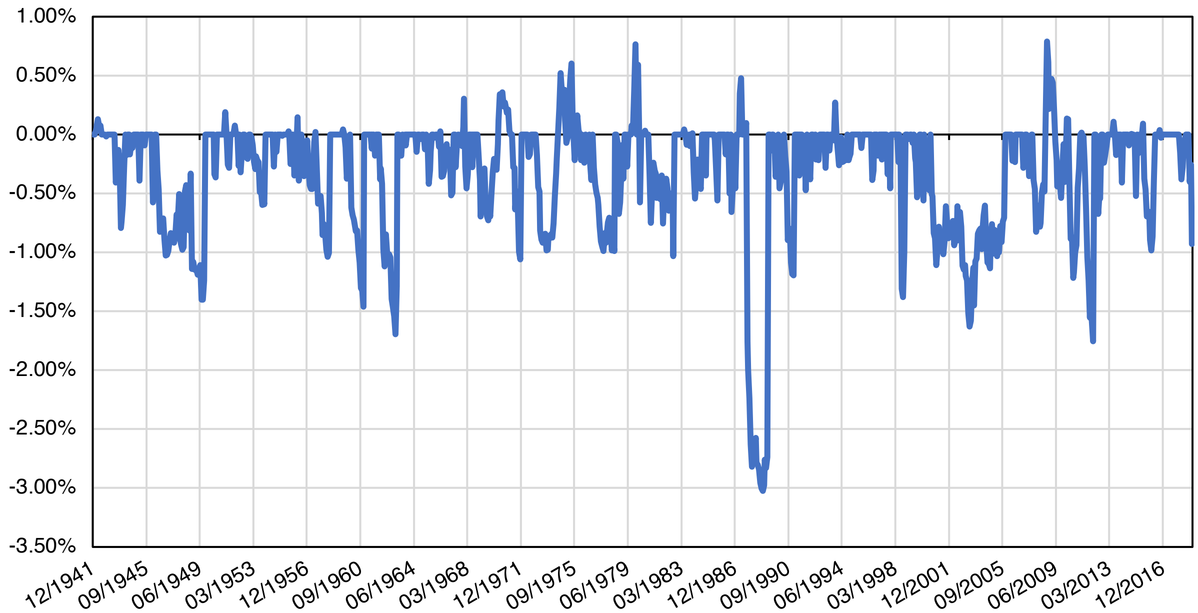


While we can see that there are periods where the 12/78/10 mix exhibited higher drawdowns (i.e. values below the 0% line), during major drawdown periods, the 12/78/10 mix historically provided relative relief. This is in line with our philosophy that risk cannot be destroyed, only transformed: the historical benefits that trend following has exhibited to avoiding significant and prolonged drawdowns have often come at the cost of increased realized drawdowns due to a slightly increased average allocation to equities as well as self-incurred drawdowns due to trading whipsaws.

On the opposite end of the spectrum, we can see that those investors sensitive to failing slowly were able to increase annualized returns without a significant increase to maximum drawdown. We should note, however, an increase in the Ulcer index, indicating more frequent and deeper drawdowns.

This makes sense, as we would expect the 78/12/10 mix to be on average over-allocated to equities, making it more sensitive to quick and sudden declines (e.g. 1987). Furthermore, the most defensive the mix can tilt is towards a 78/22 blend, leaving little wiggle-room in its ability to mitigate downside exposure. Nevertheless, we can see below that during periods of more prolonged drawdowns (e.g. 1975, 1980, and 2008), the 78/12/10 mix was able to reduce the drawdown profile slightly.

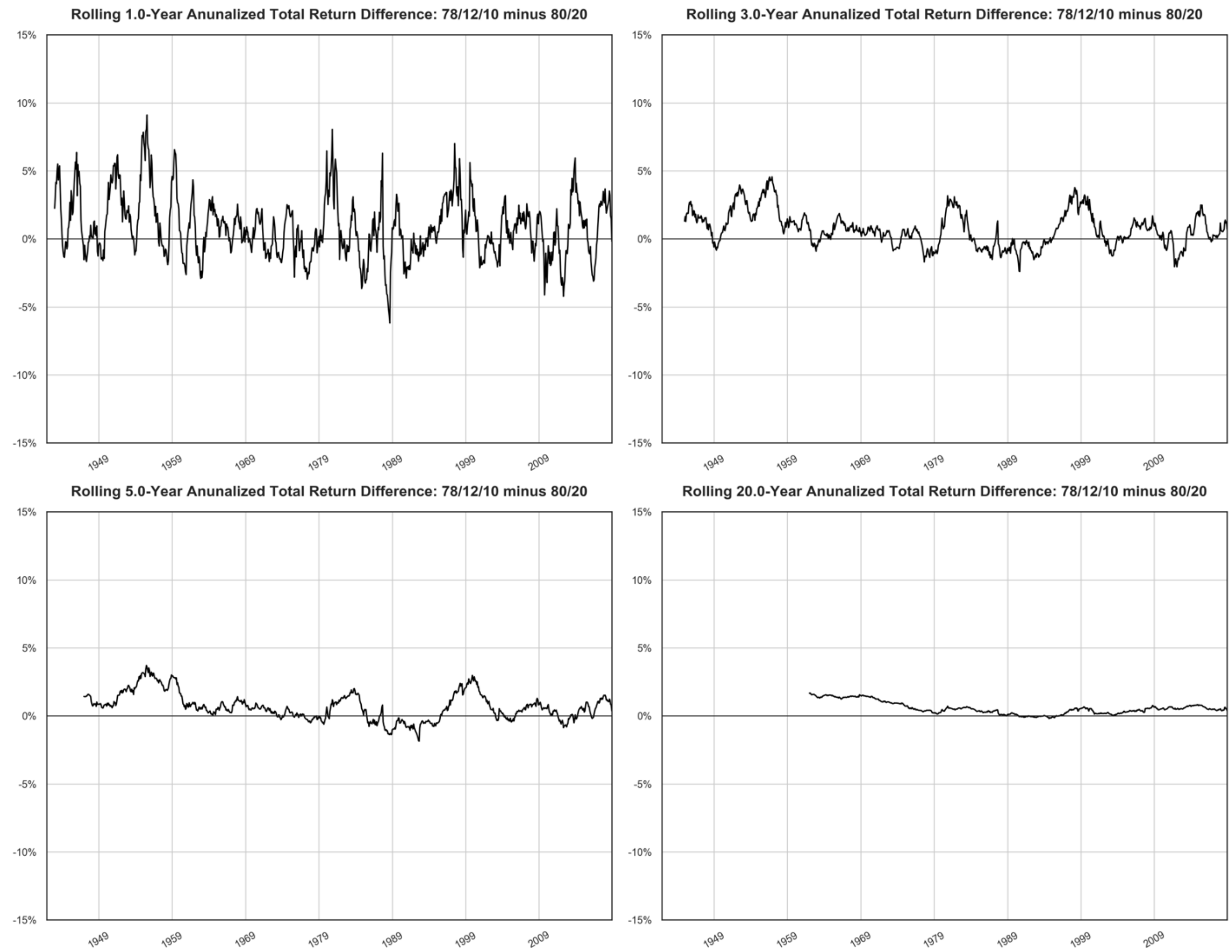
Differece in Drawdowns: 78/12/10 versus 80/20



In these backtests we see that investors sensitive to failing fast can fund a larger proportion of trend equity exposure from their traditional equity allocation in an effort to reduce risk while maintaining their return profile. Conversely, investors sensitive to failing slow can fund a larger proportion of their trend equity exposure from bonds, hoping to increase their annualized return while maintaining the same risk exposure.

Of course, long-term annualized return statistics can belie short-term experience. Examining rolling return periods, we can gain a better sense as to our confidence as to the time horizon over which we might expect, with confidence, that a strategy should contribute to our portfolio.

Below we plot rolling 1-to-10-year annualized return differences between the 78/12/10 and the 80/20 mixes.



We can see that in the short-term (e.g. 1-year), there are periods of both significant out- and under-performance. Over longer periods (5- and 10-years), which tend to capture "full market cycles," we see more consistent out-performance.

Of course, this is not always the case: the 78/12/10 mix underperformed the 80/20 portfolio for the 10 years following the October 1987 market crash. Being over-allocated to equities at that time had a rippling effect and serves to remind us that our default assumption should be that "risk cannot be destroyed, only transformed." But when we have the option to adjust our exposure to these risks, the benefit of avoiding slow failure may outweigh the potential to underperform slightly.

This evidence suggests that funding an allocation to trend equity in a manner that is in line with an investor's risk sensitivities may be beneficial. Nevertheless, we should also acknowledge that the potential benefits are rarely realized in a smooth, continuous manner and that the implementation should be considered a long-term allocation, not a trade.

Conclusion

Investors must navigate a significant number of risks throughout their lifecycle. At Newfound, we like to think of the two driving risks that investors face as the risk of failing fast and the risk of failing slow. Much like Odysseus navigating between Scylla and Charybdis, these risks are at direct odds with one another and trying to avoid one increases the risk of the other.

Fortunately, which of these risks an investor cares about evolves throughout their lifecycle. Young investors typically can afford to fail fast, as they have both future earning potential and time on their side. By not saving adequately, or investing too conservatively, however, a young investor can invite the risk of slow failure and find themselves woefully underfunded for future real liabilities. Hence investors at this stage or typically aggressively allocated towards growth assets.

As investors age, time and earning potential dwindle and the risk of fast failure increases. At this point, large and prolonged drawdowns can permanently impair an investor's lifestyle. So long as real liabilities are sufficiently funded, the risk of slow failure dwindles. Thus, investors often de-risk their portfolios towards stable return sources such as high-quality fixed income.

We believe this dual-risk framework is a useful model for determining how any asset or strategy should fit within a particular investor's plan. We demonstrate this concept with a simple trend equity strategy. For an investor sensitive to slow failure, we fund the allocation predominately from bond exposure; for an investor sensitive to fast failure, we fund the allocation predominately form equities.

Ultimately – and consistent with findings in our other commentaries – a risk-based mindset makes it obvious that allocation choices are really all about trade-offs in opportunity ("no pain, no premium") and risk ("risk cannot be destroyed, only transformed.")

TREND FOLLOWING IN CASH BALANCE PLANS

March 18, 2019

SUMMARY

- Cash balance plans are retirement plans that allow participants to save higher amounts than in traditional 401(k)s and IRAs and are quickly becoming more prevalent as an attractive alternative to defined benefit retirement plans.

- The unique goals of these plans (specified contributions and growth credits) often dictate modest returns with a very low volatility, which often results in conservative allocations.

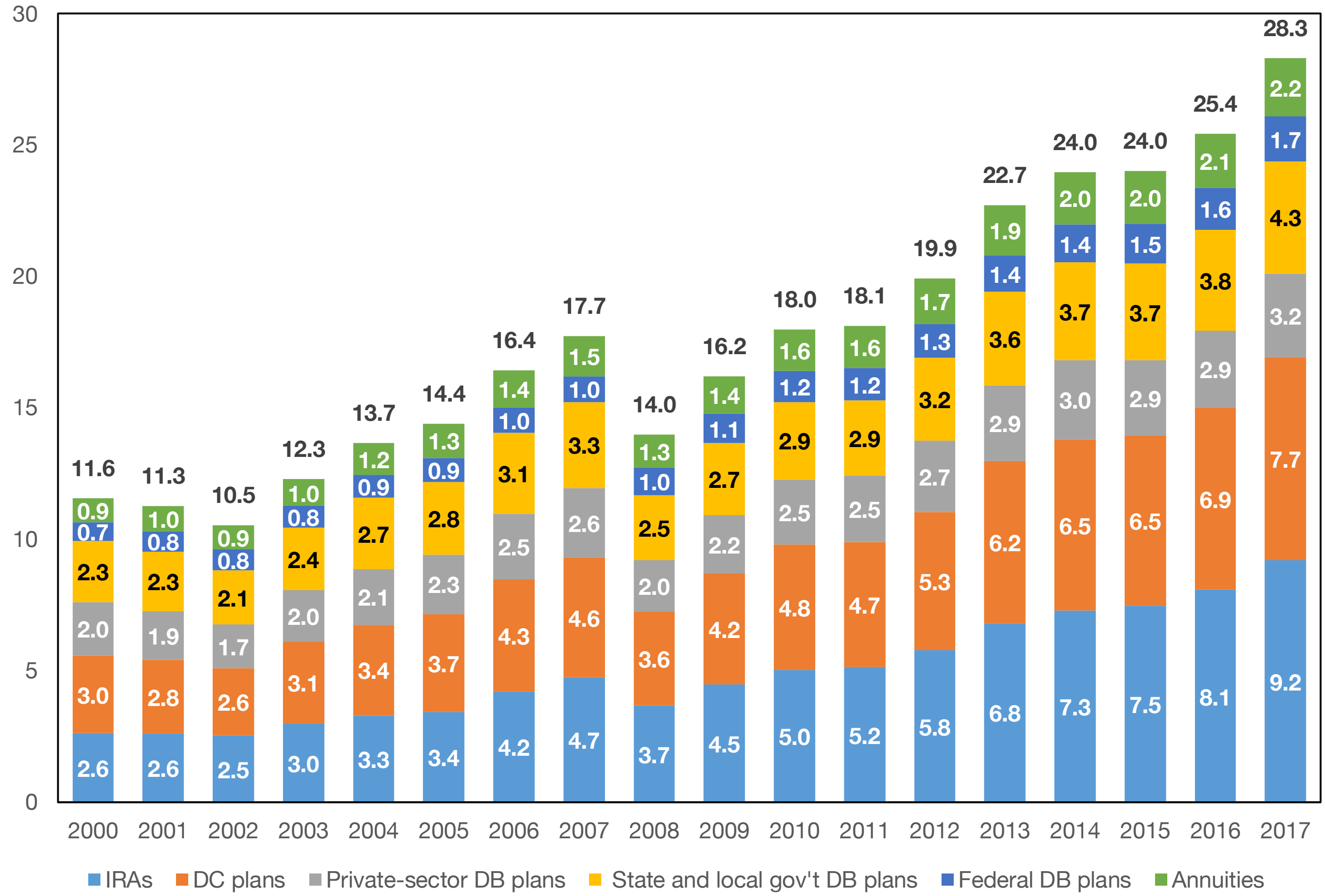
- However, at closely held companies, there is a balance between the tax-deferred amount that can be contributed by partners and the returns that the plan earns. If returns are too low, the company must make up the shortfall, but if the returns are too high the partners cannot maximize their tax-deferred contributions.

- By allocating to risk-managed strategies like trend equity, a cash balance plan can balance the frequency and size of shortfalls based on how the trend following strategy is incorporated within the portfolio.

- Trend following strategies have historically reduced the exposure to large shortfalls in exchange for more conservative performance during periods where the plan is comfortably hitting its return target.

Retirement assets have grown each year since the Financial Crisis, exhibiting the largest gains in the years that were good for the market such as 2009, 2013, and 2017.

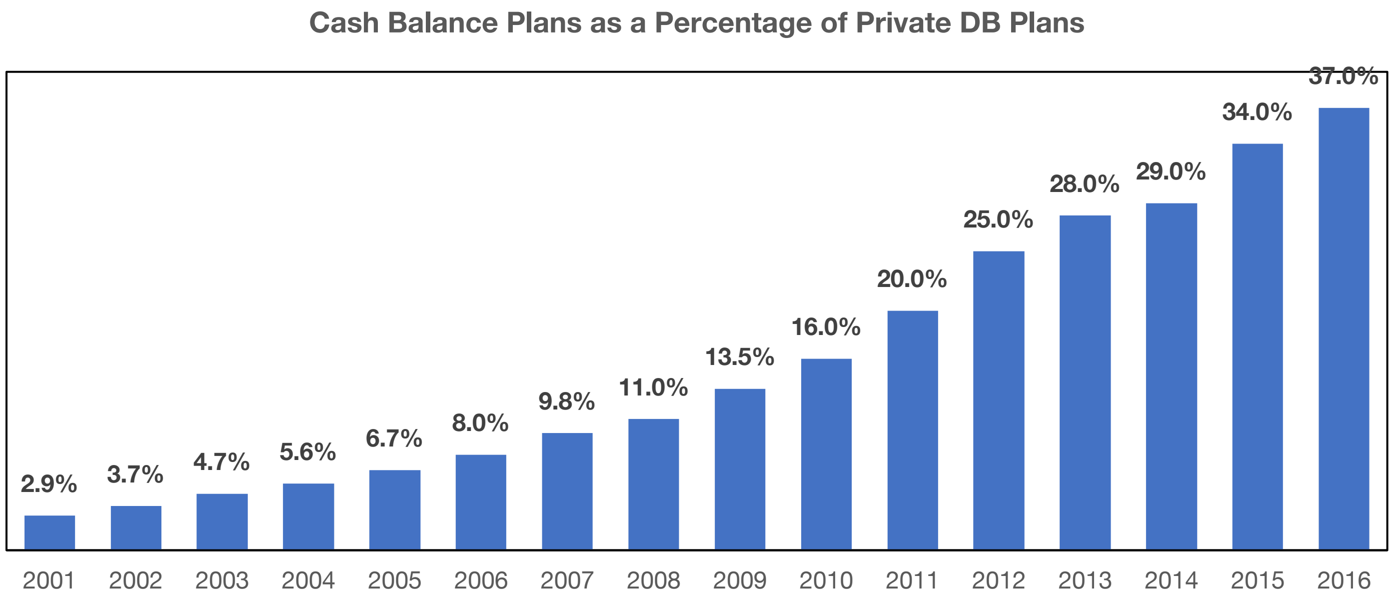
U.S. Total Retirement Market Assets (Trillions of $)



Source: Investment Company Institute (ICI).

With low interest rates, an aging workforce, and continuing pressure to reduce expected rates of return going forward, many employers have shifted from the defined benefit (DB) plans used historically to defined contribution (DC) models, such as 401(k)s and 403(b)s. While assets within DB plans have still grown over the past decade, the share of retirement assets in IRAs and DC plans has grown from around 50% to 60%.

But even with this shift toward more employee directed savings and investment, there is a segment of the private DB plan space that has seen strong growth since the early 2000s: cash balance plans.



Source: Kravitz. 2018 National Cash Balance Research Report.

What is a cash balance plan?

It's sort of a hybrid retirement plan type. Employers contribute to it on behalf of their employees or themselves, and each participant is entitled to those assets plus a rate of return according to a prespecified rule (more on that in a bit).

Like a defined contribution plan, participants have an account value rather than a set monthly payment.

Like a defined benefit plan, the assets are managed professionally, and the actual asset values do not affect the value of the participant benefits. Thus, as with any liability-driven outcome, the plan can be over- or under-funded at a given time.

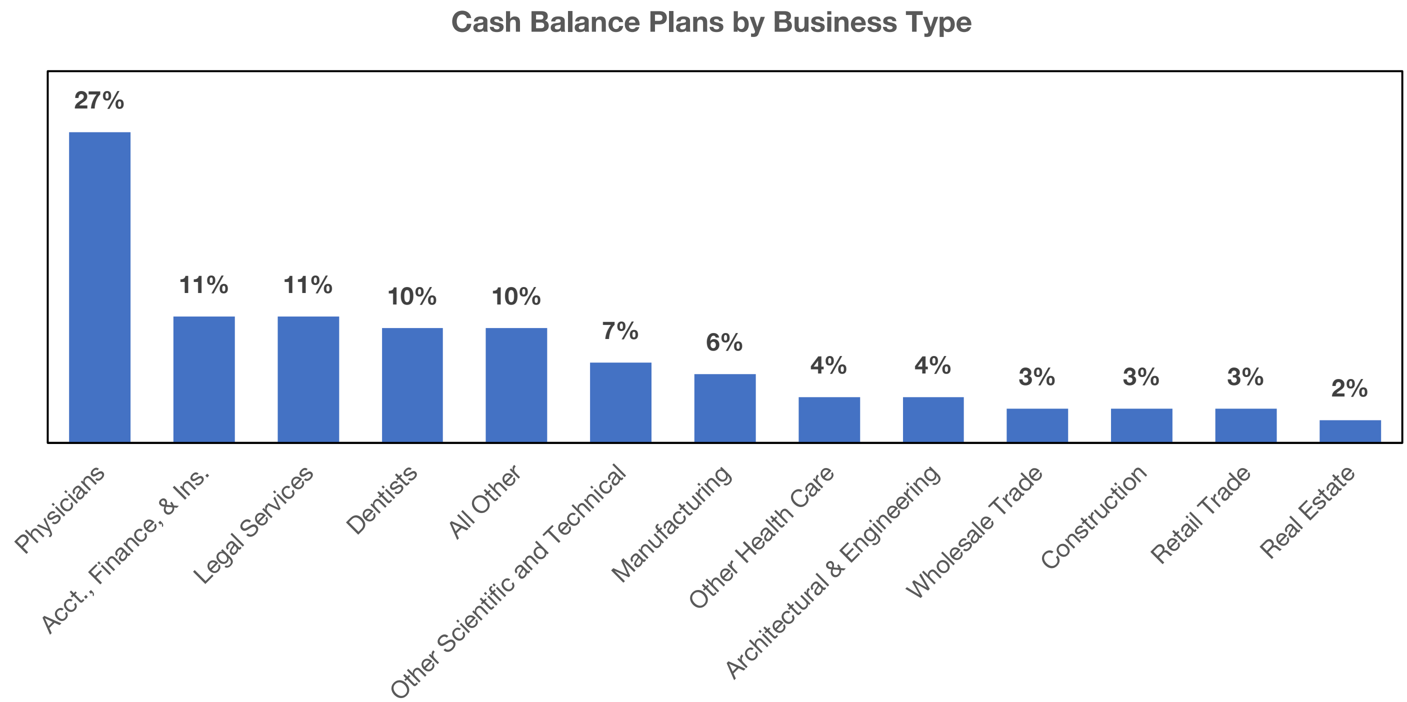
What's the appeal?

According to Kravitz, (2018)[17] over 90% of cash balance plans are in place at companies with fewer than 100 participants. These companies tend to be white-collar professionals, where a significant proportion of the employees are highly compensated (e.g. groups of doctors, dentists, lawyers, etc.).

Many of these professionals likely had to spend a significant amount of time in professional school and building up practices. Despite higher potential salaries, they may have high debt loads to pay down. Similarly, entrepreneurs may have deferred compensating themselves for the sake of building a successful business.

[17] https://www.cashbalancedesign.com/wp-content/uploads/2018/08/NationalCashBalanceResearchReport2018.pdf

Thus, by the time these professionals begin earning higher salaries, the amount of time that savings can compound for retirement has been reduced.



Source: Kravitz. 2018 National Cash Balance Research Report.

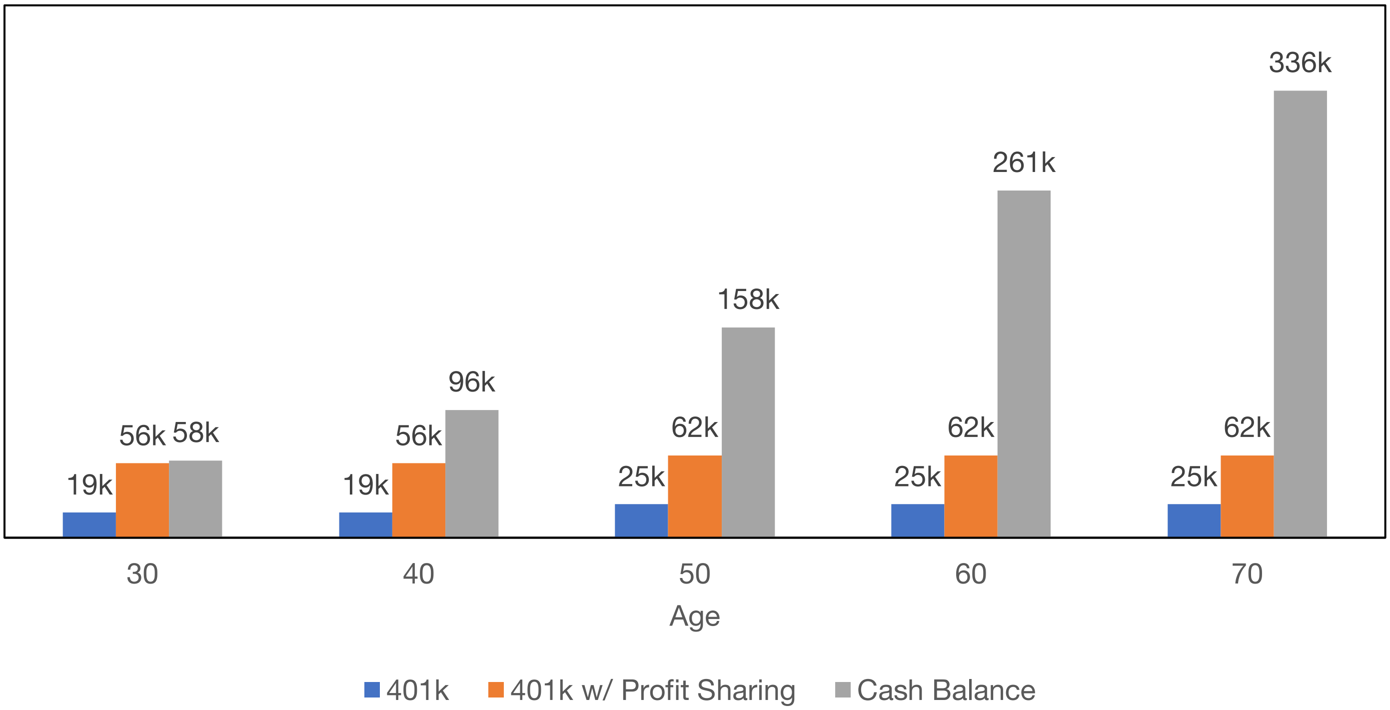
One option for these types of investors is to simply save more income in a traditional brokerage account, but this foregoes any benefit of deferring taxes until retirement.

Furthermore, even if these investors begin saving for retirement at the limit for 401(k) contributions, it is possible that they could end up with a lower account balance than a counterpart saving half as much per year but starting 10 years earlier. Time lost is hard to make up.

This, of course, depends on the sequence and level of investment returns, but an investor who is closer to retirement has less ability to bear the risk of failing fast. Not being able to take as much investment risk necessitates having a higher savings rate.

Cash balance plans can help solve this dilemma through significantly higher contribution limits.

Retirement Plan Contribution Limits



Source: Kravitz.

An extra $6,000 in catch-up contributions starting for a 401(k) at age 50 seems miniscule compared to what a cash balance plan allows.

Now that we understand why cash balance plans are becoming more prevalent in the workplace, let's turn to the investment side of the picture to see how a plan can make good on its return guarantees.

The Return Guarantee

Aside from the contribution schedule for each plan participant, the only other piece of information needed to determine the size of the cash balance plan liability in a given year is the annual rate at which the participant accounts grow.[18] There are a few common ways to set this rate:

[18] We are assuming that the goal of the plan is to remain current on liabilities. Since cash balance plans are portable, this assumption could reflect total uncertainty about the dates when participants will leave the company.

1. A fixed rate of return per year, between 2% and 6%.
2. The 30-year U.S. Treasury rate.
3. The 30-year U.S. Treasury rate with a floor of between 3% and 5%.
4. The actual rate of return of the invested assets, often with a ceiling between 3% and 6%.

The table below shows that of the plans surveyed by Kravitz (2018), the fixed rate of return was by far the most common and the actual rate of return credit was the least common.

Interest Crediting Rate	Percentage*
Fixed Rate of Return (ranging from 2% to 6%)	70.2%
30-year Treasury Rate	10.6%
30-year Treasury Rate with a floor (ranging from 3% to 5%)	13.2%
Actual Rate of Return*	6%
*80% of Actual Rate of Return (ARR) plans have a ceiling of 3%-6%	

The Actual Rate of Return option is actually becoming more popular, especially with large cash balance plans, now that federal regulations allow plan sponsors to offer multiple investments in a single plan to better serve the participants who may have different retirement goals. This return option removes much of the investment burden from the plan sponsor since what the portfolio earns is what the participants get, up to the ceiling. Anything earned above the ceiling increases the plan's asset value above its liabilities. Actual rate of return guarantees make it so that there is less risk of a liability shortfall when large stakeholders in the cash balance plan leave the company unexpectedly.

In this commentary, we will focus on the cases where the plan may become underfunded if it does not hit the target rate of return.

We often say, "No Pain, No Premium." Well, in the case of cash balance plans, plan sponsors typically only want to bear the minimal amount of pain that is necessary to hit the premium.

With large firms that can rely more heavily on actuarial assumptions for participant turnover, much of this risk can be borne over multiyear periods. A shortfall in one year can be replenished by a combination of extra contributions from the company according to IRS regulations and (hopefully) more favorable portfolio gains in subsequent years. Any excess returns can be used to offset how much the company must contribute annually for participants.

In the case of closely held firms, things change slightly.

At first glance, it should be a good thing for a plan sponsor to earn a higher rate of return than the committed rate. But when we consider that many cash balance plans are in place at firms where the participants desire to contribute as much as the IRS allows to defer taxation, then earning *more* than the guaranteed rate of return actually represents a risk. At closely held firms, "the company" and "the participants" are essentially one in the same. The more the plan earns, the less *you* can contribute.

And with higher return potential comes a higher risk of earning below the guaranteed rate. When a company is small, making up shortfalls out of company coffers or stretching for higher returns in subsequent years may not be in the company's best interest.

Investing a Cash Balance Plan

Because of the aversion to both high returns and high risk, many cash balance plans are generally invested relatively conservatively, typically in the range of a 20% stock / 80% bond portfolio (20/80) to a 40/60.

To put some numbers down on paper, we will examine the return profile of three different portfolios: a 20/80, 30/70, and 40/60 fixed mix of the S&P 500 and a constant maturity 10-year U.S. Treasury index.

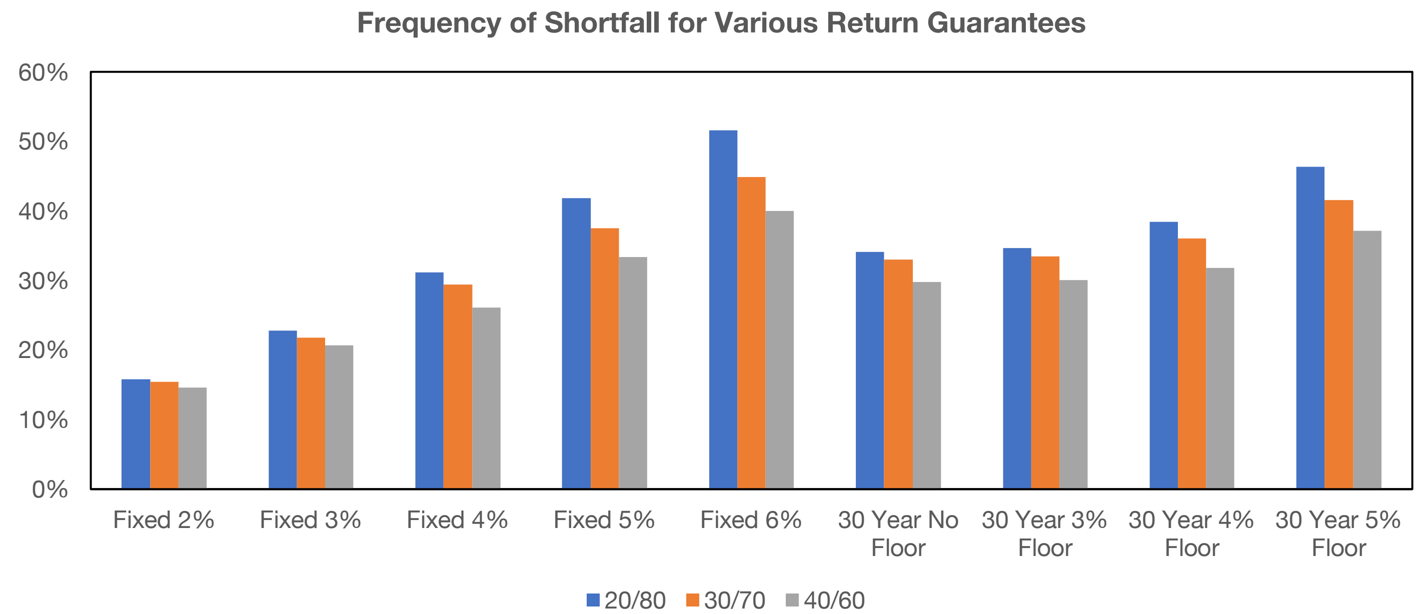
We will also calculate the rate of return guarantees described above each year from 1871 to 2018.

Starting each January, if the return of one of the portfolio profiles meets hits the target return for the year, then we will assume it is cashed out. Otherwise, the portfolio is held the entire year.

As the 30-year U.S. Treasury bond came into inception in 1977 and had a period in the 2000s where it was not issued, we will use the 10-year Treasury rate as a proxy for those periods.

The failure rate for the portfolios are shown below. [19]

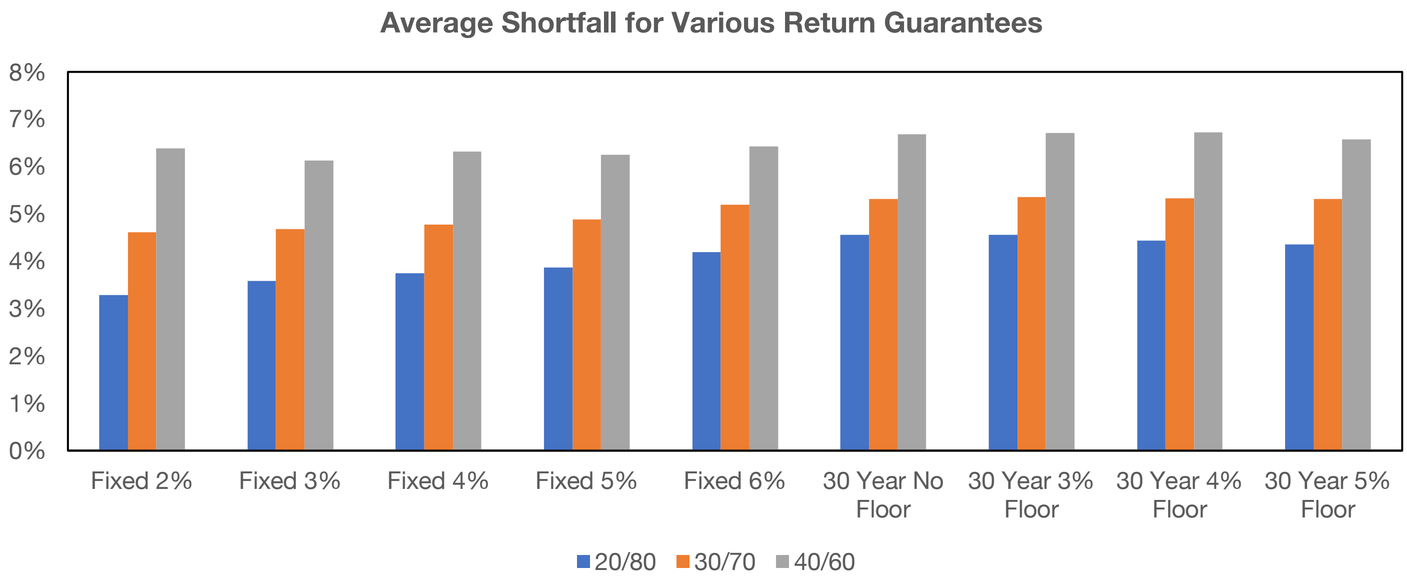
[19] The metrics shown here using monthly data are likely overestimates since portfolios could hit the return targets intramonth. However, we presume that the general trends in the data would remain even if this were taken into account.



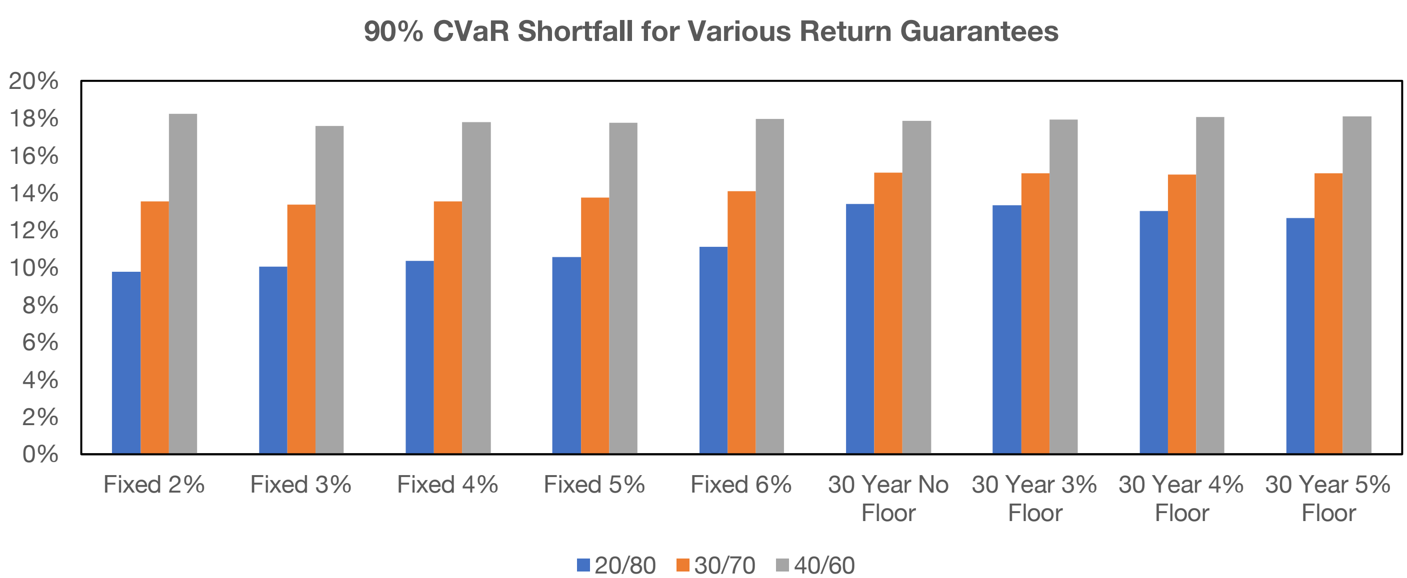
Source: Robert Shiller Data Library, St. Louis Fed. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

We can see that as the rate of return guarantee increases, either through the fixed rate or the floor on the 30-year rate, the rate of shortfall increases for all allocations, most notably for the conservative 20/80 portfolio.

In these failure scenarios, the average shortfall and the average shortfall in the 90% of the worst cases (similar to a CVaR) are relatively consistent.



Source: Robert Shiller Data Library, St. Louis Fed. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.



Source: Robert Shiller Data Library, St. Louis Fed. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

These shortfall numbers may not be a big deal for new plans when the contributions represent a significant percentage of the asset base. For example, for a $1M plan with $500k in contributions per year, a 15% shortfall is only $150k, which can be amortized over a number of years. Higher returns in the subsequent years can offset this, or partners could agree to reduce their personal contributions so that the company can have free cash to make up for the shortfall.

The problem is more pressing for plans where the asset base is significantly larger than the yearly contributions. For a $20M plan with $500k in yearly contributions, a 15% shortfall is $3M. Making up this shortfall from company assets may be more difficult, even with amortization.

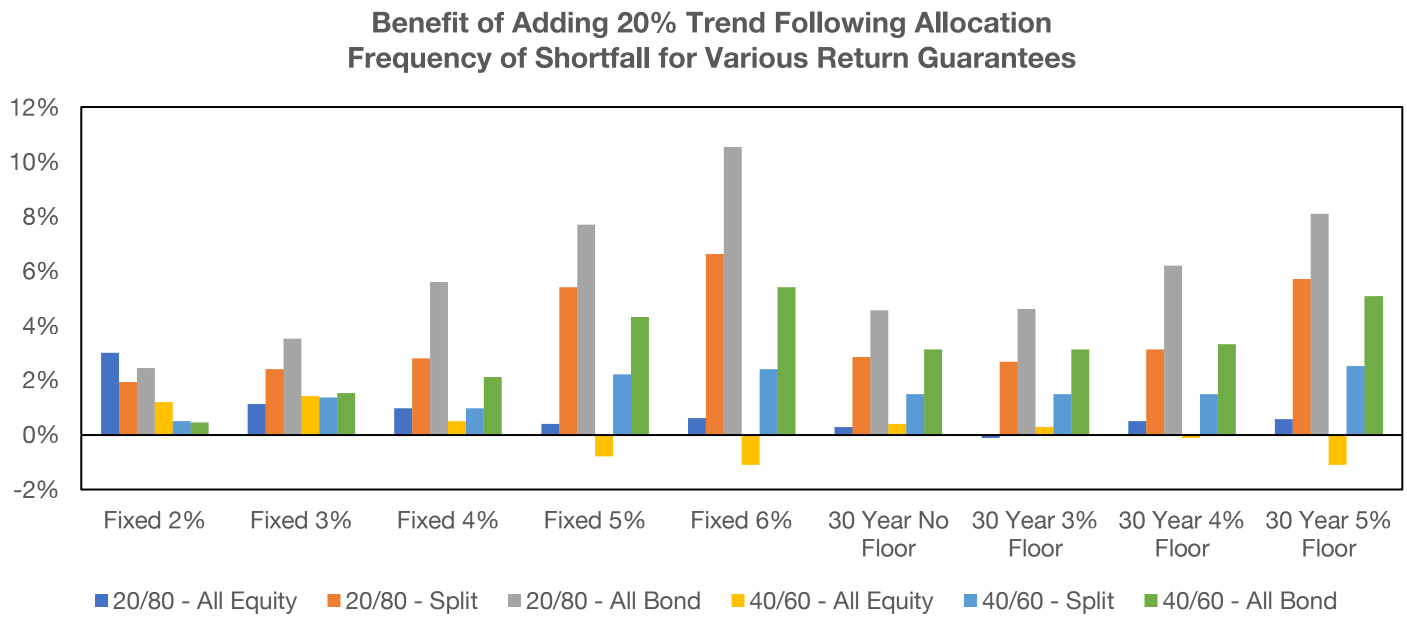
Waiting for returns from the market can also be difficult in this case when there have been historical drawdowns in the market lasting 2-3 years from peak to trough (e.g. 1929-32, 2000-02, and 1940-42).

Risk-managed strategies can be a natural way to mitigate these shortfalls, both in their magnitude and frequency.

Using Trend Following in a Cash Balance Plan

Along the lines of our Three Uses of Trend Equity, we will look at adding a 20% allocation to a simple trend-following equity ("trend equity") strategy in a cash balance plan. By taking the allocation either from all equities, all bonds, or an equal share of each.

For ease of illustration, we will only look at the 20/80 and 40/60 portfolios. The following charts show the benefit (i.e. reduction in shortfall) or detriment (i.e. increase in shortfall) of adding the 20% trend equity sleeve to the cash balance plan based on the metrics from the previous section.

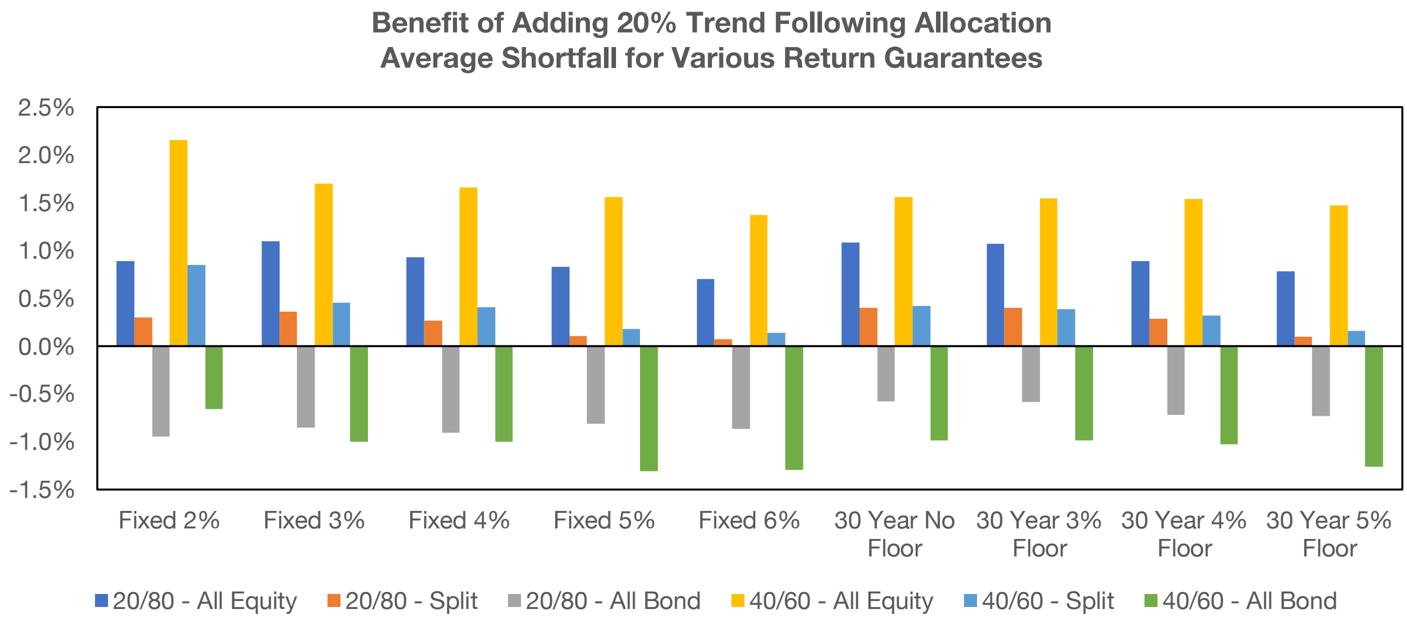


Source: Robert Shiller Data Library, St. Louis Fed. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

For most of these return guarantees, substituting a greater proportion of bonds for trend equity reduced the frequency of shortfalls. This makes sense over a period where equities generally did well and a trend equity strategy increased participation during the up-markets.

Substituting in trend equity solely from the equity allocation was detrimental for a few of the return guarantees, especially the higher ones.

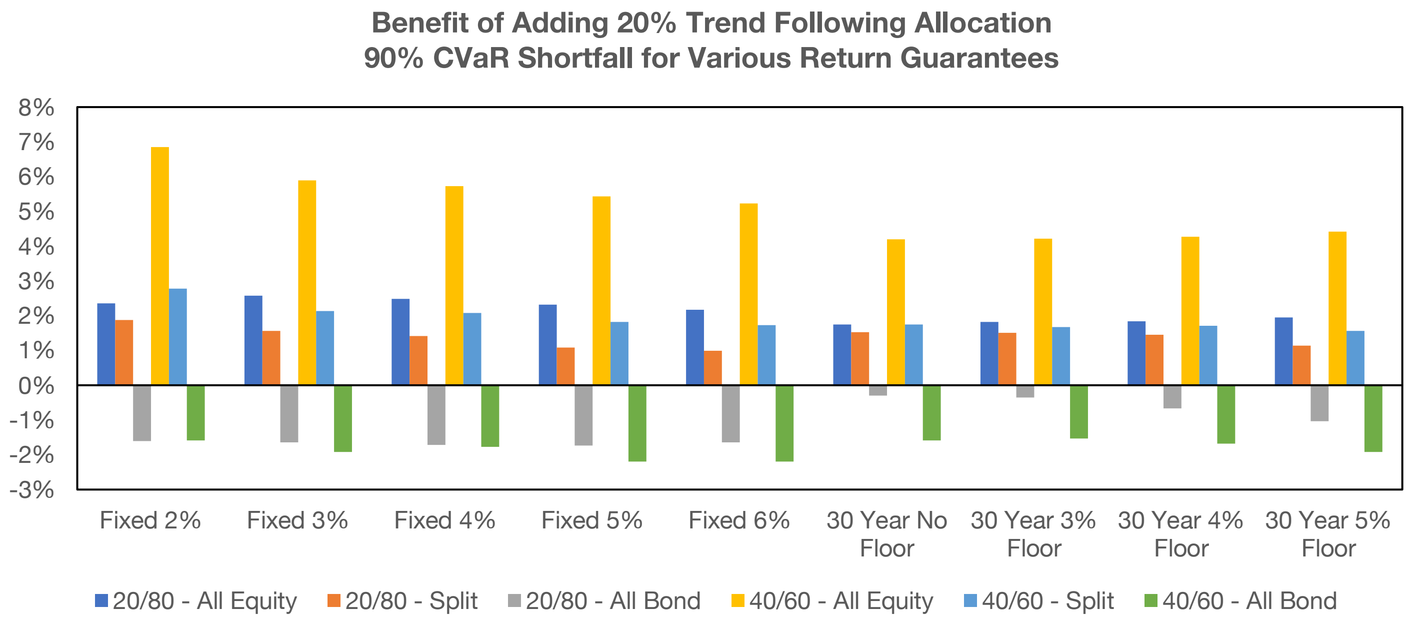
But the frequency of shortfalls is only one part of the picture.



Source: Robert Shiller Data Library, St. Louis Fed. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

Many of the cases that showed a benefit from a frequency of shortfall perspective sacrifice the average shortfall or average shortfall in the most extreme scenarios. Conversely, case that sacrifice on the frequency of shortfalls generally saw a meaningful reduction in the average shortfalls.

This is in line with our philosophy that risks are not destroyed, only transformed.



Source: Robert Shiller Data Library, St. Louis Fed. Calculations by Newfound Research. Past performance is not a guarantee of future results. All returns are hypothetical and backtested. Returns are gross of all fees. This does not reflect any investment strategy offered or managed by Newfound Research and was constructed exclusively for the purposes of this commentary. It is not possible to invest in an index.

So which risks should a cash balance plan bear?

This can be answered by determining the balance of the plan to be exposure to failing fast and failing slow.

If a cash balance plan is large, even a moderate shortfall can be very large in dollar terms. These plans are at risk of failing fast. Mitigating the size of the shortfalls is definitely a primary concern.

If a cash balance plan is new or relatively small, it is somewhat like an investor early in their working career. Larger losses from a percentage perspective are smaller in dollar terms compared to a larger plan. These plans can stand to have larger shortfalls. If the shortfalls occur less frequently, there is the ability to generate higher returns in years after a loss to recoup some of the losses.

However, these small plans should still be concerned mostly about fast failure. The yearly reckoning of the liability to the participants skews the risks more heavily in the direction of fast failure. This is especially true when we factor in the demographic of the workforce. When employees leave, they are entitled to their account value based on the guaranteed return, not the underlying asset value. If a participant cashes out at a time when the assets are down, then the remaining participant are less funded based on the assets that are left.

Therefore, allocating to the trend strategy out of the equity sleeve or an equal split between equities and bonds is likely more in line with the goals of a cash balance plan.

Conclusion

Cash balance plans are quickly becoming more prevalent as an attractive alternative to defined benefit retirement plans. They are desirable both from an employer and employee perspective and can be a way to accelerate retirement savings, especially for highly compensated workers at small companies.

The unique goals of these plans (e.g. guaranteed returns, maximizing tax-deferred contributions, etc.) often dictate modest returns with a very low volatility. Since some risk must be borne in order to generate returns, these portfolios are typically allocated very conservatively.

Even so, there is a risk they will not hit their return targets.

By allocating to risk-managed strategies like trend equity, a cash balance plan can balance the frequency and size of shortfalls based on how the trend following strategy is incorporated within the portfolio.

Allocating to a trend equity strategy solely from bonds can reduce the frequency of shortfalls in exchange for larger average shortfalls. Allocating to a trend following equity strategy solely from equities can increase the frequency of shortfalls but reduce the average size of shortfalls and the largest shortfalls.

The balance for a specific plan depends on its size, the demographic of the participants, the company's willingness and ability to cover shortfalls, and the guaranteed rate of return.

As with most portfolio allocation problems the solution exists on a sliding scale based on what risks the portfolio is more equipped to bear. For cash balance plans, managing the size of shortfalls is likely a key issue, and trend following strategies can be a way to adjust the exposure to large shortfalls in exchange for more conservative performance during periods where the plan is comfortably hitting its return target.

TIME DILATION

March 25, 2019

SUMMARY

- Information does not flow into the market at a constant frequency or with constant magnitude.
- By sampling data using a constant time horizon (e.g. “200-day simple moving average”), we may over-sample during calm market environments and under-sample in chaotic ones.
- As an example, we introduce a highly simplified price model and demonstrate that trend following lookback periods should be a dynamic function of trend and volatility in the time domain.
- By changing the sampling domain slightly, we are able to completely eliminate the need for the dynamic lookback period.
- Finally, we demonstrate a more complicated model that samples market prices based upon cumulative log differences, creating a dynamic moving average in the time domain.
- We believe that there are other interesting applications of this line of thinking, many of which may already be in use today by investors who may not be aware of it (e.g. tracking-error-based rebalancing techniques).

In the 2014 film *Interstellar*, Earth has been plagued by crop blights and dust storms that threaten the survival of mankind. Unknown, interstellar beings have opened a wormhole near Saturn, creating a path to a distant galaxy and the potential of a new home for humanity.

Twelve volunteers travel into the wormhole to explore twelve potentially hospitable planets, all located near a massive black hole named Gargantua. Of the twelve, only three reported back positive results.

With confirmation in hand, the crew of the spaceship Endurance sets out from Earth with 5,000 frozen human embryos, intent on colonizing the new planets.

After traversing the wormhole, the crew sets down upon the first planet – an ocean world – and quickly discovers that it is actually inhospitable. A gigantic tidal wave kills one member of the crew and severely delays the lander’s departure.

The close proximity of the planet to the gravitational forces of the supermassive black hole invites exponential time dilation effects. The positive beacon that had been tracked had perhaps been triggered just minutes prior on the planet. For the crew, the three hours spent on the planet amounted to over *23 years* on Earth. The crew can only watch, devastated, as their loved ones age before their eyes in the video messages received – and never responded to – in their multi-decade absence.

Our lives revolve around the clock, though we do not often stop to reflect upon the nature of time.

Some aspects of time tie to corresponding natural events. A day is simply reckoned from one midnight to the next, reflecting the Earth's full rotation about its axis. A year, which reflects the length of time it takes for the Earth to make a full revolution around the Sun, will also correspond to a full set of a seasons.

Others, however, are seemingly more arbitrary. The twenty-four-hour day is derived from ancient Egyptians, who divided day-time into 10 hours, bookended by twilight hours. The division of an hour into sixty minutes comes from the Babylonians, who used a sexagesimal counting system.

We impose the governance of the clock upon our financial system as well. Public companies prepare quarterly and annual reports. Economic data is released at a scheduled monthly or quarterly pace. Trading days for U.S. equity markets are defined as between the hours of 9:30am and 4:30pm ET.

In many ways, our imposition of the clock upon markets creates a natural cadence for the flow of information.

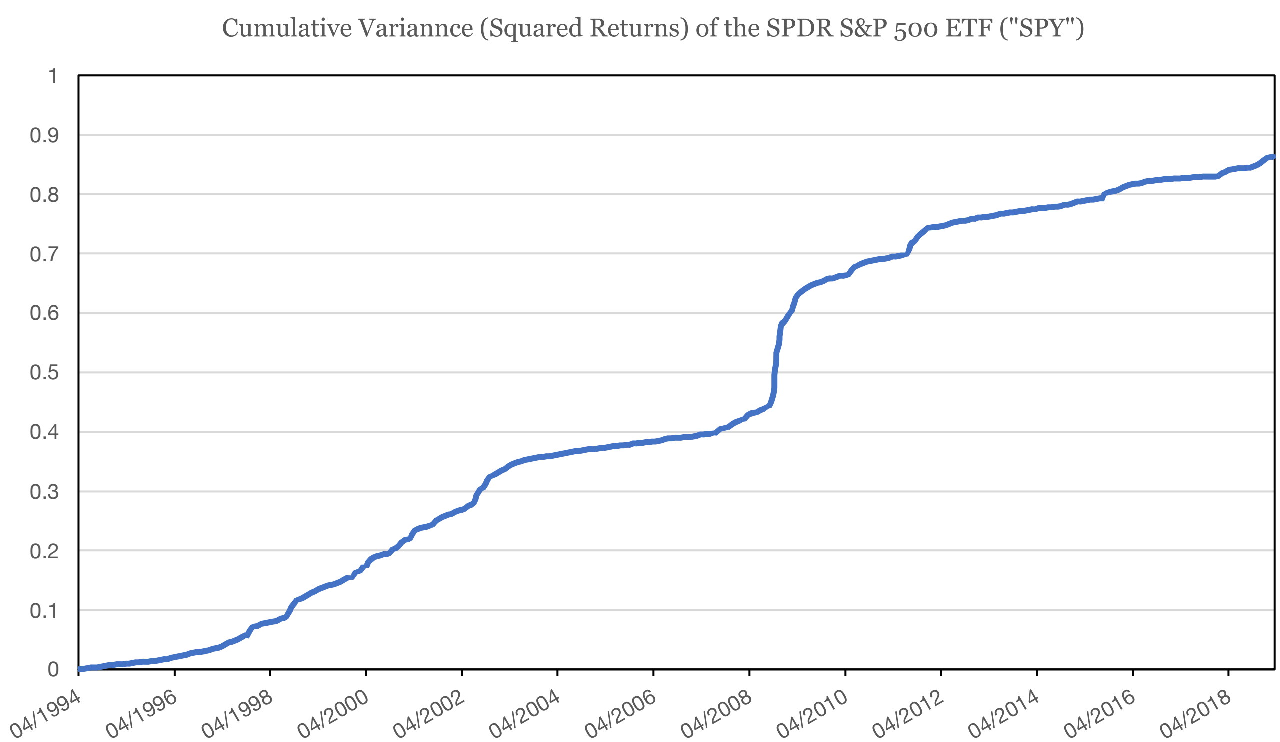
Yet, despite our best efforts to impose order, information most certainly does not flow into the market in a constant or steady manner.

New innovations, geopolitical frictions, and errant tweets all represent idiosyncratic events that can reshape our views in an instant. A single event can be of greater import than all the cumulative economic news that came before it; just consider the collapse of Lehman Brothers.

And much like the time dilation experienced by the crew of Endurance, a few, harrowing days of 2008 may have felt longer than the entirety of a tranquil year like 2017.

One way of trying to visualize this concept is by looking at the cumulative variance of returns. Given the clustered nature of volatility, we would expect to see periods where the variance accumulates slowly ("calm markets") and periods where the variance accumulates rapidly ("chaotic markets").

When we perform this exercise – by simply summing squared daily returns for the S&P 500 over time – we see precisely this. During market environments that exhibit stable economic growth and little market uncertainty, we see very slow and steady accumulation of variance. During periods when markets are seeking to rapidly reprice risk (e.g. 2008), we see rapid jumps.



Source: CSI Data. Calculations by Newfound Research.

If we believe that information flow is not static and constant, then sampling data on a constant, fixed interval will mean that during calm markets we might be *over-sampling* our data and during chaotic markets we might be *under-sampling.*

Let's make this a bit more concrete.

Below we plot the –adjusted closing price of the S&P 500– and its –200-day simple moving average–. Here, the simple moving average aims to estimate the trend component of price. We can see that during the 2005-2007 period, it estimates the underlying trend well, while in 2008 it dramatically lags price decline.

Source: CSI Data. Calculations by Newfound Research.

The question we might want to ask ourselves is, *why are looking at the prior 200 days?* Or, more specifically, why is a day a meaningful unit of measure? We already demonstrated above that it very well may not be: one day might be packed with economically-relevant information and another entirely devoid.

Perhaps there are other ways in which we might think about sampling data. We could, for example, sample data based upon cumulative volume intervals. Another might be on a fixed number of cumulative ticks or trades. Yet another might be on a fixed cumulative volatility or variance.

As a firm which makes heavy use of trend-following techniques, we are particularly partial to the latter approach, as the volatility of an asset's trend versus its price should inform the trend lookback horizon. If we think of trend following as being the trading strategy that replicates the payoff profile of a straddle, increased volatility levels will decrease the delta of the option positions, and therefore decrease our position size. An interpretation of this effect is that the increased volatility decreases our certainty of where price will fall at expiration, and therefore we need to decrease our sensitivity to price movements.

If that all sounds like Greek, consider this simple example. Assume that price follows a highly simplified model as a function of time:

$$p_t = mt + \mathrm{a} * \sin(t)$$

There are two components of this model: the linear trend and the noise.

Now let's assume we are attempting to identify whether the linear trend is positive or negative by using a simple moving average ("SMA") of price:

$$s_n(t) = \frac{1}{n} \sum_{i=t-n}^{t} p_i$$

To determine if there is a positive or a negative trend, we simply ask if our current SMA value is greater or less than the prior SMA value. For a positive trend, we require:

$$s_n(t) > s_n(t-1)$$

Substituting our above definition of the simple moving average:

$$\frac{1}{n} \sum_{i=t-n}^{t} p_i > \frac{1}{n} \sum_{i=t-1-n}^{t-1} p_i$$

When we recognize that most of the terms on the left also appear on the right, we can re-write the whole comparison as the new price in the SMA being greater than the old price dropping out of the SMA:

$$p_t > p_{t-1-n}$$

Which, through substitution of our original definition, leaves us with:

$$mt + \mathrm{a} * \sin(t) > m(t-1-n) + \mathrm{a} * \sin\,(t-1-n)$$

$$mt + \mathrm{a} * \sin(t) > mt - m - mn + \mathrm{a} * \sin\,(t-1-n)$$

Re-arranging a bit, we get:

$$m + mn + \mathrm{a}(\sin(t) - \sin\,(t-1-n)) > 0$$

Here we use the fact that sin(x) is bounded between -1 and 1, meaning that:

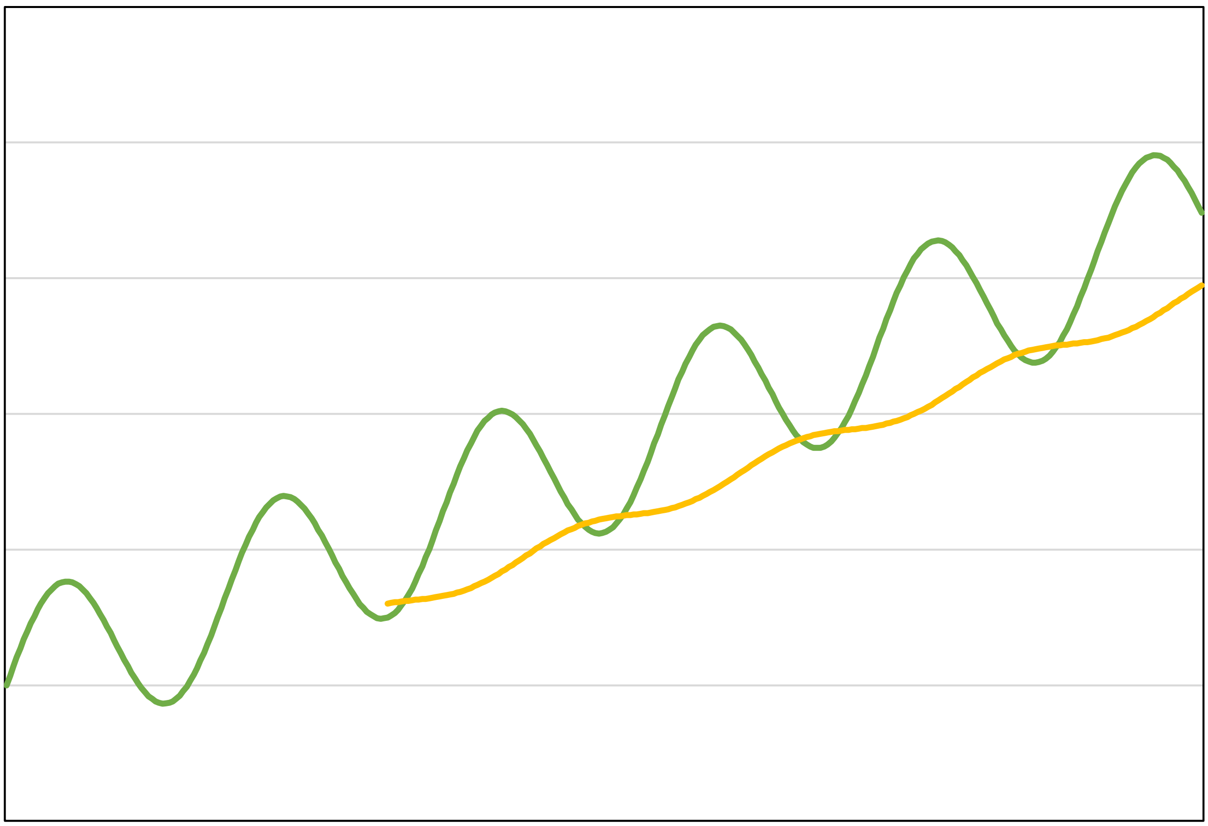
$$2 \geq (\sin(t) - \sin(t-1-n) \geq -2$$

Assuming a positive trend (m > 0), we can replace with our worst-case scenario,

$$m + mn - 2a > 0$$

$$n > \frac{2a}{m} - 1$$

To quickly test this result, we can construct a simple time series where we assume a=3 and m=0.5, which implies that our SMA length should be greater than 11. We plot the –time series– and –SMA– below. Note that the –SMA– is always increasing.



Despite being a highly simplified model, it illuminates that our lookback length should be a function of noise versus trend strength. The higher the ratio of noise to trend, the longer the lookback required to smooth out the noise. On the other hand, when the trend is very strong and the noise is week, the lookback can be quite short.[20]

[20] In practice, estimating parameters for "noise" and "trend" can be rather complicated. However, strategies that employ a volatility targeting approach may have an advantage, as one parameter becomes fixed.

Thus, if trend and noise change over time (which we would expect them to), the optimal lookback will be a dynamic function. When trend is much weaker than noise, we our lookback period will be extended; when trend is much stronger than noise, the lookback period shrinks.

But what if we transform the sampling domain? Rather than sampling price every time step, what if we sample price as a function of cumulative noise? For example, using our simple model, we could sample when cumulative noise sums back to zero (which, in this example, will be the equivalent of sampling every 2π time-steps).[21]

Sampling at that frequency, how many of data points would we need to estimate our trend? We need not even work out the math as before; a bit of analytical logic will suffice. In this case, because we know the cumulative noise equals zero, we know that a point-to-point comparison will be affected *only* by the trend component. Thus, we only need n=1 in this new domain.

And this is true regardless of the parameterization of trend or noise. *Goodbye!* dynamic lookback function.

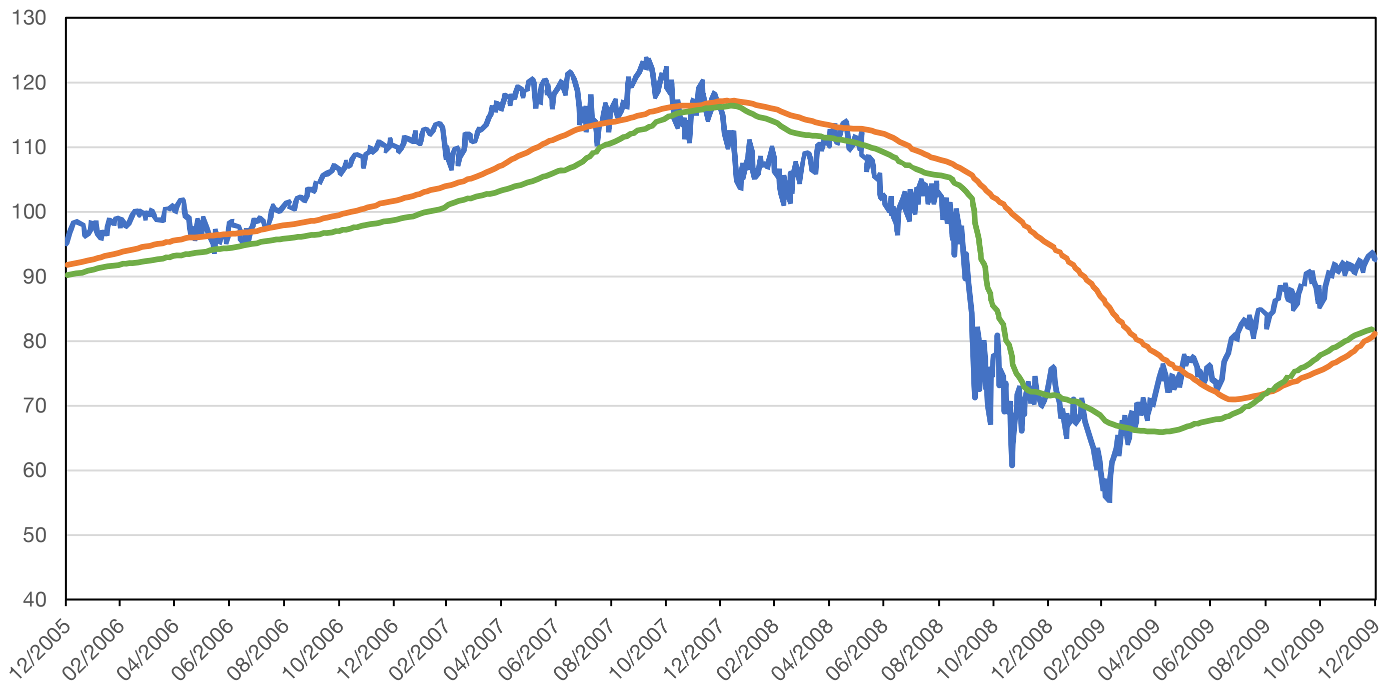
Of course, this is a purely hypothetical – and dramatically over-simplified – model. Nevertheless, it may illuminate why time-based sampling may not be the most efficient practice if we do not believe that information flow is constant.

Below, we again plot the –S&P 500– as well as a standard –200-day simple moving average–.

We also sample prices of the S&P 500 based upon cumulative magnitude of log differences, approximating a cumulative 2.5% volatility move. When the market exhibits low volatility levels, the process samples price less frequently. When the market exhibits high volatility, it samples more frequently. Finally, we plot a –200 period moving average– based upon these samples.

[21] If this isn't immediately obvious, consider that the integral of sin(x) from 0 to 2π is zero.

Log Difference-Based Moving Averages in Time Domain



Source: CSI Data. Calculations by Newfound Research.

We can see that sampling in a different domain – in this case, the log difference space – we can generate a process that reacts dynamically in the time domain. During the calm markets of 2006 and early 2007, the –200 period moving average– behaves like the –200-day simple moving average–, whereas during the 2008 crisis it adapts to the changing price level far more quickly.

By changing the domain in which we sample, we may be able to create a model that is dynamic in the time domain, avoiding the time-dilation effects of information flow.

Conclusion

Each morning the sun rises and each evening it sets. Every year the Earth travels in orbit around the sun. What occurs during those time spans, however, varies dramatically day-by-day and year-by-year. Yet in finance – and especially quantitative finance – we often find ourselves using time as a measuring stick.

We find the notion of time almost everywhere in portfolio construction. Factors, for example, are often defined by measurements over a certain lookback horizon and reformed based upon the decay speed of the signal.

Even strategic portfolios are often rebalanced based upon the calendar. As we demonstrated in our paper *Rebalance Timing Luck: The Difference Between Hired and Fired*, fixed-schedule rebalancing can invite tremendous random impact in our portfolios.

Information does not flow into the market at a constant rate. While time may be a convenient measure, it may actually cause us to sample too frequently in some market environments and not frequently enough in others.

One answer may be to transform our measurements into a different domain. Rather than sampling price based upon the market close of each day, we might sample price based upon a fixed amount of cumulative volume, trades, or even variance. In doing so, we might find that our measures now represent a more consistent amount of information flow, despite representing a dynamic amount of data in the time domain.

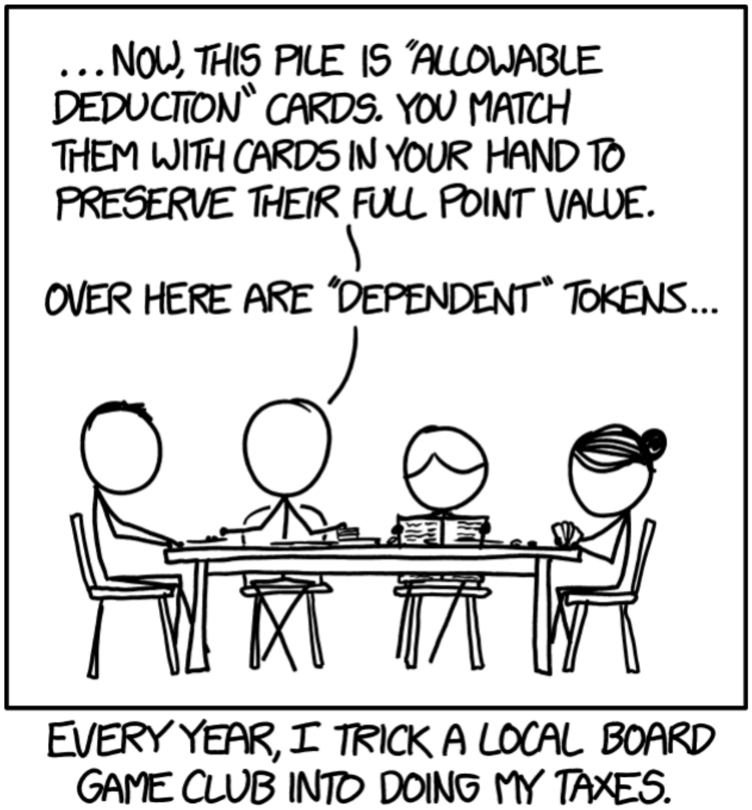
TAXES AND TREND EQUITY

April 1, 2019

SUMMARY

- Due to their highly active nature, trend following strategies are generally assumed to be tax inefficient.
- Through the lens of a simple trend equity strategy, we explore this assertion to see what the actual profile of capital gains has looked like historically.
- While a strategic allocation may only realize small capital gains at each rebalance, a trend equity strategy has a combination of large long-term capital gains interspersed with years that have either no gains or short-term capital losses.
- Adding a little craftsmanship to the trend equity strategy can potentially improve the tax profile to make it less lumpy, thereby balancing the risk of having large unrealized gains with the risk of getting a large unwanted tax bill.
- We believe that investors who expect to have higher tax rates in the future may benefit from strategies like trend equity that systematically lock in their gains more evenly through time.

Tax season for the year is quickly coming to a close, and while taxes are not a topic we cover frequently in these commentaries, it has a large impact on investor portfolios.



Source: xkcd

One of the primary reasons we do not cover it more is that it is investor-specific. Actionable insights are difficult to translate across investors without making broad assumptions about state and federal tax rates, security location (tax-exempt, tax deferred, or taxable), purchase time and holding period, losses or gains in other assets, health and family situation, etc.

Some sweeping generalizations can be made, such as that it is better to realize long-term capital gains than short-term ones, that having qualified dividends is better than having non-qualified ones, and that it is better to hold bonds in tax-deferred or tax-exempt accounts. But even these assertions are nuanced and depend on a variety of factors specific to an individual investor.

Trend equity strategies – and tactical strategies, in general – get a bad rap for being tax-inefficient. As assets are sold, capital gains are realized, often with no regard as to whether they are short-term or long-term. Wash sales are often ignored and holding periods may exclude dividends from qualifying status.

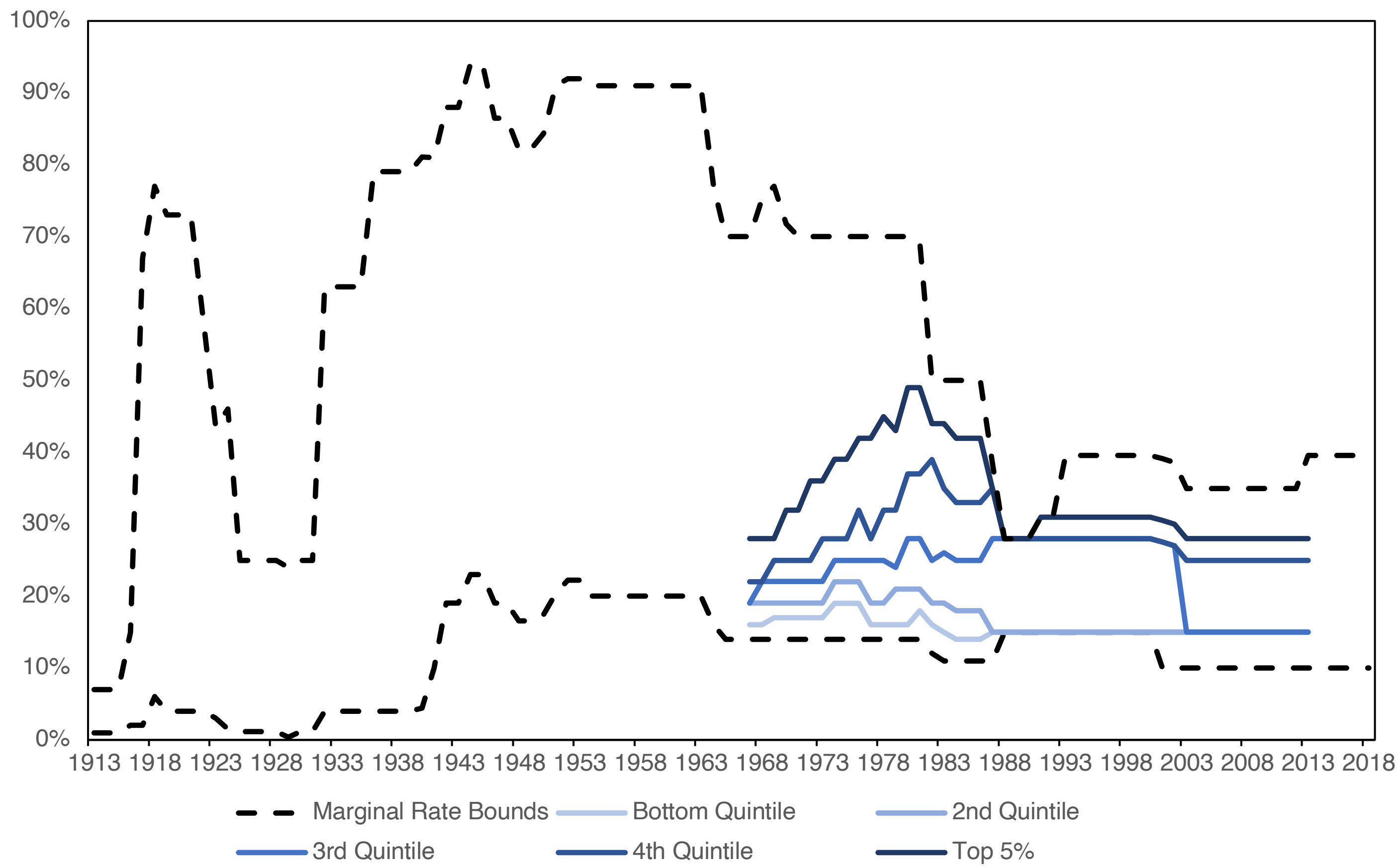
However, taxes represent yet another risk in a portfolio, and as you can likely guess if you are a frequent reader of these commentaries, reducing one risk is often done at the expense of increasing another.

December 30th, 2019

The Risk in Taxes

Tax rates have been constant for long periods of time historically, especially in recent years, but they can change very rapidly depending on the overall economic environment.

U.S. Marginal Income Tax Rates with Cutoffs Based on Income Distribution



Source: IRS, U.S. Census Bureau, and Tax Foundation. Calculations by Newfound Research. Series are limited by historical data availability.

The history shows a wide array of scenarios.

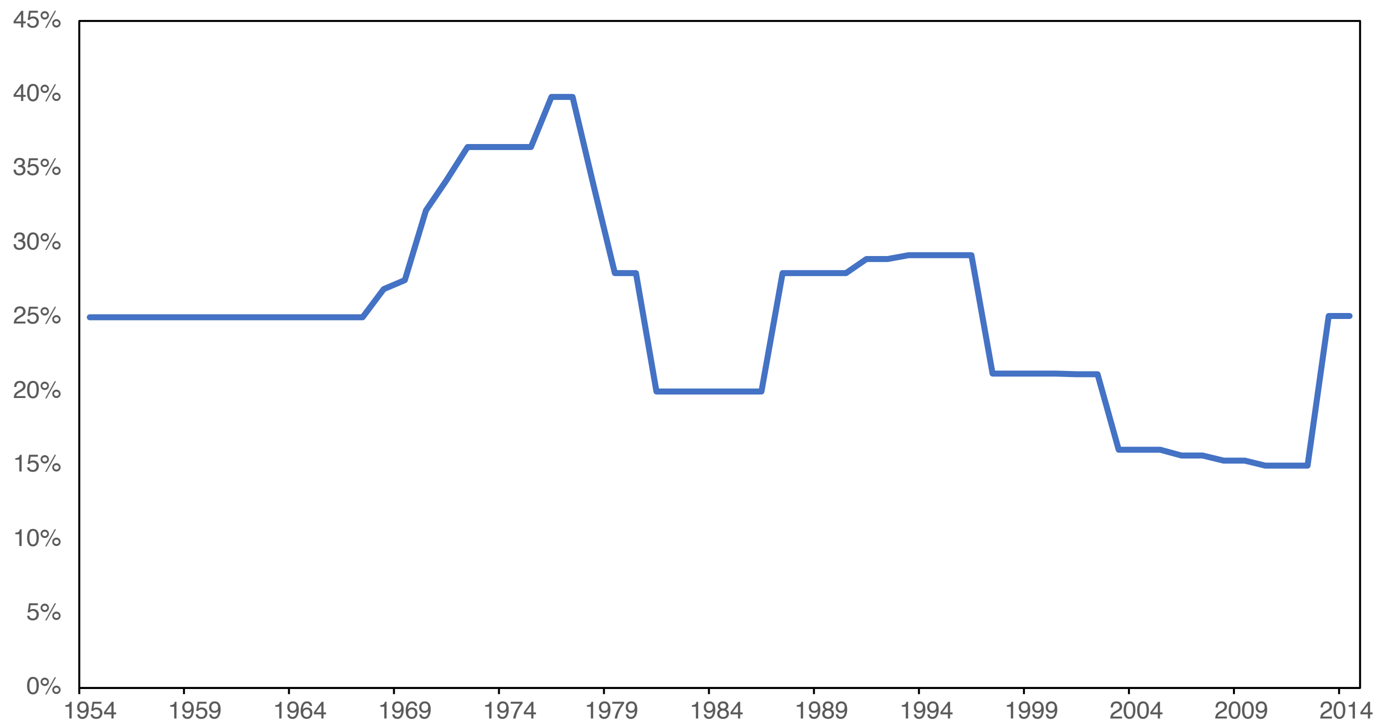
Prior to the 1980s, marginal tax rates spanned an extremely wide band, with the lowest tier near 0% and the top rate approaching 95%. However, this range has been much narrower for the past 30 years.

In the late 1980s when tax policy became much less progressive, investors could fall into only two tax brackets.

While the income quantile data history is limited, even prior to the narrowing of the marginal tax range, the bulk of individuals had marginal tax rates under 30%.

Turning to long-term capital gains rates, which apply to asset held for more than a year, we see similar changes over time.

Maximum Long-term Capital Gains Rate



Source: U.S. Department of the Treasury, Office of Tax Analysis and Tax Foundation.

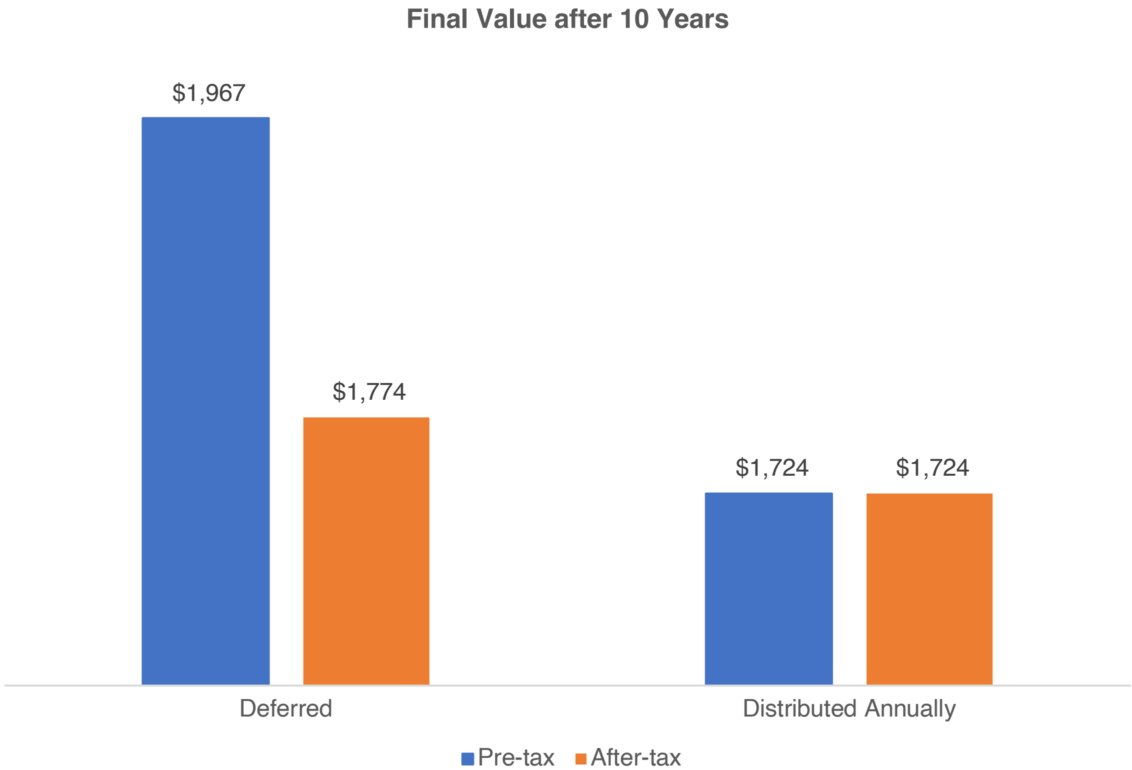
For all earners, these rates are less than their marginal rates, which is currently the tax rate applied to short-term capital gains. While there were times in the 1970s when these long-term rates topped out at 40%, the maximum rate has dipped down as low as 15%. And since the Financial Crisis in 2008, taxpayers in the lower tax brackets pay 0% on long-term capital gains.

It is these large potential shifts in tax rates that introduce risk into the tax-aware investment planning process.

To see this more concretely, consider a hypothetical investment that earns 7% every year. Somehow – how is not relevant for this example – you have the choice of having the gains distributed annually as long-term capital gains or deferred until the sale of the asset.

Which option should you choose?

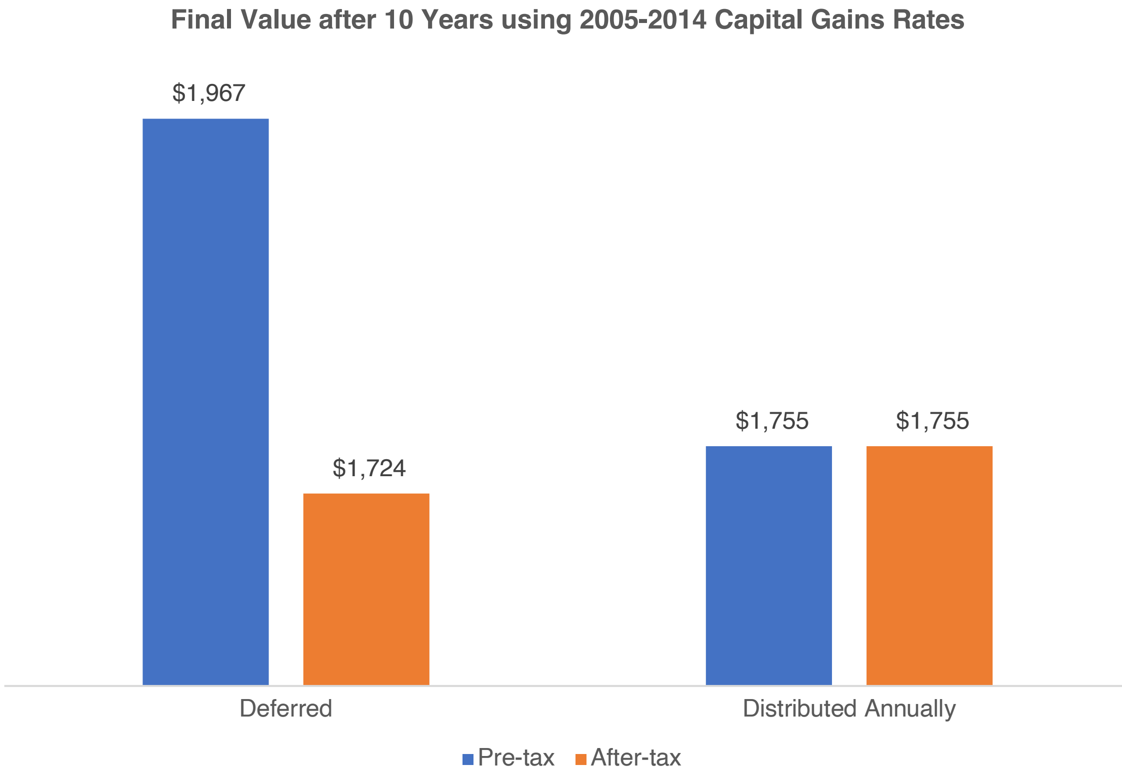
The natural choice is to have the taxes deferred until the sale of the asset. For a 10-year holding period where long-term capital gains are taxed at 20%, the pre-tax and after-tax values of a $1,000 investment are shown below.



The price return only version had a substantially higher pre-tax value as the full 7% was allowed to compound from year to year without the hinderance of an annual tax hit.

At the end of the 10-year period, the tax basis of the approach that distributed gains annually had increased up to the pre-tax amount, so it owed no additional taxes once the asset was sold. However, the approach that deferred taxes still ended up better after factoring in the tax on the embedded long-term capital gains that were realized upon the sale.

Now let's consider the same assets but this time invested from 2004 to 2014 when the maximum long-term capital gains rate jumped to 25% in 2013 after being around 15% for the first 8 years.



The pre-tax picture is still the same: the deferred approach easily beat the asset that distributed capital gains annually.

But the after-tax values have changed order. Locking in more of the return when long-term capital gains tax rates were lower was advantageous.

The difference in this case may not be that significant. But consider a more extreme – yet still realistic – example where your tax rate on the gains jumps by more than ten percentage points (e.g. due to a change in employment or family situation or tax law changes), and the decision over which type of asset you prefer is not as clear cut.

Moving beyond this simple thought experiment, we now turn to the tax impacts on trend equity strategies.

Tax Impacts in Trend Equity

We will begin with a simple trend equity strategy that buys the U.S. stock market (the ETF VTI) when it has a positive 9-month return and buys short-term U.S. Treasuries (the ETF SHV) otherwise. Prior to ETF inception, we will rely on data from the Kenneth French Data Library to extend the analysis back to the 1920s. We will evaluate the strategy monthly and, for simplicity, will treat dividends as price returns.

With taxes now in the mix, we must track the individual tax lots as the strategy trades over time based on the tactical model. For deciding which tax lots to sell, we will select the ones with the lowest tax cost, making the assumption that short-term capital gains are taxed 50% higher than long-term capital gains (approximately true for investors with tax rates of 22% and 15%, respectively, in the current tax code).

We must address the question of when an investor purchases the trend equity strategy as a long bull market with a consistent positive trend would have very different tax costs for an investor holding all the way through versus one who bought at end.

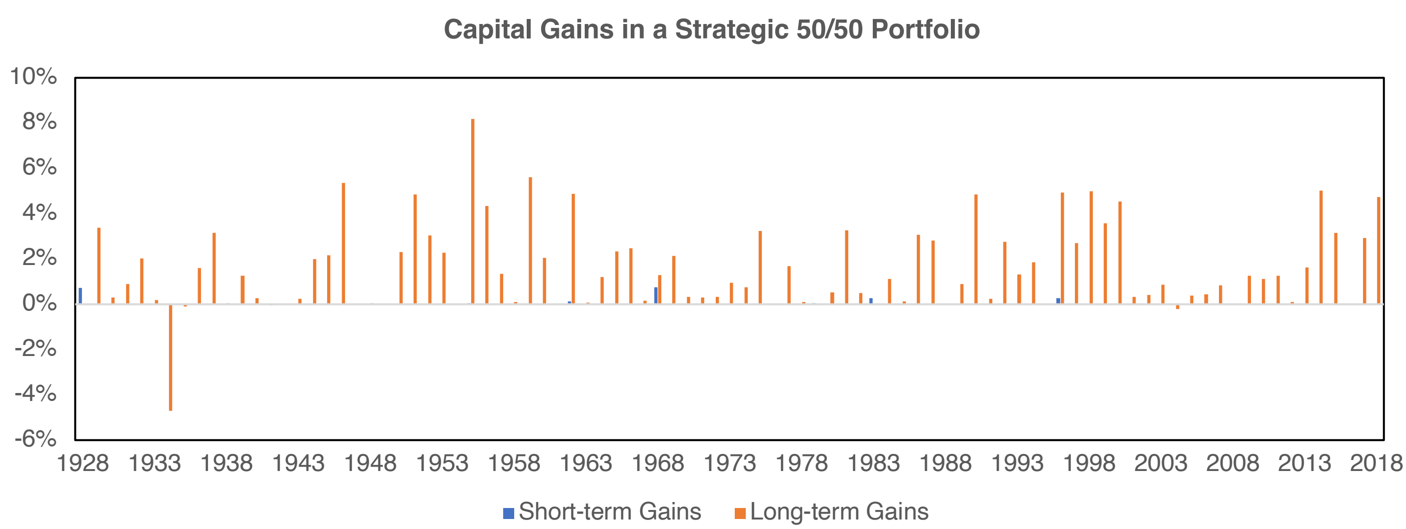
To keep the analysis as simple as possible given the already difficult specification, we will look at an investment that is made at the very beginning, assume that taxes are paid at the end of each year, and compare the average annualized pre-tax and after-tax returns. Fortunately, for this type of trend strategy that can move entirely in and out of assets, the tax memory will occasionally reset.

To set some context, first, we need a benchmark.

Obviously, if you purchased VTI and held it for the entire time, you would be sitting on some large embedded capital gains.[22]

Instead, we will use a more appropriate benchmark for trend equity: a 50%/50% blend of VTI and SHV. We will rebalance this blend annually, which will lead to some capital gains.

The following chart shows the capital gains aggregated by year as a percentage of the end of the year account value.



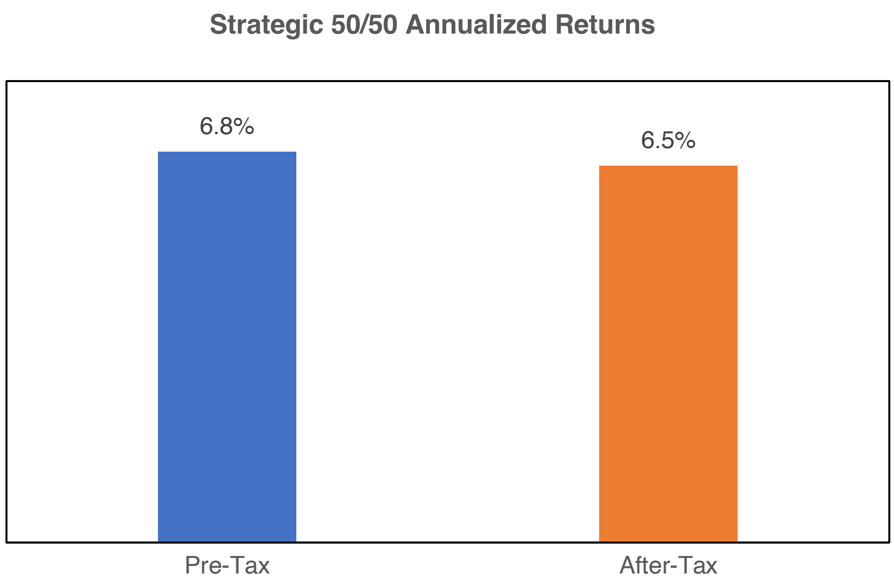
[22] You would also be in your early 90s, so congratulations!

Source: CSI Data and Kenneth French Data Library. Calculations by Newfound.

As expected with the annual rebalancing, all of the capital gains are long-term. Any short-term gains are an artifact of the rigidity of the rebalancing system where the first business day of subsequent years might be fewer than 365 days apart. In reality, you would likely incorporate some flexibility in the rebalances to ensure all long-term capital gains.

While this strategy incurs some capital gains, they are modest, with none surpassing 10%. Paying taxes on these gains is a small price to pay for maintaining a target allocation, supposing that is the primary goal.

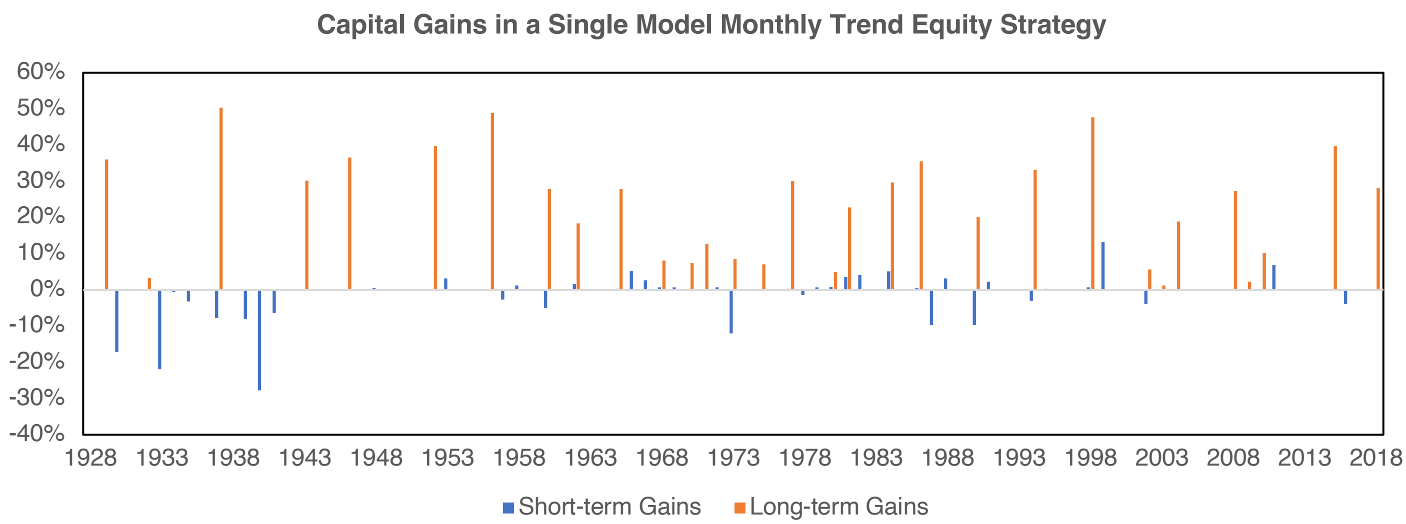
Assuming tax rates of 15% for long-term gains and 25% for short-term gains, the annualized returns of the strategic allocation pre-tax and after-tax are shown below. The difference is minor.



Source: CSI Data and Kenneth French Data Library. Calculations by Newfound.

Now on to the trend equity strategy.

The historical capital gains look very different than those of the strategic portfolio.



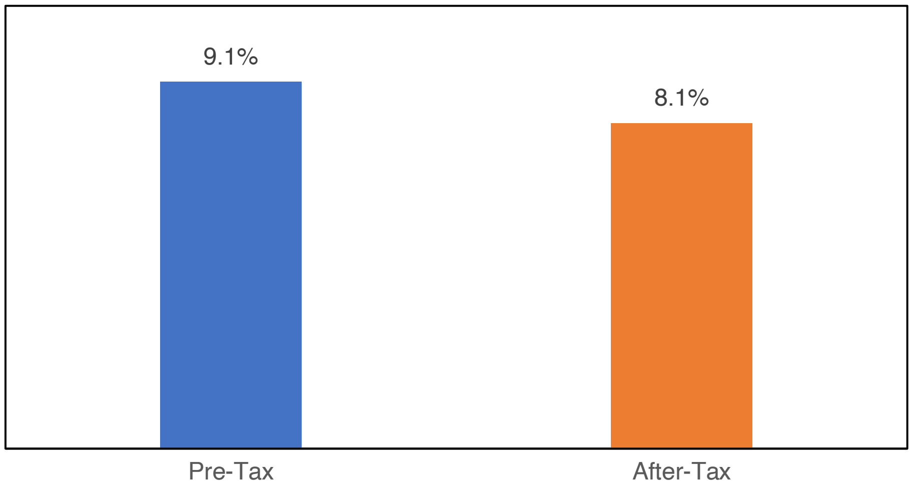
Source: CSI Data and Kenneth French Data Library. Calculations by Newfound.

In certain years, the strategy locks in long-term capital gains greater than 50%. The time between these years is interspersed with larger short-term capital losses from whipsaws or year with essentially no realized gains or losses, either short- or long-term.

In fact, 31 of the 91 years had absolute realized gains/losses of less than 1% for both short- and long-term.

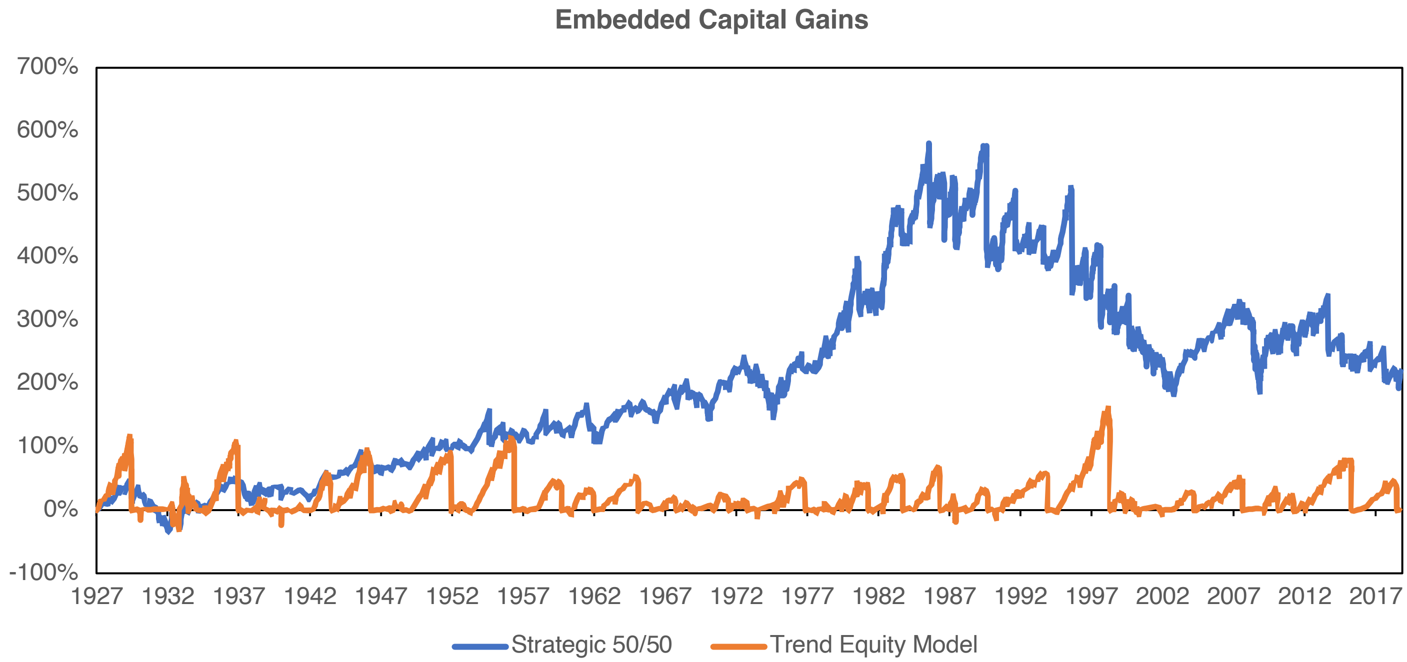
Now the difference between pre-tax and after-tax returns is 100 bps per year using the assumed tax rates (15% and 25%). This is significantly higher than with the strategic allocation.

Single Model Monthly Annualized Returns



Source: CSI Data and Kenneth French Data Library. Calculations by Newfound.

It would appear that trend equity is far less tax efficient than the strategic benchmark. As with all things taxes, however, there are nuances. As we mentioned in the first section of this commentary, tax rates can change at any time, either from a federal mandate or a change in an individual's situation. If you are stuck with a considerable unrealized capital gain, it may be too late to adjust the course.



Source: CSI Data and Kenneth French Data Library. Calculations by Newfound.

The median unrealized capital gain for the trend equity strategy is 10%. This, of course, means that you must *realize* the gains periodically and therefore pay taxes.

But if you are sitting with a 400% unrealized gain in a different strategy, behaviorally, it may be difficult to make a prudent investment decision knowing that a large tax bill will soon follow a sale. And a 10 percentage point increase in the capital gains tax rate can have a much larger impact in dollar terms on the large unrealized gain than missing out on some compounding when rates were lower.

Even so, the thought of paying taxes intermediately and missing out on compound growth can still be irksome. Some small improvement to the trend equity strategy design can prove beneficial.

Improving the Tax Profile Within Trend Equity

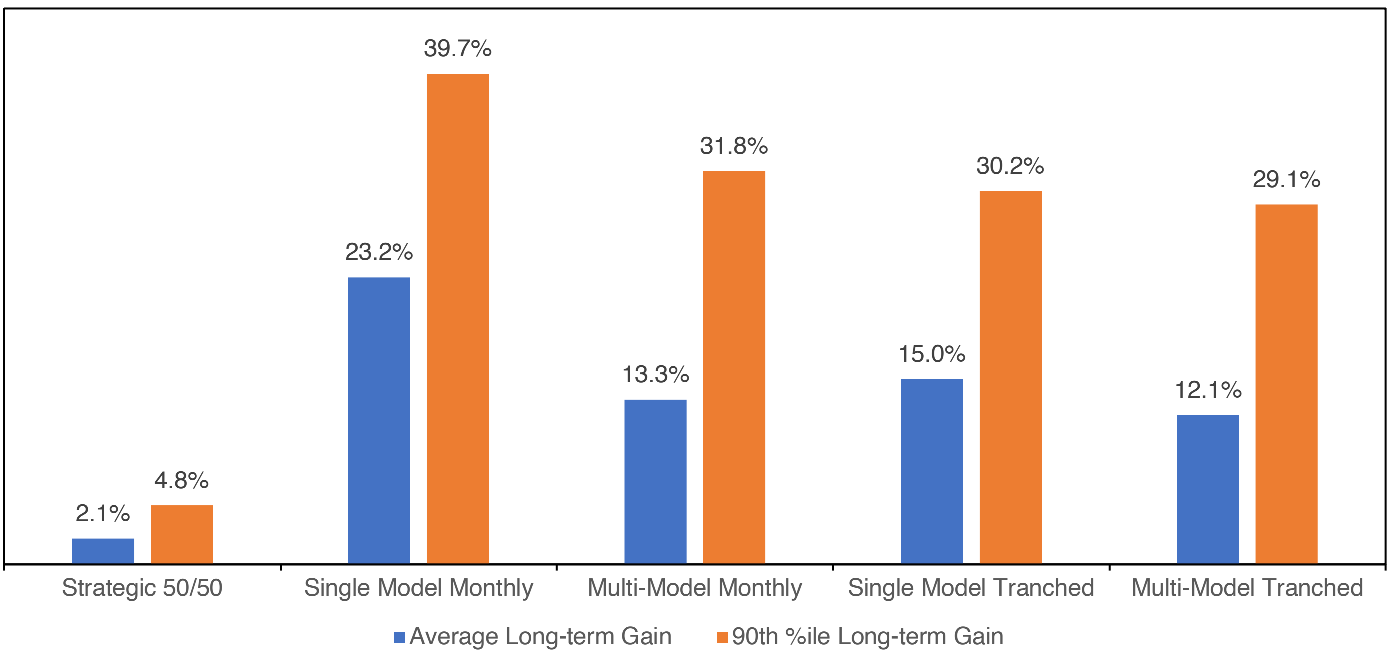
This commentary would be incomplete without a further exploration down some of the axes of diversification.

We can take the simple 9-month trend following strategy and diversify it along the "how" axis using a multi-model approach with multiple lookback periods. Specifically, we will use price versus moving average and moving average cross-overs in addition to the trailing return signal and look at windows of data ranging from 6 to 12 months.[23]

We can also diversify along the "when" axis by tranching the monthly strategy over 20 days. This has the effect of removing the luck – either good or bad – of rebalancing on a certain day of the month.

Below, we plot the characteristics of the long-term capital gains for the strategies in years in which a long-term gain was realized.

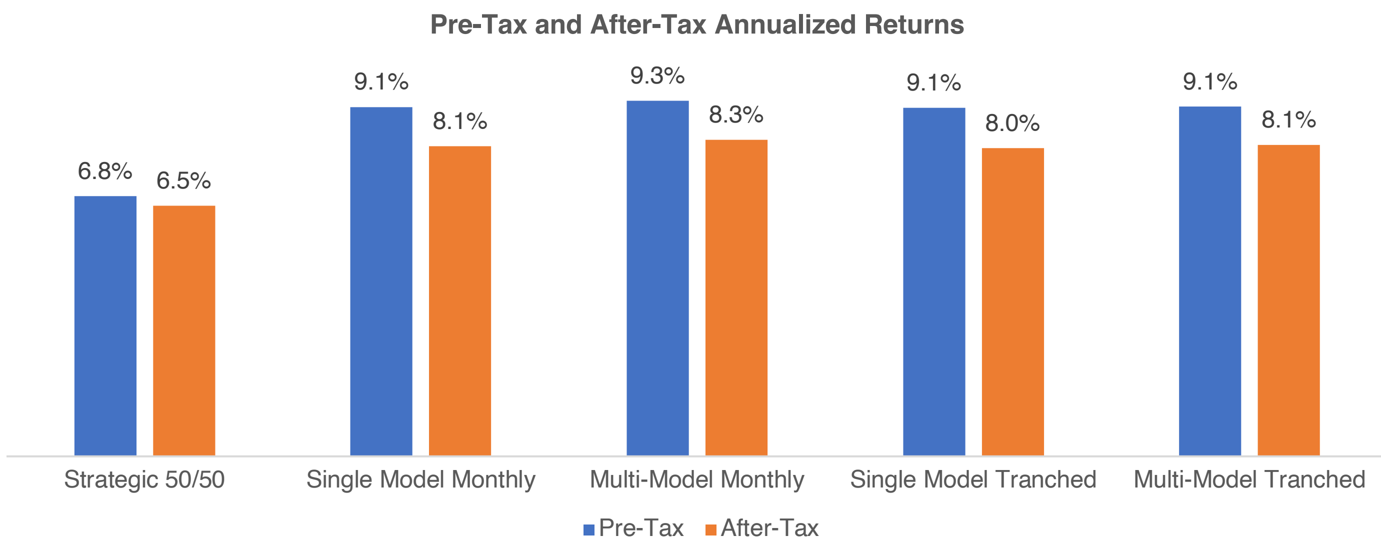
Long-term Capital Gains Characteristics for Trend Equity in Years When There are Gains



Source: CSI Data and Kenneth French Data Library. Calculations by Newfound.

The single monthly model had about a third of the years with long-term gains. Tranching it took that fraction to over a half. Moving to a multi-model approach brought the fraction to 60%, and tranching *that* upped it to 2 out of every 3 years.

[23] This is an ensemble approach. For a thorough discussion of this approach, albeit in the context of a different strategy, see ReSolve Asset Management's Paper entitled, Global Equity Momentum: A Craftsman's Perspective.



Source: CSI Data and Kenneth French Data Library. Calculations by Newfound.

From an annualized return perspective, all of these trend equity strategies exhibited similar return differentials between pre-tax and after-tax.

In previous commentaries, we have illustrated how tranching to remove timing luck and utilizing multiple trend following models can remove the potential dispersion in realized terminal wealth. However, in the case of taxes, these embellishments did not yield a reduction in the tax gap.

While these improvements to trend equity strategies reduce specification-based whipsaw, they often result in similar allocations for large periods of time, especially since these strategies only utilize a single asset.

But to assume that simplicity trumps complexity just because the return differentials are not improved misses the point.[24]

With similar returns among within the trend-following strategies, using an approach that realizes more long-term capital gains could still be beneficial from a tax perspective.

In essence, this can be thought of as akin to dollar-cost averaging.

Dollar-cost averaging to invest a lump sum of capital is often not optimal if the sole goal is to generate the highest return.[25] However, it is often beneficial in that it reduces the risk of bad outcomes (i.e. tail events).

Having a strategy – like trend equity – that has the potential to generate strong returns while taking some of those returns as long-term capital gains can be a good hedge against rising tax rates. And having a diversified trend equity strategy that can realize these capital gains in a smoother fashion is icing on the cake.

---

24 See our commentary entitled When Simplicity Met Fragility.
25 See our commentary entitled Should You Dollar-Cost Average?

Conclusion

Taxes are a tricky subject, especially from the asset manager's perspective. How do you design a strategy that suits all tax needs of its investors?

Rather than trying to develop a one-size-fits-all strategy, we believe that a better approach to the tax question is education. By more thoroughly understanding the tax profile of a strategy, investors can more comfortably deploy it appropriately in their portfolios.

As highly active strategies, trend equity mandates are generally assumed to be highly tax-inefficient. We believe it is more meaningful to represent the tax characteristics an exchange of risks: capital gains are locked in at the current tax rates (most often long-term) while unrealized capital gains are kept below a reasonable level. These strategies have also historically exhibited occasional periods with short-term capital losses.

These strategies can benefit investors who expect to have higher tax rates in the future without the option of having a way to mitigate this risk otherwise (e.g. a large tax-deferred account like a cash balance plan, donations to charity, a step-up in cost basis, etc.).

Of course, the question about the interplay between tax rates and asset returns, which was ignored in this analysis, remains. But in an uncertain future, the best course of investment action is often the one that diversifies away as much uncompensated risk as possible and includes a comprehensive plan for risk management.

REVISITING THE WEIRD PORTFOLIO

April 8, 2019

SUMMARY

- A few years ago, we blindly applied mean-variance optimization to a set of capital market assumptions, and The Weird Portfolio was born.
- This portfolio is weird because it does not look like typical investor portfolios since it tilts heavily toward credit-based and alternative asset classes.
- Despite having weird allocations, the portfolio actually performed in line with the iShares Core Moderate Allocation ETF (AOM), which has entirely different holdings.
- By decomposing the investment universe of the strategies into their underlying independent risk factors, we explore how even different allocations can lead to closely shared risks.
- While mathematical exercise has its limitations, we believe that investors can learn a great deal about their portfolio performance by looking at the performance of the underlying risk factors and the portfolio's exposure to those factors.

Back in August 2017, I joined Meb Faber on his podcast for a wide-ranging conversation on the market outlook, tactical asset allocation, and a variety of research commentaries we had recently written.

Early in the conversation, we discussed *Portfolios in Wonderland*, a presentation I had put together that highlighted the unique nature of the current market outlook versus prior periods, its implications for financial planning, and ideas for asset allocation that might increase an investor's odds of achieving retirement success.

In particular, we highlighted JP Morgan's capital market assumptions and blindly ran them through a mean-variance optimization. Unlike us, the optimizer has no attachment to a particular asset class and blindly seeks to maximize return for a stated risk target.

The resulting portfolio was could only be described as *weird.*

And thus The Weird Portfolio was born.

The idea behind The Weird Portfolio is to generate a portfolio that blindly follows the recommendations of a portfolio optimizer over time, with no ad hoc constraints or regard for tracking error. The ultimate question being, "if we did not care about the optics of our portfolio and had full confidence in our capital market assumptions, how might we invest?"

To achieve this goal, we employ intermediate-term (7-to-10 year) capital market assumptions from JP Morgan, BlackRock, and BNY Mellon.

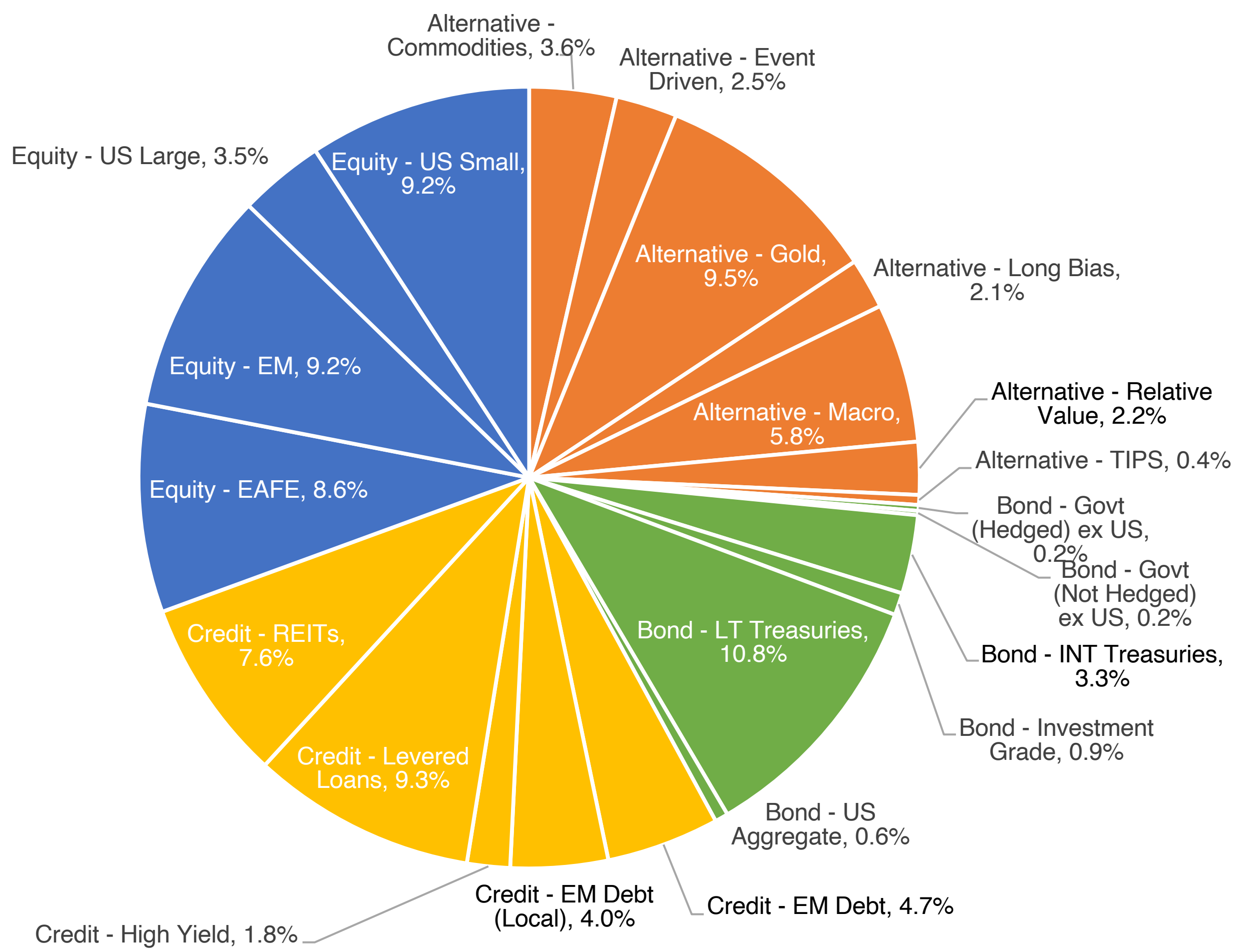
For each set of capital market assumptions, we run several thousand 10-year return simulations, generating random monthly return samples using the provided expected return and covariance matrices. For a bit of spice, we also randomly shock the covariance matrix to appropriately reflect the time-varying, crashing nature of cross-asset correlations.

For each simulation generated, we construct the portfolio that maximizes return while having the same realized volatility profile as a 60/40 global equity / U.S. aggregate bond portfolio.

After generating the thousands of different portfolios, we average them together to create our final allocation profile.

It should be noted that this sort of stochastic optimization has a particular weakness. Namely, because the lowest possible weight is 0%, there is a potential upward bias in recommended allocations for small positions and a symmetric downward bias in larger positions.

Using this methodology, the target allocations for Q2 2019 are shown below.



By almost any standard, the allocations look rather odd. We can see almost no allocation to U.S. large-cap equities, a meaningful overweight towards credit-based assets, and large allocations to long-term U.S. Treasuries and gold.

While a couple years of data is not particularly meaningful, we thought it would be interesting to evaluate the performance of The Weird Portfolio to see if any insights could be gleaned. To track the performance, we allocate to low-cost ETFs that track each asset class. For alternatives, we utilize category index data from HFRI.

Below we plot the growth of $1 in the –Vanguard Balanced Index Fund (VBINX)–, the –iShares Core Moderate Allocation ETF (AOM)–, and –The Weird Portfolio–.

Source: CSI Data and HFRI. Calculations by Newfound Research. Returns for The Weird Portfolio are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees except for underlying ETF expense ratios. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

Well, this is a head scratcher. For all its *weirdness*, the performance of The Weird Portfolio really is not all that weird at all. What gives?

After all, the –iShares Core Moderate Allocation ETF (AOM)– has no exposure to gold or alternatives, its exposure to U.S. large-cap is 5.5x larger and credit exposures are largely non-existent.

We'll attack this question two ways. The first is philosophically and the second is quantitatively.

From a philosophical perspective, we'll let Harley Bassman do all the talking:

> *"In a nutshell, all financial risk vectors are related. The shape of the Yield Curve, the level of Credit Spreads, the correlation of various points on the yield curve and the level of Implied Volatility should all move in tandem since the risk premium embedded in the Duration, Credit, and Convexity risk vectors should correlate in some grand manner.*
>
> *...*
>
> *A simpler explanation may be that the net carry (profit) across risk categories should equilibrate as 'alpha seekers' allocate capital across the various risky assets in search of excess return."*
>
> *--Harley Bassman, "A Guide for the Perplexed"*

From a quantitative perspective, we can try to distill this by not looking at allocations to asset classes, but by looking at allocations to *eigen portfolios.*

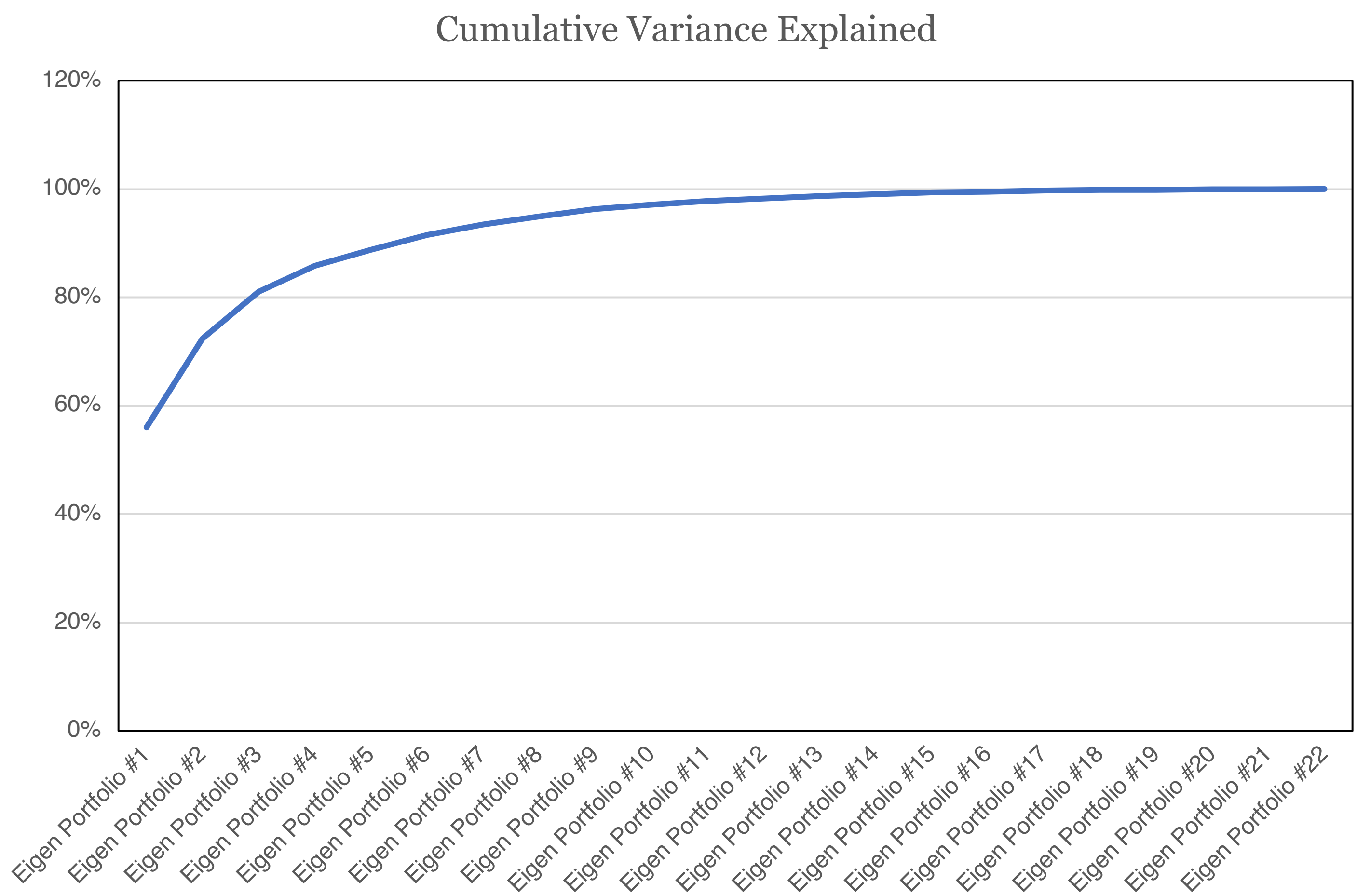
What, pray tell, is an *eigen portfolio*?

The simple answer is that eigen portfolios give us a way of looking at our investment universe through the lens of statistical factors, not assets. The factors, in this case, are each created as a portfolio of the investments in our universe. Through a special mathematical process, we can ensure that the factors all have zero correlation to one another.

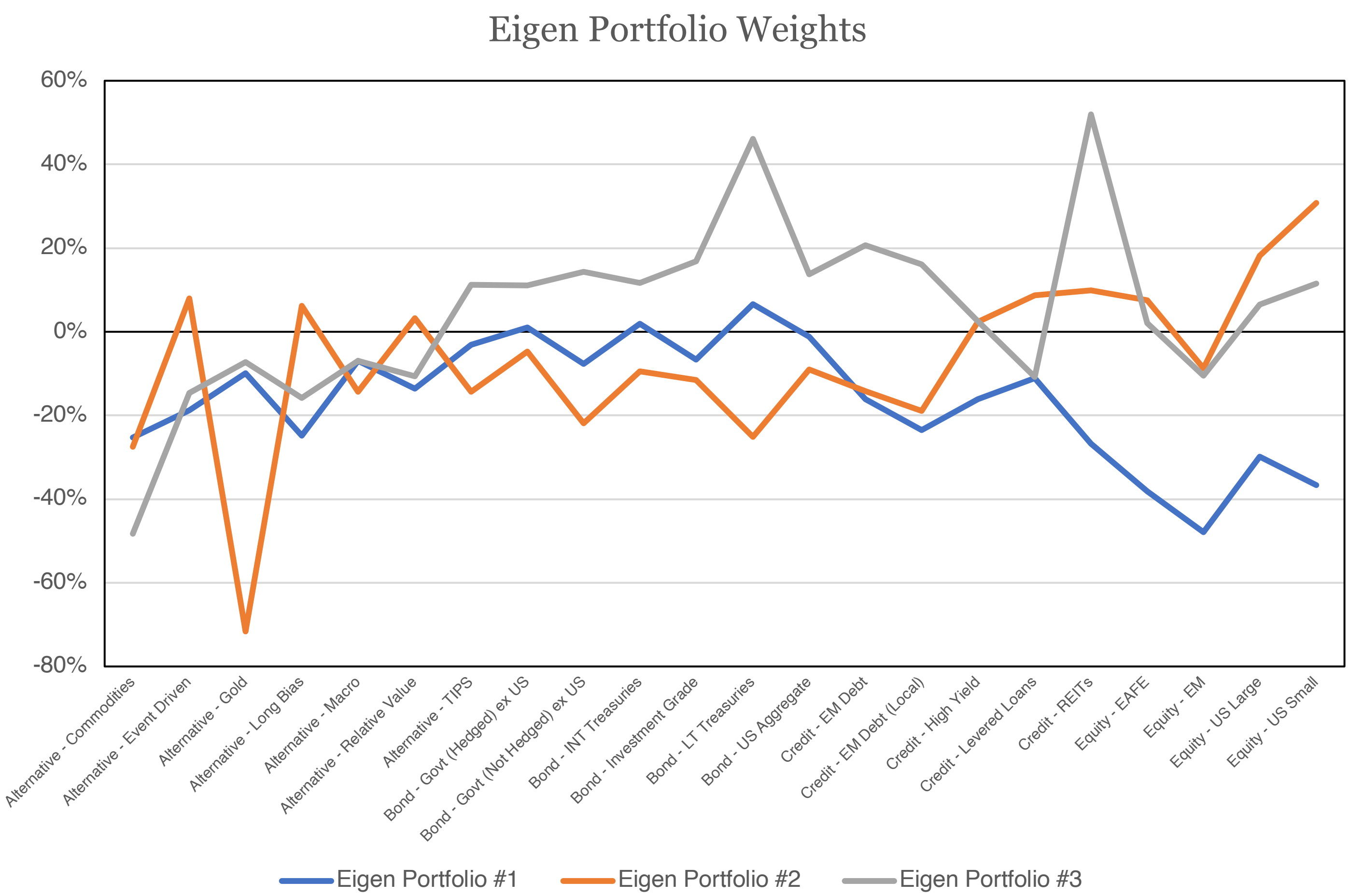
Here is a more nuanced, quantitative answer. We can think of our NxN covariance matrix as an N-dimensional cloud of points. By performing an eigendecomposition, we can factorize the matrix into N linearly independent vectors. As each of these vectors will have a loading on each asset, we can think of them as portfolios (hence *eigen portfolios*). Furthermore, since the vectors are linearly independent, we know that they will have zero correlation to one another. Therefore, we can think of them as a set of basis portfolios that describe our investment universe.

Generally, we sort eigen portfolios by the proportion of variance they explain of our data. The first eigen portfolio will explain the most variance, the second the second most, et cetera. In most cases, there is a steep drop-off after the first several eigen portfolios, with 95% of the total variance being explained by just a handful of eigen portfolios in most asset class universes. These are often interpreted as the driving risk factors such as equity risk, interest rates, etc.

Note that with an N assets in our universe, there will always be N eigen portfolios. From a statistical perspective, not all of these eigen portfolios will necessarily be significant. Random matrix theory provides a number of ideas on identifying which eigen portfolios are meaningful and which are not, but we will leave that to another article on another day.



Let's take a look at the weights of the first three eigen portfolios, which combined explained 80% of the cumulative variance.



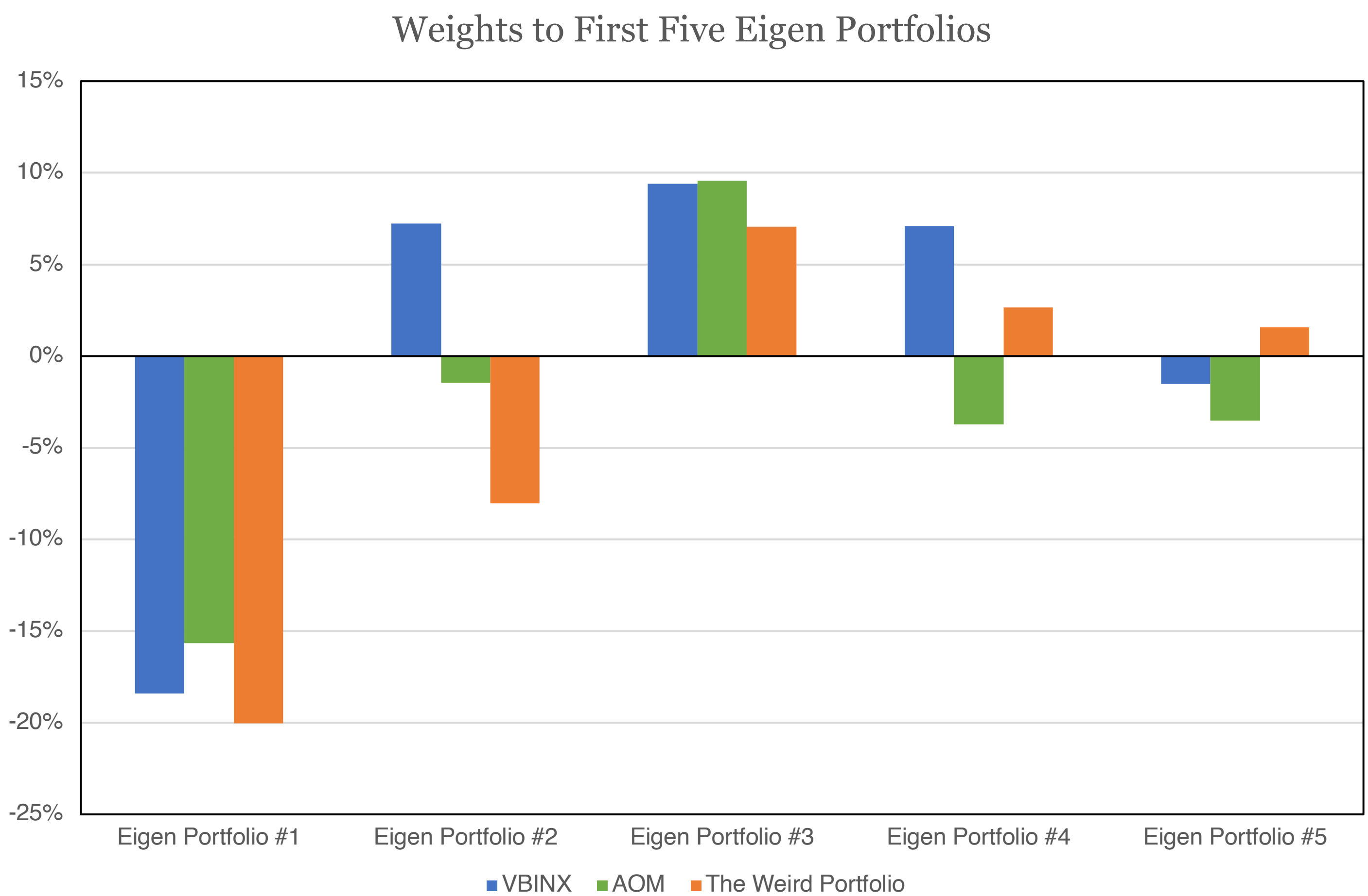
While eigen portfolios are mathematically derived, it is sometimes possible to ascribe some economic intuition. For example, –Eigen Portfolio #1– has significant negative loadings on equities, credit exposures, and some alternatives. We might therefore interpret this as an equity beta factor. –Eigen Portfolio #2–, on the other hand, has slightly negative rate exposure, a significant positive loading on small-caps and a negative loading on gold. Finally, –Eigen Portfolio #3– has a significant loading on bonds and REITs, indicating that it is likely a duration factor of sorts.

Because these are portfolios, we can even go so far as to create their historical performance!

Note how –Eigen Portfolio #1– appears to be almost the mirror image of equity markets over the last 21 months. This makes sense, as that portfolio had significant *negative* weights towards equities.

Knowing that our eigen portfolios are independent from one another allows us to take one last step: we can examine our portfolios not as a set of asset classes, but rather as their implicit exposure to these eigen portfolios.

If we take this step for the –Vanguard Balanced Index Fund (VBINX)–, the –iShares Core Moderate Allocation ETF (AOM)– , and –The Weird Portfolio–, we see a very interesting result. Below we plot the loadings for the first five eigen portfolios.



We see that all three portfolios have significant negative weights on the first eigen portfolio (i.e. have *positive* exposure to equities) and positive weights to the third eigen portfolio. The biggest difference appears to be in the 2nd eigen portfolio, where the –Vanguard Balanced Index Fund (VBINX)– retains a positive loading, the –iShares Core Moderate Allocation ETF (AOM)– has almost no loading, and –The Weird Portfolio– has a negative loading.

Unfortunately, the Eigen Portfolio #2 was one of the more difficult to interpret due to the significant weight spread between small-caps and gold, but it clearly provides a wedge between the –Vanguard Balanced Index Fund (VBINX)– and –The Weird Portfolio–. This may seem somewhat odd, as it implies a rather significant negative gold bias in the –Vanguard Balanced Index Fund (VBINX)–, but this bias is largely neutralized by the fourth eigen portfolio (not plotted) which has a heavy positive loading on gold, REITs, and U.S. small-caps with a large negative loading on emerging market equity. Taken together, these portfolios net out to negative emerging market exposure, which makes more economic sense since –Vanguard Balanced Index Fund (VBINX)– does not have any EM equity while the other two strategies do.

Conclusion

What is the takeaway to all of this analysis? We believe there are a handful worth noting.

First of all, extracting performance guidance from an asset allocation can be misleading. What looks weird on its face may exhibit little-to-no meaningful performance difference from more traditionally allocated portfolios in most environments. This is because what matters more than the asset allocation profile is the underlying risk factor profile. As we demonstrated in this commentary, while holdings may look meaningfully different, the embedded risk factors can still be quite similar.

Secondly, we should recognize that market environments can go through periods where certain risk factors are more important than others. As a naïve example, consider that stocks and bonds have largely exhibited negative correlations over the last several decades as economic growth shocks dominated market risk pricing. In prior decades, however, when inflation shocks were top of mine, stocks and bonds exhibited positive correlations. Over the time horizon evaluated, equity growth appears to be the dominant statistical risk factor, but that need not necessarily be the case going forward. The difference between the exposure of a standard 60/40 and The Weird Portfolio to the second eigen portfolio may manifest more significantly in the future.

Finally, we must acknowledge that 21 months is not a particularly meaningful horizon (despite this particular sample exhibiting some rather tumultuous periods), either from an economic regime perspective or a statistical perspective. Trying to draw too many conclusions from this small slice of history may be entirely misleading going forward.

The Path-Dependent Nature of Perfect Withdrawal Rates

April 22, 2019

Summary

- The Perfect Withdrawal Rate (*PWR*) is the rate of regular portfolio withdrawals that leads to a zero balance over a given time frame.
- 4% is the commonly accepted lower bound for safe withdrawal rates, but this is only based on one realization of history and the actual risk investors take on by using this number may be uncertain.
- Using simulation techniques, we aim to explore how different assumptions match the historical experience of retirement portfolios.
- We find that simple assumptions commonly used in financial planning Monte Carlo simulations do not seem to reflect as much variation as we have seen in the historical *PWR*.
- Including more stress testing and utilizing richer simulation methods may be necessary to successfully gauge that risk in a proposed *PWR*, especially as it pertains to the risk of failure in the financial plan.

Financial planning for retirement is a combination of art and science. The problem is highly multidimensional, requiring estimates of cash flows, investment returns and risk, taxation, life events, and behavioral effects. Reduction along the dimensions can simplify the analysis, but introduces consequences in the applicability and interpretation of the results. This is especially true for investors who are close to the line between success and failure.

One of the primary simplifying assumptions is the 4% rule. This heuristic was derived using worst-case historical data for portfolio withdrawals under a set of assumptions, such as constant inflation adjusted withdrawals, a fixed mix of stock and bonds, and a set time horizon.

Below we construct a monthly-rebalanced, fixed-mix 60/40 portfolio using the S&P 500 index for U.S. equities and the Dow Jones Corporate Bond index for U.S. bonds. Using historical data from 12/31/1940 through 12/31/2018, we can evaluate the margin for error the 4% rule has historically provided and how much opportunity for higher withdrawal rates was sacrificed in "better" market environments.

Historical Perfect Withdrawal Rate

Source: Global Financial Data and Shiller Data Library. Calculations by Newfound Research. Returns are backtested and hypothetical. Past performance is not a guarantee of future results. Returns are gross of all fees. Returns assume the reinvestment of all distributions. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

But the history is only a single realization of the world. Risk is hard to gauge.

Perfect Withdrawal Rates

The formula (in plain English) for the perfect withdrawal rate ("*PWR"*) in a portfolio, assuming an ending value of zero, is relatively simple since it is just a function of portfolio returns:

$$PWR = \frac{Portfolio\,Value}{Sequence\,Risk}$$

The portfolio value in the numerator is the final value of the portfolio over the entire period, assuming no withdrawals. The sequence risk in the denominator is a term that accounts for both the order and magnitude of the returns.

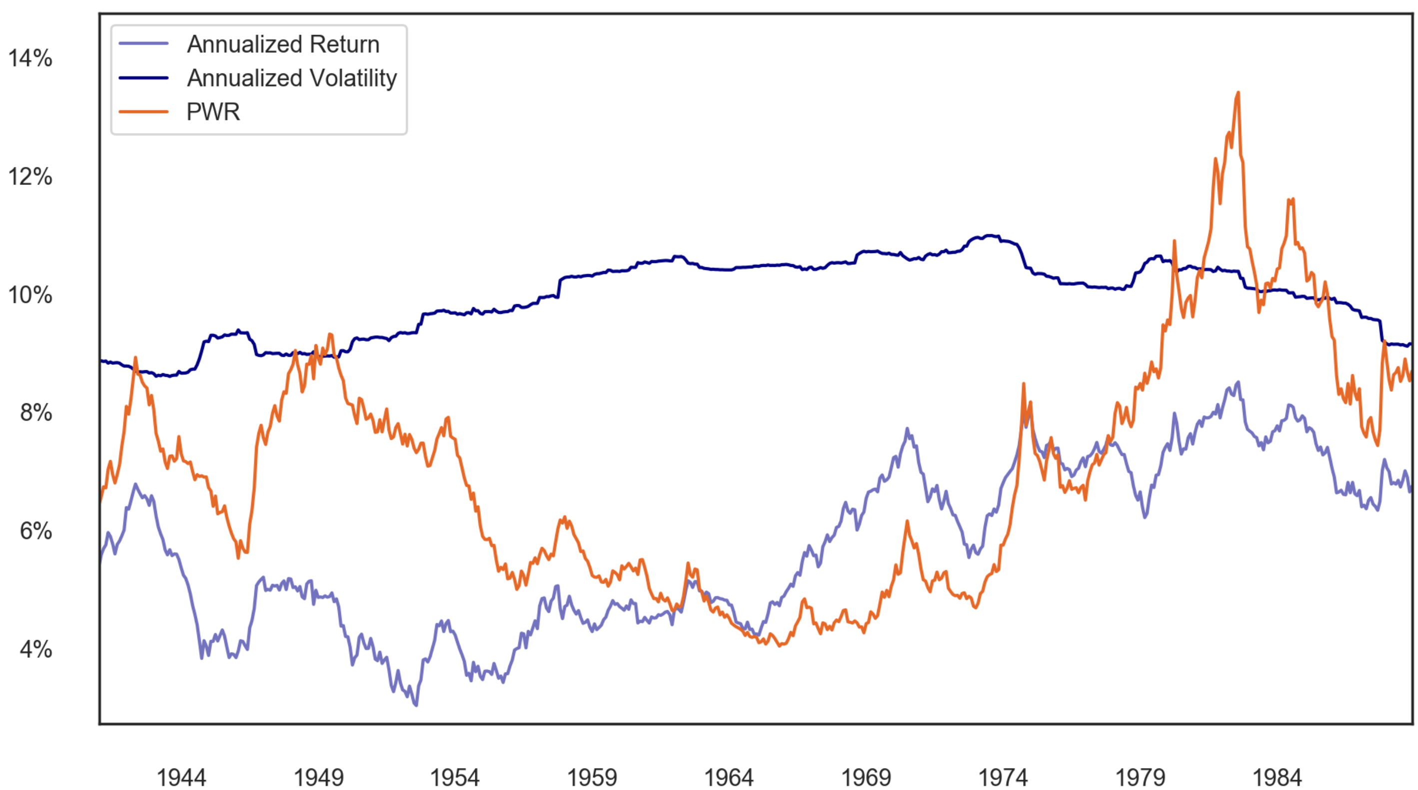
Larger negative returns earlier on in the period increase the sequence risk term and therefore reduce the *PWR*.

From a calculation perspective, the final portfolio value in the equation is typically described (e.g. when using Monte Carlo techniques) as a log-normal random variable, i.e. the log-returns of the portfolio are assumed to be normally distributed. This type of random variable lends itself well to analytic solutions that do not require numerical simulations.

The sequence risk term, however, is not so friendly to closed-form methods. The path-dependent, additive structure of returns within the sequence risk term means that we must rely on numerical simulations.

To get a feel for some features of this equation, we can look at the *PWR* in the context of the historical portfolio return and volatility.

Perfect Withdrawal Rate, Return, and Volatility



Source: Global Financial Data and Shiller Data Library. Calculations by Newfound Research. Returns are backtested and hypothetical. Past performance is not a guarantee of future results. Returns are gross of all fees. Returns assume the reinvestment of all distributions. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

The relationship is difficult to pin down.

As we saw in the equation shown before, the –annualized return of the portfolio– does appear to impact the –PWR– (correlation of 0.51), but there are periods (e.g. those starting in the 1940s) that had higher PWRs with lower returns than in the 1960s. Therefore, investors beginning withdrawals in the 1960s must have had higher sequence risk.

Correlation between –annualized volatility– and –PWR– was slightly negative (-0.35).

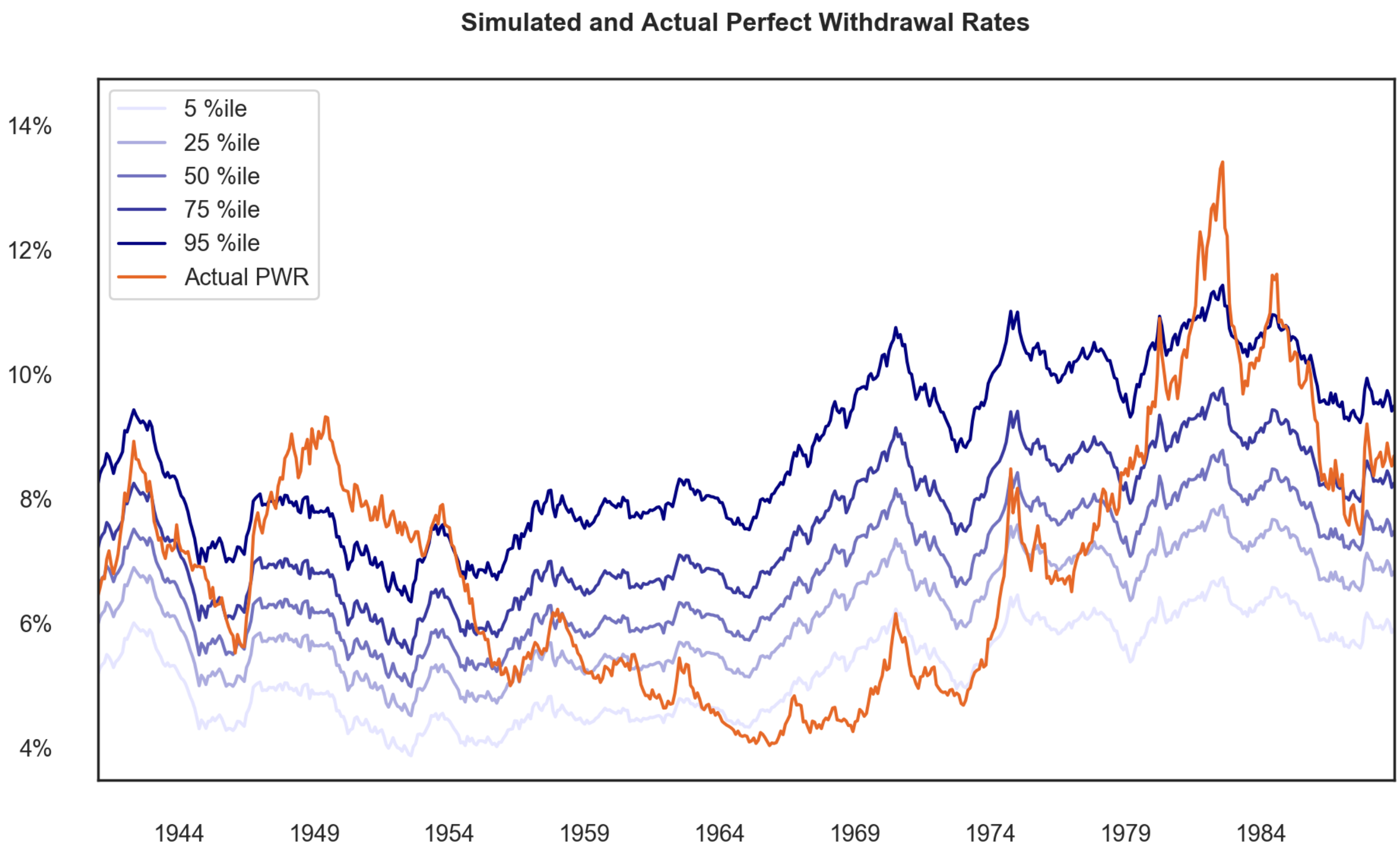
The Risk in Withdrawal Rates

Since our goal is to assess the risk in the historical *PWR* with a focus on the sequence risk, we will use the technique of Brownian Bridges to match the return of all simulation paths to the historical return of the 60/40 portfolio over rolling 30-year periods. We will use the historical full-period volatility of the portfolio over the period for the simulation.

This is essentially a conditional *PWR* risk based on assuming we know the full-period return of the path beforehand.

To more explicitly describe the process, consider a given 30-year period. We begin by computing the full-period annualized return and volatility of the 60/40 portfolio over that period. We will then generate 10,000 simulations over this 30-year period but using the Brownian Bridge technique to ensure that all of the simulations have the exact same full-period annualized return and intrinsic volatility. In essence, this approach allows us to vary the path of portfolio returns without altering the final return. As PWR is a path-dependent metric, we should gain insight into the distribution of PWRs.

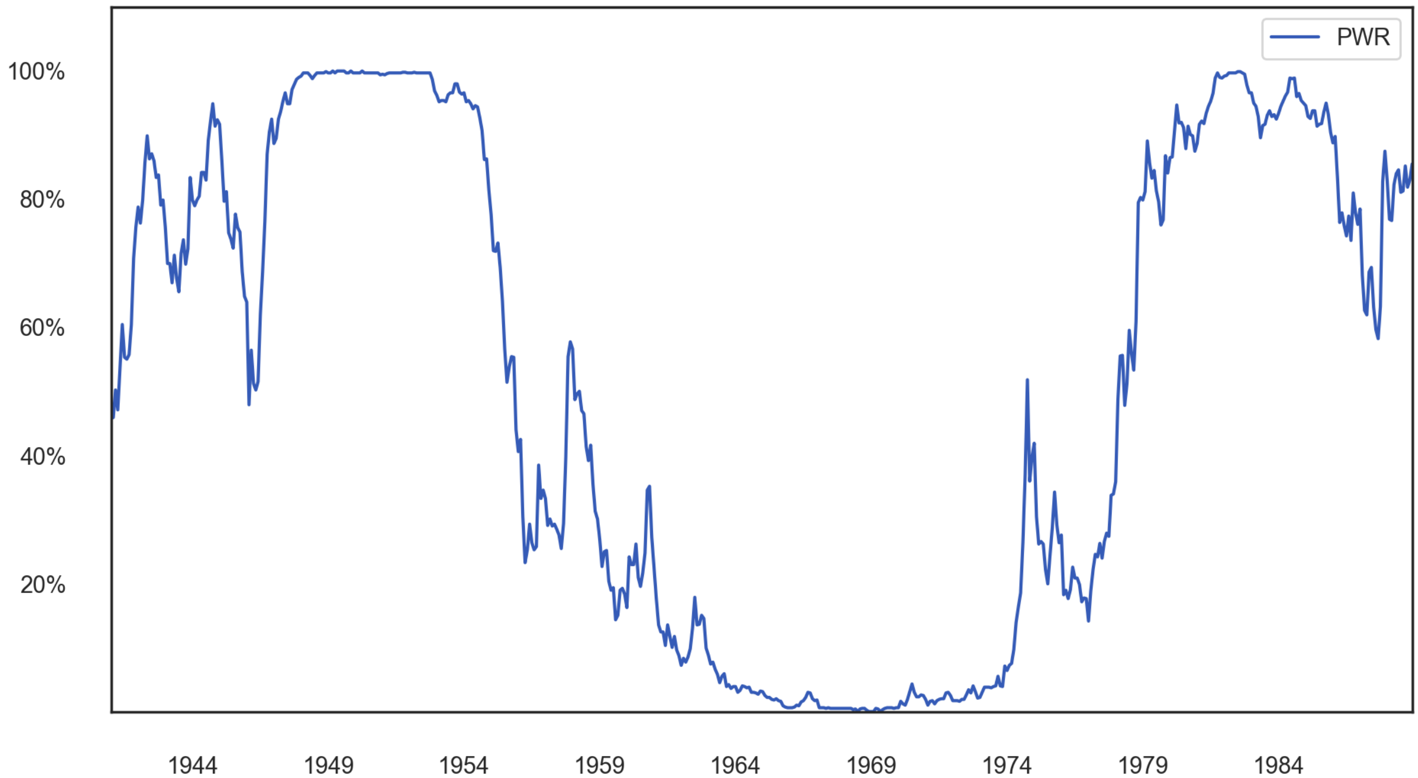
The percentile bands for the simulations using this method are shown below with the actual PWR in each period overlaid.



Source: Global Financial Data and Shiller Data Library. Calculations by Newfound Research. Returns are backtested and hypothetical. Past performance is not a guarantee of future results. Returns are gross of all fees. Returns assume the reinvestment of all distributions. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

From this chart, we see two items of note: The percentile bands in the distribution roughly track the historical return over each of the periods, and the actual PWR fluctuates into the left and right tails of the distribution rather frequently. Below we plot where the *actual* PWR actually falls within the simulated PWR distribution.

Perfect Withdrawal Rate Percentile



Source: Global Financial Data and Shiller Data Library. Calculations by Newfound Research. Returns are backtested and hypothetical. Past performance is not a guarantee of future results. Returns are gross of all fees. Returns assume the reinvestment of all distributions. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

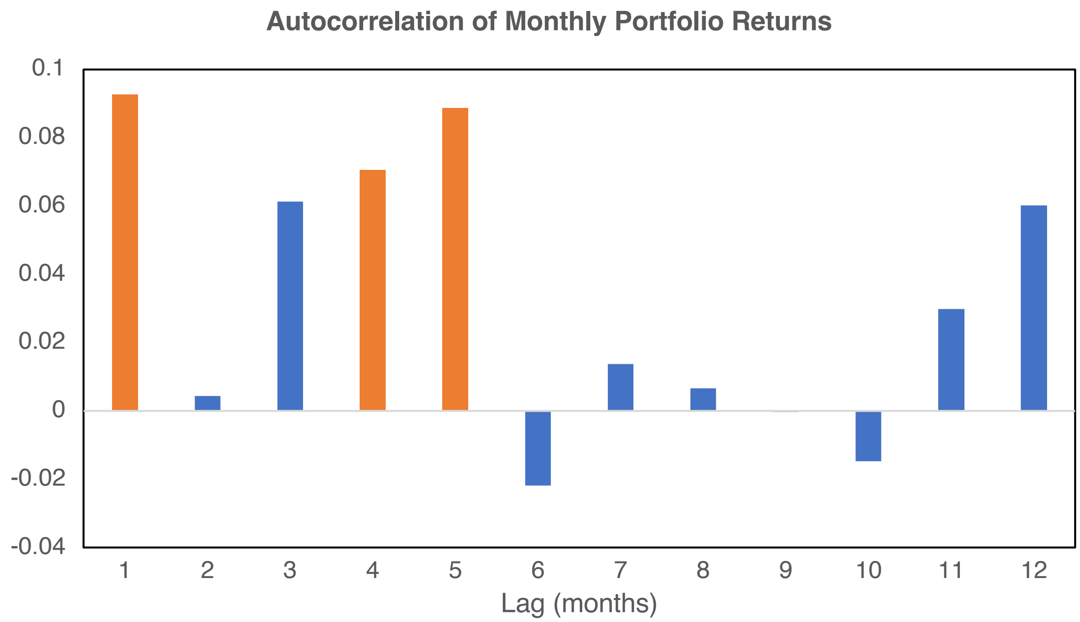
The actual PWR is below the 5th percentile 12% of the time, below the 1st percentile 4% of the time, above the 95th percentile 11% of the time, and above the 99th percentile 7% of the time. Had our model been more well calibrated, we would expect the percentiles to align; e.g. the PWR should be below the 5th percentile 5% of the time and above the 99th percentile 1% of the time.

This seems odd until we realize that our model for the portfolio returns was likely too simplistic. We are assuming Geometric Brownian Motion for the returns. And while we are fixing the return over the entire simulation path to match that of the actual portfolio, the path to get there is assumed to have constant volatility and independent returns from one month to the next.

In reality, returns do not always follow these rules. For example, the skew of the monthly returns over the entire history is -0.36 and the excess kurtosis is 1.30. This tendency toward larger magnitude returns and returns that are skewed to the left can obscure some of the risk that is inherent in the PWRs.

Additionally, returns are not totally independent. While this is good for trend following strategies, it can lead to an understatement of risk as we explored in our previous commentary on Accounting for Autocorrelation in Assessing Drawdown Risk.

Over the full period, monthly returns of lags 1, 4, and 5 exhibit autocorrelation that is significant at the 95% confidence level.



Source: Global Financial Data and Shiller Data Library. Calculations by Newfound Research. Returns are backtested and hypothetical. Past performance is not a guarantee of future results. Returns are gross of all fees. Returns assume the reinvestment of all distributions. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

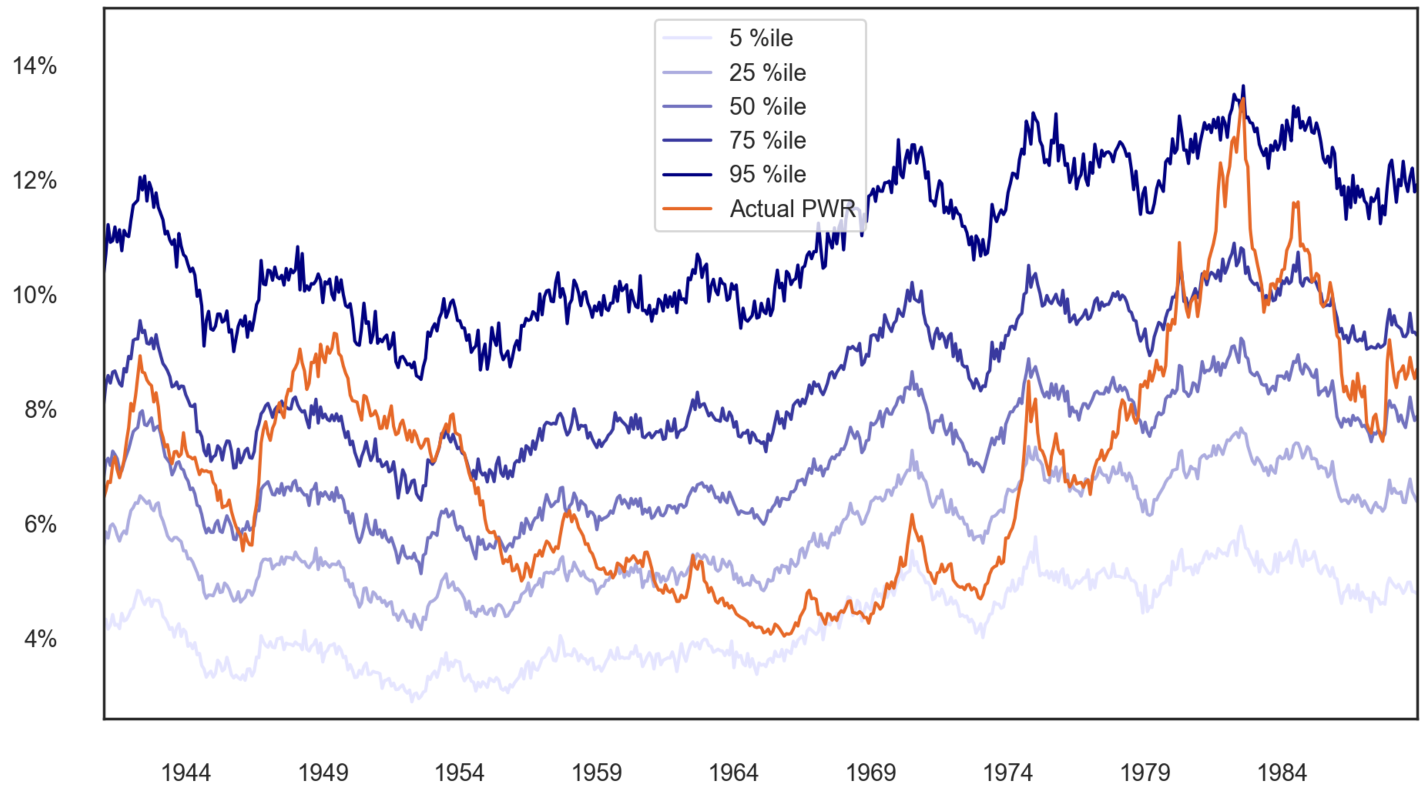
To incorporate some of these effects in our simulations, we must move beyond the simplistic assumption of normally distributed returns.

First, we will fit a skewed normal distribution to the rolling historical data and use that to draw our random variables for each period. This is essentially what was done in the previous section for the normally distributed returns.

Then, to account for some autocorrelation, we will use the same adjustment to volatility as we used in the previously reference commentary on autocorrelation risk. For positive autocorrelations (which we saw in the previous graphs), this results in a higher volatility for the simulations (typically around 10% - 25% higher).

The two graphs below show the same analysis as before under this modified framework.

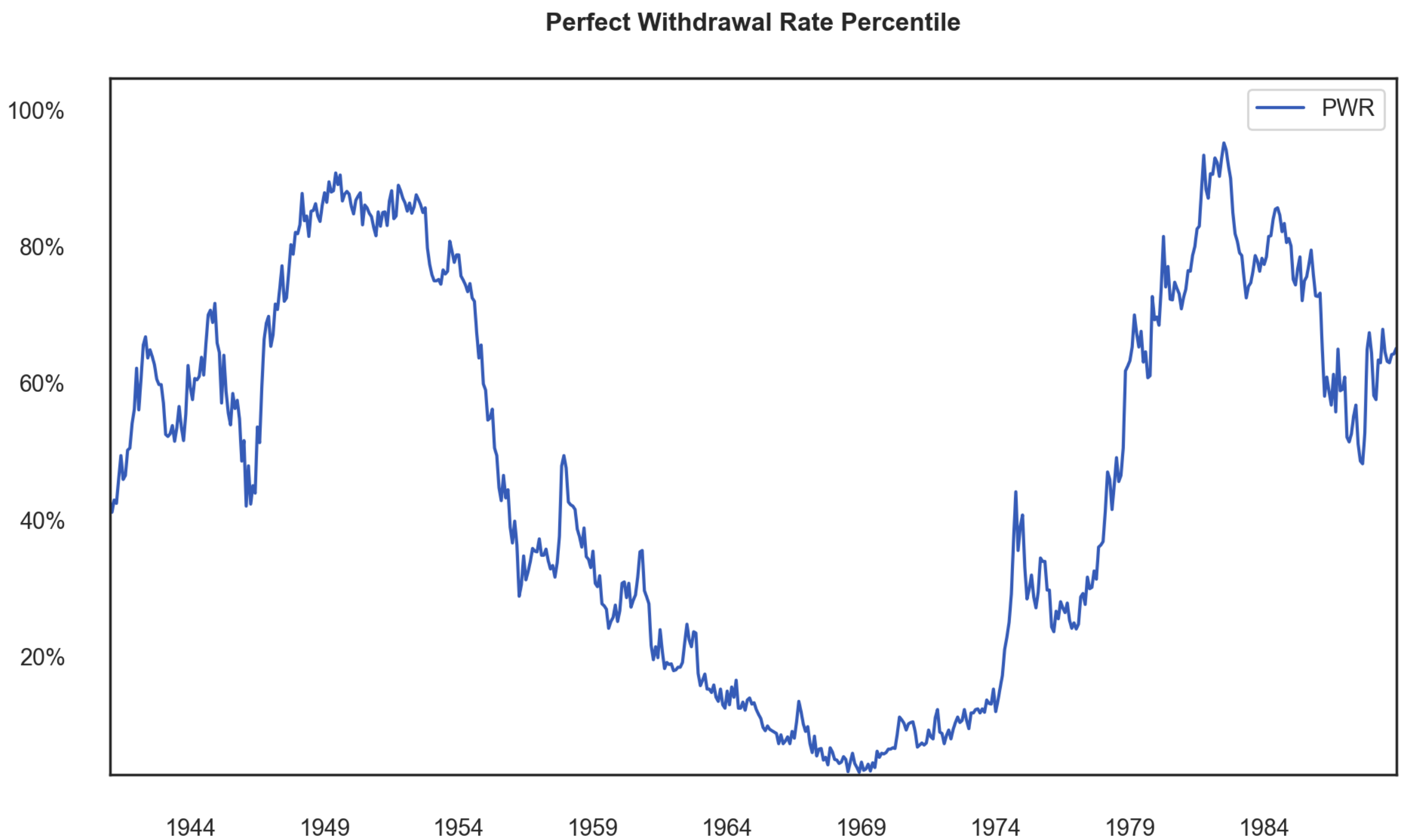
Simulated and Actual Perfect Withdrawal Rates



Source: Global Financial Data and Shiller Data Library. Calculations by Newfound Research. Returns are backtested and hypothetical. Past performance is not a guarantee of future results. Returns are gross of all fees. Returns assume the reinvestment of all distributions. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

The historical PWR now fall more within the bounds of our simulated results.

Additionally, the 5th percentile band now shows that there were periods where a 4% withdrawal rule may not have made the cut.



Source: Global Financial Data and Shiller Data Library. Calculations by Newfound Research. Returns are backtested and hypothetical. Past performance is not a guarantee of future results. Returns are gross of all fees. Returns assume the reinvestment of all distributions. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

Conclusion

Heuristics can be a great way to distill complex data into actionable insights, and the perfect withdrawal rate in retirement portfolios is no exception.

The 4% rule is a classic example where we may not be aware of the risk in using it. It is the commonly accepted lower bound for safe withdrawal rates, but this is only based on one realization of history.

The actual risk investors take on by using this number may be uncertain.

Using simulation techniques, we explored how different assumptions match the historical experience of retirement portfolios.

The simple assumptions (expected return and volatility) commonly used in financial planning Monte Carlo simulations do not seem to reflect as much variation as we have seen in the historical PWR. Therefore, relying on these assumptions can be risky for investors who are close to the “go-no-go” point; they do not have much room for failure and will be more likely to have to make cash flow adjustments in retirement.

Utilizing richer simulation methods (e.g. accounting for negative skew and autocorrelation like we did here or using a downside shocking method like we explored in A Shock to the Covariance System) may be necessary to successfully gauge that risk in a proposed PWR, especially as it pertains to the risk of failure in the financial plan.

Having a number to base planning calculations on makes life easier in the moment, but knowing the risk in using that number makes life easier going forward.

STYLE SURFING THE BUSINESS CYCLE

April 29, 2019

SUMMARY

- In this commentary, we ask whether we should consider rotating factor exposure based upon the business cycle.
- To eliminate a source of model risk, we assume perfect knowledge of future recessions, allowing us to focus only on whether prevailing wisdom about which factors work during certain economic phases actually adds value.
- Using two models of factor rotation and two definitions of business cycles, we construct four timing portfolios and ultimately find that rotating factor exposures does not add meaningful value above a diversified benchmark.
- We find that the cycle-driven factor rotation recommendations are extremely close to data-mined optimal results. The similarity of the recommendations coupled with the lackluster performance of conventional style timing recommendations may highlight how fragile the rotation process inherently is.

Just as soon as the market began to meaningfully adopt factor investing, someone had to go and ask, "yeah, but can they be timed?" After all, while the potential opportunity to harvest excess returns is great, who wants to live through a decade of relative drawdowns like we're seeing with the value factor?

And thus the great valuation-spread factor timing debates of 2017 were born and from the ensuing chaos emerged new, dynamic factor rotation products.

There is no shortage of ways to test factor rotation: valuation-spreads, momentum, and mean-reversion to name a few. We have even found mild success using momentum and mean reversion, though we ultimately question whether the post-cost headache is worth the potential benefit above a well-diversified portfolio.

Another potential idea is to time factor exposure based upon the state of the economic or business cycle.

It is easy to construct a narrative for this approach. For example, it sounds logical that you might want to hold higher quality, defensive stocks during a recession to take advantage of the market's flight-to-safety. On the other hand, it may make sense to overweight value during a recovery to exploit larger mispricings that might have occurred during the contraction period.

An easy counter-example, however, is the performance of value during the last two recessions. During the dot-com fall-out, cheap out-performed expensive by a wide margin. This fit a wonderful narrative of value as a defensive style of investing, as we are buying assets at a discount to intrinsic value and therefore establishing a margin of safety.

Of course, we need only look towards 2008 to see a very different scenario. From peak to trough, AQR's HML Devil factor had a drawdown of nearly 40% during that crisis.

Two recessions with two very different outcomes for a single factor. But perhaps there is still hope for this approach if we diversify across enough factors and apply it over the long run.

The problem we face with business cycle style timing is really two-fold. First, we have to be able to identify the factors that will do well in a given market environment. Equally important, however, is our ability to predict the future economic environment.

Philosophically, there are limitations in our ability to accurately identify both simultaneously. After all, if we could predict both perfectly, we could construct an arbitrage.

If we believe the markets are at all efficient, then being able to identify the factors that will out-perform in a given state of the business cycle should lead us to conclude that we cannot predict the future state of the business cycle. Similarly, if we believe we can predict the future state of the business cycle, we should not be able to predict which factors will necessarily do well.

Philosophical arguments aside, we wanted to test the efficacy of this approach.

Which Factors and When?

Rather than simply perform a data-mining exercise to determine which factors have done well in each economic environment, we wanted to test prevalent beliefs about factor performance and economic cycles. To do this, we identified marketing and research materials from two investment institutions that tie factor allocation recommendations to the business cycle.

Both models expressed a view using four stages of the economic environment: a slowdown, a contraction, a recovery, and an economic expansion.

Model #1

- Slowdown: Momentum, Quality, Low Volatility
- Contraction: Value, Quality, Low Volatility
- Recovery: Value, Size
- Expansion: Value, Size, Momentum

Model #2

- Slowdown: Quality, Low Volatility
- Contraction: Momentum, Quality, Low Volatility
- Recovery: Value, Size
- Expansion: Value, Size, Momentum

Defining the Business Cycle

Given these models, our next step was to build a model to identify the current economic environment. Rather than build a model, however, we decided to dust off our crystal ball. After all, if business-cycle-based factor rotation does not work with perfect foresight of the economic environment, what hope do we have for when we have to predict the environment?

We elected to use the National Bureau of Economic Research's ("NBER") listed history of US business cycle expansions and contractions. With the benefit of hindsight, they label recessions as the peak of the business cycle prior to the subsequent trough.

Unfortunately, NBER only provides a simple indicator as to whether a given month is in a recession or not. We were left to fill in the blanks around what constitutes a slowdown, a contraction, a recovery, and an expansionary period. Here we settled on two definitions:

Definition #1

- Slowdown: The first half of an identified recession
- Contraction: The second half of an identified recession
- Recovery: The first third of a non-recessionary period
- Expansion: The remaining part of a non-recessionary period

Definition #2

- Slowdown: The 12-months leading up to a recession
- Contraction: The identified recessionary periods
- Recovery: The 12-months after an identified recession
- Expansion: The remaining non-recessionary period

For definition #2, in the case where two recessions were 12 or fewer months apart (as was the case in the 1980s), the intermediate period was split equivalently into recovery and slowdown.

Implementing Factor Rotation

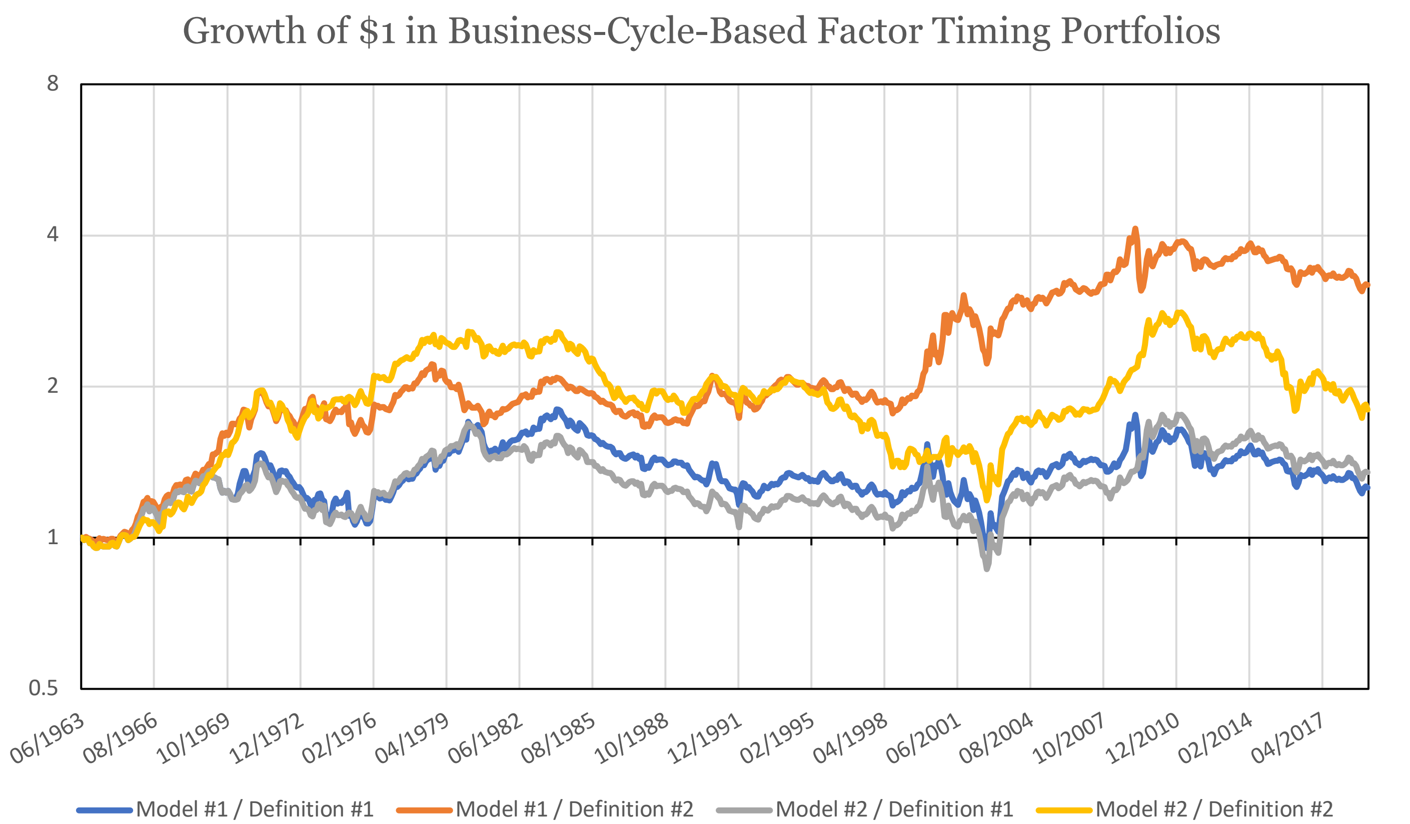
After establishing the rotation rules and using our crystal ball to identify the different periods of the business cycle, our next step was to build the factor rotation portfolios.

We first sourced monthly long/short equity factor returns for size, value, momentum, and quality from AQR's data library. To construct a low-volatility factor, we used portfolios sorted on variance from the Kenneth French library and subtracted bottom-quintile returns from top-quintile returns.

As the goal of our study is to identify the benefit of factor *timing*, we de-meaned the monthly returns by the average of all factor returns in that month to identify relative performance.

We constructed four portfolios using the two factor rotation definitions and the two economic cycle definitions. Generically, at the end of each month, we would use the next month's economic cycle label to identify which factors to hold in our portfolio. Identified factors were held in equal weight.

Below we plot the four equity curves. Remember that these series are generated using de-meaned return data, so reflect the out-performance against an equal-weight factor benchmark.



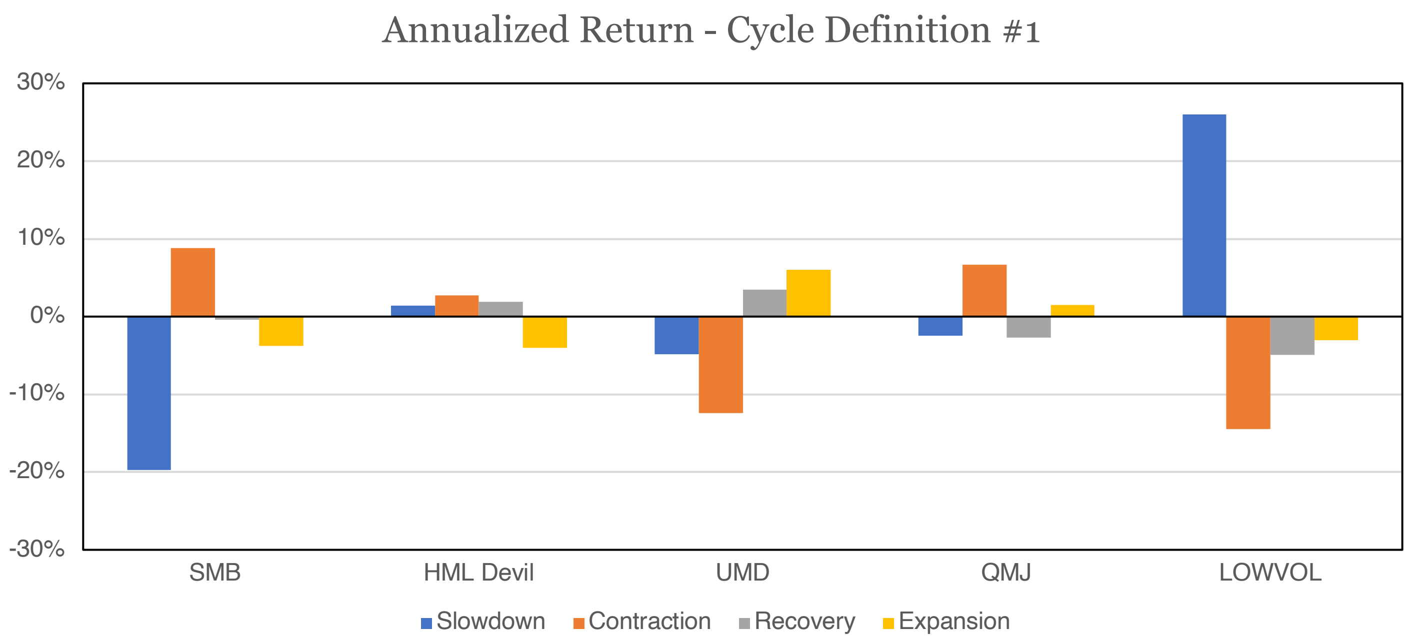
Source: NBER, AQR, and Kenneth French Data Library. Calculations by Newfound Research. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

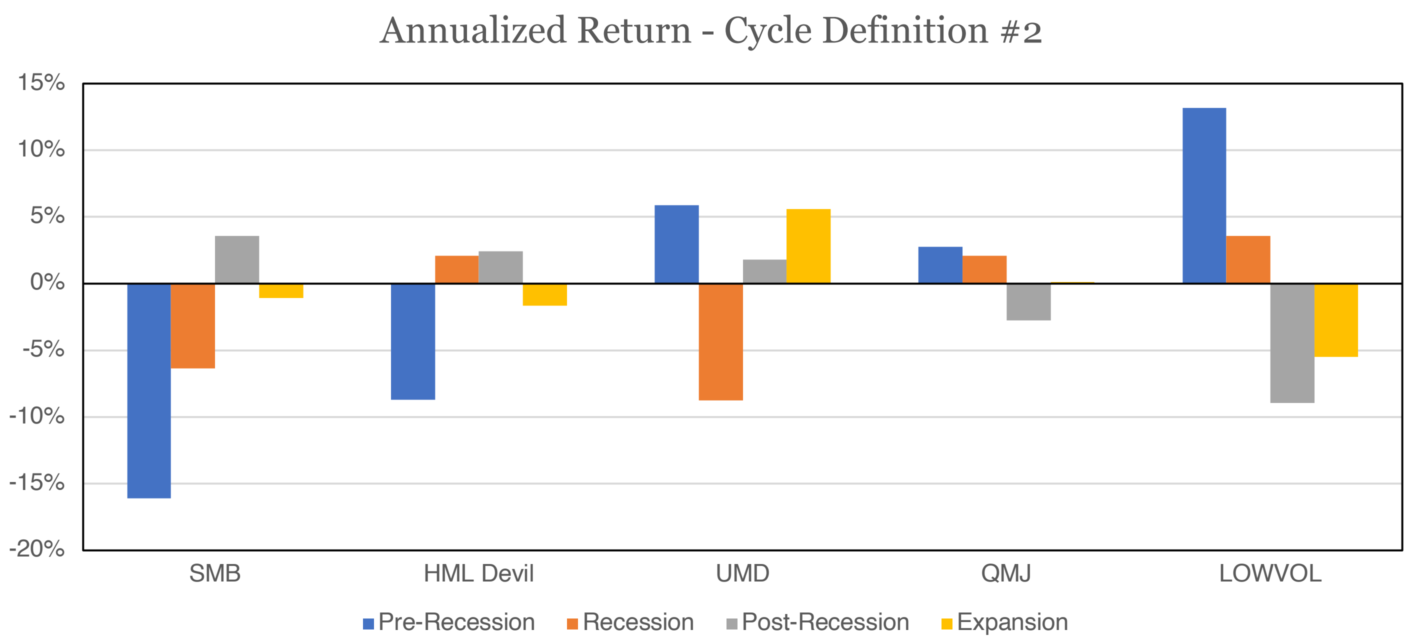
It would appear that even with a crystal ball, conventional wisdom about style rotation and business cycles may not hold. And even where it might, we can see multi-decade periods where it adds little-to-no value.

Data-Mining Our Way to Success

If we are going to use a crystal ball, we might as well just blatantly data-mine our way to success and see what we learn along the way.

To achieve this goal, we can simply look at the annualized de-meaned returns of each factor during each period of the business cycle.



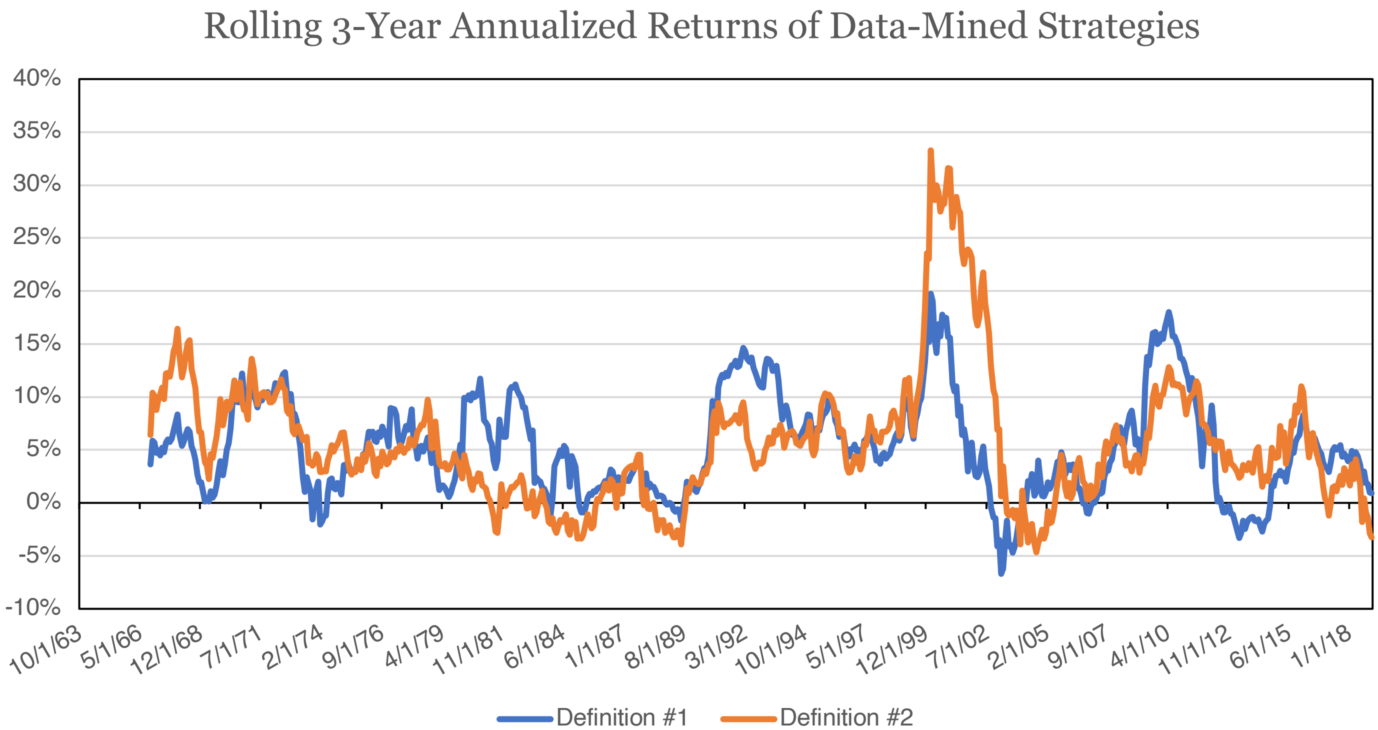


Source: NBER, AQR, and Kenneth French Data Library. Calculations by Newfound Research. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

Despite two different definitions of the business cycle, we can see a strong alignment in which factors work when. Slow-downs / pre-recessionary periods are tilted towards momentum and defensive factors like quality and low-volatility. Momentum may seem like a curious factor, but its high turnover may give it a chameleon-like nature that can tilt it defensively in certain scenarios.

In a recession, momentum is replaced with value while quality and low-volatility remain. In the initial recovery, small-caps, value, and momentum are favored. In this case, while value may actually be benefiting from multiple expansion, small-caps may simply be a way to play higher beta. Finally, momentum is strongly favored during an expansion.

Yet even a data-mined solution is not without its flaws. Below we plot rolling 3-year returns for our data-mined timing strategies. Again, remember that these series are generated using de-meaned return data, so reflect the out-performance against an equal-weight factor benchmark.



Source: NBER, AQR, and Kenneth French Data Library. Calculations by Newfound Research. Returns are backtested and hypothetical. Returns assume the reinvestment of all distributions. Returns are gross of all fees. None of the strategies shown reflect any portfolio managed by Newfound Research and were constructed solely for demonstration purposes within this commentary. You cannot invest in an index.

Despite a crystal ball telling us what part of the business cycle we are in *and* completely data-mined results, there are still a number of 3-year periods with low-to-negative results. And we have not even considered manager costs, transaction costs, or taxes yet.

A few more important things to note.

Several of these factors exhibit strong *negative* performance during certain parts of the market cycle, indicating a potential for out-performance by taking the opposite side of the factor. For example, value appears to do poorly during pre-recession and expansion periods. One hypothesis is that during expansionary periods, markets tend to over-extrapolate earnings growth potential, favoring growth companies that appear more expensive.

We should also remember that our test is on long/short portfolios and may not necessarily be relevant for long-only investors. While we can think of a long-only portfolio as a market-cap portfolio plus a long/short portfolio, the implicit long/short is not necessarily identical to academic factor definitions.

Finally, it is worth considering that these results are data-mined over a 50+ year period, which may allow outlier events to dramatically skew the results. Momentum, for example, famously exhibited dramatic crashes during the Great Depression and in the 2008-crisis, but may have actually relatively out-performed in other recessions.

Conclusion

In this commentary we sought to answer the question, "can we use the business cycle to time factor exposures?" Assuming access to a crystal ball that could tell us where we stood precisely in the business cycle, we found that conventional wisdom about factor timing did not add meaningful value over time. We do not hold out much hope, based on this conventional wisdom, that someone without a crystal ball would fare much better.

Despite explicitly trying to select models that reflected conventional wisdom, we found a significant degree of similarity in these recommendations with those that came from blindly data-mining optimal results. Nevertheless, even slight recommendation differences lead to lackluster results.

The similarities between data-mined results and conventional wisdom, however, should give us pause. While the argument for conventional wisdom is often a well-articulated economic rationale, we have to wonder whether we have simply fooled ourselves with a narrative that has been inherently constructed with the benefit of hindsight.

TACTICAL PORTABLE BETA

May 6, 2019

SUMMARY

- In this commentary, we revisit the idea of portable beta: utilizing leverage to overlay traditional risk premia on existing strategic allocations.

- While a 1.5x levered 60/40 portfolio has historically out-performed an all equity blend with similar risk levels, it can suffer through prolonged periods of under-performance.

- Positive correlations between stocks and bonds, inverted yield curves, and rising interest rate environments can make simply adding bond exposure on top of equity exposure a non-trivial pursuit.

- We rely on prior research to introduce a tactical 90/60 model, which uses trend signals to govern equity exposure and value, momentum, and carry signals to govern bond exposure.

- We find that such a model has historically exhibited returns in-line with equities with significantly lower maximum drawdown.

In November 2017, I was invited to participate in a Bloomberg roundtable discussion with Barry Ritholtz, Dave Nadig, and Ben Fulton about the future of ETFs. I was quoted as saying,

> *"Most of the industry agrees that we are entering a period of much lower returns for stocks and fixed income. That's a problem for younger generations. The innovation needs to be around efficient use of capital. Instead of an ETF that holds intermediate-term Treasuries, I would like to see a U.S. Treasury ETF that uses Treasuries as collateral to buy S&P 500 futures, so you end up getting both stock and bond exposure. By introducing a modest amount of leverage, you can take $1 and trade it as if the investor has $1.50. After 2008, people became skittish around derivatives, shorting, and leverage. But these aren't bad things when used appropriately."*

Shortly after the publication of the discussion, we penned a research commentary titled *Portable Beta* which extolled the potential virtues of employing prudent leverage to better exploit diversification opportunities. For investors seeking to enhance returns, increasing beta exposure may be a more reliable approach than the pursuit of alpha.

In August 2018, WisdomTree introduced the 90/60 U.S. Balanced Fund (ticker: NTSX), which blends core equity exposure with a U.S. Treasury futures ladder to create the equivalent of a 1.5x levered 60/40 portfolio. On March 27, 2019, NTSX was awarded ETF.com's Most Innovative New ETF of 2018.

The idea of portable beta was not even remotely uniquely ours. Two anonymous Twitter users – "Jake" (@EconomPic) and "Unrelated Nonsense" (@Nonrelatedsense) – had discussed the idea several times prior to my round-table in 2017. They argued that such a product could be useful to free up space in a portfolio for alpha-generating ideas. For example, an investor could hold 66.6% of their wealth in a 90/60 portfolio and use the other 33.3% of their portfolio for alpha ideas. While the leverage is technically applied to the 60/40, the net effect would be a 60/40 portfolio with a set of alpha ideas overlaid on the portfolio. Portable beta becomes portable alpha.

Even then, the idea was not new. After NTSX launched, Cliff Asness, co-founder and principal of AQR Capital Management, commented on Twitter that even though he had a "22-year head start," WisdomTree had beat him to launching a fund. In the tweet, he linked to an article he wrote in 1996, titled *Why Not 100% Equities*, wherein Cliff demonstrated that from 1926 to 1993 a 60/40 portfolio levered to the same volatility as equities achieved an excess return of 0.8% annualized above U.S. equities. Interestingly, the appropriate amount of leverage utilized to match equities was 155%, almost perfectly matching the 90/60 concept.

EXHIBIT 3
Effect of Leverage (%)

Portfolio	Compound Return	Standard Deviation
100% Stocks	10.3	20.0
100% Bonds	5.6	6.8
60% Stocks, 40% Bonds	8.9	12.9
Levered 60/40	11.1	20.0

Source: Asness, Cliff. Why Not 100% Equities. Journal of Portfolio Management, Winter 1996, Volume 22 Number 2.

Following up on Cliff's Tweet, Jeremy Schwartz from WisdomTree extended the research out-of-sample, covering the quarter century that followed Cliff's initial publishing date. Over the subsequent 25 years, Jeremy found that a levered 60/40 outperformed U.S. equities by 2.6% annualized.

Extended, 1/31/1926–7/31/2018

Portfolio	Annualized Total Return	Annualized Volatility
100% Stocks	10.1%	18.7%
100% Bonds	5.9%	7.5%
100% Cash	3.3%	0.9%
60% Stocks, 40% Bonds	8.9%	12.1%
Levered 60/40	11.4%	18.8%

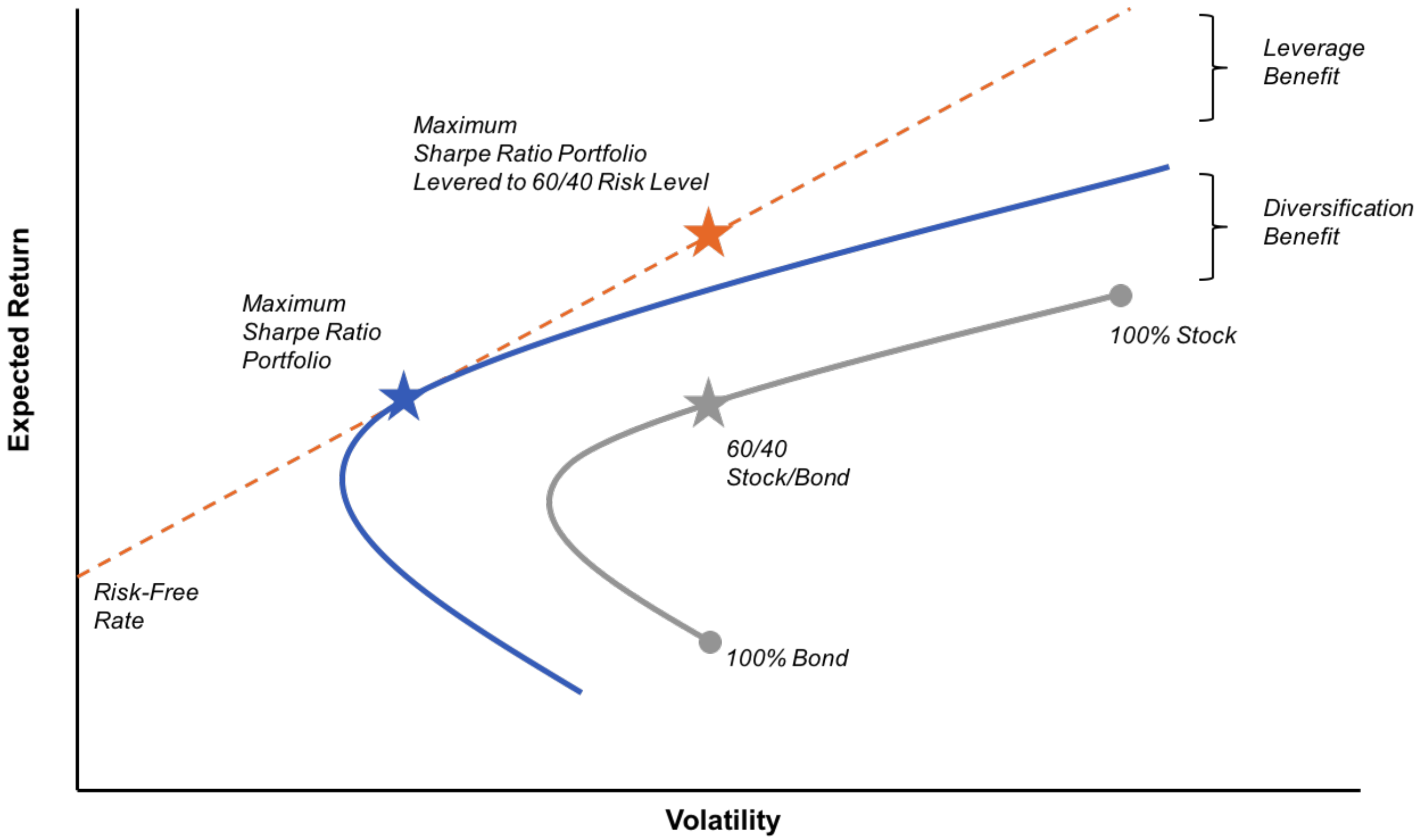
Extension, 1/1/1994–7/31/2018

Portfolio	Annualized Total Return	Annualized Volatility
100% Stocks	9.6%	14.3%
100% Bonds	7.0%	9.1%
100% Cash	2.4%	0.6%
60% Stocks, 40% Bonds	8.9%	9.6%
Levered 60/40	12.2%	14.9%

Sources: WisdomTree, Ibbotson, Morningstar Direct, as of 7/31/18. Past performance is not indicative of future results. You cannot invest directly in an index. Index performance does not represent actual fund or portfolio performance. A fund or portfolio may differ significantly from the securities included in the index. Index performance assumes reinvestment of dividends but does not reflect any management fees, transaction costs or other expenses that would be incurred by a portfolio or fund, or brokerage commissions on transactions in fund shares. Such fees, expenses and commissions could reduce returns. WisdomTree, its affiliates and their independent providers are not liable for any informational errors, incompleteness or delays or for any actions taken in reliance on information contained herein. The research shown is not a back test or hypothetical representation of NTSX.

NTSX is not the first product to try to exploit the idea of diversification and leverage. These ideas have been the backbone of managed futures and risk parity strategies for decades. The entire PIMCO's StocksPLUS suite – which traces its history back to 1986 – is built on these foundations. The core strategy combines an actively managed portfolio of fixed income with 100% notional exposure in S&P 500 futures to create a 2x levered 50/50 portfolio.

The concept traces its roots back to the earliest eras of modern financial theory. Finding the maximum Sharpe ratio portfolio and gearing it to the appropriate risk level has *always* been considered to be the theoretically optimal solution for investors.



Nevertheless, after 2008, the words "leverage" and "derivatives" have largely been terms *non gratis* in the realm of investment products. But that may be to the detriment of investors.

90/60 Through the Decades

While we are proponents of the foundational concepts of the 90/60 portfolio, frequent readers of our commentary will not be surprised to learn that we believe there may be opportunities to enhance the idea through tactical asset allocation. After all, while a 90/60 may have out-performed over the long run, the short-run opportunities available to investors can deviate significantly. The prudent allocation at the top of the dot-com bubble may have looked quite different than that at the bottom of the 2008 crisis.

To broadly demonstrate this idea, we can examine the how the realized efficient frontier of stock/bond mixes has changed shape over time. In the table below, we calculate the Sharpe ratio for different stock/bond mixes realized in each decade from the 1920s through present.

	100% Bonds	20/80	40/60	60/40	80/20	100% Stocks
1920s	0.33	0.71	0.79	0.76	0.71	0.66
1930s	0.58	0.39	0.21	0.11	0.03	-0.03
1940s	0.86	0.81	0.68	0.61	0.55	0.51
1950s	-0.51	0.58	1.12	1.32	1.41	1.46
1960s	-0.39	-0.11	0.1	0.22	0.28	0.31
1970s	-0.06	-0.04	-0.03	-0.03	-0.03	-0.04
1980s	0.33	0.47	0.57	0.59	0.55	0.5
1990s	0.45	1.07	0.97	0.84	0.53	0.43
2000s	0.45	0.38	0.16	-0.03	-0.15	-0.22
2010s*	0.27	0.82	1.19	1.2	1.12	1.04

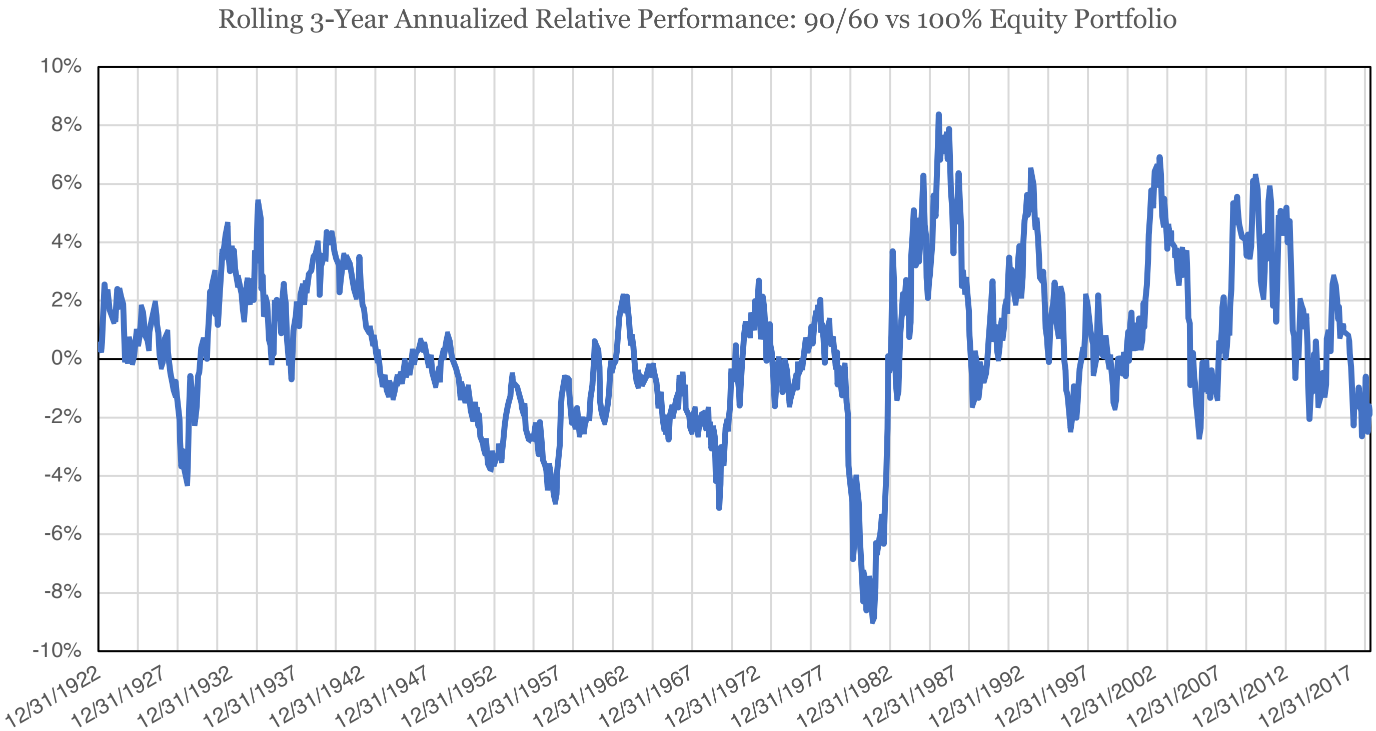
Source: Global Financial Data. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. Bonds are the GFD Indices USA 10-Year Government Bond Total Return Index and Stocks are the S&P 500 Total Return Index (with GFD Extension). Sharpe ratios are calculated with returns excess of the GFD Indices USA Total Return T-Bill Index. You cannot invest in an index. 2010s reflect a partial decade through 4/2019.

We should note here that the original research proposed by Asness (1996) assumed a bond allocation to an Ibbotson corporate bond series while we employ a constant maturity 10-year U.S. Treasury index. While this leads to lower total returns in our bond series, we do not believe it meaningfully changes the conclusions of our analysis.

We can see that while the 60/40 portfolio has a higher realized Sharpe ratio than the 100% equity portfolio in eight of ten decades, it has a lower Sharpe ratio in two *consecutive* decades from 1950 – 1960. And the 1970s were not a ringing endorsement.

In theory, a higher Sharpe ratio for a 60/40 portfolio would imply that an appropriately levered version would lead to higher realized returns than equities at the same risk level. Knowing the appropriate leverage level, however, is non-trivial, requiring an estimate of equity volatility. Furthermore, leverage requires margin collateral and the application of borrowing rates, which can create a drag on returns.

Even if we conveniently ignore these points and assume a constant 90/60, we can still see that such an approach can go through lengthy periods of relative under-performance compared to buy-and-hold equity. Below we plot the annualized rolling 3-year returns of a 90/60 portfolio (assuming U.S. T-Bill rates for leverage costs) minus 100% equity returns. We can clearly see that the 1950s through the 1980s were largely a period where applying such an approach would have been frustrating.

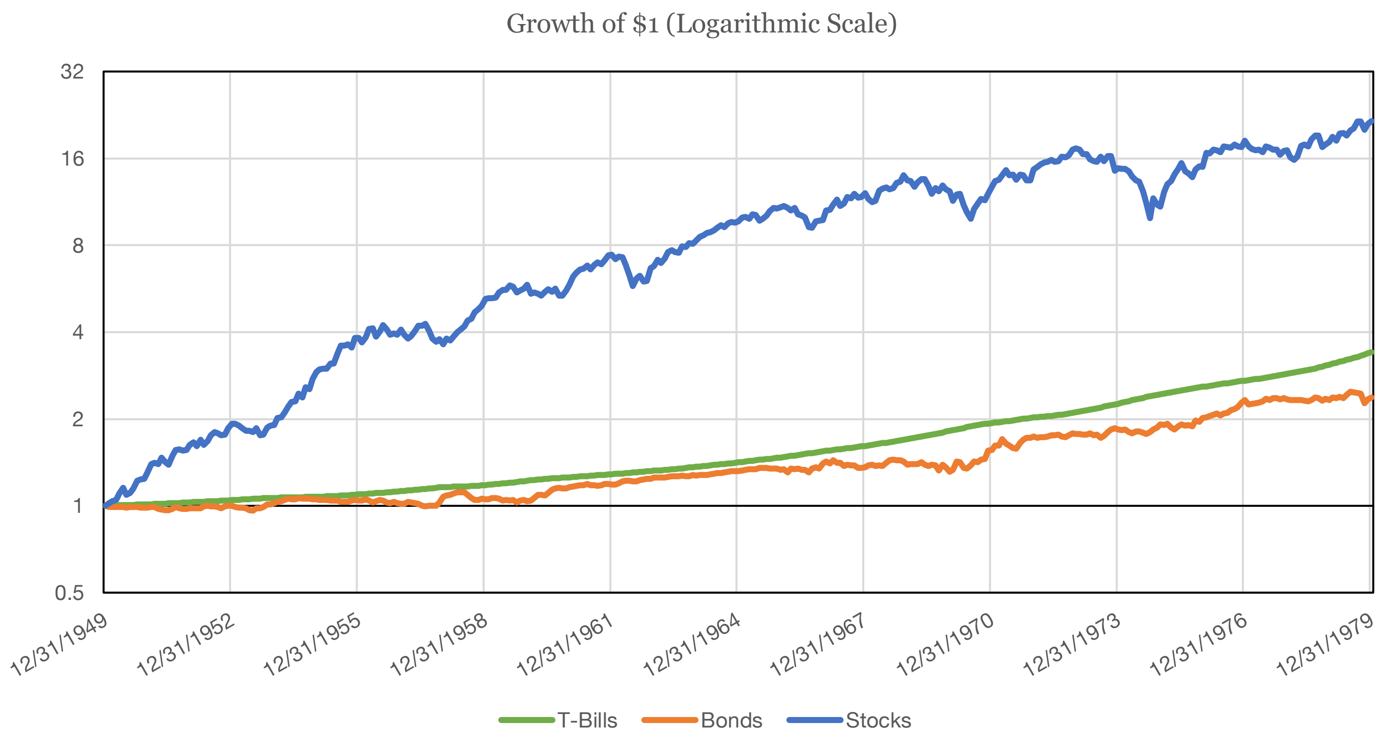


Source: Global Financial Data. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Bonds are the GFD Indices USA 10-Year Government Bond Total Return Index and Stocks are the S&P 500 Total Return Index (with GFD Extension). The 90/60 portfolio invests 150% each month in the 60/40 portfolio and -50% in the GFD Indices USA Total Return T-Bill Index. You cannot invest in an index.

Poor performance of the 90/60 portfolio in this era is due to two effects.

First, 10-year U.S. Treasury rates rose from approximately 4% to north of 15%. While a constant maturity index would constantly roll into higher interest bonds, it would have to do so by selling old holdings at a loss. Constantly harvesting price losses created a headwind for the index.

This is compounded in the 90/60 by the fact that the yield curve over this period spent significant time in an inverted state, meaning that the cost of leverage exceeded the yield earned on 40% of the portfolio, leading to negative carry. This is illustrated in the chart below, with –T-Bills– realizing a higher total return over the period than –Bonds–.

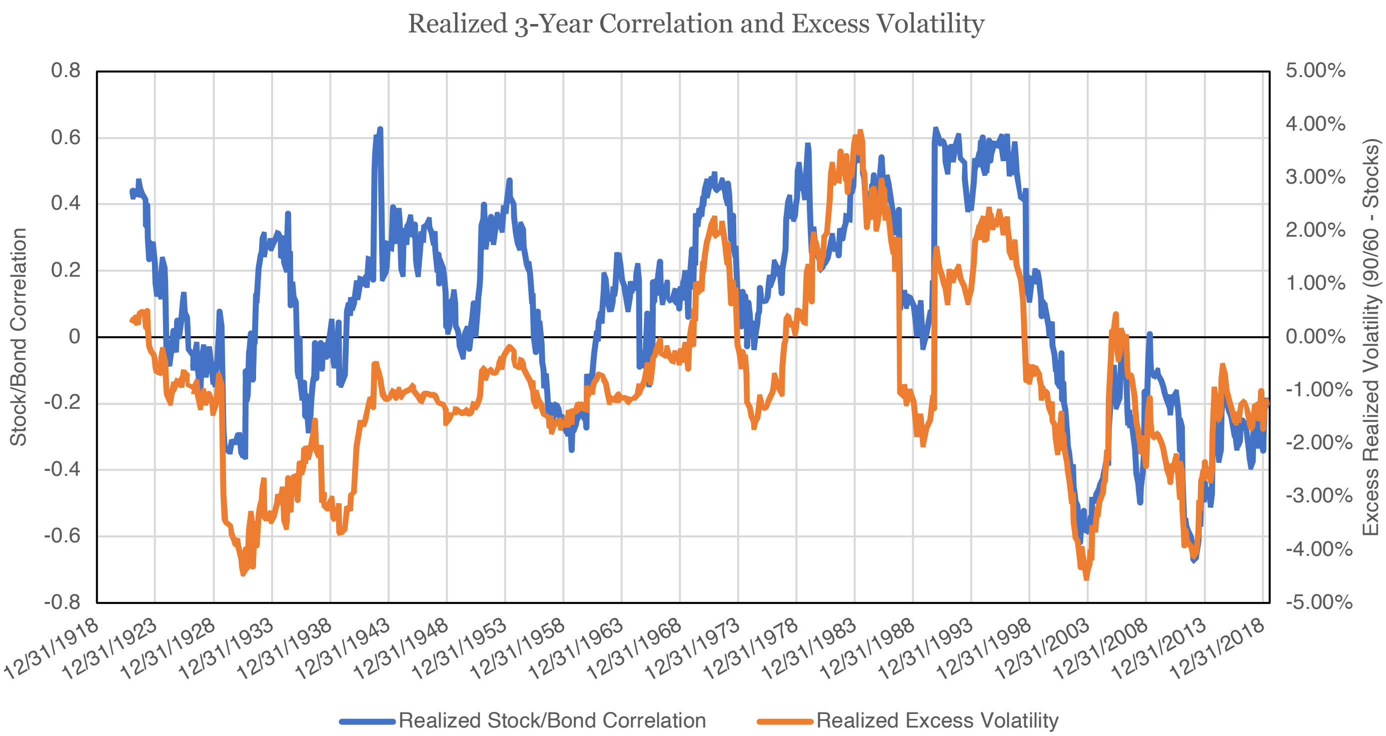


Source: Global Financial Data. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. T-Bills are the GFD Indices USA Total Return T-Bill Index, Bonds are the GFD Indices USA 10-Year Government Bond Total Return Index, and Stocks are the S&P 500 Total Return Index (with GFD Extension). You cannot invest in an index.

This is all arguably further complicated by the fact that while a 1.5x levered 60/40 may closely approximate the risk level of a 100% equity portfolio over the long run, it may be a far cry from it over the short-run. This may be particularly true during periods where stocks and bonds exhibit positive realized correlations as they did during the 1960s through 1980s. This can occur when markets are more pre-occupied with inflation risk than economic risk. As inflationary fears abated and economic risk become the foremost concern in the 1990s, correlations between stocks and bonds flipped.

Thus, during the 1960s-1980s, a 90/60 portfolio exhibited realized volatility levels in excess of an all-equity portfolio, while in the 2000s it has been below.

This all invites the question: should our levered allocation necessarily be static?



Getting Tactical with a 90/60

We might consider two approaches to creating a tactical 90/60.

The first is to abandon the 90/60 model outright for a more theoretically sound approach. Specifically, we could attempt to estimate the maximum Sharpe ratio portfolio, and then apply the appropriate leverage such that we either hit a (1) constant target volatility or (2) the volatility of equities. This would require us to not only accurately estimate the expected excess returns of stocks and bonds, but also their volatilities and correlations. Furthermore, when the Sharpe optimal portfolio is highly conservative, notional exposure far exceeding 200% may be necessary to hit target volatility levels.

In the second approach, equity and bond exposure would each be adjusted tactically, without regard for the other exposure. While less theoretically sound, one might interpret this approach as saying, “we generally want exposure to the equity and bond risk premia over the long run, and we like the 60/40 framework, but there might be certain scenarios whereby we believe the expected return does not justify the risk.” The downside to this approach is that it may sacrifice potential diversification benefits between stocks and bonds.

Given the original concept of portable beta is to increase exposure to the risk premia we’re already exposed to, we prefer the second approach. We believe it more accurately reflects the notion of trying to provide long-term exposure to return-generating risk premia while trying to avoid the significant and prolonged drawdowns that can be realized with buy-and-hold approaches.

Equity Signals

To manage exposure to the equity risk premium, our preferred method is the application of trend following signals in an approach we call trend equity. We will approximate this class of strategies with our Newfound Research U.S. Trend Equity Index.

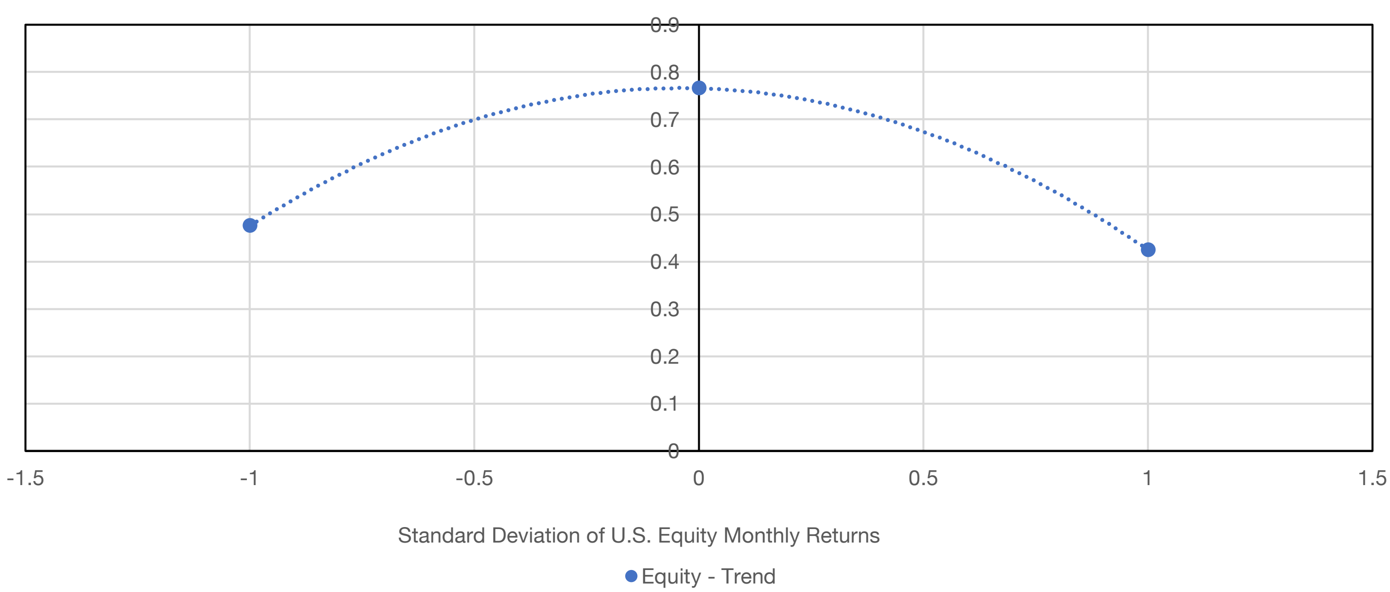
To determine whether our signals are able to achieve their goal of "protect and participate" with the underlying risk premia, we will plot their regime-conditional betas. To do this, we construct a simple linear model:

$$r_{strategy} = \{r\ in\ bear\ regime\}\beta_{bear}r_{index} + \{r\ in\ normal\ regime\}\beta_{normal}r_{index} + \{r\ in\ bull\ regime\}\beta_{bull}r_{index}$$

We define a bear regime as the worst 16% of monthly returns, a bull regime as the best 16% of monthly returns, and a normal regime as the remaining 68% of months. Note that the bottom and top 16th percentiles are selected to reflect one standard deviation.

Below we plot the strategy conditional betas relative to U.S. equity

Conditional Bear, Normal, and Bull Betas to U.S. Equity Returns



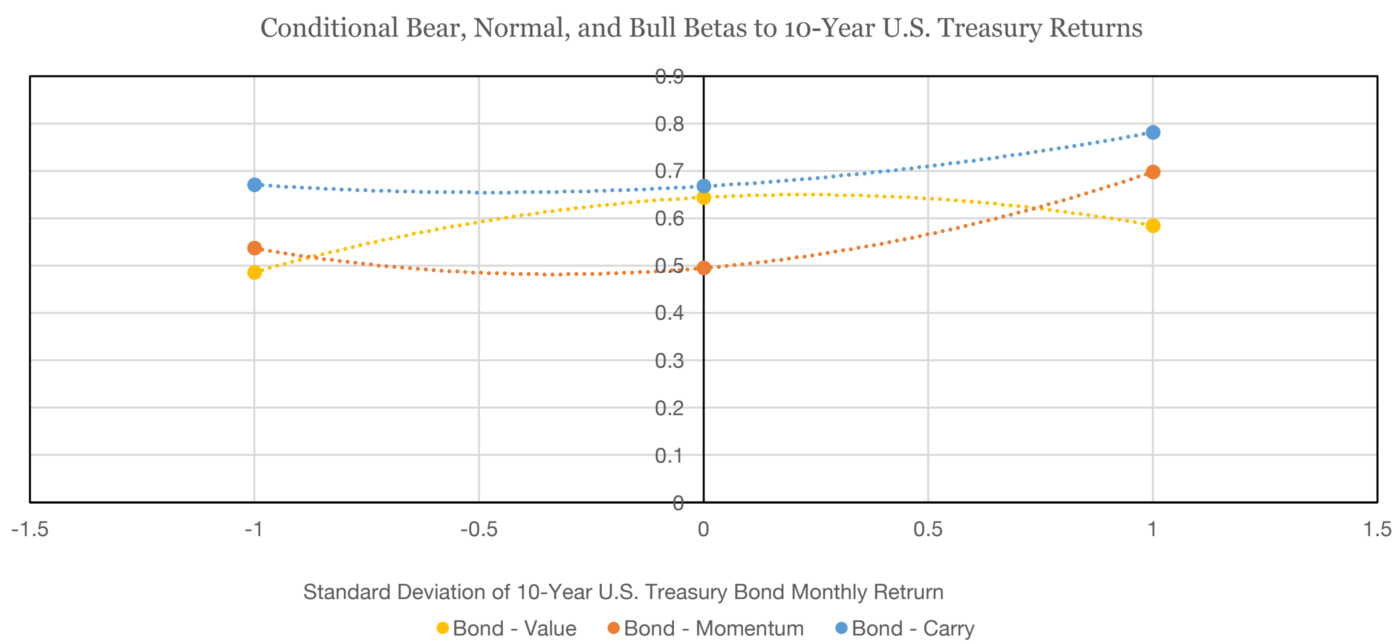
We can see that trend equity has a normal regime beta to U.S. equities of approximately 0.75 and a bear market beta of 0.5, in-line with expectations that such a strategy might capture 70-80% of the upside of U.S. equities in a bull market and 40-50% of the downside in a prolonged bear market. Trend equity beta of U.S. equities in a bull regime is close to the bear market beta, which is consistent with the idea that trend equity as a style has historically sacrificed the best returns to avoid the worst.

Bond Signals

To govern exposure to the bond risk premium, we prefer an approach based upon a combination of quantitative, factor-based signals. We've written about many of these signals over the last two years; specifically in *Duration Timing with Style Premia (June 2017)*, *Timing Bonds with Value, Momentum, and Carry (January 2018)*, and *A Carry-Trend-Hedge Approach to Duration Timing (October 2018).* In these three articles we explore various mixes of value, momentum, carry, flight-to-safety, and bond risk premium measures as potential signals for timing duration exposure.

We will not belabor this commentary unnecessarily by repeating past research. Suffice it to say that we believe there is sufficient evidence that value (deviation in real yield), momentum (prior returns), and carry (term spread) can be utilized as effective timing signals and in this commentary are used to construct bond indices where allocations are varied between 0-100%. Curious readers can pursue further details of how we construct these signals in the commentaries above.

As before, we can determine conditional regime betas for strategies based upon our signals.



We can see that our value, momentum, and carry signals all exhibit an asymmetric beta profile with respect to 10-year U.S. Treasury returns. Carry and momentum exhibit an increase in bull market betas while value exhibits a decrease in bear market beta.

Combining Equity and Bond Signals into a Tactical 90/60

Given these signals, we will construct a tactical 90/60 portfolio as being comprised of 90% trend equity, 20% bond value, 20% bond momentum, and 20% bond carry. When notional exposure exceeds 100%, leverage cost is assumed to be U.S. T-Bills. Taken together, the portfolio has a large breadth of potential configurations, ranging from 100% T-Bills to a 1.5x levered 60/40 portfolio.

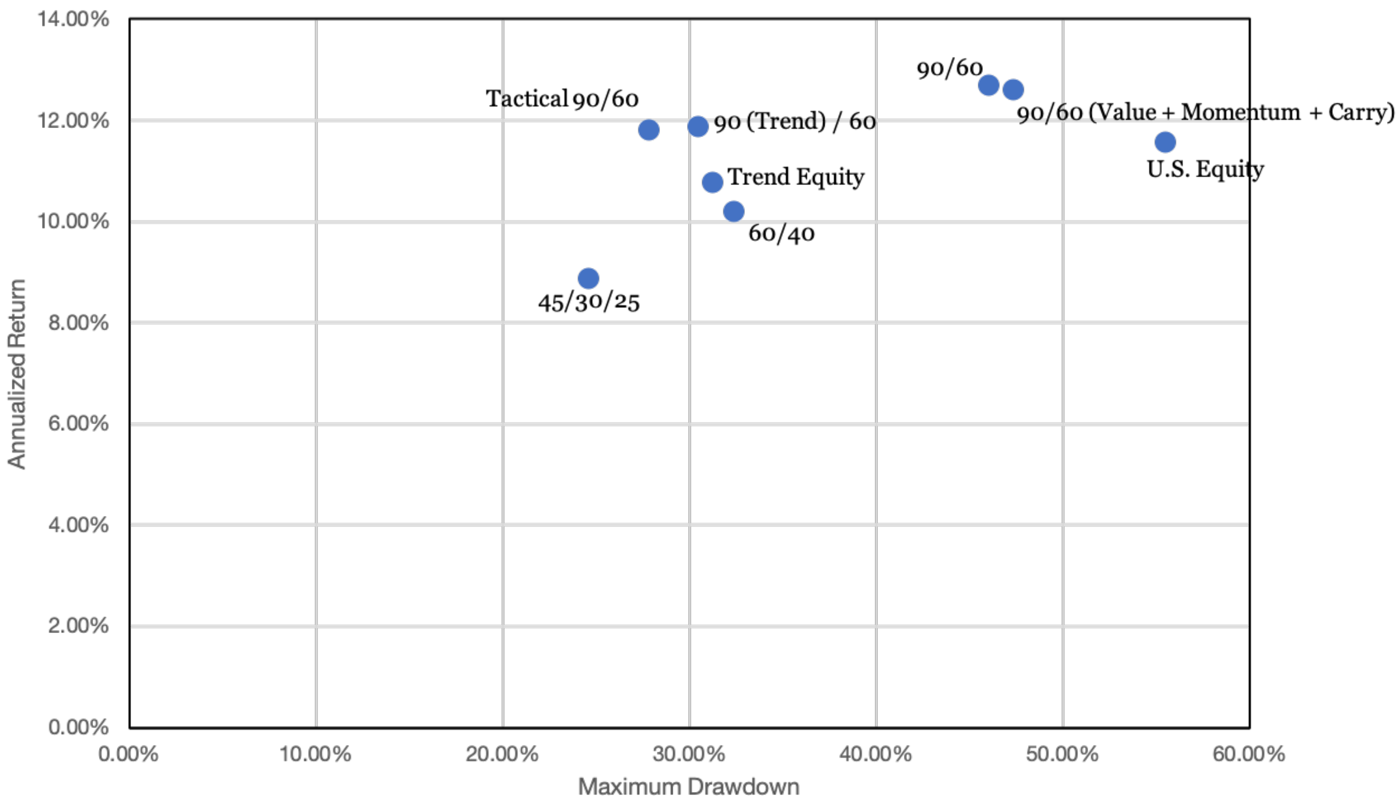
But what is the appropriate benchmark for such a model?

In the past, we have argued that the appropriate benchmark for trend equity is a 50% stock / 50% cash benchmark, as it not only reflects the strategic allocation to equities empirically seen in return decompositions, but it also allows both positive and negative trend calls to contribute to active returns.

Similarly, we would argue that the appropriate benchmark for our tactical 90/60 model is not a 90/60 itself – which reflects the upper limit of potential capital allocation – but rather a 45% stock / 30% bond / 25% cash mix. Though, for good measure we might also consider a bit of hand-waving and just use a 60/40 as a generic benchmark as well.

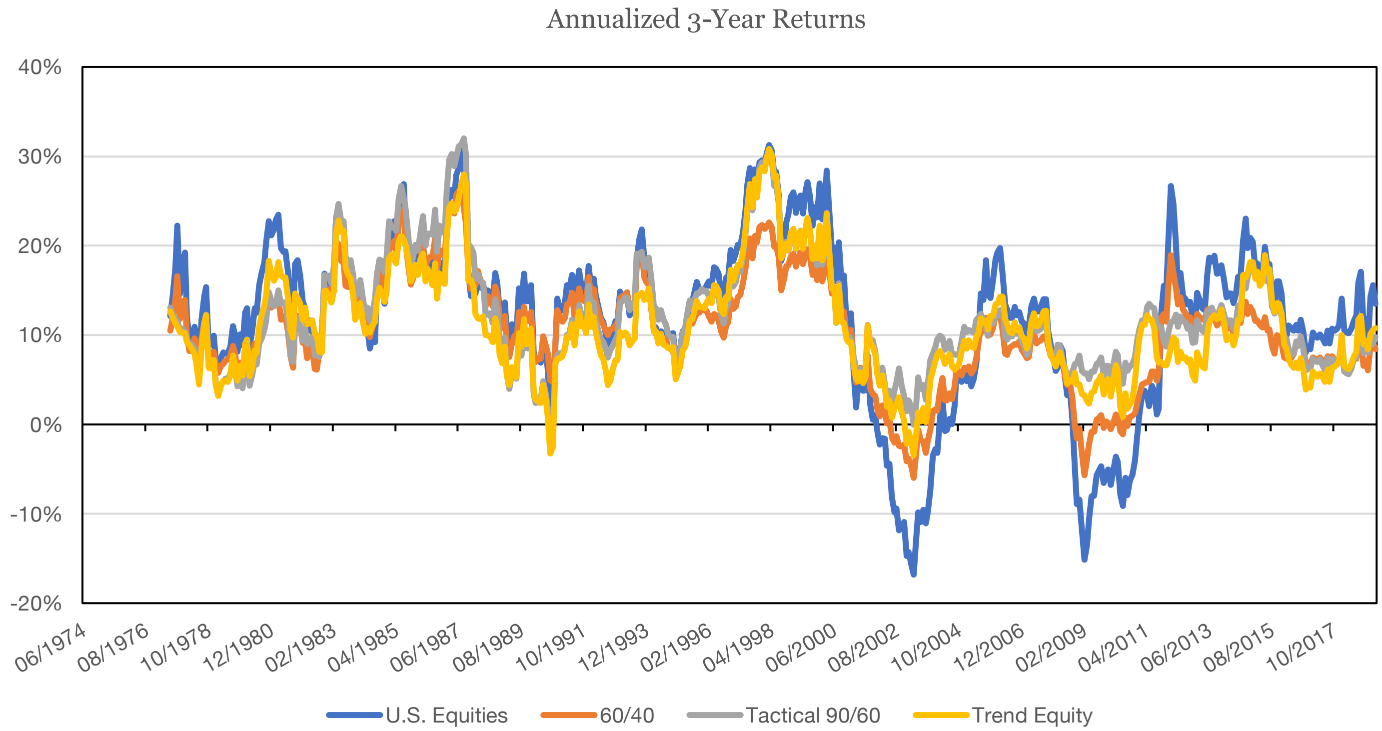
Below we plot the annualized returns versus maximum drawdown for different passive and active portfolio combinations from 1974 to present (reflecting the full period of time when strategy data is available for all tactical signals). We can see that not only does the tactical 90/60 model (with both trend equity and tactical bonds) offer a return in line with U.S. equities over the period, it does so with significantly less drawdown (approximately half). Furthermore, the tactical 90/60 exceeded trend equity and 60/40 annualized returns by 102 and 161 basis points respectively.

These improvements to the return and risk were achieved with the same amount of capital commitment as in the other allocations. That's the beauty of portable beta.



Source: Federal Reserve of St. Louis, Kenneth French Data Library, and Newfound Research. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

Of course, full-period metrics can deceive what an investor's experience may actually be like. Below we plot rolling 3-year annualized returns of U.S. equities, the 60/40 mix, trend equity, and the tactical 90/60.



Source: Federal Reserve of St. Louis, Kenneth French Data Library, and Newfound Research. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

The tactical 90/60 model out-performed a 60/40 in 68% of rolling 3-year periods and the trend equity model in 71% of rolling 3-year periods. The tactical 90/60, however, only out-performs U.S. equities in 35% of rolling 3-year periods, with the vast majority of relative out-performance emerging during significant equity drawdown periods.

For investors already allocated to trend equity strategies, portable beta – or portable tactical beta – may represent an alternative source of potential return enhancement. Rather than seeking opportunities for alpha, portable beta allows for an overlay of more traditional risk premia, which may be more reliable from an empirical and academic standpoint.

The potential for increased returns is illustrated below in the rolling 3-year annualized return difference between the tactical 90/60 model and the Newfound U.S. Trend Equity Index.

Rolling Anualized 3-Year Return Difference: Tactical 90/60 - Trend Equity

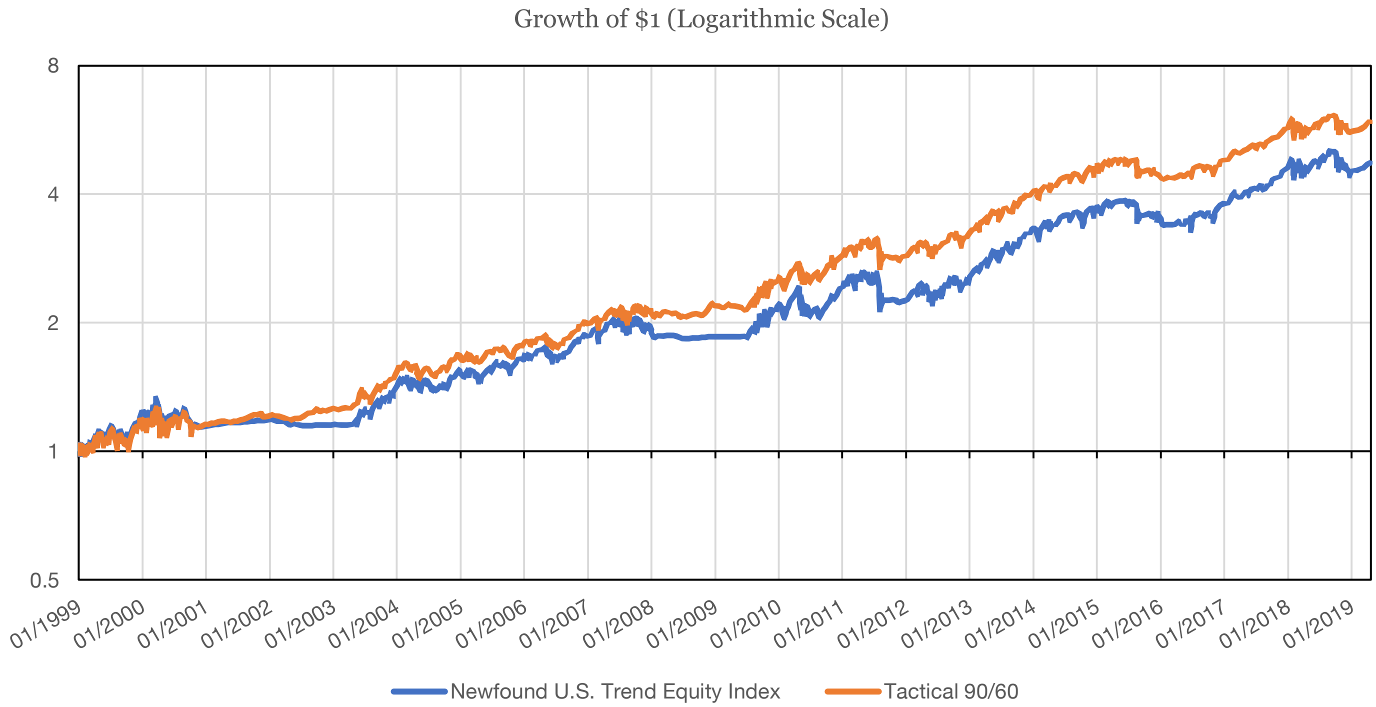
Source: Federal Reserve of St. Louis, Kenneth French Data Library, and Newfound Research. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

From Theory to Implementation

In practice, it may be easier to acquire leverage through the use of futures contracts. For example, applying portable bond beta on-top of an existing trend equity strategy may be achieved through the use of 10-year U.S. Treasury futures.

Below we plot the growth of $1 in the Newfound U.S. Trend Equity Index and a tactical 90/60 model implemented with Treasury futures. Annualized return increases from 7.7% to 8.9% and annualized volatility declines from 9.7% to 8.5%. Finally, maximum drawdown decreases from 18.1% to 14.3%.

We believe the increased return reflects the potential return enhancement benefits from introducing further exposure to traditional risk premia, while the reduction in risk reflects the benefit achieved through greater portfolio diversification.



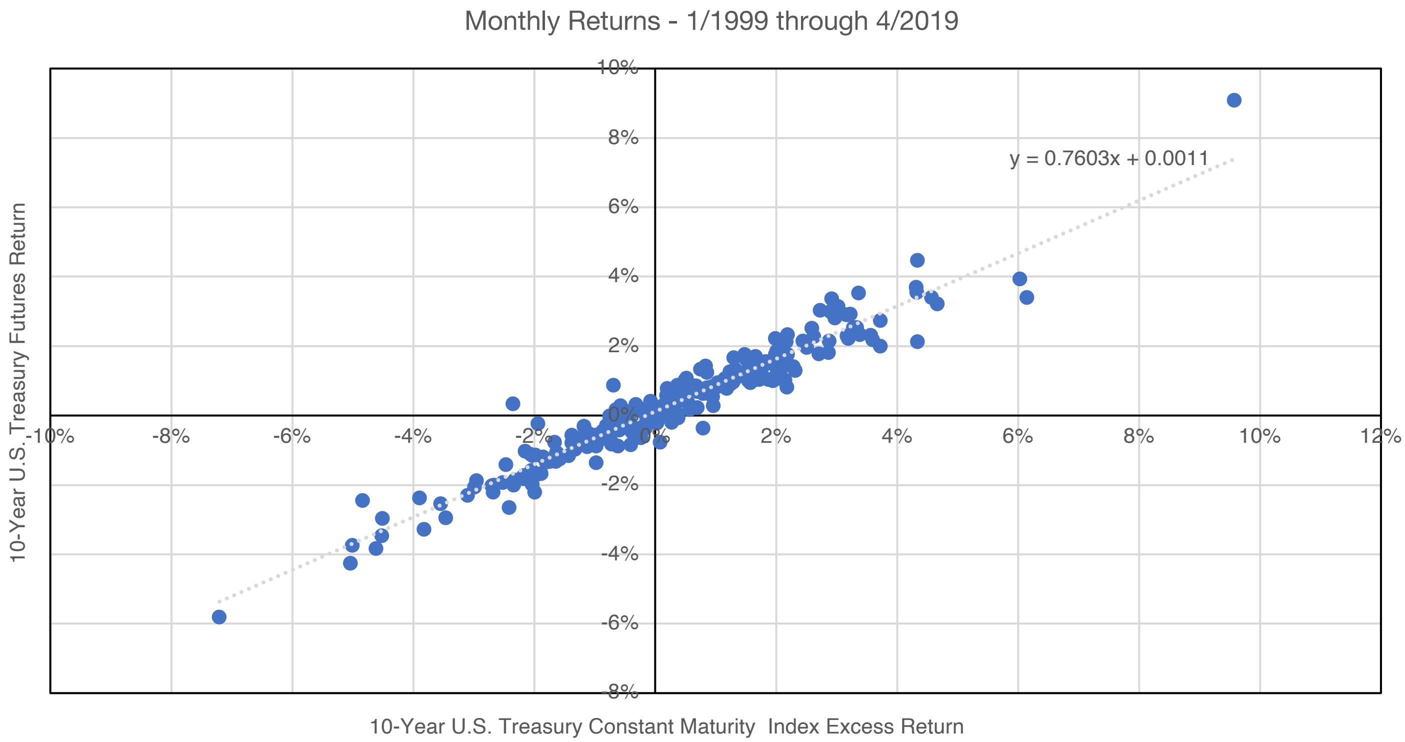
Source: Quandl and Newfound Research. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

It should be noted, however, that a levered constant maturity 10-year U.S. Treasury index and 10-year U.S. Treasury futures are not the same. The futures contracts are specified such that eligible securities for delivery include Treasury notes with a remaining term to maturity of between 6.5 and 10 years. This means that the investor short the futures contract has the option of which Treasury note to deliver across a wide spectrum of securities with potentially varying characteristics.

In theory, this investor will always choose to deliver the bond that is cheapest. Thus, Treasury futures prices will reflect price changes of this so-called *cheapest-to-deliver* bond, which often does *not* reflect an actual on-the-run 10-year Treasury note.

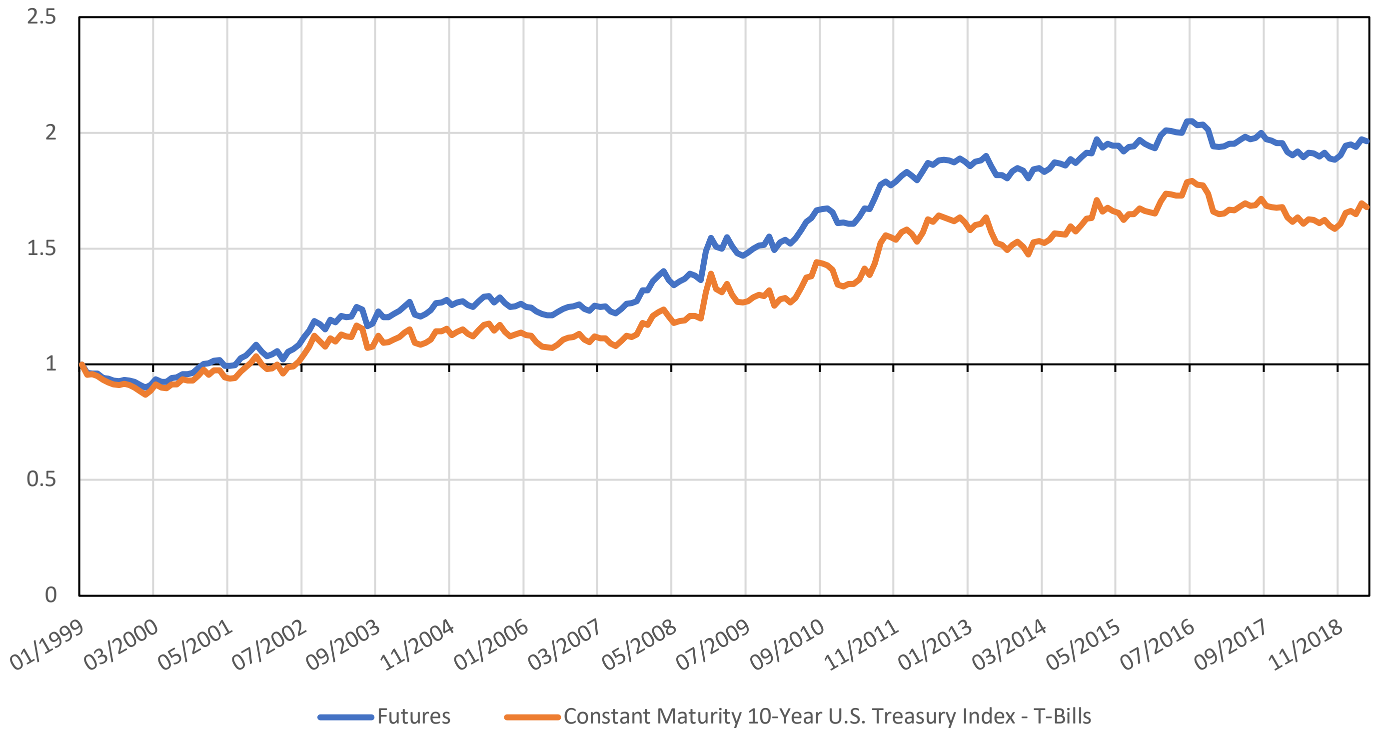
Treasury futures therefore utilize a “conversion factor” invoicing system referenced to the 6% futures contract standard. Pricing also reflects a basis adjustment that reflects the coupon income a cash bond holder would receive minus financing costs (i.e. the cost of carry) as well as the value of optionality provided to the futures seller.

Below we plot monthly returns of 10-year U.S. Treasury futures versus the excess returns of a constant maturity 10-year U.S. Treasury index. We can see that the futures had a beta of approximately 0.76 over the nearly 20-year period, which closely aligns with the conversion factor over the period.



Source: Quandl and the Federal Reserve of St. Louis. Calculations by Newfound Research.

Despite these differences, futures can represent a highly liquid and cost-effective means of implementing a portable beta strategy. It should be further noted that having a lower "beta" over the last two decades has not necessarily implied a lower return as the basis adjustment can have a considerable impact. We demonstrate this in the graph below by plotting the returns of continuously-rolled 10-year U.S. Treasury futures (rolled on open interest) and the excess return of a constant maturity 10-year U.S. Treasury index.



Source: Quandl and Newfound Research. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

Conclusion

In a low return environment, portable beta may be a necessary tool for investors to generate the returns they need to hit their financial goals and reduce their risk of failing slow.

Historically, a 90/60 portfolio has outperformed equities with a similar level of risk. However, the short-term dynamics between stocks and bonds can make the volatility of a 90/60 portfolio significantly higher than a simple buy-and-hold equity portfolio. Rising interest rates and inverted yield curves can further confound the potential benefits versus an all-equity portfolio.

Since constant leverage is not a guarantee and we do not know how the future will play out, moving beyond standard portable beta implementations to tactical solutions may augment the potential for risk management and lead to a smoother ride over the short-term.

Getting over the fear of using leverage and derivatives may be an uphill battle for investors, but when used appropriately, these tools can make portfolios work harder. Risks that are known and compensated with premiums can be prudent to take for those willing to venture out and bear them.

If you are interested in learning how Newfound applies the concepts of tactical portable beta to its mandates, please reach out (info@thinknewfound.com).

COUNTRY ROTATION WITH GROWTH/VALUE SENTIMENT

May 13, 2019

SUMMARY

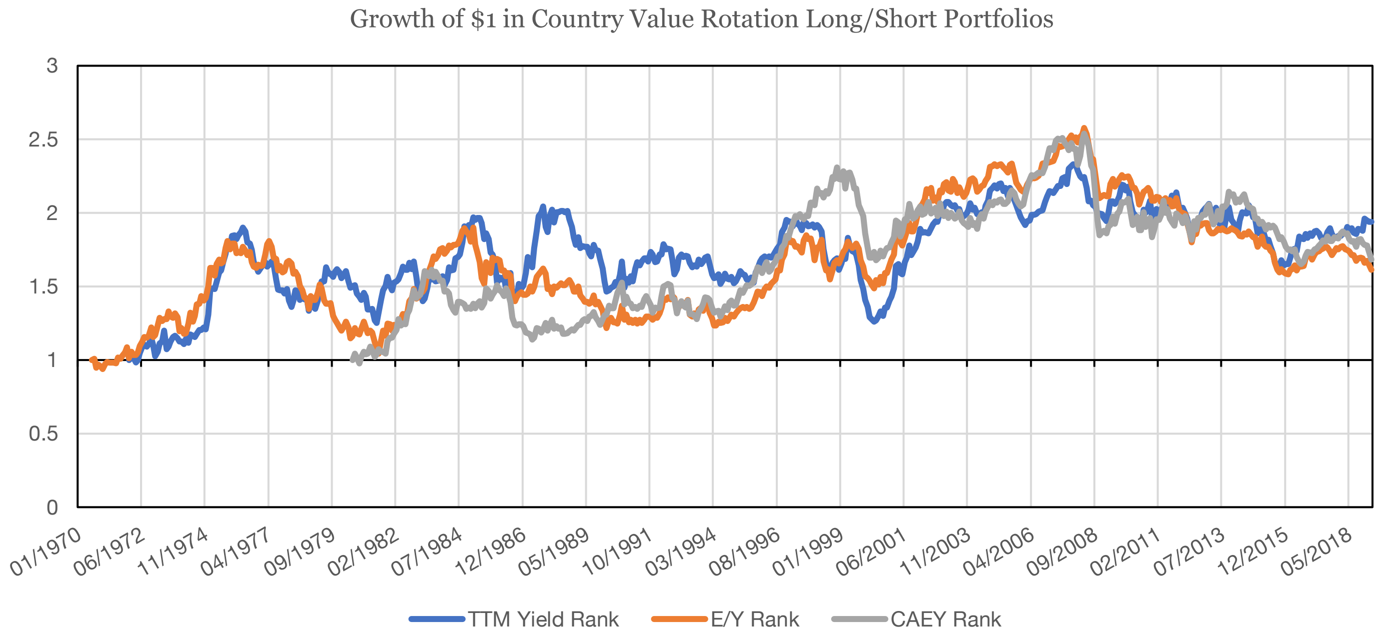
- Value investing has not only underperformed with regard to security selection, but also country selection over the last decade.
- In an effort to avoid country value traps, we set out to design two signals that might better confirm when a country is likely to exhibit positive re-valuation.
- We find that one of the signals exhibits curious results, leading us to develop an entirely new metric for country rotation.
- Initial tests indicate that the signal appears distinct from both traditional value and momentum models.
- From 1976 – 2019, a dollar-neutral long/short portfolio that implements this signal exhibits a gross-of-cost 3.7% annualized return with a 9.6% annualized volatility, implying a Sharpe ratio of 0.39.

As has been well publicized, it has been a tough decade for value. Not just in the realm of stock picking, either: pretty much value *anywhere* has been a tough go.

For example, despite the ample evidence suggesting that valuation-driven country rotation works, it certainly has not worked as of late. To demonstrate, below we plot dollar-neutral long/short portfolios that capture "cheap minus expensive" for developed global markets.[26,27]

[26] Countries include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Israel, Italy, Japan, Netherlands, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, United Kingdom, and USA; inclusion is based upon available data.

[27] The portfolio employs twelve overlapping sub-indexes, each which rebalances monthly on a different month of the year. Weight for a given country, w_i, with rank r_i is calculated as: $w_i = c(r_i - \text{average}(r))$, where c is a scaling coefficient set so that the portfolio is dollar neutral.

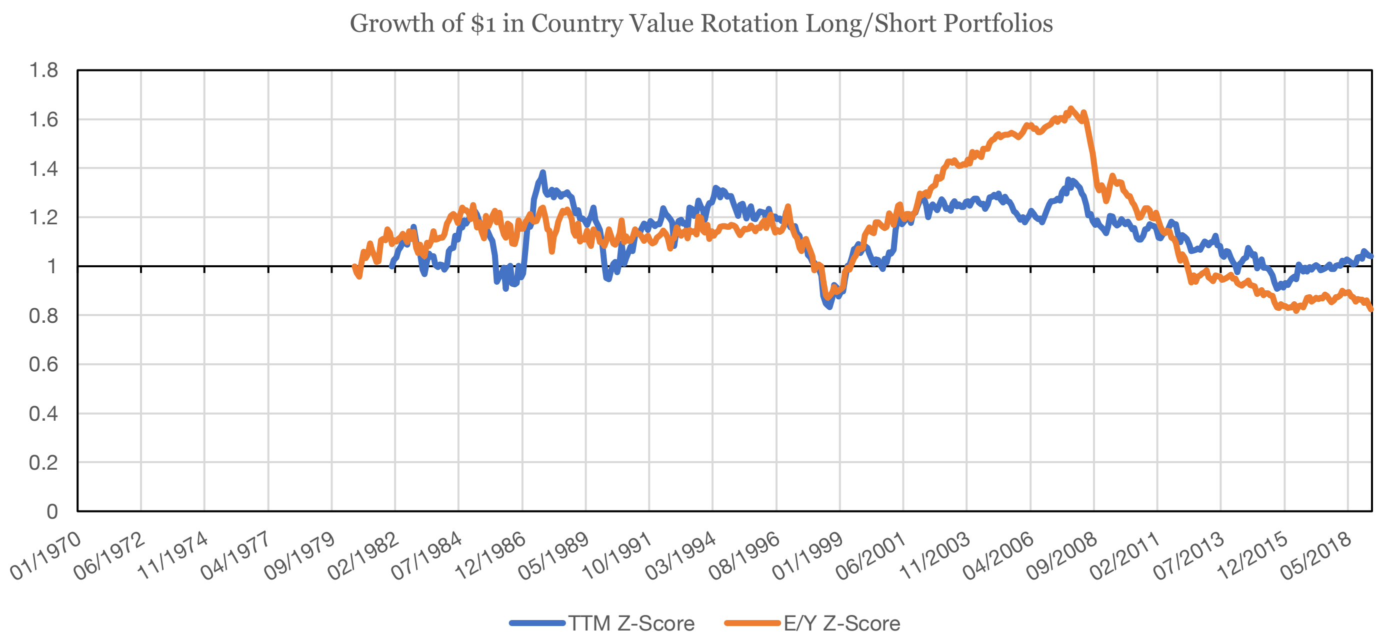


Source: MSCI and Bloomberg. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

(It is worth noting that none of these long/shorts are actually statistically significant at the 5% level, even if we rewind the clock back to 12/31/2006. Nevertheless, we will charge blindly forward.)

Lawrence Hamtil has written quite a bit about how relative valuations may be distorted by sector-based differences in country index composition. One idea for correcting for this is by first normalizing each country's valuation to its own historical valuation distribution. This might help adjust for structural level differences in country valuations that may emerge for sector-driven, sentiment, demographic, or other reasons.

Therefore, before using our value signal to rank across countries, we first transform it by calculating a z-score using the country's prior valuation data. In theory, this should help better normalize the metric and identify meaningful deviations from that country's own "normal" valuation levels.



Source: MSCI and Bloomberg. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

We can see that self-normalization does little to help the problem (and, arguably, made valuation-driven country rotation worse).

The problem here might simply be more trivial: expensive just has out-performed cheap.

Fighting the Zeitgeist

Share prices fall when market participants adopt a negative view on a security, industry, or country's fundamental outlook.

The relationship is clear if we adopt a simple dividend growth model for share prices; i.e. $P = D_1 / (R - G)$, where P is price, D_1 is the annual dividend level, R is the required return, and G is the growth rate. As G declines, the denominator increases, and price decreases.

Note that if Dividend_1 is held constant and the growth rate declines, we will expect price to also decline, which will lead to an increase in yield. And if we replace D_1 with Payout Ratio x Earnings_1, we can also see a clear picture of how earnings yield and growth rates are similarly related.

Re-arranging the formula to solve for return, we find that $R = (D_1 / P) + G$. This formula tells us that return should be equal to dividend yield plus our long-term dividend growth rate. Thus, if our required rate of return does not change and the outlook for a company's growth declines, we need to see a decrease in a price to create a commensurate increase in yield.

Taken together, we can see that as the market outlook for a given security deteriorates, valuation multiples (e.g. P/E) should decline. Should earnings growth realities prove to be less gloomy than forecasts imply, however, price will appreciate and valuation multiples will expand.

We should also note that a change in aggregate risk preferences among market participants should have a similar effect. A decrease in risk appetite should manifest in an increase in the required rate of return. In our first formula, this informs us that price will decrease. In our re-written formula, if our outlook on growth does not change but our appetite for risk does, then price must change to affect yield.

However, given that fundamental measures tend to be less volatile than price, an investor who purchases a security at a low valuation multiple and sells at a high valuation multiple will earn a return not necessarily due to higher realized yield, but rather a re-rating of future fundamentals that materializes in a price increase.

For an investor purchasing a stock today, we might say that their future return is: $R = (D_1 / P) + G + V$, where V capturers re-valuation over the holding period.

What does this all have to do with country rotation? Re-valuation will be driven not only by changes in growth outlook, but also changes in risk *sentiment*. For example, an aggregate decrease in risk appetite among market participants will appear as a contraction in valuation multiples, while an aggregate increase in risk appetite will appear as an expansion.

Thus, our question: can we avoid country-level value traps by understanding country-level sentiment?

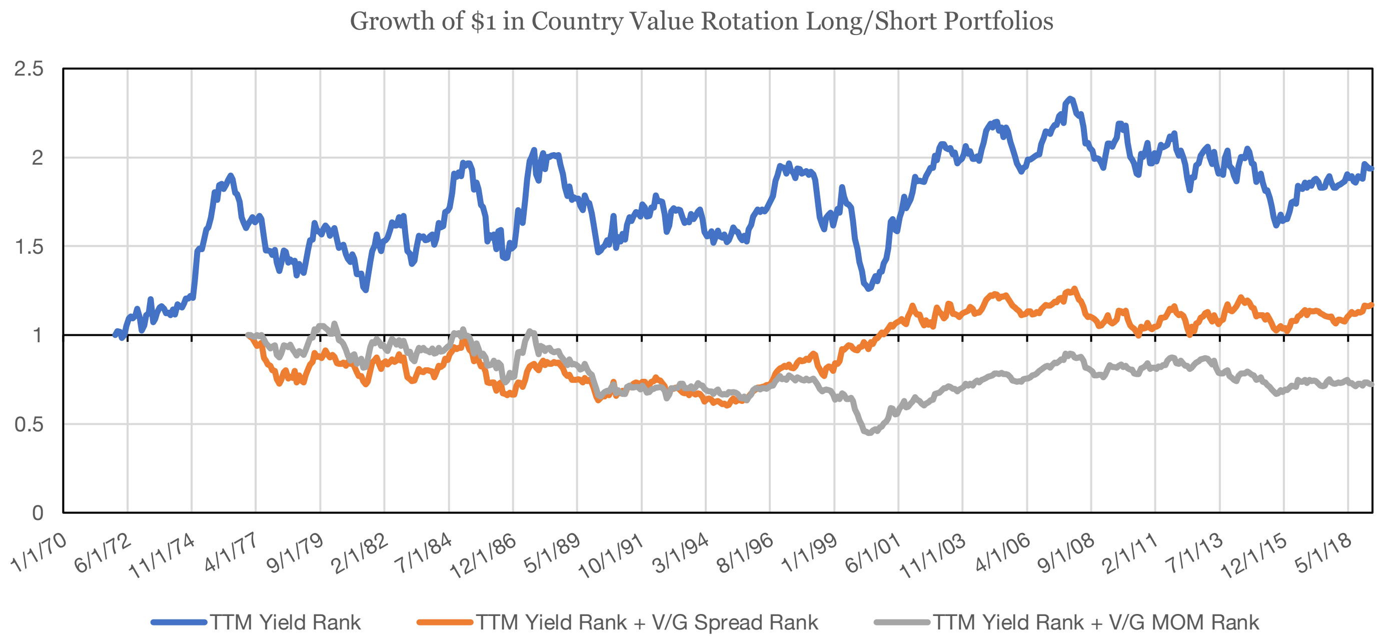
Measuring Sentiment with Value Dispersion

To measure evolving sentiment / risk appetite for a given country, we will examine two different metrics:

1. *Value/Growth Valuation Spreads*: High valuation dispersion between growth and value indices may reflect a better entry point for buying cheap countries. Specifically, for each country we calculate the trailing 12-month yield of that country's value index minus the trailing 12-month yield of the corresponding growth index. We then subtract the long-term mean (using an expanding window) to better identify cyclical changes.

2. *Value/Growth Momentum*: When value is out-performing growth within a country, it may reflect a better opportunity for capturing re-valuation changes in cheap countries. Specifically, for each country we will calculate the trailing 12-month return of that country's value index minus the trailing 12-month return of the corresponding growth index.

To construct our portfolios, we will first rank each country on the value metric. We will then rank each country on the sentiment metric. We will then add these ranks to create a composite score, and re-rank on this composite score and construct our dollar-neutral long/short as before.



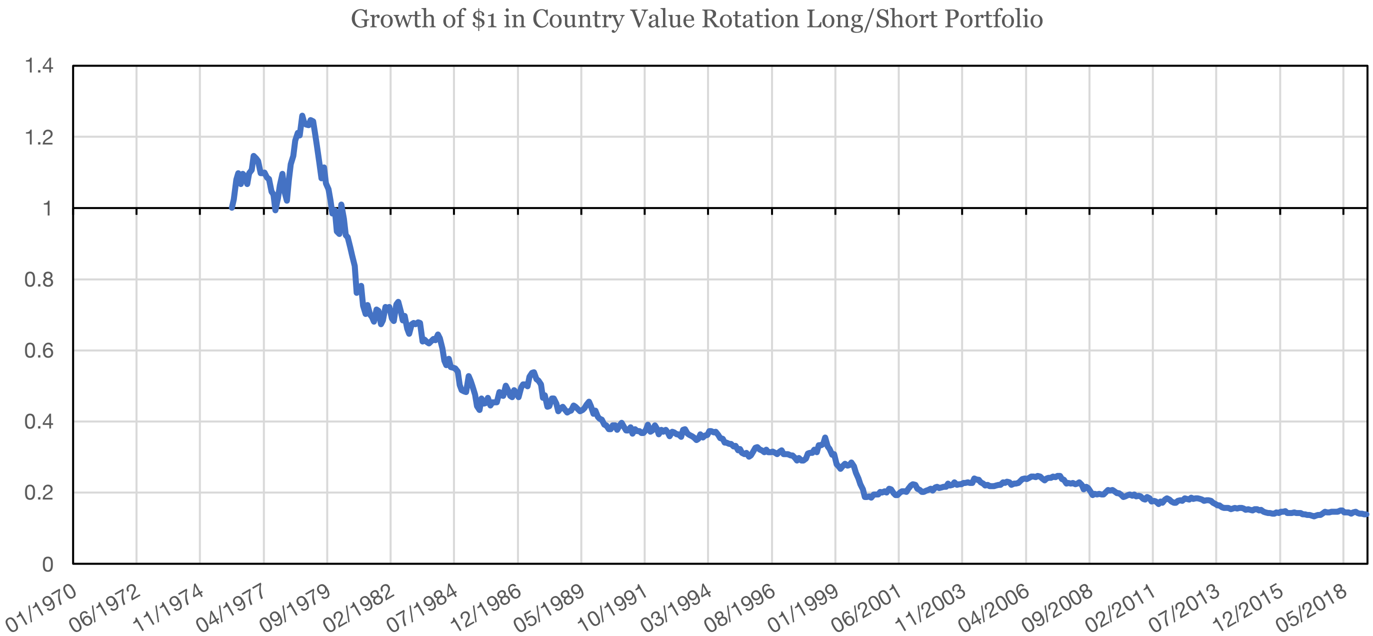
Source: MSCI. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

Well, that's tragic.

Growing Sentiment

Something is a bit odd though. The addition of the value/growth momentum metric turned a positive (albeit statistically insignificant) return stream negative. This is a bit of a head-scratcher and perhaps worth exploring more deeply. What does this signal look like on its own?

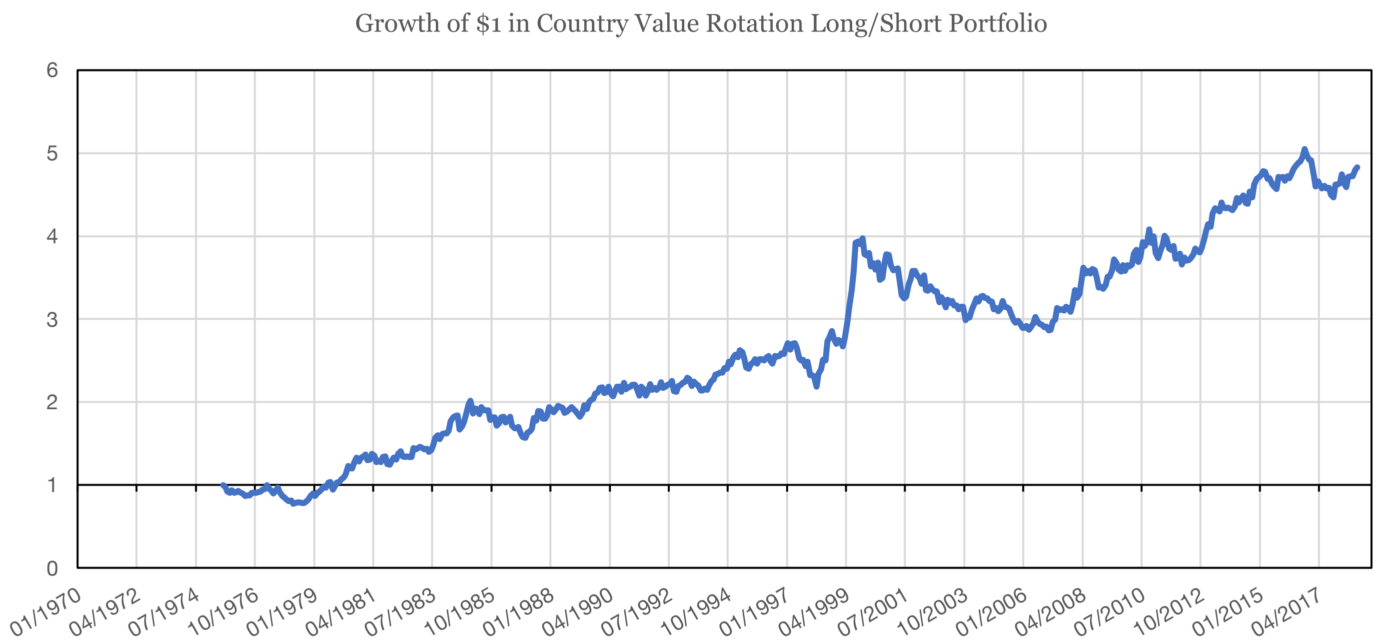
Below we plot a dollar-neutral long/short constructed on this signal alone (i.e. rank countries based upon prior return spread of respective growth and value indices).



Source: MSCI. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

That return stream is so bad, it might just be good.

After all, if we can *short* this return stream, we could have a pretty good result on our hands. And in this case, shorting the return stream is effectively the same as just inverting our signals. In other words, instead of ranking on the prior return of value minus growth, we should rank on the prior return of growth minus value.



Source: MSCI. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

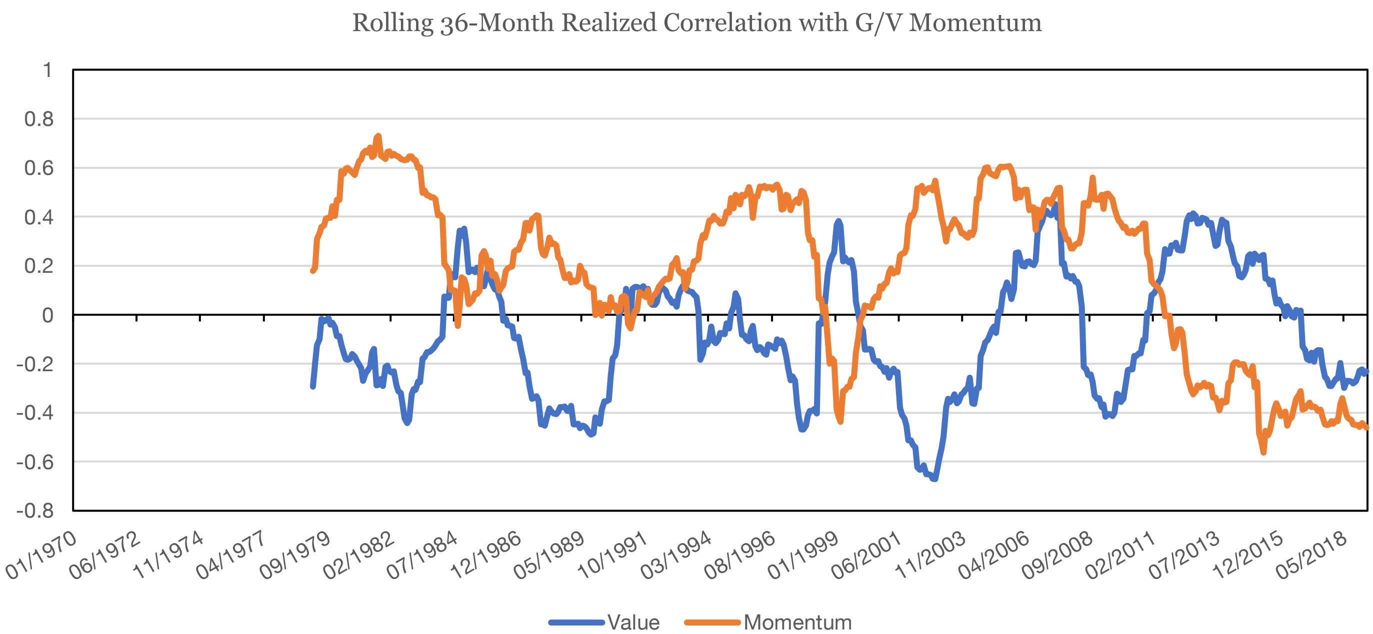
But is this signal just another way to extract momentum? After all, we are using a measure of prior return. To answer this question, we computed a prior 12-month return signal on country indices explicitly and constructed the corresponding long/short index. Below we plot the 36-month realized correlation between the indices constructed on TTM yield signals ("Value"), prior 12-month return signals ("Momentum"), and the 12-month return spread between growth and value indices ("G/V Momentum").

While both Value and Momentum have sporadic periods of strongly positive or negative correlation to G/V Momentum (with the mirror-like relationship reflecting the well-documented negative relationship between Value and Momentum), we can see that, for the most part, absolute correlations are benign. Furthermore, the correlations are meaningfully time-varying, suggesting that there might be unique information found in the G/V Momentum signal.

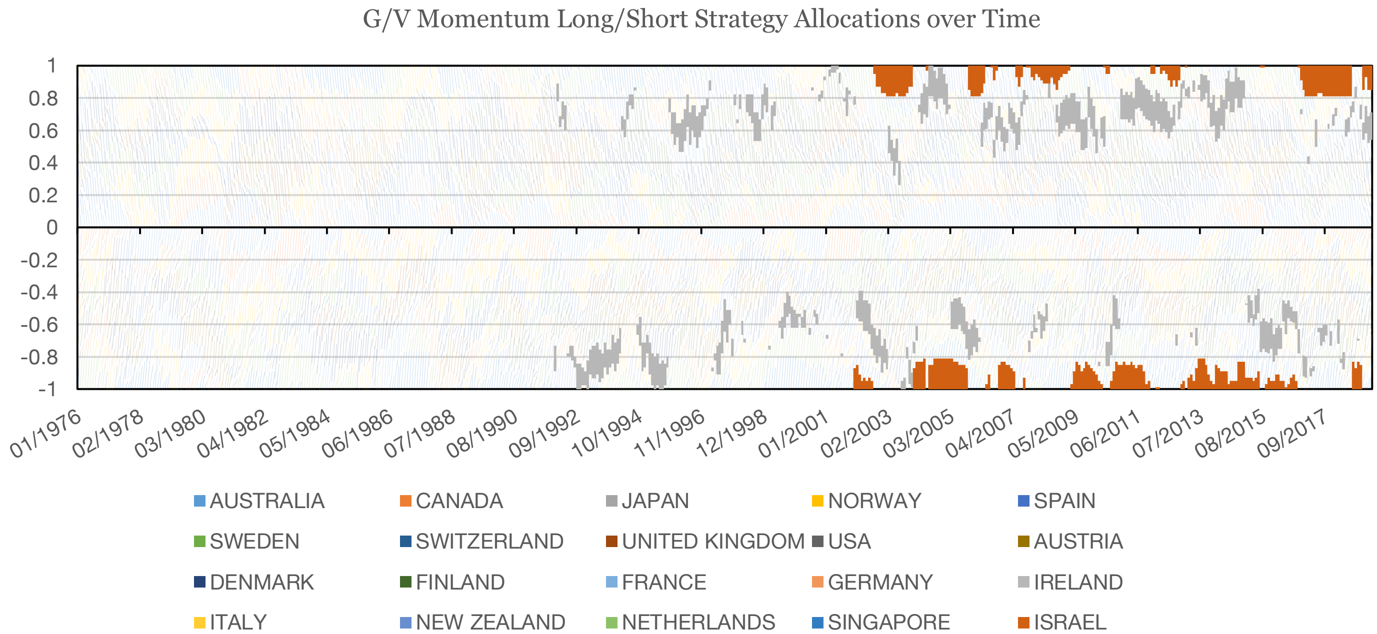
What is most curious about this signal is that the strategy does not invest in the spread itself. We're not saying, for example, "when growth is outperforming value, buy the country's growth index." Rather, a higher (lower) cross-sectional return spread between growth and value implies a higher (lower) cross-section return for a given country index.

Why that relationship exists is certainly worth pondering.

One possible explanation is that the prior return of growth minus value captures some measure of market cycle sentiment. An increasing spread may reflect an increasing intra-market risk appetite for growth stocks over value stocks, potentially signaling a positive sentiment shift. This *relative* appetite change may, therefore, foreshadow a broader, market-wide increase in risk-appetite, leading to a positive re-valuation of economic expansion.



Source: MSCI. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.



Conclusion

In this commentary, we set out to fix value-driven country rotation, an approach which has largely exhibited negative returns for the last decade. To do so, we designed two signals that aimed to capture internal market sentiment for a given country. The first signal – measured as the spread in yield between value and growth – aimed to capture intra-country value spreads. The second signal – measured as the prior total return of value minus growth – aimed to capture the intra-market risk appetite for value stocks.

In combining these signals with country value signals, our hope was to buy cheap countries with market internals that implied a greater likelihood of re-valuation.

Unfortunately, we failed miserably.

In our failure, however, we identified that as a stand-alone signal, the prior return of value minus growth lead to a long/short return profile that lost money consistently. In flipping the signal to the prior return of growth minus value, we are able to generate a long/short strategy that not only exhibited an attractive return profile but had also exhibited meaningfully positive returns over the last decade.

Interestingly, it appears that the signal is not merely a proxy for momentum and exhibits low realized correlations to both momentum and value strategies, indicating that it might contribute unique information on its own. *What*, precisely, this signal is capturing remains a mystery, but we believe it warrants further research.

Identifying the *why* behind a signal can be particularly risky, as we can fall prey to constructing a narrative to fit a signal, rather than identifying a signal from a starting hypothesis. This risk might be heightened in this case where the signal we identified was specifically the antithesis of the hypothesis we set out to test.

Even if we were convinced of the *why*, it is worth noting that this brief note only scratches the surface of research that would need to be performed before even considering implementing a signal like this. For example:

- We should determine how sensitive the strategy is to changes in the investment universe. Not only should we see if we can replicate this signal in an entirely different universe (e.g. timing sectors with intrra-sector growth vs value), but also determine how sensitive the results are to changes in the universe examined (e.g. robustness tests via subset resampling).

- We would want to determine strategy sensitivity to model specification. This would require changing not only the return model employed (e.g. risk-adjusted returns, regression slopes, or short-minus-long moving averages) as well as the formation period.

- We would want to determine strategy sensitivity to rebalance timing luck, as the model examined in this brief note only evaluated end-of-month rebalancing.

- We would want to quantify the decay speed of the signal to better determine the appropriate holding period.

- We would need to identify how much return is available after cost and tax adjustments, as well as appropriate discounting for data-mining risk.

- We would want to explore the shape of quintile returns and volatility profiles to determine what type of signal we're working with. For example, is the excess return potential generated from a monotonic increase in returns across quintiles, or is it simply from avoiding the worst quintile? This will have important implications for portfolio construction as well as whether this approach can be ported into a long-only model.

It is only after we are comfortable that this signal is reasonably robust across specifications and survives after conservative cost assumptions are applied should we consider employing it.

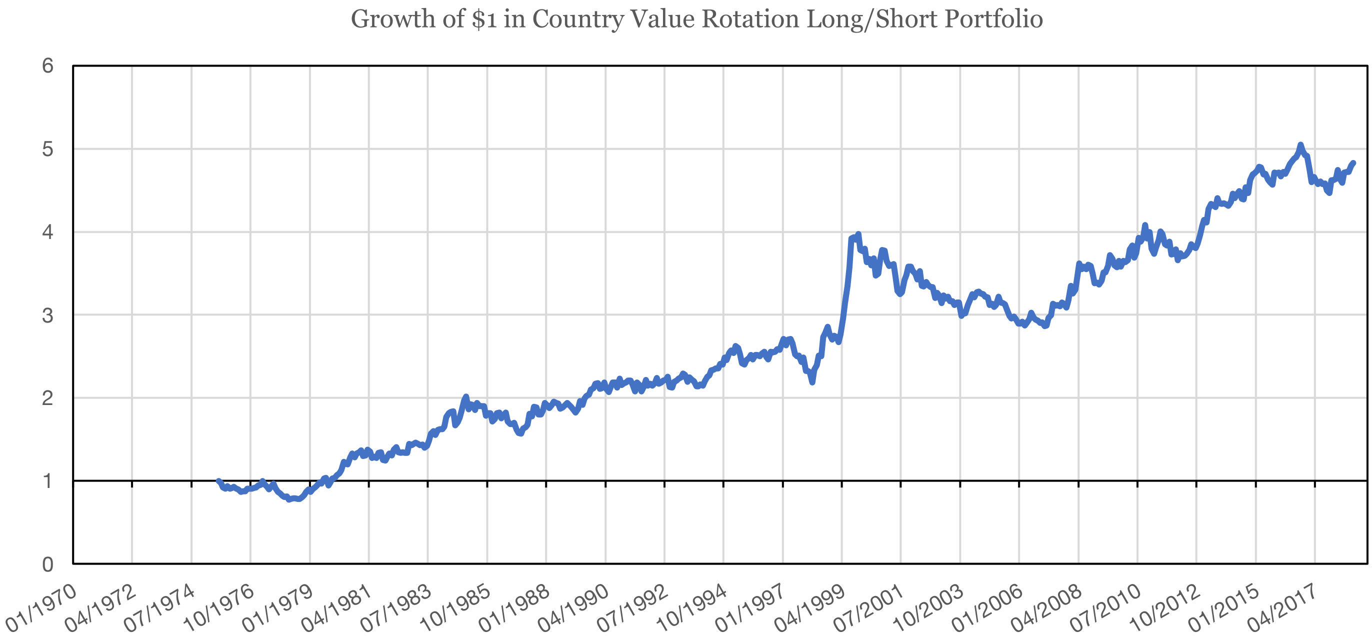
DISPROVING A SIGNAL

May 20, 2019

SUMMARY

- Last week we introduced a signal that appeared to generate statistically significant performance results for performing country rotation.
- This week, we walk through the steps taken to explore the robustness of the signal.
- We first explore out-of-sample data with sector and emerging market country indices. Unfortunately, definitional differences and limited data impact our ability to pass judgement.
- We then use simulation methods to test the robustness of the original strategy.
- We find that just four countries are responsible for nearly 100% of the original strategy's performance: Australia, Austria, Denmark, and Spain. Removing these countries from the eligible universe makes the signal become insignificant.
- In performing a similar analysis for a standard momentum signal, we do not find the same impact from a limited subset of countries, indicating that our signal is likely an artifact of data-mining.

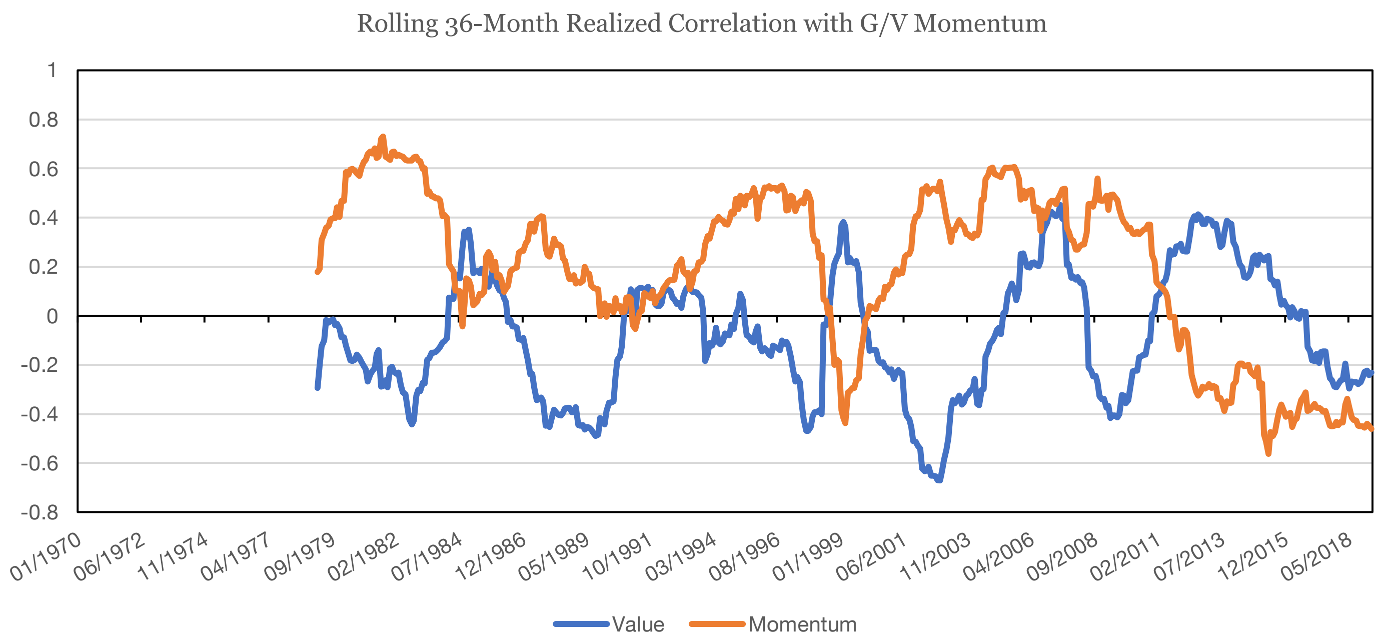
In last week's commentary (*Country Rotation with Growth/Value Sentiment*) we introduced a signal we had stumbled upon for country rotation that had a pretty attractive backtest.



Source: MSCI. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

The signal was generated by calculating the prior return spread between a country's growth and value indices (which we'll call "GMV" for the remainder of this commentary) and using the cross-sectional rank of the return spread to allocate to countries. Countries for which growth had dramatically out-performed value were likely to receive an overweight, while countries for which value had dramatically out-performed growth would receive an underweight.

What made this signal even more interesting was that it seemed to exhibit little-to-no meaningful correlation with traditional factors like momentum and value, perhaps indicating that it captured some additional source of information.



Source: MSCI. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

The danger of such a signal was that we had discovered it by accident. Well, worse. We discovered it by inverting a hypothesis that had failed spectacularly. This left us grasping for narratives that might confirm our data bias. We postulated – in-line with the initial hypothesis – that the growth-minus-value signal was some sort of sentiment indicator.

In publishing our commentary, we hoped to invite responses from the community as to whether there was prior literature about similar signals or how this signal might be capturing other effects.

"It's Just Momentum"

By and large the most consistent feedback we heard was, "it has to be related to momentum."

The connection seems apparent, given that we are calculating the prior return spread of growth minus value, much like a momentum signal which relies upon the prior return of the country itself.

The most likely way we hypothesize that this could occur was if growth indices had higher country beta coefficients than their value counterparts. If this were the case, a dollar-neutral growth-minus-value calculation would create a positive beta residual. Prior total return of this spread, then, would be highly correlated to the prior total return of the beta residual, and therefore momentum.

Unfortunately, this was not the case. The relative betas of growth-minus-value indices were not only meaningfully time-varying, but also exhibited both positive and negative signals, indicating that growth-minus-value may, in some cases, be a negative beta signal.

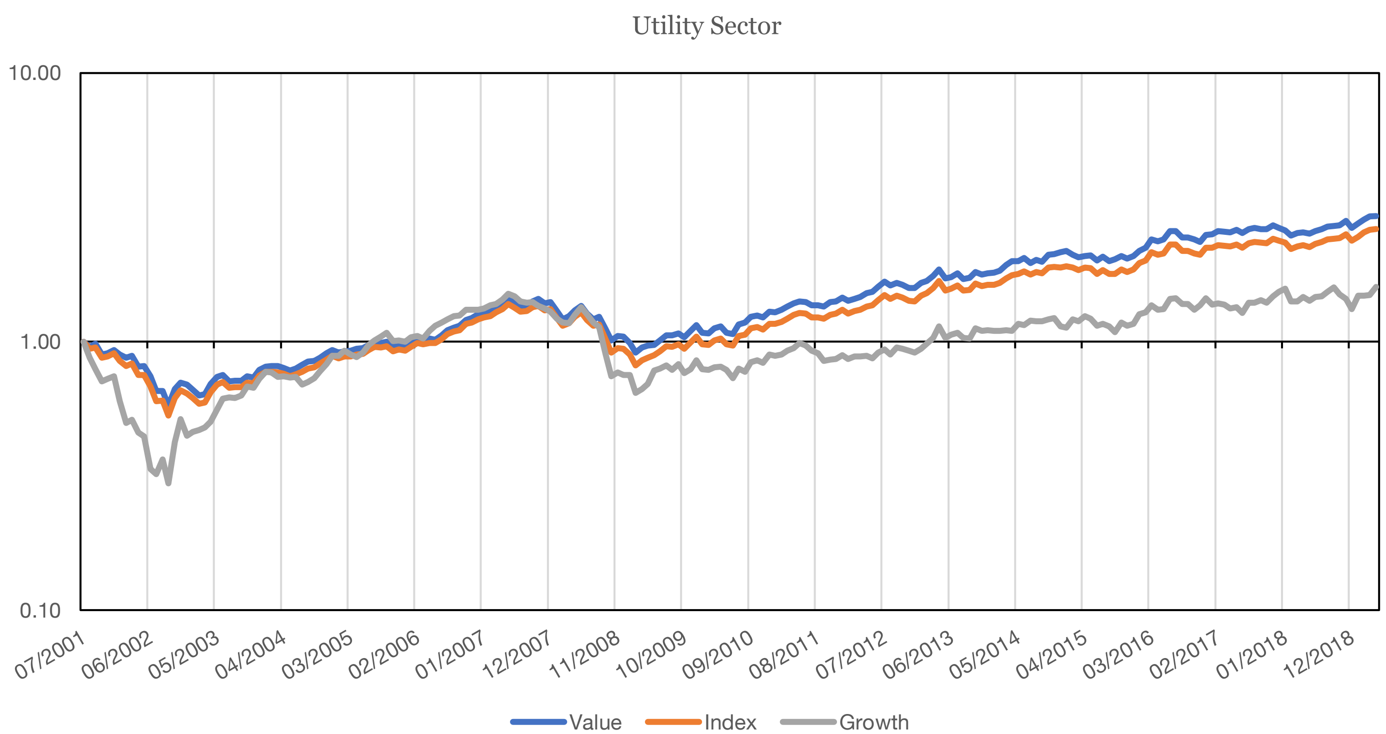
Given the low realized correlations and the lack of evidence of consistent beta effects, we did not feel that this signal was just cloaked momentum.

Out-of-Sample Tests

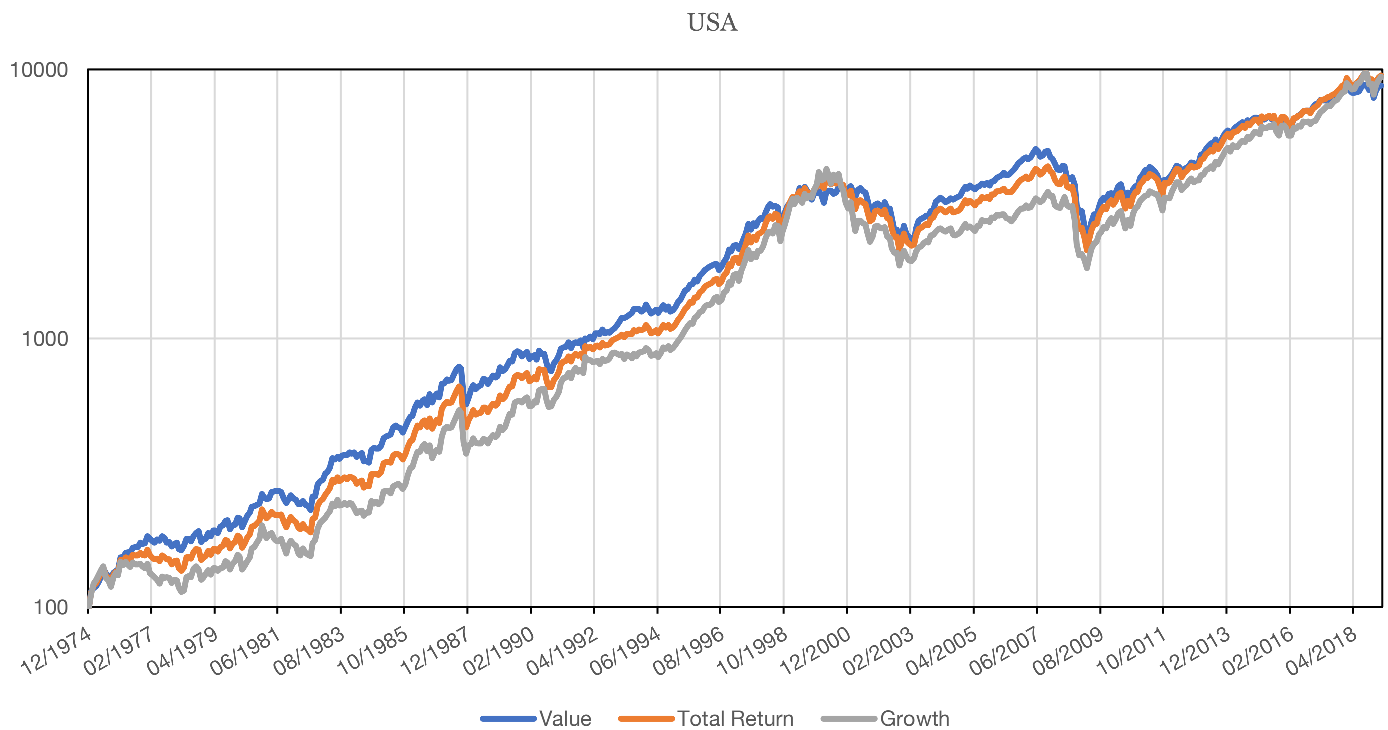
Our first test for robustness was to perform out-of-sample tests. We evaluated two different datasets: U.S. sectors and emerging market countries.

To cut right to the chase: the signal did not appear to work on the sector data. Unfortunately, it was difficult to tell whether it did not work because the signal did not work, or because the construction methodology of the growth and value indices from Russell was meaningfully different from the country data from MSCI we were employing.

While we would hope our signal would be robust to minor definitional tweaks of growth and value, what we observed in sector data was significant disparities in how sectors were decomposed versus countries. For example, below we plot total return, growth, and value data for the Russell U.S. Utility Sector and the MSCI USA indices.



Source: Morningstar and Russell. Returns are hypothetical. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.



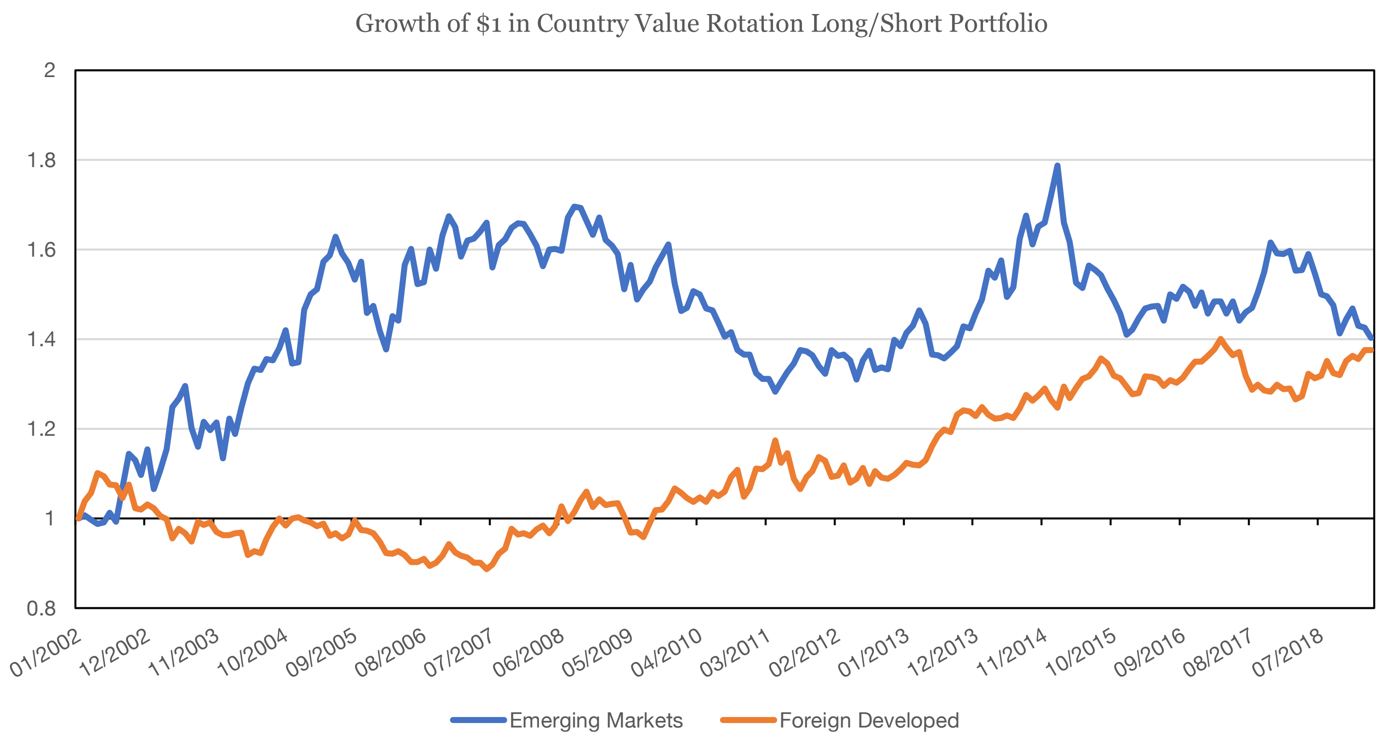
Source: MSCI. Returns are hypothetical. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

We can see that while the Utility sector tends to behave more consistently like its value definition, the USA total return index more evenly split its growth and value counter-parts. This was a consistent artifact in the data: sectors tended to skew towards either growth or value, while the country data bisected the two.

We would hypothesize that this is because the Russell methodology likely defines the value and growth indices of a sector as those securities that are in a market-wide value or growth index that fall in that sector, rather than bisecting the sector itself. Country data, on the other hand, represents a more even market-capitalization split, such that growth plus value should approximately equal the total index.

Whether the failure of the GMV signal on sectors was a failure of the signal or an artifact of the data is difficult to determine.

Fortunately, we could look towards another data set: emerging market country data from MSCI. This data would avoid the potential definitional issues of growth and value. Dropping the data in, we calculated the same long/short index returns.



Source: MSCI. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

Despite the positive total return over the period for, the realized volatility for the Emerging Market strategy was so high, we cannot reject the null hypothesis that the positive return was due to luck alone.

Unfortunately, over this same period, the same can be said for the original Foreign Developed index data we calculated. Which leaves us in a bit of a tough spot. The signal failed to work in two out-of-sample scenarios, but with two highly plausible excuses: (1) a data mis-match and (2) too short a time-horizon for evaluation.

Deriving the Big Muscle Movements

Left somewhat high and dry by our out-of-sample tests, we wanted to turn our attention to evaluating the robustness of the original signal we developed.

One of the key questions to ask when evaluating signals such as these is, "where is the return coming from?"

Not only should we ask this from a qualitative perspective, but we should also evaluate the return impact of underlying holdings. We can already see from the time series that the results are not due to a single trade, but it is possible that they are due to a structural overweight to just a few holdings.

While it sounds like a trivial exercise to back out return attribution, it turns out to be rather complicated when you have to take into account the effects of compounding (look up "Frongello linking" if you want to go down that rabbit hole). Fortunately, this is a case where we need directional guidance and not exact precision.

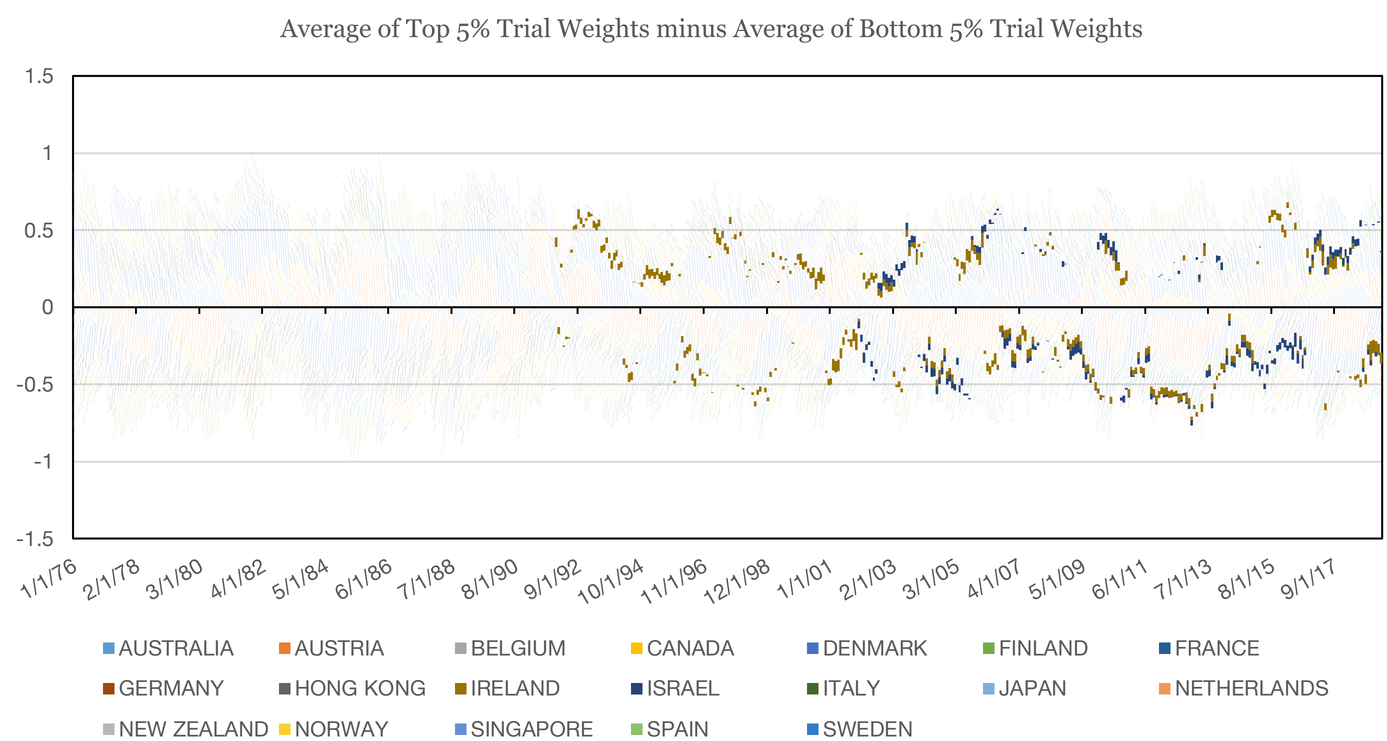
Thus, we elected to just blindly throw some computing horsepower at the problem.

Specifically, we performed the following exercise:

1. Of the 16 possible countries, select 10 randomly
2. Run the strategy on the subset of 10 countries
3. Calculate and store annualized return of the strategy as well as associated country weights over time
4. Repeat 10,000 times

We then sorted trials by their annualized returns. Trials falling in the worst 5% of realizations had their corresponding weights averaged together, as did those trials falling in the top 5%.

The purpose of this exercise is to look at the weight differences of trials falling in the best- and worst-case scenarios. The *differences* in weights will tell us how these scenarios differ.



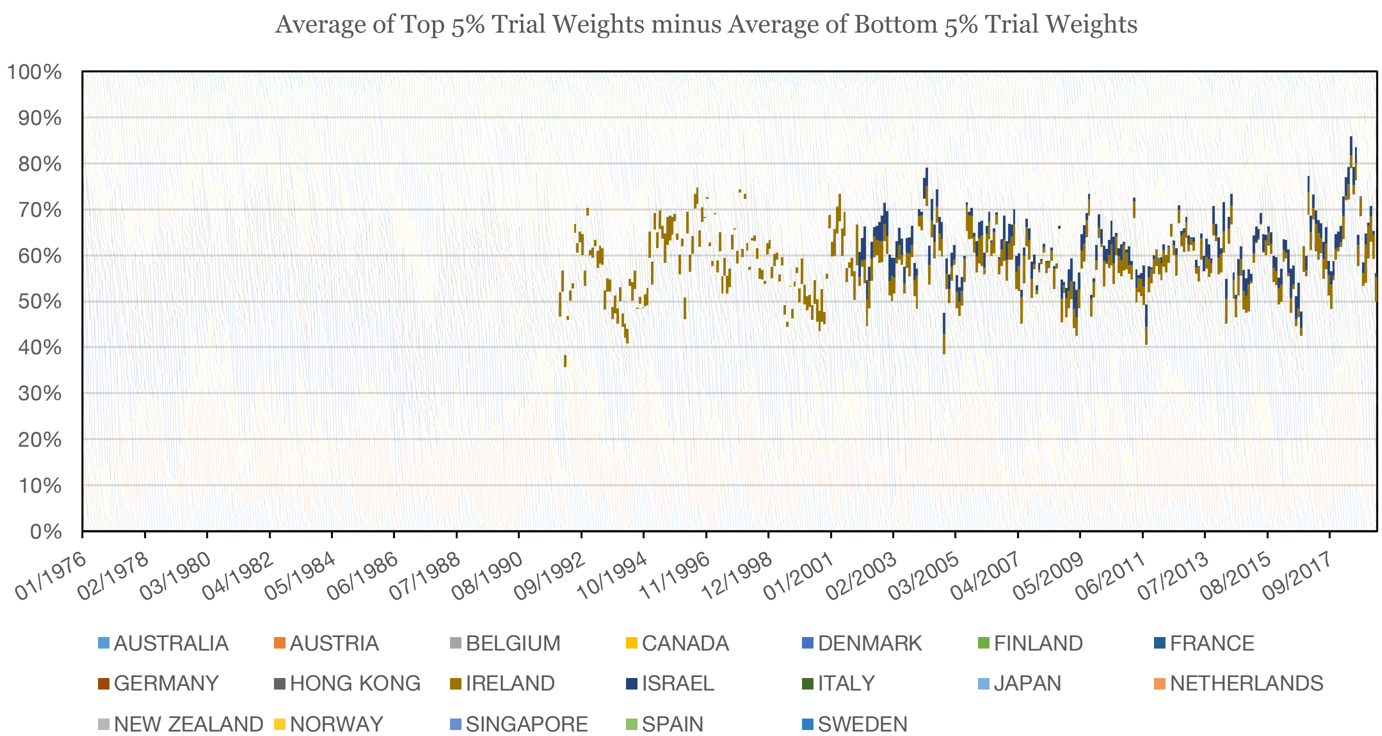
Source: MSCI. Calculations by Newfound Research.

We'll be the first to admit that this graph is not the easiest to interpret. Generally speaking, it says, "weights above the zero line imply larger allocations in top 5% trials, while weights below the zero line imply larger allocations in the bottom 5% of trials."

Two things are apparent.

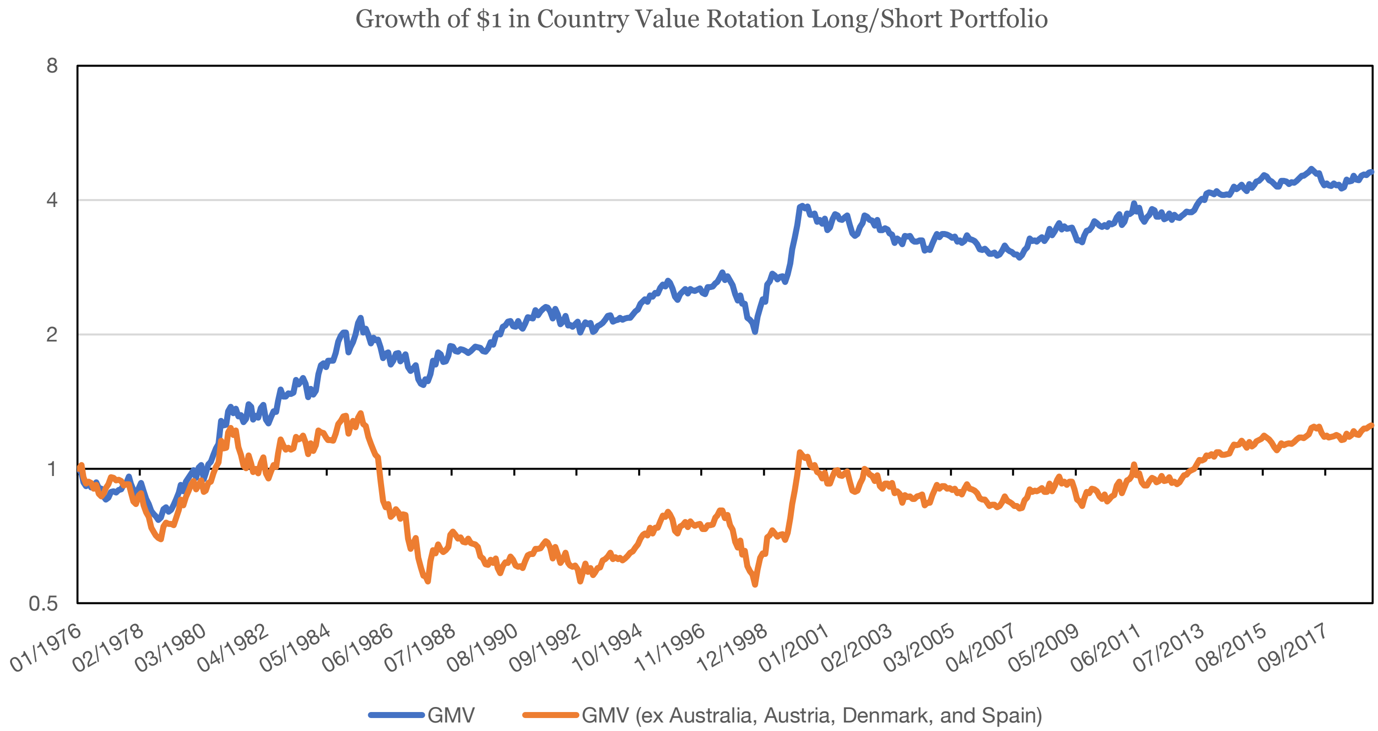
First, the relative weight in the top 5% versus the bottom 5% is highly time varying, with countries spending time in both. This implies that success of the strategy was not likely due to a structural over- or under-weight towards certain countries.

Certain countries do appear, however, to maintain large absolute weights, implying that they might have an out-sized impact on performance. We can get a better idea of this by looking at the normalized absolute value of weights over time.



Source: MSCI. Calculations by Newfound Research.

We can see that Australia, Austria, Denmark, and Spain appear to maintain consistently high weights relative to other countries. What happens, then, if we simply remove them from the original 16 and re-run our strategy?

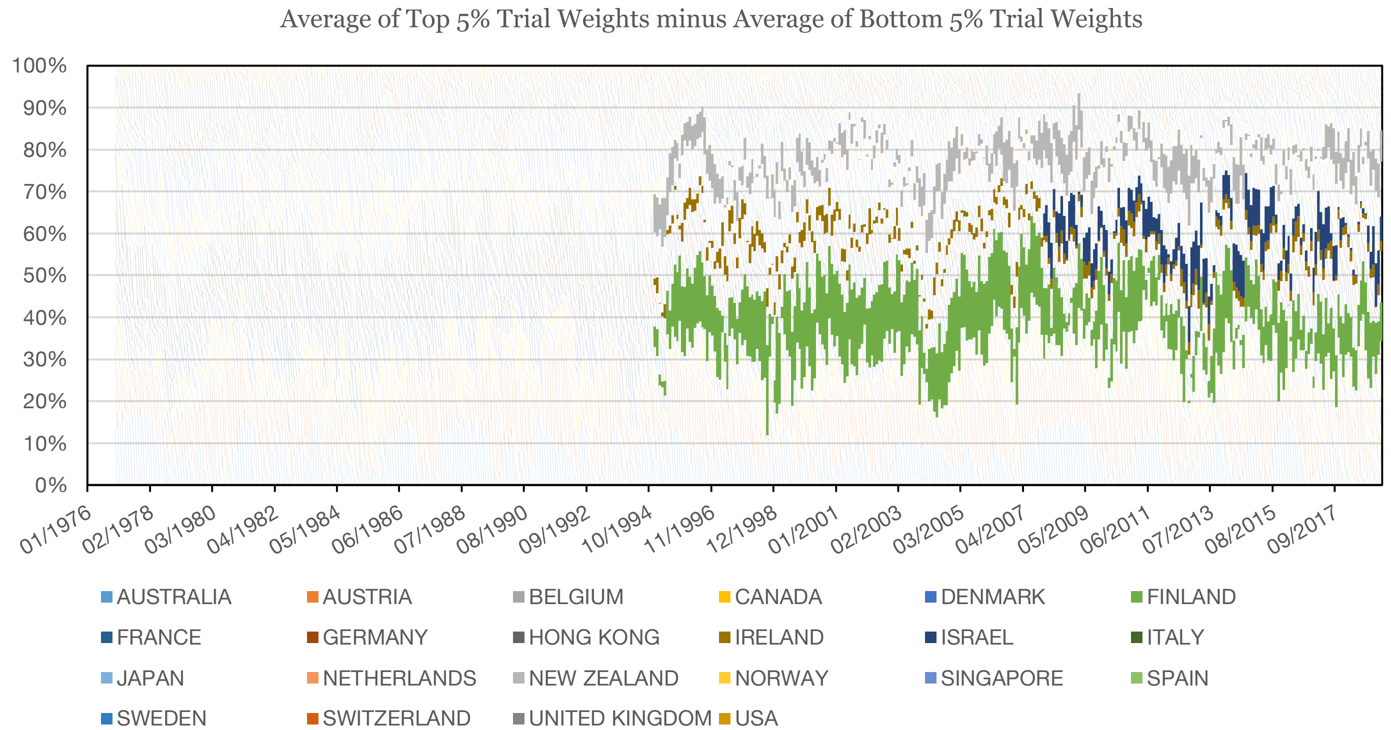


Source: MSCI. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

The performance entirely disappears. This indicates that the entirety of the result in the original strategy can be attributed to allocation decisions on just four countries: Australia, Austria, Denmark, and Spain.

The fact that just four countries drove the entire performance should certainly give us reason for significant pause. Not least of which is the fact that even if we did believe this signal was robust, implementing it in a long-only manner now becomes almost entirely impossible, as “under-weighting” these countries relative to an ACWI benchmark would have little-to-no impact.

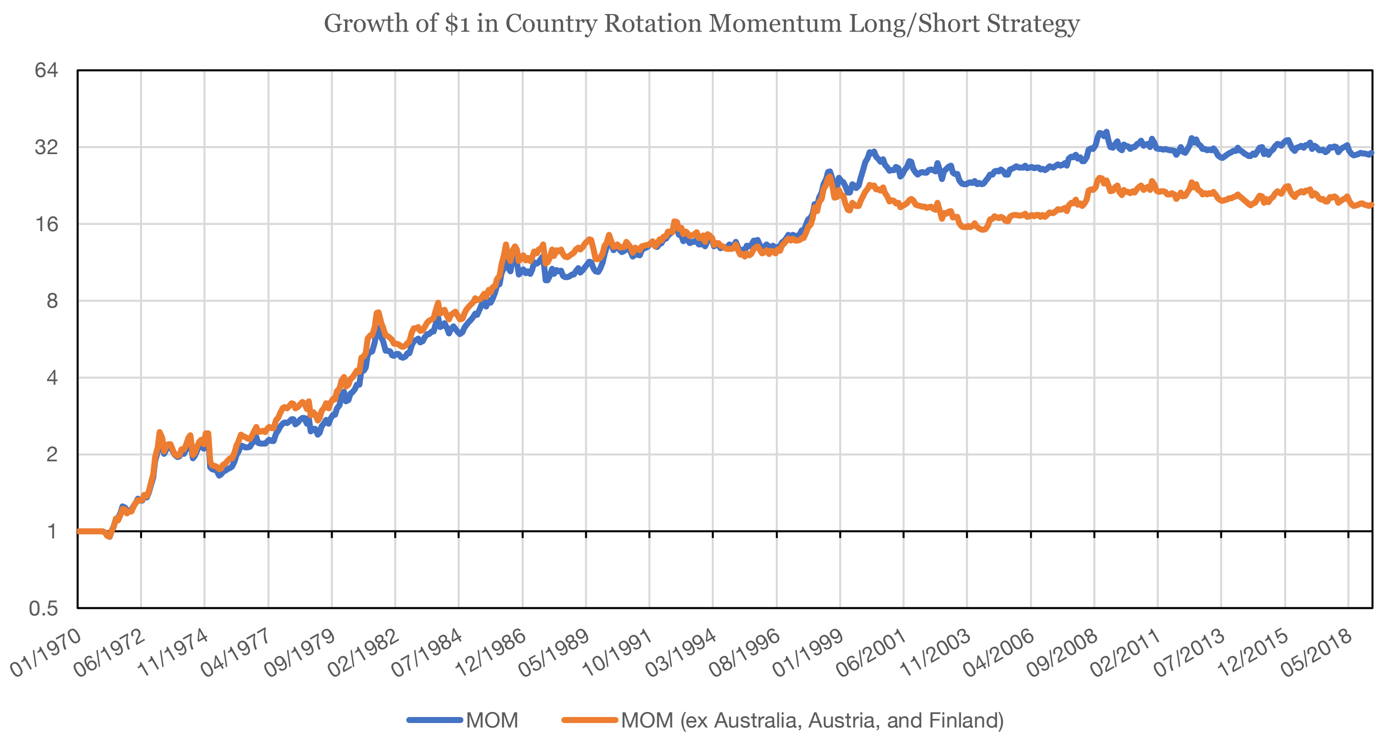
But is it all that unusual? Consider the same analysis for a standard momentum signal.



Source: MSCI. Calculations by Newfound Research.

We can see that Australia, Austria, Finland and the United Kingdom (at least in early years) are stand-out weights[28]. If we remove them from the eligible universe, what happens to the momentum factor?

[28] Interestingly, the overlap of Australia and Austria might lend credence to the notion that the GMV signal was picking up on some momentum effects.



Source: MSCI. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

We can see that the momentum factor largely maintains its mojo. This indicates that our GMV signal is likely just an artifact of data-mining, while MOM has a higher likelihood of being an anomaly (though, 20 years of sideways performance does give cause for concern; especially when we consider that not only does this analysis not consider cost, but prior to 20 years ago, implementing this sort of strategy would have been highly difficult for most investors.)

Conclusion

If you stare at the same data long enough, you're bound to find spurious patterns. In this case, the pattern we found gave us extra pause as we discovered it by doing the exact opposite of our initial hypothesis. That alone is not enough to dismiss a signal, but we should be aware of trying to fit a narrative to a signal after we discover it, particular when the signal flies in the face of the original narrative we were testing.

In this case, our country growth-minus-value signal proved to be an artifact of data-mining. While out-of-sample testing left us with no solid conclusion, using simulation techniques to back out the driving allocations of performance allowed us to identify that only 25% of the investible universe was responsible for 100% of the strategy returns.

While we would expect *some* degradation in performance when we remove exposures identified as being high impact, the elimination of return tells us that the remaining 75% of the universe had zero meaningful contribution. By comparison, when we perform a similar analysis on a standard momentum signal, we find that removing the three most impactful countries only reduces annualized returns from 7.3% to 6.3%.

Even if we did believe in the efficacy of the signal, deriving the meaningful drivers of performance from an allocation perspective is important because it informs us about implementation. In this case, the large drivers all have relatively small market-capitalizations, which would make it difficult to create meaningful "underweights" in a long-only portfolio. Thus, if we were to pursue this idea further, it would likely have to be in a long/short capacity.

Of course, there is no need to pursue it further. Forcing something to work over a backtest is not likely to end up well on a live basis. Like most ideas we explore, the lack of robustness lands it in our research graveyard.

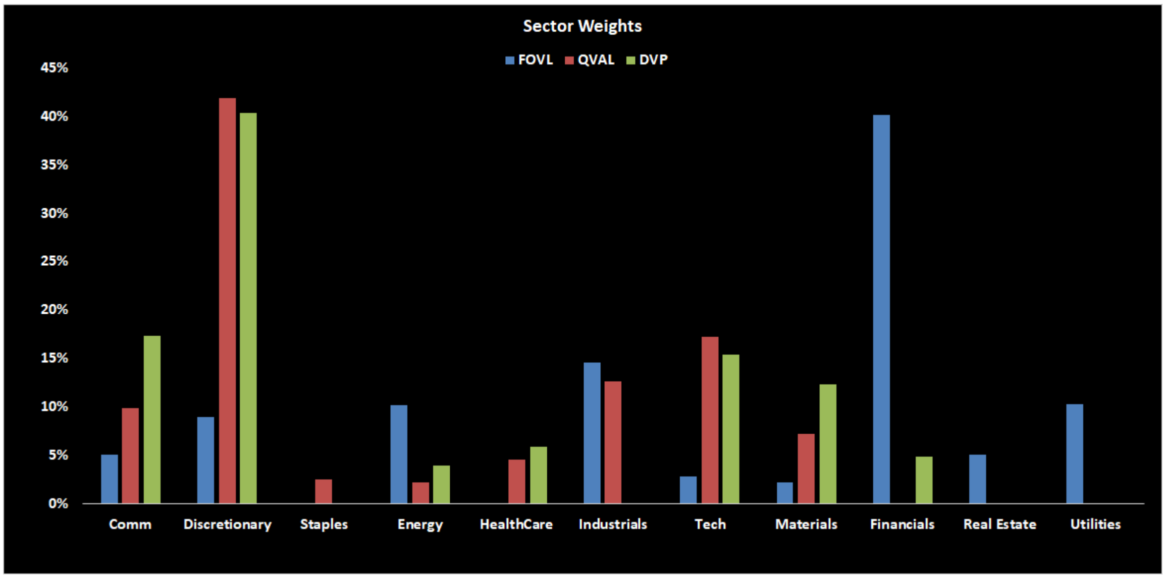
OUR SYSTEMATIC VALUE PHILOSOPHY

May 28, 2019

SUMMARY

- As a firm, Newfound Research focuses on tactical allocation strategies.
- However, we also spend time researching other mandates – such as systematic value – in an effort to introduce lateral thinking to our process.
- Three years ago, we built a systematic value portfolio that seeks to create a "style pure" investment result, attempting to diversify process specification and timing luck risk.
- This commentary introduces our process and the philosophy behind it.
- We conclude by comparing three years of returns against popular systematic value ETF implementations and find that our performance results closely track the median monthly performance, indicating that the ensemble approach may accrue the potential diversification benefits of a multi-manager implementation.

Earlier this week, Tom Psarofagis, ETF Analyst at Bloomberg, highlighted the holdings difference between the new iShares Focused Value ETF (FOVL) versus other concentrated value funds such as the Alpha Architect Quantitative Value ETF (QVAL) and the Deep Value ETF (DVP). Embedded in his analysis was the following graph, which highlights the dramatic sector weight differences between the three products. While QVAL and DVP have over 40% of their portfolio weight in Consumer Discretionary stocks, FOVL has over 40% in Financials.



This analysis raises a very reasonable question: if these ETFs all claim to pursue the same objective, how can they have such different portfolios? We might even take a full step back and ask ourselves, "what *is* value?"[29]

Following on Tom's analysis, Justin Carbonnneau at Validea further highlighted this conundrum with his recent post *Differences in Value*. In the post, Justin evaluates nine different value portfolios that seek to track the methods of famous investors such as Benjamin Graham, Ken Fisher, Jim O'Shaughnessy, and Joel Greenblatt.

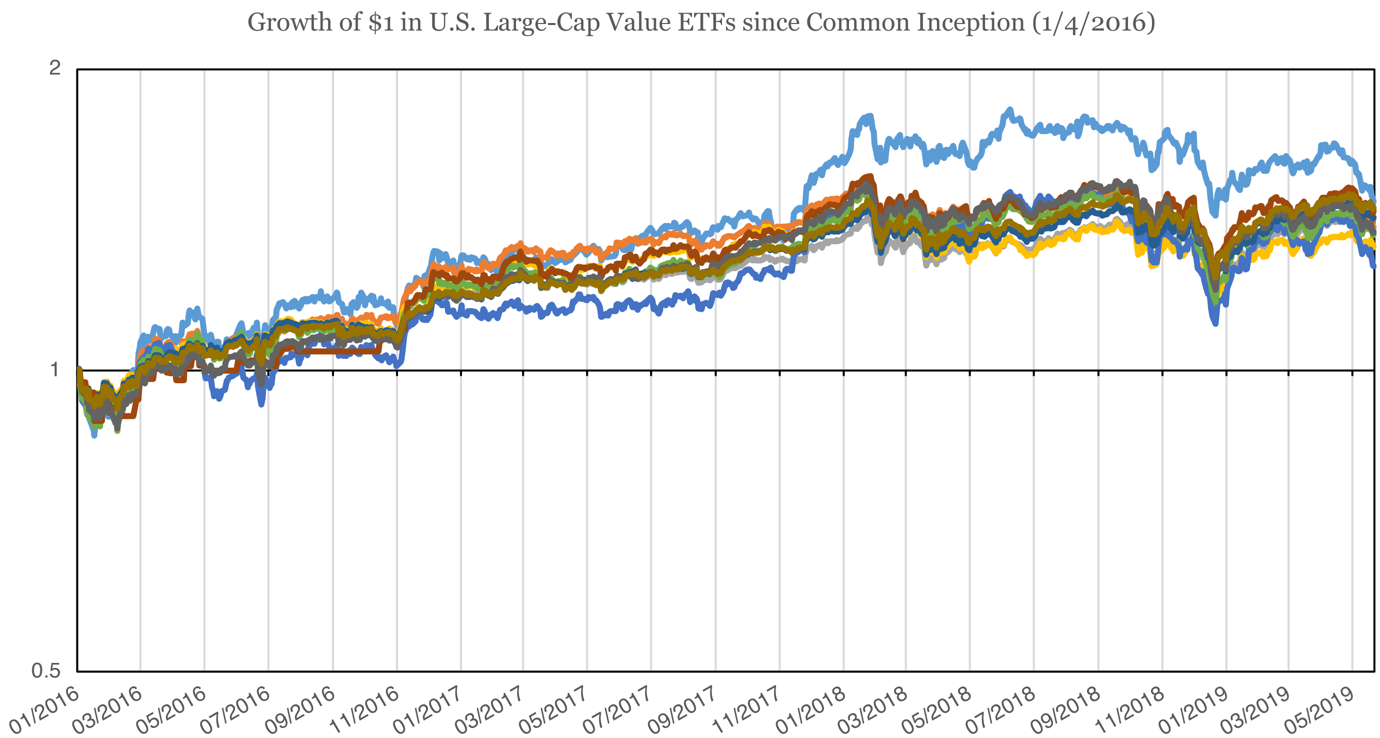
Justin walks through the portfolios using a six question framework for evaluating a value strategy proposed earlier by his co-founder Jack Forehand. The six questions are as follows:

1. What is the universe?
2. Which metrics are used?
3. How many stocks are held?
4. How often is the portfolio rebalanced?
5. How does the portfolio handle industry concentration?
6. How are position sizes determined?

We believe that these questions capture a sufficient cross-section of the portfolio construction choices that lead to different portfolio results. Ultimately, Tom and Justin's analysis captures the notion of what we at Newfound call "specification versus style." While all the strategies aim to capture the style of value investing, the specific means by which they go about pursuing that objective will create tracking error around the median result.

[29] In this commentary, I am specifically talking about a subset of "value" strategies implemented by quants; i.e. "systematic value."

Below we plot the growth of $1 in 10 different U.S. large-cap value ETFs since their common inception date. We can see that while there are significant relative deviations, the portfolios track a common beta, driven both by broader equity market exposure and the general value style. We would argue that the unpredictable performance dispersion justifies the exploration of a multi-manager approach. This is especially true when we consider that it is impossible to know which, if any, method actually captures the "true" value premium.



Source: CSI. Calculations by Newfound Research. Returns are gross of all fees, transaction costs, and taxes with the exception of underlying ETF expense ratios. Returns assume the reinvestment of all distributions.

At Newfound, we believe that investors are best served by first focusing on the management of risk rather than the pursuit of excess returns. Further, we believe the best opportunities for creating differentiated returns is at the asset allocation level.

Nevertheless, we've spent quite a bit of time writing about value in the past. In fact, three years ago we even asked ourselves the question, "if we were tasked to build a systematic value portfolio, how would we do it?"

So, we thought about it. Then we did it.

While this is a meaningful departure from our usual tactical asset allocation mandates, we believe that projects such as these can inspire creativity and insights through lateral thinking. By building and managing a value portfolio, we may solve problems that can directly influence the way we manage other mandates.

Our objective in this project was to simultaneously create a deep value portfolio, but do so in a "style pure" manner that reduced specification risk potentially associated with employing a single measure or methodology. In line with our past research, we aim to do this by diversifying across the three axes of *what, how*, and *when.*

In the remainder of this commentary, we will explore how we thought about building and managing the portfolio using the framework of questions outlined by Validea.

What is the Universe?

While we believe that there is a larger opportunity to harvest style premia in small- and micro-cap securities, we elected to use the S&P 500 as our starting point. We did this for two primary reasons.

First, implementation in the small- and micro-cap universes requires far more effort in trade execution. Trading can have a significant negative impact on less liquid securities and illiquidity can even serve as a limit to arbitrage, potentially causing phantom value signals. As the objective of this research project was to create a style pure portfolio, we wanted to limit the potential negative impacts of execution and reduce operational burden of tracking this portfolio over time.

Secondly, as a firm that focuses on tactical asset allocation, we expect that our data for individual securities is dirtier than that of a firm that focuses specifically on building equity portfolios. We expected that data for well-followed large-cap securities was likely to be cleaner than for small- and micro-cap stocks. Therefore, while focusing on the large-cap universe likely reduces the potential premium earned by a systematic value approach, it also potentially limits risk from dirty data.

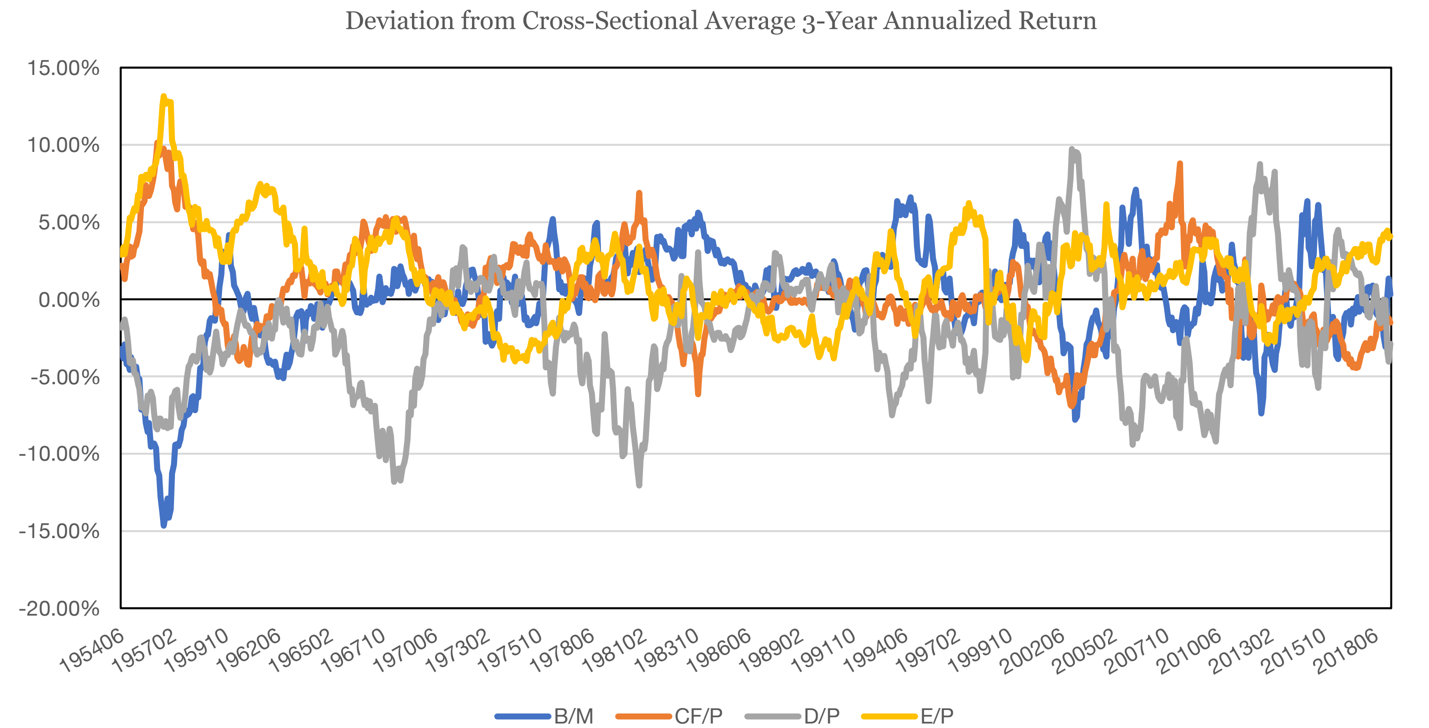
Which Metrics are Used?

There is no shortage of systematic value measures. A quick survey of popular value indices and ETFs demonstrates the breadth of measures used in practice, including:

- Book value to price
- Forward earnings to price
- Dividend to price
- Operating cash flow to price
- EBITDA to enterprise value
- Sales to price

Historically, value investing has worked regardless of the metric used. In the short run, however, specific choices can lead to significant performance deviations.

Using top-quintile performance data from the Kenneth French data library, we plot rolling 3-year annualized returns for portfolios formed on book-to-market, cash-flow-to-price, dividend yield, and earnings yield versus the cross-sectional average 3-year return. We can see that each metric spends time in-and-out of favor relative to the other metrics.



Source: Kenneth French Data Library. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all fees, transaction costs, and taxes. Returns assume the reinvestment of all distributions. You cannot invest in an index.

Without a particular view as to which metric will necessarily perform best going forward, we elected to use a composite of metrics. Much like asset diversification, ensemble approaches can help us hedge our uncertainty related to specification risk.

We categorize metrics into four broad categories: book, earnings, sales, and cash flow. Based upon our research, we viewed each of these categories as providing a different, albeit biased, perspective as to whether a security was cheap or expensive.

	Example	Strengths	Weaknesses
Book	Price-to-Book	Can be used on companies with negative earnings.	Not applicable for companies with few tangibles on their balance sheet.
Earnings	Forward P/E	Accounts for past and forward growth expectations of a company.	Affected by capital structure; not applicable for firms with negative earnings outlooks; manipulation risk.
Sales	Price-to-Sales	Helps account for high reinvestment and highly cyclical industries; tied to growth.	Does not account for capital structure; need earnings at *some* point.
Cash Flow	EV-to-EBITDA	Capital-structure neutral; focuses on cash-flow.	Not necessarily applicable across industries with different capital expenditure needs.

For each category, we identify a number of metrics that may apply. We then construct a composite score for each security by z-scoring each individual metric, averaging the z-scores at the category level, and then averaging the category z-scores.

While there is room for improvement here (e.g. removing price-to-book based upon evidence that it is no longer a meaningful value signal or exploring a sector-specific weighting of signals), we believe that the broad cross-section of signals helps meaningfully reduce single-signal bias without introducing additional biases by selectively weighting certain signals.

"Average of Portfolios" versus "Portfolio of Average Signals"

Historically, we have argued that ensembles should be implemented as averages of portfolios and not averages of signals. This argument arises from the math of Jensen's Inequality, which says that the expectation of a nonlinear function applied is not necessarily equal to the function applied to the expectation.

As a simple example, consider two signals that are used to make all-in or all-out calls on the market. The signals range from 0 to 100 and if the score is above 50, we invest. If each of these signals is used to build a portfolio and we then average the portfolio weights, the resulting allocations can be all-out, all-in, or fifty-fifty exposed. However, if we first average the signals together and then apply our rule, we can only be all-out or all-in.

We would argue for the former approach rather than the latter as we believe it represents diversification of signals whereas the latter simply represents the construction of a single, new signal.

This logic would lead us to believe that the appropriate way to construct our value portfolio would be to build portfolios based upon each metric and then average them together (or average them within each category and then average the categories). So why do we not take this approach?

To be clear, we do not have any empirical evidence to support our decision. Our hypothesis, however, is that the individual bias introduced by each signal may overwhelm the process if the "average of portfolios" approach is used rather than the "portfolio of average signals" approach. The former method may simply lead to a portfolio of securities that are identified as cheap only by one signal, whereas we are specifically trying to employ a composite approach to benefit from multiple definitional perspectives of value.

Cross-Market versus Cross-Industry

The academic implementation of the value-factor ranks stocks based upon their book-to-price metric and market-capitalization-weights the cheapest decile (or quintile) of securities. Doing so, however, can lead to significant sector tilts, as certain sectors may have structurally lower valuation metrics.

One solution to this problem is to measure valuation relative to other like securities. For example, instead of calculating cross-market scores, cross-industry-group scores could be employed. In *Predicting Stock Returns Using Industry-Relative Firm Characteristics*, Asness, Porter, and Stevens find that evaluating a firm's characteristics versus industry average characteristics provides "more precise estimates" than the traditional cross-market approach.

In *Are Factor Investors Getting Paid to Take on Industry Risk?*, Bryan and McCullough find that while the traditionally built value portfolio expresses meaningful sector tilts, they are not additive to performance and can be neutralized by investors to reduce risk.

This is further confused by the fact that the sector tilts created by unconstrained value implementation can vary significantly by the valuation metric utilized. Price-to-book, for example, tends to dramatically overweight Financials, whereas EV-to-EBITDA overweights Consumer Discretionary.

This may all point to an argument supporting an industry-neutral implementation (or, at the very least, potentially support an ensemble approach to help neutralize structural industry biases that arise with a given measure).

Further, one strong argument towards cross-industry scoring is that it can serve as a mechanism for regularization of the already fuzzy heuristic measures we are employing.

The downside to a purely industry-neutral implementation, however, is that it can be highly susceptible to industry bubbles. During the late 1990s, for example, an industry-neutral value portfolio would still hold a significant allocation to technology companies. In most environments, this may not matter. In some environments, however, an unconstrained industry approach may be a significant boon to risk management.

With strong arguments supporting both approaches, we calculate both cross-market and a cross-industry score for each security and combine them. Note that this approach can still leave us with significant industry tilts. We'll address this in a later section.

Cross Market Scores

- Aligns with traditional value research (i.e. HML factor).
- Empirically higher annualized growth rate.
- Valuation-based sector rotation "works."
- Can potentially help defuse sector-based bubbles.

Cross Industry Scores

- Style bets are made only *within* industries.
- Empirically higher Information Ratio.
- In theory, sector bets should be uncompensated.
- Helps prevent structural underweight to growth sectors.
- Empirically higher turnover.

How Many Stocks are Held?

Once valuation scores are calculated, we screen our universe down to the cheapest quintile (approximately 100 securities) with a goal of eventually creating a deep value portfolio from approximately 1/10th of the securities in our universe.

Securities falling in the cheapest are quality scored. Our goal in using quality scores is to focus the portfolio on mis-pricings; i.e. identify deep-value stocks whose valuations are not justified by their poor quality. We believe that such an approach is supported by the empirical evidence suggesting that the value premium emerges from the re-valuation of mis-priced securities.

Like our valuation model, our quality score model is a composite of categories. The categories are informed by the elements of a Gordon Growth Model of security pricing.

Quality categories informed by Gordon Growth Model: Price = EPS x Payout Ratio / (k – g)

	Why?	Example
Profitability	More profitable companies should command a higher stock price.	ROA, ROE
Payout	Investors should pay a higher price for a higher payout ratio.	Dividend Growth, Payout-to-Profitability
Safety	Investors should demand a higher return for increased risk.	Ohlson O-Score, Altman-Z Score, Beta
Growth	Investors should pay a higher price for faster growing companies.	Change in ROA, Change in ROE
Momentum	Seek to defer purchase of securities exhibiting negative price momentum.	12-1 Month Return

We combine value and quality scores and re-rank, buying a number of stocks equal to 10% of our investable universe (approximately 50). However, we do not simply buy the top ranked stocks, as our selection simultaneously seeks to manage industry concentration risk (more on this point below).

We should note that the momentum measure is not incorporated in the overall quality score. Rather, if a stock has been selected for purchase but is exhibiting extreme negative relative momentum, we will defer its purchase to a later date.

How Often is the Portfolio Rebalanced?

Frequent readers will know that rebalance timing luck is an obsession of ours here at Newfound. So much so, we believe when a portfolio rebalances represents a unique axis of diversification.

To distill the potential problem, consider two deep value portfolios that only rebalance on an annual basis. If one portfolio rebalances in January and one rebalances in July, the holdings may be dramatically different even if the portfolios are constructed with the same process. This represents timing risk due to the opportunities present at the time of rebalancing.

This can lead to non-trivial performance dispersion. For example, in *Fundamental Indexation: Rebalancing Assumptions and Performance,* Blitz, van der Grient, and van Vliet demonstrate that a fundamental index rebalanced every March outperformed a capitalization-weighted benchmark in 2009 by over 10%, while the same methodology rebalanced in September underperformed.

In building our systematic value portfolio, we wanted to specifically address this risk. To do so, we implement a method we call "tranching." Each month, we take 1/60th of our capital and purchase a deep value portfolio. We then hold that tranche for 5 years, at which point it is liquidated. At any given point we hold 60 tranches.

The purpose of this approach is two-fold.

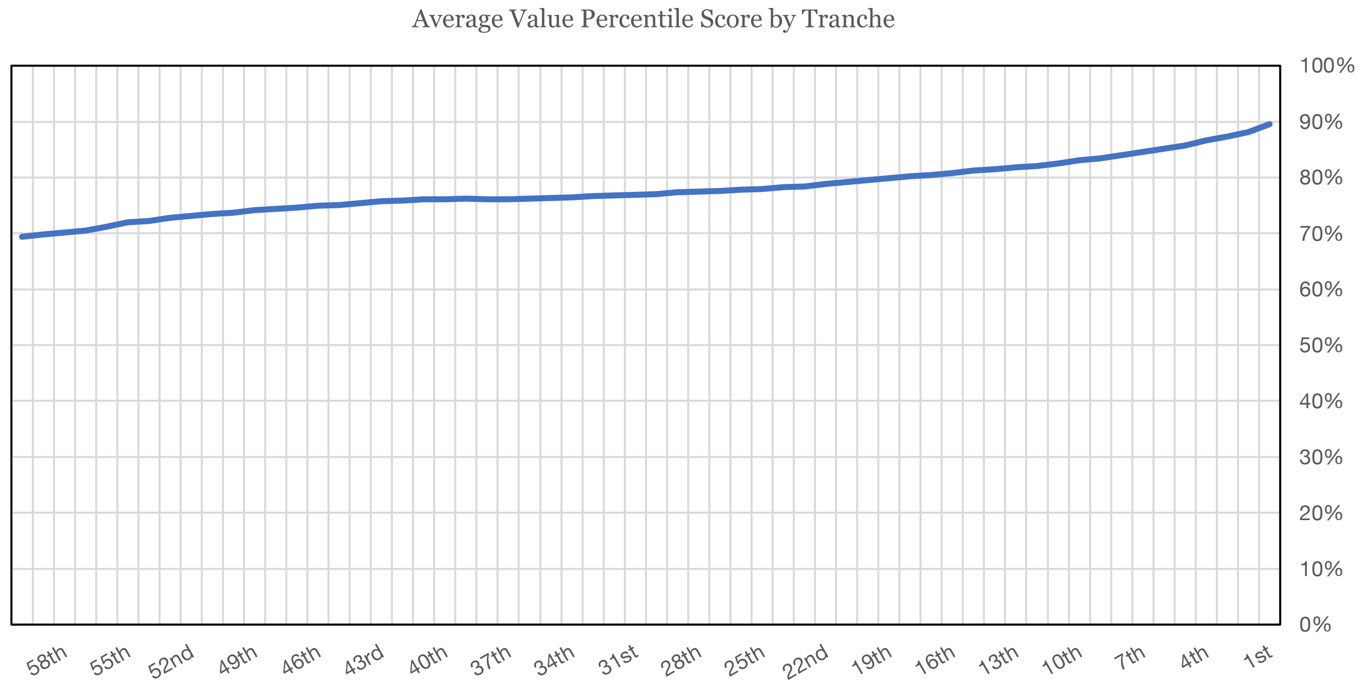
First, we are able to significantly reduce the timing luck associated with a deep value signal sampled at a given time. In many ways, our approach is akin to how many investors think about managing market cycle risk to private equity funds. Without any particular view as to where we sit in the market cycle, investors often try to deploy their capital to multiple private equity funds over time, implicitly diversifying market cycle risk. Here we adopt the same approach, but do so to diversify value signal timing risk.

Secondly, we align our holding period with the decay speed of our value signal, giving our holdings ample time to not only be re-valued by the market, but also to benefit from the underlying growth potential of the companies, which we purchased at a hopefully unreasonably cheap price.

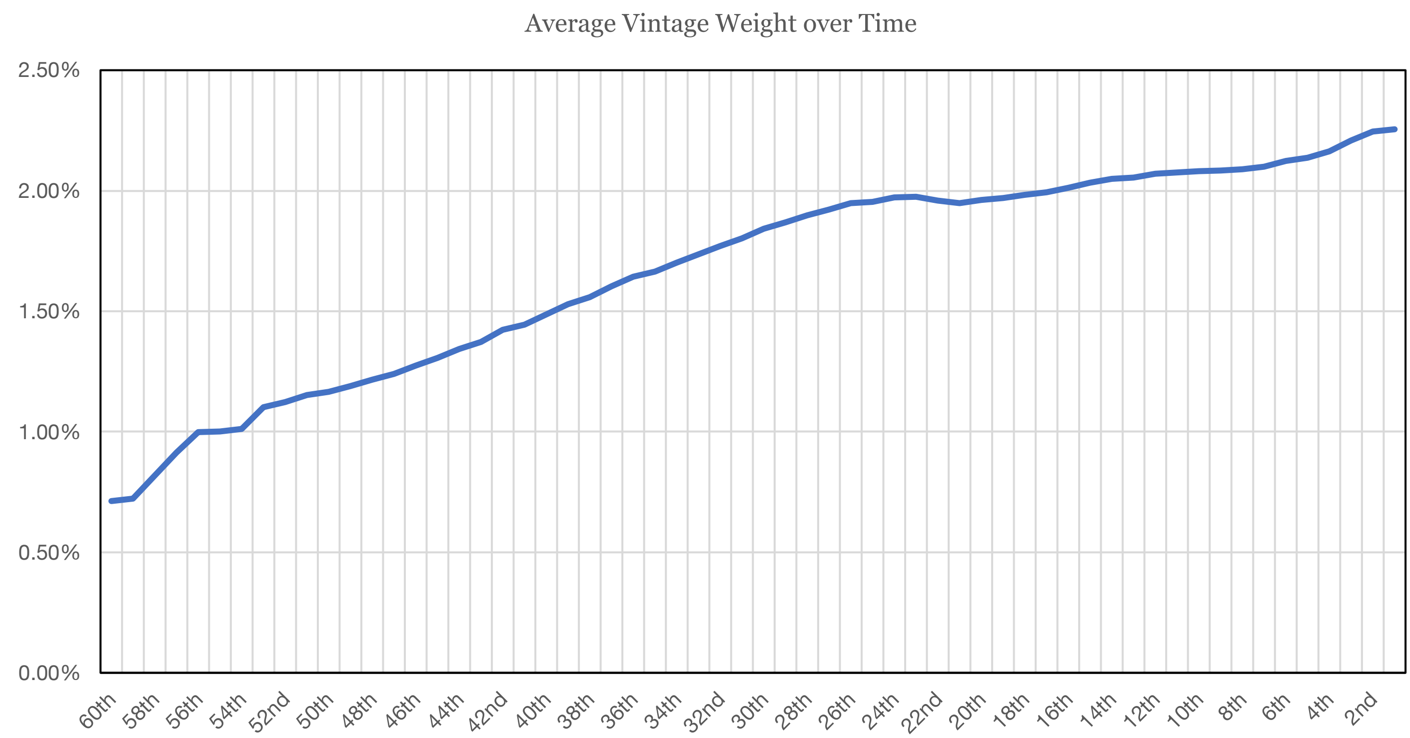
A potential downside of this approach is that securities which are deeply undervalued only for a brief period will only ever represent a small contribution to returns. On the other hand, securities that are undervalued for a longer period of time will have position sizes built up within multiple tranches. This allows us the potential benefit of increasing our exposure over time if a position continues to cheapen. The net effect is that we are patient allocators, with both signal strength and signal duration serving as meaningful contributors to position size.

Blindly holding each tranche for 5 years, however, may lead to a dilution of value characteristics within the portfolio. For example, what happens if valuations of securities in older tranches normalize faster than expected? To account for this, each month we evaluate all the securities in our portfolio and sell anything with a valuation score above the median score for the universe.

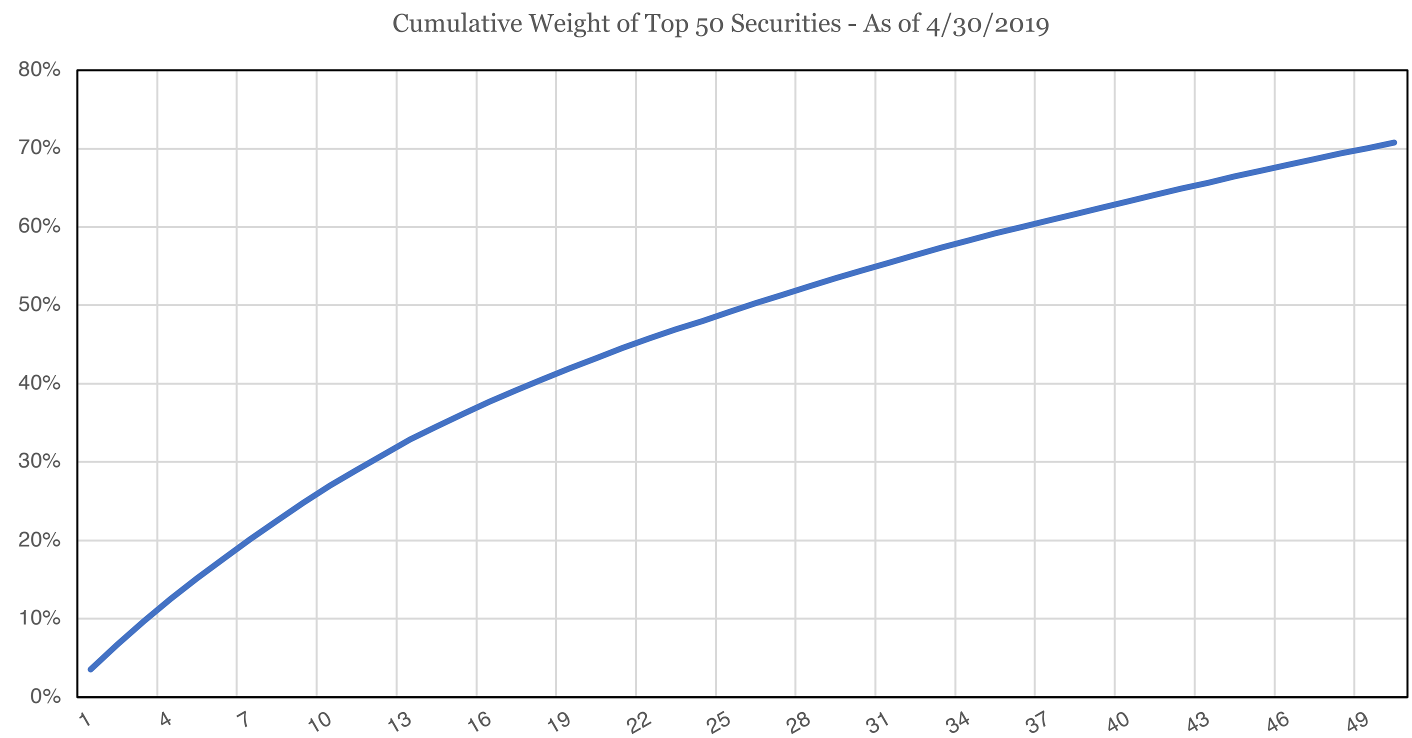
Below we plot the average value percentile score by tranche. We can see that older tranches have, on average over time, had lower average valuation scores than newer tranches. This implies that securities purchased are indeed re-valuing upward relative to peers (though we should acknowledge that this may not necessarily be due to price appreciation but could simply be due to deteriorating fundamentals).



We can also plot the average portfolio allocation to each tranche over time. We can see that the portfolio tends to heavily tilt towards more recently created tranches, with securities in older tranches being removed over time as they exceed the median universe valuation score.



Using this "portfolio-of-portfolios" approach to manage timing luck, the number of holdings at any given time can exceed our target of 10% of universe securities. In fact, it has historically hovered around 140, which is far higher than the approximate 50 we would expect. However, as we do not equally weight securities and tranches tend to get smaller over time, the concentration coefficient of the portfolio hovers closer to 55, indicating a more concentrated portfolio from an allocation perspective. The top 10 holdings currently account for over 25% of the portfolio weight, and the top 50 holdings account for just over 70% of the portfolio weight.



How Does the Portfolio Handle Industry Concentration?

As discussed above, we believe that industry bets represent a largely non-compensated source of risk within a value portfolio during most market environments. "Most" being the operative word in the prior sentence, as meaningfully avoiding technology during the dot-com era or financials during the 2008 crisis would have created meaningful relative outperformance.

Therefore, our goal was to retain the optionality for industry divergence while remaining mostly industry neutral during most periods. Note that we specifically seek to be industry neutral, not sector neutral. While industry neutrality implies sector neutrality, the reverse is not true as some sectors may have very diverse industries within them.

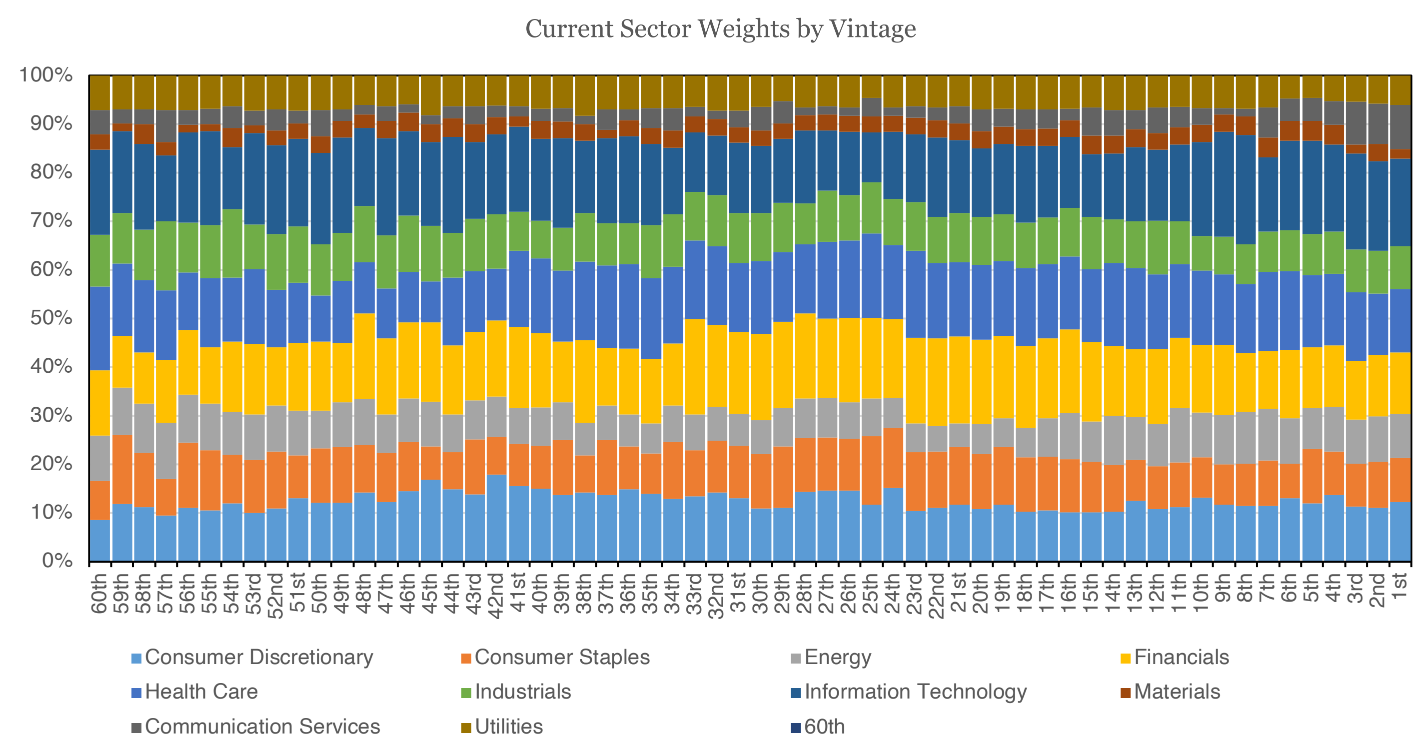
We begin by calculating the target relative industry weights based upon current market capitalizations. Assuming we will pick N securities for our portfolio, we run the following algorithm:

1. Identify the industry that has both (1) selectable securities remaining and (2) is currently furthest away from its target weight.
2. Select the highest ranked security in this industry.
    a. Increase the industry weight by 1/N.
    b. Remove the security from the list of eligible securities.
3. Repeat until N securities have been identified.

With this algorithm, there are two particular things to note.

First, this process will only approximate an industry neutral implementation assuming N is sufficiently large and we actually implement with an equal-weight portfolio. In our case, N is approximately 50 and, as we will discuss in the next section, we do not equally-weight our portfolio. Thus, even when all industries are fairly represented, we should still expect some deviation.

Second, the algorithm will always target an industry neutral implementation assuming there are sufficient eligible securities to build one. The flexibility to avoid an industry enters into play when an industry does not have a sufficient number of constituents falling in the top quintile of value scores.



Source: Bloomberg. Calculations by Newfound Research.

How are Position Sizes Determined?

Once we have selected our securities, the final step of the process is to construct our portfolio. Each technique for portfolio construction will inherently introduce its own assumptions, biases, estimation risks, and implicit resulting portfolio characteristics. It should come as no surprise that we embrace an ensemble approach.

Specifically, we create six different portfolios using different construction techniques and average their results together. Each technique makes various assumptions about what we know as well as our model of the relationship between risk and return. The techniques are:

- **Equal Weight:** Assumes we know nothing about an individual security's expected return or covariance with any other security. Optimal if returns and covariances are identical.
- **Inverse Volatility:** Assumes we have a view on security volatilities but not correlations. Optimal if excess returns are proportional to volatility (i.e. equal Sharpe ratios) and correlations are homogenous.
- **Minimum Variance Portfolio:** Assumes we have a view on volatility and correlation but no view on returns. Assumes markets are not risk-efficient and is optimal if expected returns are similar.
- **Maximum Sharpe Portfolio:** Assumes a view on volatility, correlation, and returns. Specifically, we assume returns are proportional to downside deviation.
- **Quality-Score Weighted:** Optimal if Sharpe ratios are linearly proportional to the return and covariances are homogenous.
- **Value-Score Weighted:** Optimal if Sharpe ratios are linearly proportional to the return and covariances are homogenous.

In equally-weighting the portfolios constructed by these six techniques, we aim to diminish strategy-specific risks.

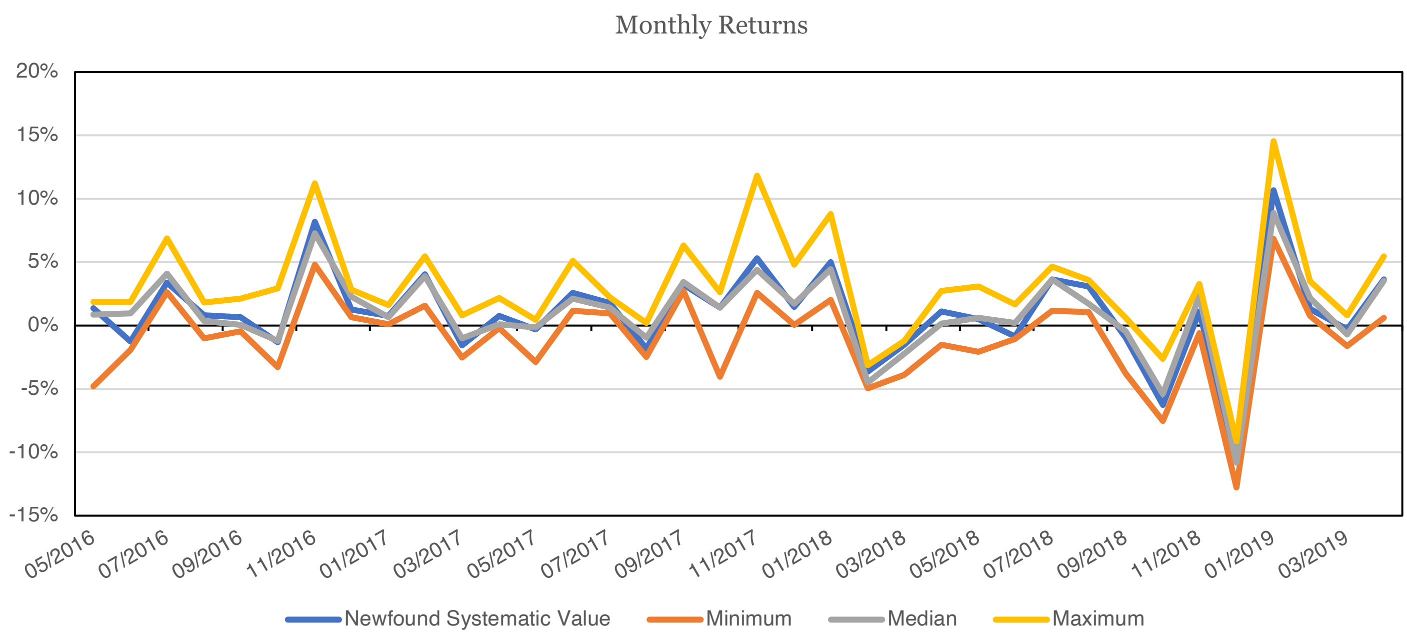
Conclusion

We set out on this project to develop a "style pure" value strategy that would diversify across process and timing specification risk. As we approach three years of live performance, we can now attempt to quantify the success of our project.

Success, in this case, is *not* relative outperformance. We did not set out to "fix" value; rather, we set out to avoid the potential downside risks that might come from selecting an individual measure or methodology. Of course, avoiding this risk also means knowingly foregoing any potential benefit from that selection.

This means we explicitly do not want to see outlier performance, whether positive or negative, relative to a distribution of other value funds. The purpose of an ensemble approach is that we should specifically reduce the impact of specification-driven outlier events. In many ways, we can think of it as a virtual fund-of-funds.

Using a sample of U.S. value ETFs, we plot the minimum, maximum, and median monthly return as well as the monthly return for our Systematic Value strategy.



Source: CSI. Calculations by Newfound Research. Returns represent live strategy results. Returns for the Newfound Systematic Value strategy are gross of all management fees and taxes, but net of execution fees. Returns for ETFs included in study are gross of any management fees, but net of underlying ETF expense ratios. Returns assume the reinvestment of all distributions. ETFs included in the study are (in alphabetical order): DVP, FTA, IWD, PWV, QVAL, RPV, SCHV, SPVU, VLUE, and VTV.

We can see that our portfolio closely tracks the median result[30], indicating that the process appears to provide access to the style of systematic value without necessarily inviting the specification risks that might go along with picking just one process. Thus, the potential benefits that typically accrue to a multi-manager implementation may be achievable with an ensemble approach.

Ultimately, if you feel confident about a specific measure of value, this type of strategy will not be for you. There are a number of ETFs that track indices constructed with much more concentrated approaches that can align with your philosophy.

[30] The results also very closely track the mean result.

But if you do not know which flavor of value will be favored over the short-term and want to hedge against that risk, diversifying specification and timing risk can make sense.

Earning the value premium requires bearing some risk. And as the last decade has shown, this can lead to long periods of underperformance relative to the broader market. It is the uncompensated risks that can compound this underperformance, and these are the risks that our Systematic Value portfolio seeks to mitigate.

TACTICAL CREDIT

June 3, 2019

SUMMARY

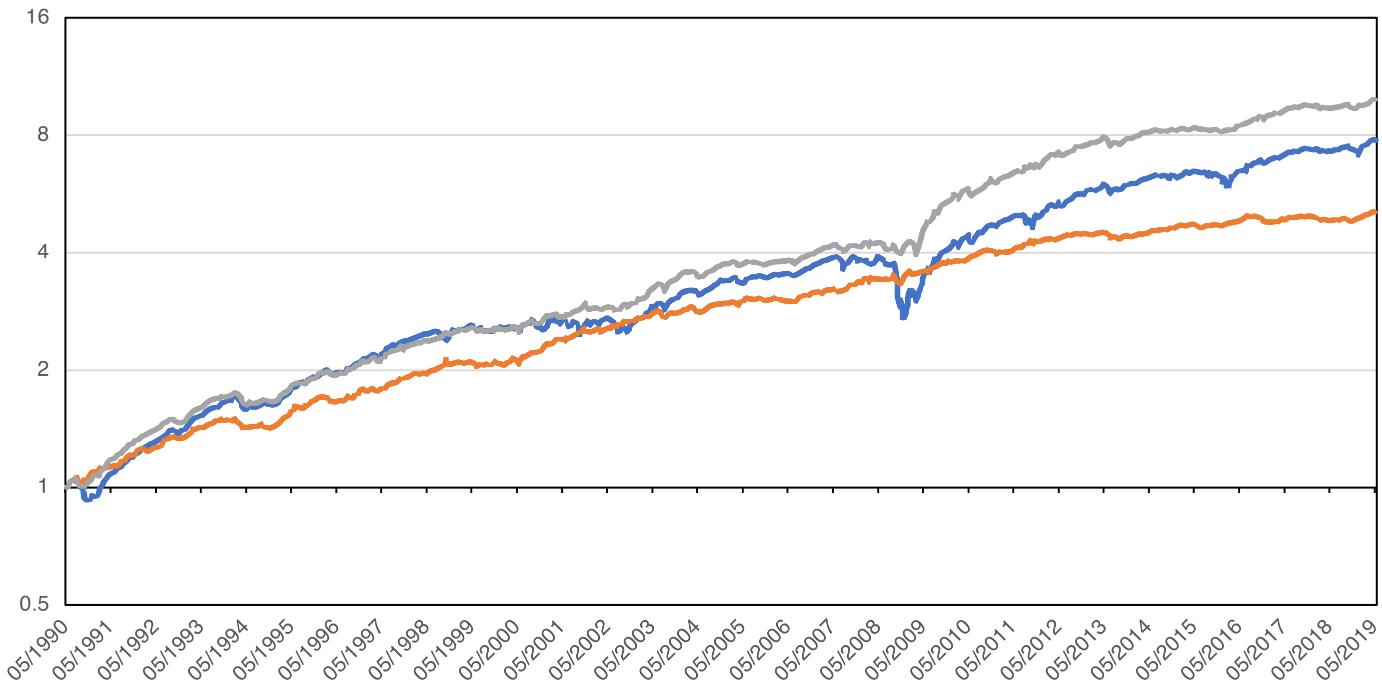
- In this commentary we explore tactical credit strategies that switch between high yield bonds and core fixed income exposures.
- We find that short-term momentum signals generate statistically significant annualized excess returns.
- We use a cross-section of statistically significant strategy parameterizations to generate an ensemble strategy. Consistent with past research, we find that this ensemble approach helps reduce idiosyncratic specification risk and dramatically increases the strategy's information ratio above the median underlying strategy information ratio.
- To gain a better understanding of the strategy, we attempt to determine the source of strategy returns. We find that a significant proportion of returns are generated as price returns occurring during periods when credit spreads are above their median value and are expanding.
- Excluding the 2000-2003 and 2008-2009 sub-periods reduces gross-of-cost strategy returns from 2.9% to 1.5%, bringing into question how effective post-of-cost implementation can be if we do not necessarily expect another crisis period to unfold.

There is a certain class of strategies we get asked about quite frequently but have never written much on: tactical credit.

The signals driving these strategies can vary significantly (including momentum, valuation, carry, macro-economic, et cetera) and implementation can range from individual bonds to broad index exposure to credit default swaps. The simplest approach we see, however, are high yield switching strategies. The strategies typically allocate between high yield corporate bonds and core fixed income (or short-to-medium-term U.S. Treasuries) predominately based upon some sort of momentum-driven signal.

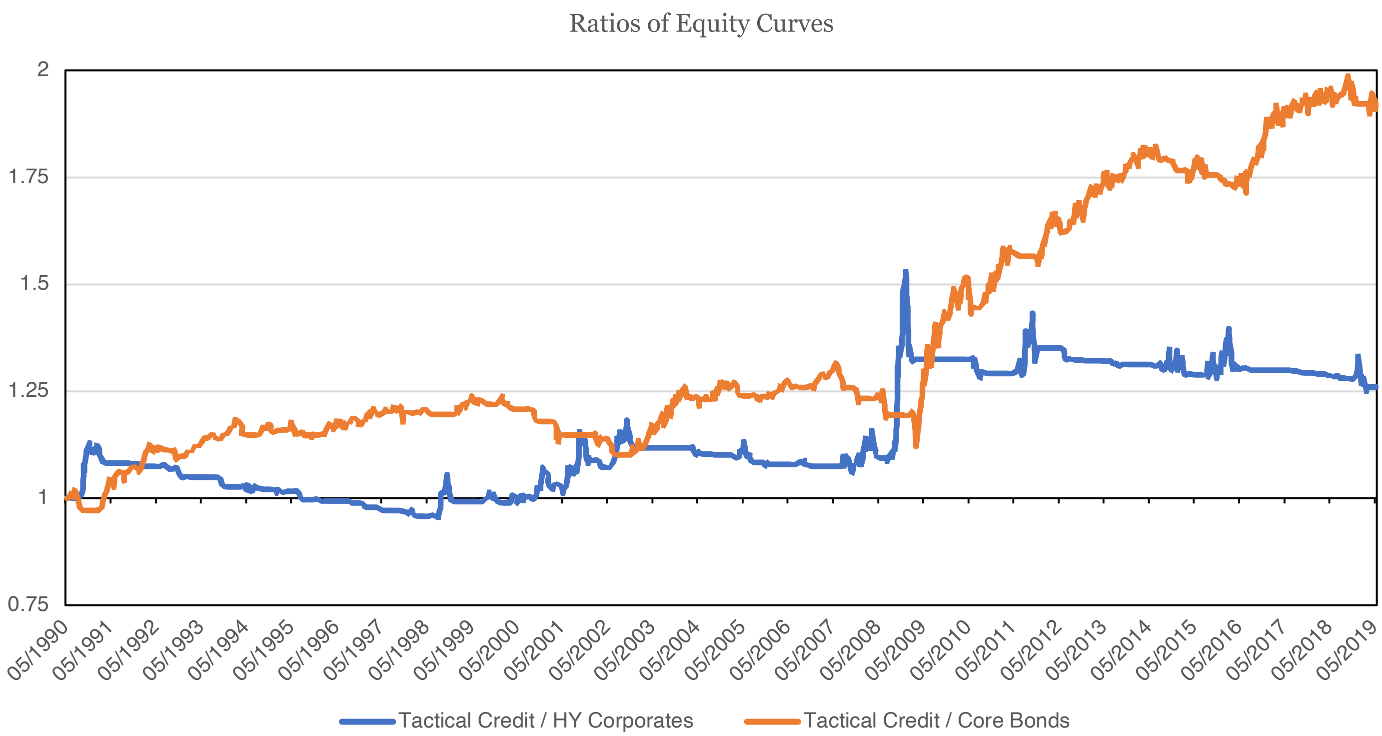
It is easy to see why this seemingly naïve approach has been attractive. Implementing a simple rotation between –high-yield corporates– and –core U.S. fixed income– with a 3-month lookback with 1-month hold creates a fairly attractive looking –tactical credit– strategy.

Growth of $1



Source: Tiingo. Calculations by Newfound Research. Tactical Credit strategy returns are hypothetical and backtested. Returns gross of all management fees and taxes, but net of underlying fund fees. HY Corporates represents the Vanguard High-Yield Corporate Fund (VWEHX). Core Bonds is represented by the Vanguard Total. Bond Market Index Fund (VBMFX). Returns assume the reinvestment of all distributions.

Visualizing the ratio of the equity curves over time, we see a return profile that is reminiscent of past writings on tactical and trend equity strategies. The tactical credit strategy tends to outperform core bonds during most periods, with the exception of periods of economic stress (e.g. 2000-2002 or 2008). On the other hand, the tactical credit strategy tends to underperform high yield corporates in most environments, but has historically added significant value in those same periods of economic stress



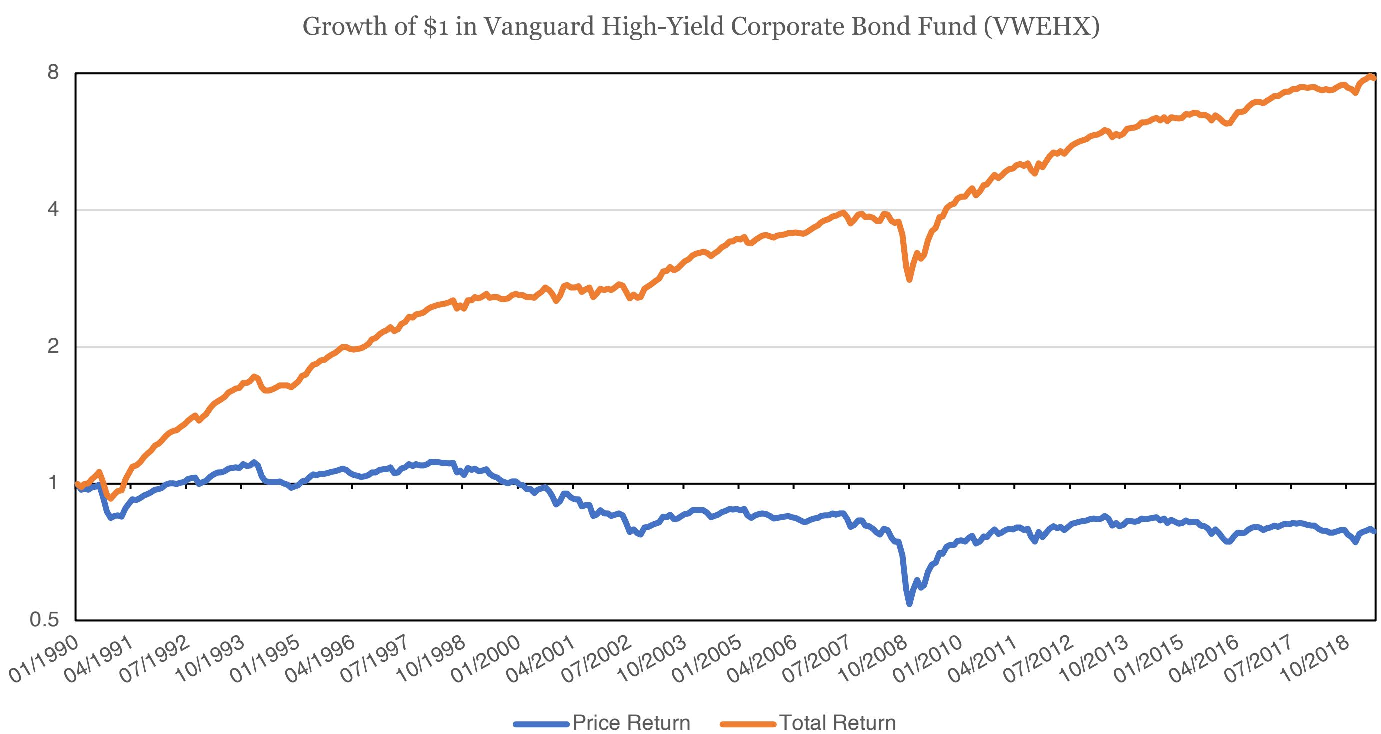
Source: Tiingo. Calculations by Newfound Research. Tactical Credit strategy returns are hypothetical and backtested. Returns gross of all management fees and taxes, but net of underlying fund fees. HY Corporates represents the Vanguard High-Yield Corporate Fund (VWEHX). Core Bonds is represented by the Vanguard Total. Bond Market Index Fund (VBMFX). Returns assume the reinvestment of all distributions.

This is akin to tactical equity strategies, which have historically out-performed the safety asset (e.g. cash) during periods of equity market tailwinds, but under-performed buy-and-hold equity during those periods due to switching costs and whipsaw. As the most aggressive stance the tactical credit strategy can take is a 100% position in high yield corporates, it would be unrealistic for us to expect such a strategy to out-perform in an environment that is conducive to strong high yield performance.

What makes this strategy different than tactical equity, however, is that the vast majority of total return in these asset classes comes from income rather than growth. In fact, since the 1990s, the price return of high yield bonds has annualized at -0.8%. This loss reflects defaults occurring within the portfolio offset by recovery rates.[31]

[31] Fixed income indices tend to have high levels of turnover, leading to other sources of gains and losses including shifts in the yield curve and credit spreads, realized roll in the yield curve or credit spreads, and other sources of additions and deletions such as bond callability.

This is potentially problematic for a tactical strategy as it implies a significant potential opportunity cost of switching *out* of high yield. However, we can also see that the price return is volatile. In years like 2008, the price return was -27%, more than offsetting the 7%+ yield you would have achieved just holding the fund.



Source: Tiingo. Calculations by Newfound Research. Returns gross of all management fees and taxes, but net of underlying fund fees.

Like trend equity, we can think of this tactical credit strategy as being a combination of two portfolios:

- A fixed-mix of 50% high yield corporates and 50% core bonds; and
- 50% exposure to a dollar-neutral long/short portfolio that captures the tactical bet.

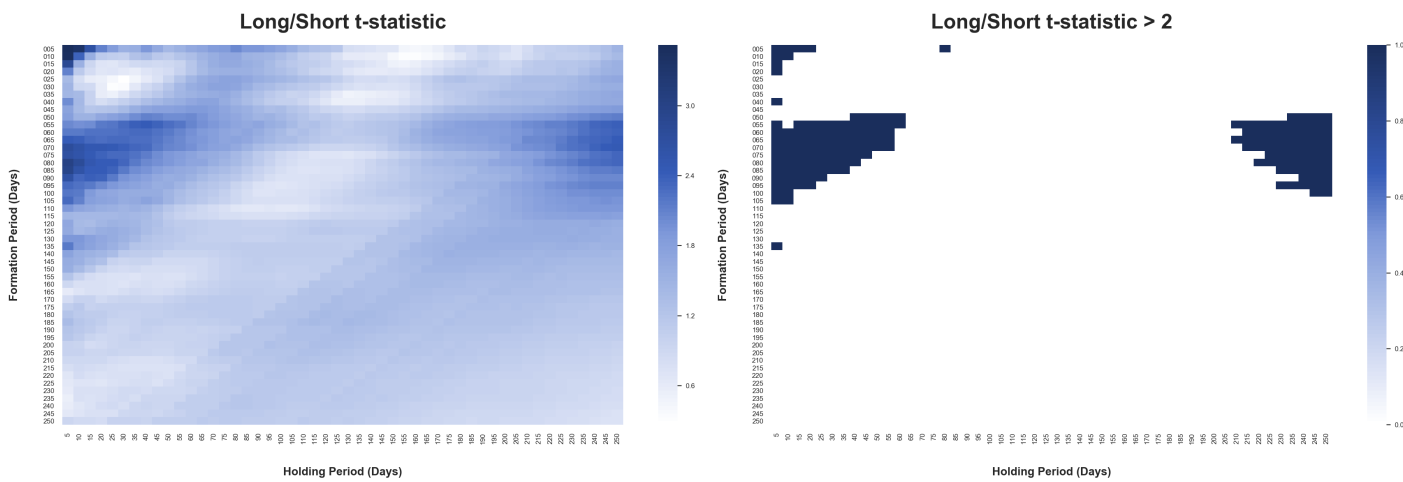
For example, when the tactical credit portfolio is 100% in high yield corporates, we can think of this as being a 50/50 strategy portfolio with a 50% overlay that is 100% long high yield corporates and 100% short core bonds, leading to a net exposure that is 100% long high yield corporates.

Thinking in this manner allows us to isolate the active returns of the portfolio actually being generated by the tactical signals and determine value-add beyond a diversified buy-and-hold core. Thus, for the remainder of this commentary we will focus our exploration on the long/short component.

Before we go any further, we do want to address that a naïve comparison between high yield corporates and core fixed income may be plagued by changing composition in the underlying portfolios as well as unintended bets. For example, without specifically duration matching the legs of the portfolio, it is likely that a dollar-neutral long/short portfolio will have residual interest rate exposure and will not represent an isolated credit bet. Thus, naïve total return comparisons will capture both interest rate and credit-driven effects.

This is further complicated by the fact that sensitivity to these factors will change over time due both to the math of fixed income (e.g. interest rate sensitivity changing over time due to higher order effects like convexity) as well as changes in the underlying portfolio composition. If we are not going to specifically measure and hedge out these unintended bets, we will likely want to rely on faster signals such that the bet our portfolio was attempting to capture is no longer reflected by the holdings.

We will begin by first evaluating the stability of our momentum signals. We do this by varying formation period (i.e. lookback) and holding period of our momentum rotation strategy and calculating the corresponding t-statistic of the equity curve's returns. We plot the t-statistics below and specifically highlight those regions were t-statistics exceed 2, a common threshold for significance.



Source: Tiingo. Calculations by Newfound Research.

It should be noted that data for this study only goes back to 1990, so achieving statistical significance is more difficult as the sample size is significantly reduced. Nevertheless, unlike trend equity which tends to exhibit strong significance across formation periods ranging 6-to-18 months, we see a much more limited region with tactical credit. Only formation periods from 3-to-5 months appear significant, and only with holding periods where the total period (formation plus holding period) is less than 6-months.

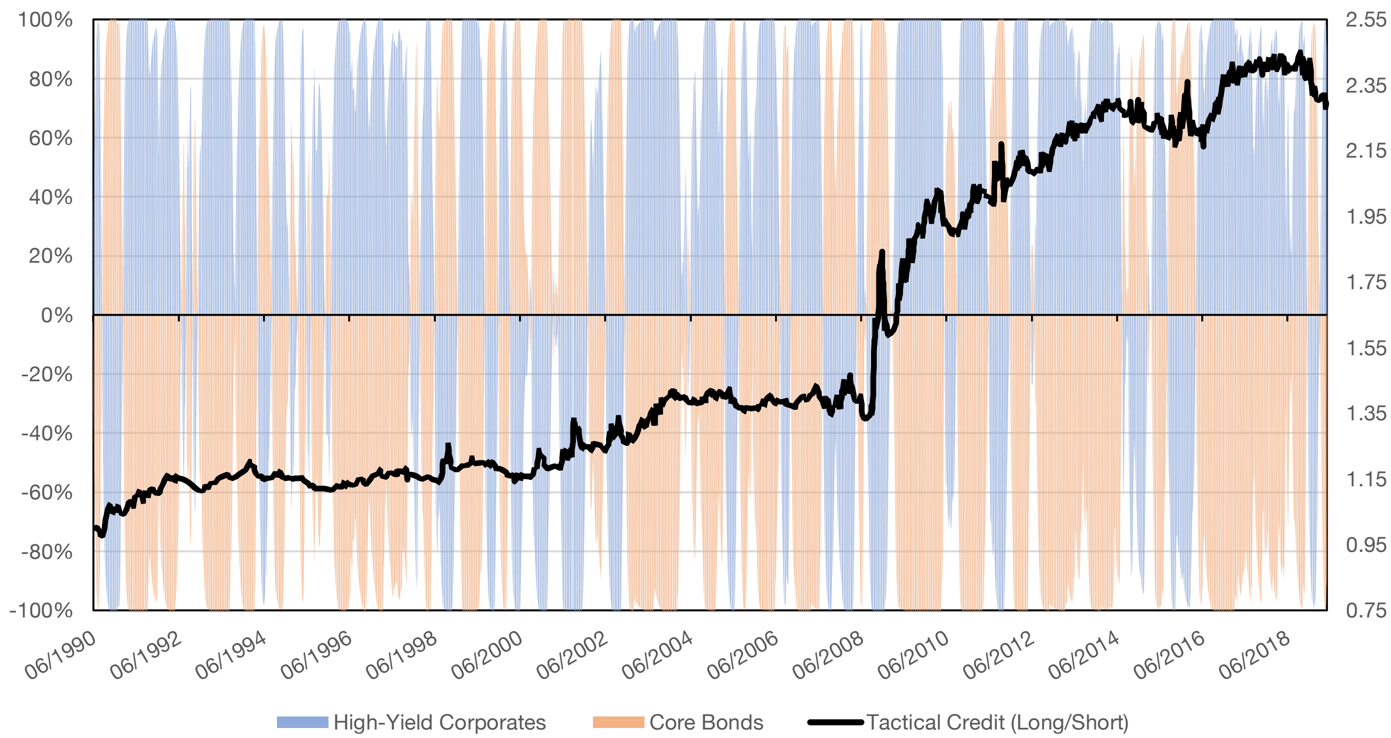
Note that our original choice of 63-day (approximately 3 months) formation and 21-day (approximately 1 month) hold falls within this region.

We can also see that very short formation and holding period combinations (e.g. less than one month) also appear significant. This may be due to the design of our test. To achieve the longest history for this study, we employed mutual funds. However, mutual funds holding less liquid underlying securities tend to exhibit positive autocorrelation. While we adjusted realized volatility levels for this autocorrelation effect in an effort to create more realistic t-statistics, it is likely that positive results in this hyper short-term region emerge from this effect.

Finally, we can see another rather robust region representing the same formation period of 3-to-5 months, but a much longer holding length of 10-to-12 months. For the remainder of this commentary, we'll ignore this region, though it warrants further study.

Assuming formation and holding periods going to a daily granularity, the left-most region represents over 1,800 possible strategy combinations. Without any particular reason for choosing one over another, we will embrace an ensemble approach, calculating the target weights for all possible combinations and averaging them together in a virtual portfolio-of-portfolios configuration.

Below we plot the long/short allocations as well as the equity curve for the ensemble long/short tactical credit strategy.

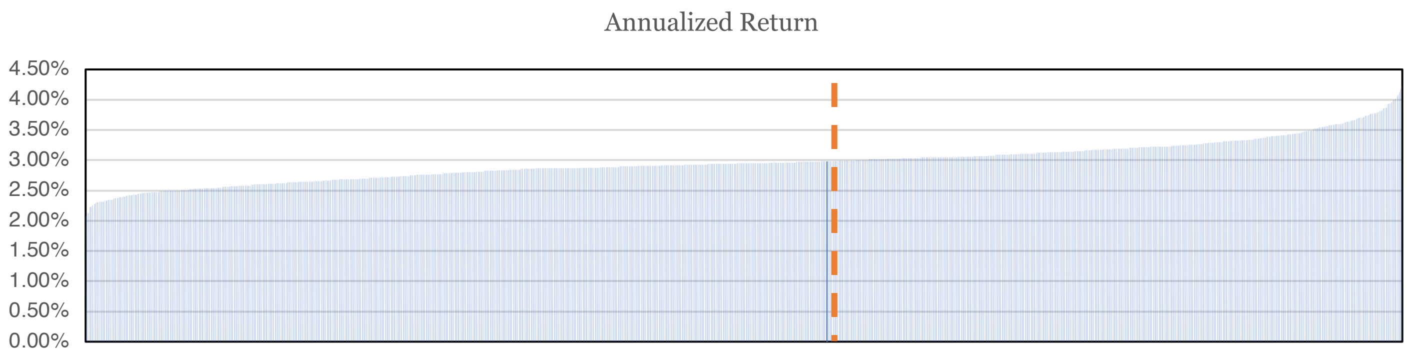


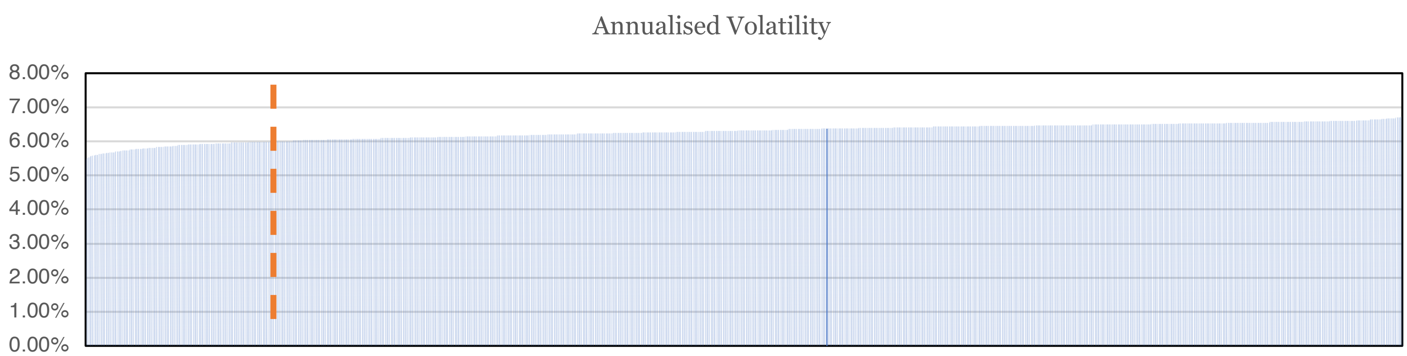
Source: Tiingo. Calculations by Newfound Research. Tactical Credit strategy returns are hypothetical and backtested. Returns gross of all management fees and taxes, but net of underlying fund fees. Returns assume the reinvestment of all distributions.

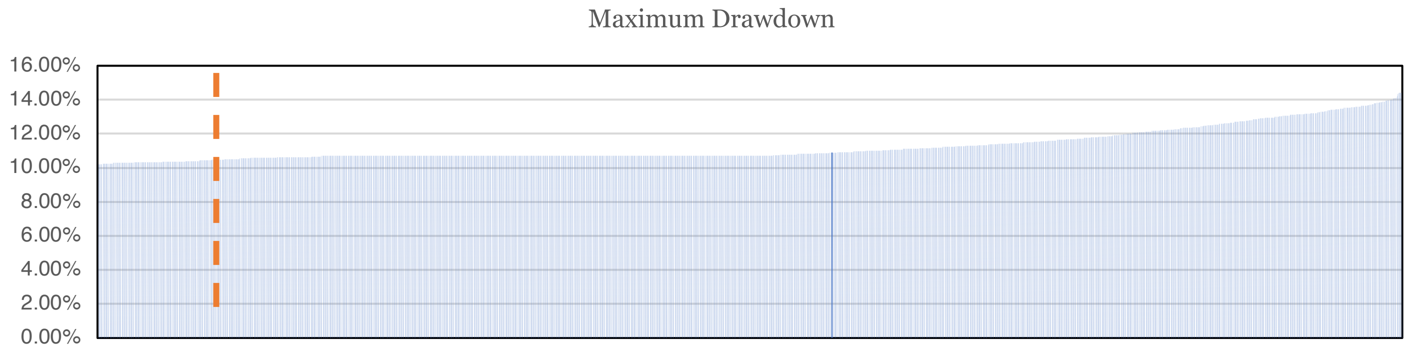
Note that each leg of the long/short portfolio does not necessarily equal 100% notional. This reflects conflicting signals in the underlying portfolios, causing the ensemble strategy to reduce its gross allocation as a reflection of uncertainty.

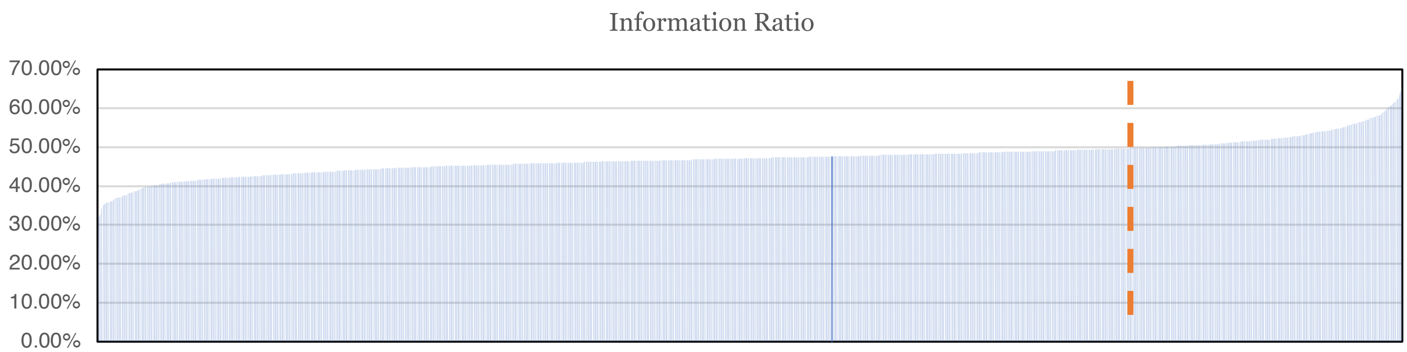
As a quick aside, we do want to highlight how the performance of the ensemble compares to the performance of the underlying strategies.

Below we plot the annualized return, annualized volatility, maximum drawdown, and information ratio of all the underlying equity curves of the strategies that make up the ensemble. We also identify the –ensemble approach–. While we can see that the ensemble approach brings the annualized return in-line with the median annualized return, its annualized volatility is in the 14th percentile and its maximum drawdown is in the 8th percentile.









Source: Tiingo. Calculations by Newfound Research. Tactical Credit strategy returns are hypothetical and backtested. Returns gross of all management fees and taxes, but net of underlying fund fees. Returns assume the reinvestment of all distributions.

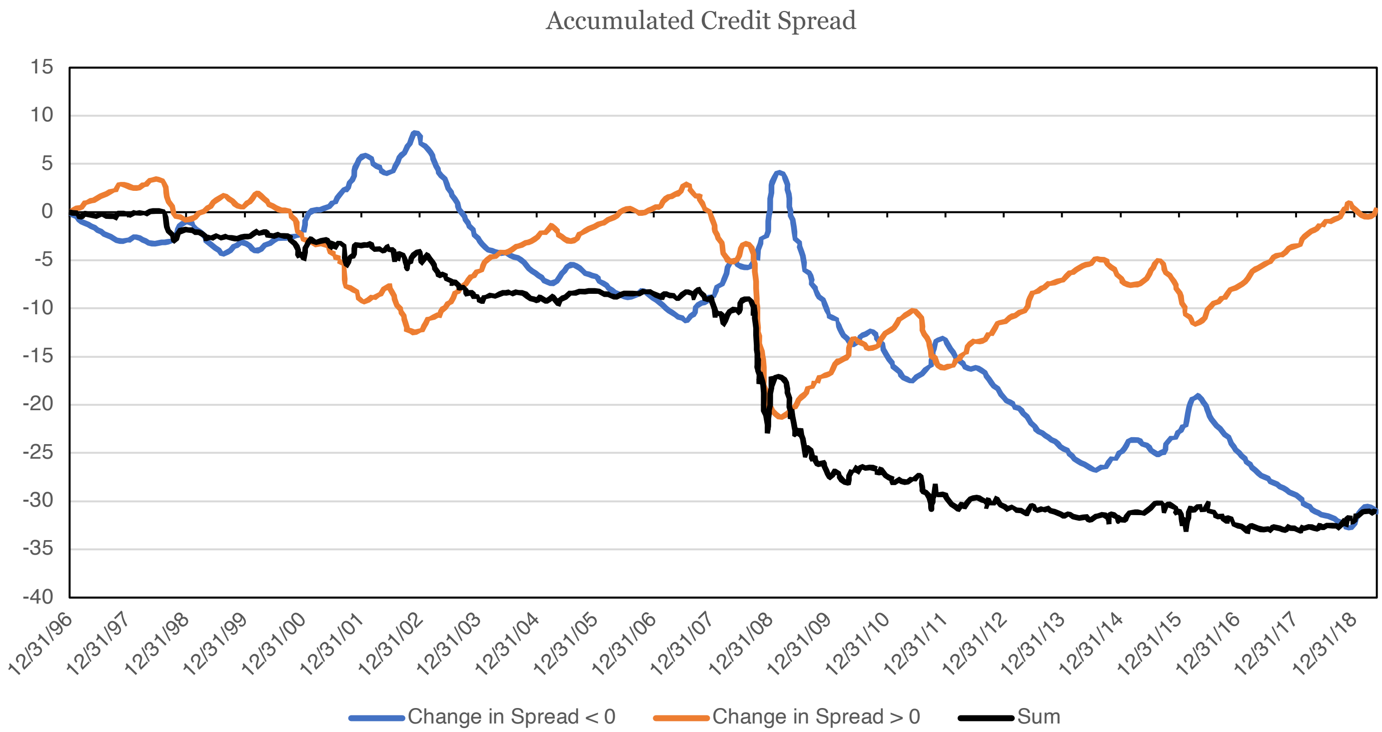
By maintaining the median annualized return and significantly reducing annualized volatility, the ensemble has an information ratio in the 78th percentile. As we've demonstrated in prior commentaries, by diversifying idiosyncratic specification risk, the ensemble approach is able to generate an information ratio significantly higher than the median without having to explicitly choose which specification we believe will necessarily outperform.

Given this ensemble implementation, we can now ask, "what is the driving force of strategy returns?" In other words, does the strategy create returns by harvesting price return differences or through carry (yield) differences?

One simple way of evaluating this question is by evaluating the strategy's sensitivity to changes in credit spreads. Specifically, we can calculate daily changes in the ICE BofAML US High Yield Master II Option-Adjusted Spread and multiply it against the strategy's exposure to high yield bonds on the prior day.

By accumulating these weighted changes over time, we can determine how much spread change the strategy has captured. We can break this down further by isolating positive and negative change days and trying to figure out whether the strategy has benefited from avoiding spread expansion or from harvesting spread contraction.

In the graph below, we can see that the strategy harvested approximately 35,000 basis points ("bps") from 12/1996 to present (the period for which credit spread data was available). Point-to-point, credit spreads actually widened by 100bps over the period, indicating that tactical changes were able to harvest significant changes in spreads.

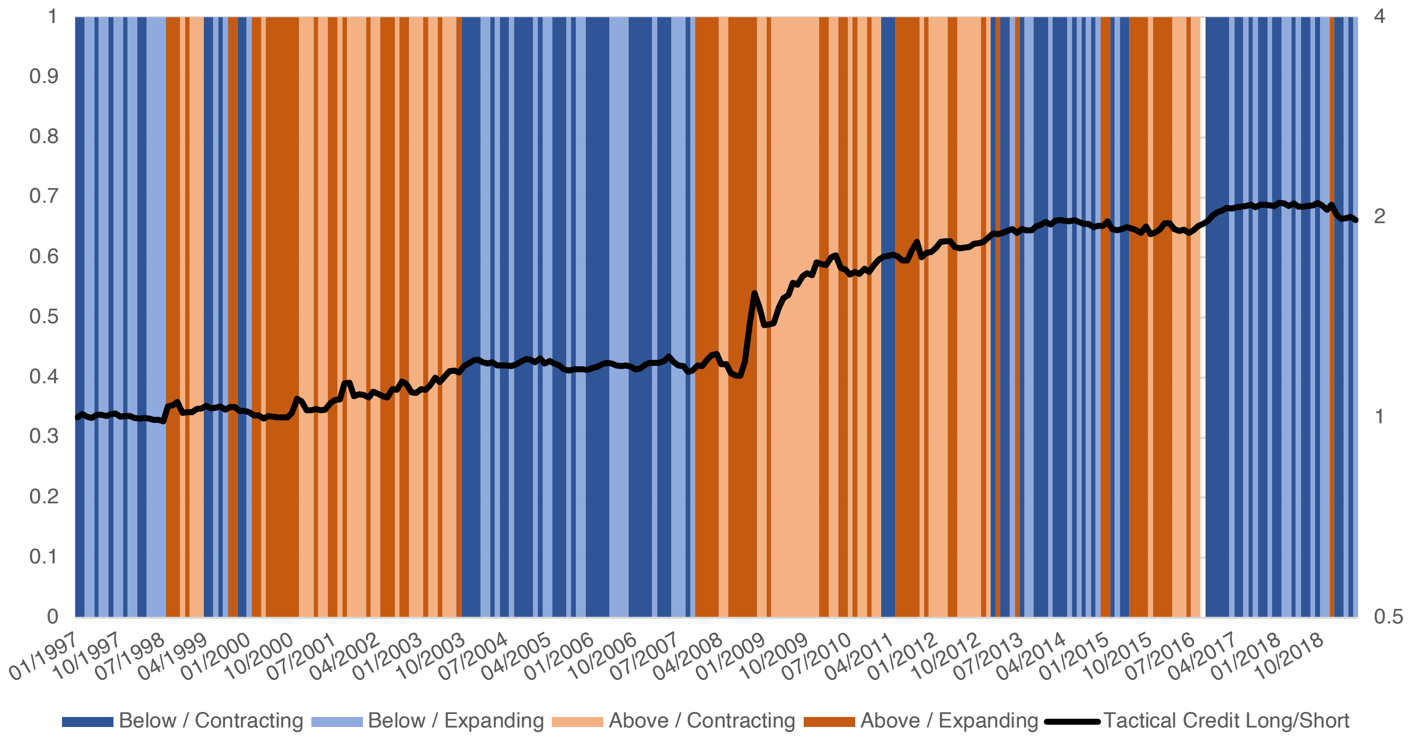


Source: St. Louis Federal Reserve. Calculations by Newfound Research.

We can see that over the full period, the strategy predominately benefited from harvesting contracting spreads, as exposure to expanding spreads had a cumulative net zero impact. This analysis is incredibly regime dependent, however, and we can see that periods like 2000-2003 and 2008 saw a large benefit from short-exposure in high yield during a period when spreads were expanding.

We can even see that in the case of post-2008, switching to long high yield exposure allowed the strategy to benefit from subsequent credit spread declines.

While this analysis provides some indication that the strategy benefits from harvesting credit spread changes, we can dig deeper by taking a regime-dependent view of performance. Specifically, we can look at strategy returns conditional upon whether spreads are above or below their long-term median, as well as whether they expand or contract in a given month.



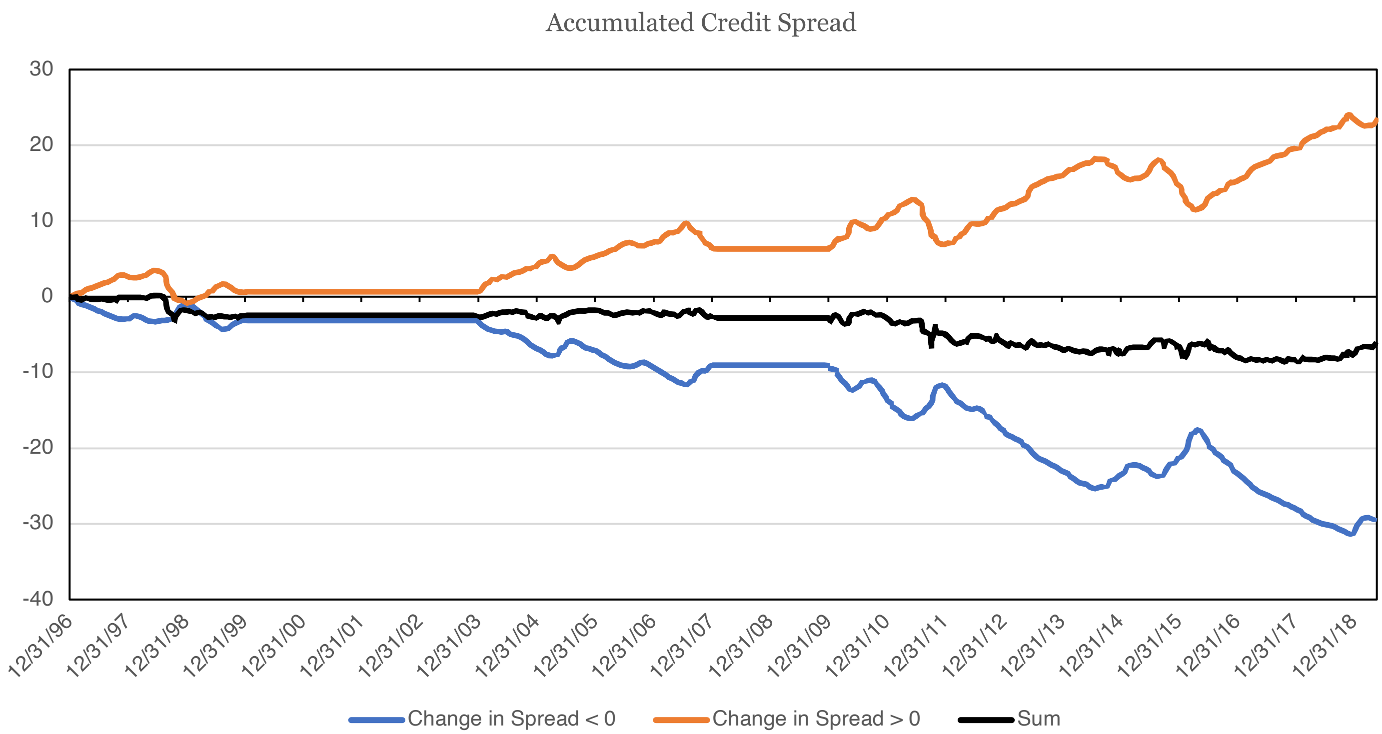
Source: St. Louis Federal Reserve. Calculations by Newfound Research. Tactical Credit strategy returns are hypothetical and backtested. Returns gross of all management fees and taxes, but net of underlying fund fees. Returns assume the reinvestment of all distributions.

Most of the strategy return appears to occur during times when spreads are above their long-term median. Calculating regime-conditional annualized returns confirms this view.

	Above	**Below**
Expanding	10.88%	-2.79%
Contracting	1.59%	4.22%

The strategy appears to perform best during periods when credit spreads are expanding above their long-term median level (e.g. crisis periods like 2008). The strategy appears to do its worst when spreads are below their median and begin to expand, likely representing periods when the strategy is generally long high yield but has not had a chance to make a tactical switch.

This all points to the fact that the strategy harvests almost all of its returns in crisis periods. In fact, if we remove 2000-2003 and 2008-2009, we can see that the captured credit spread declines dramatically.



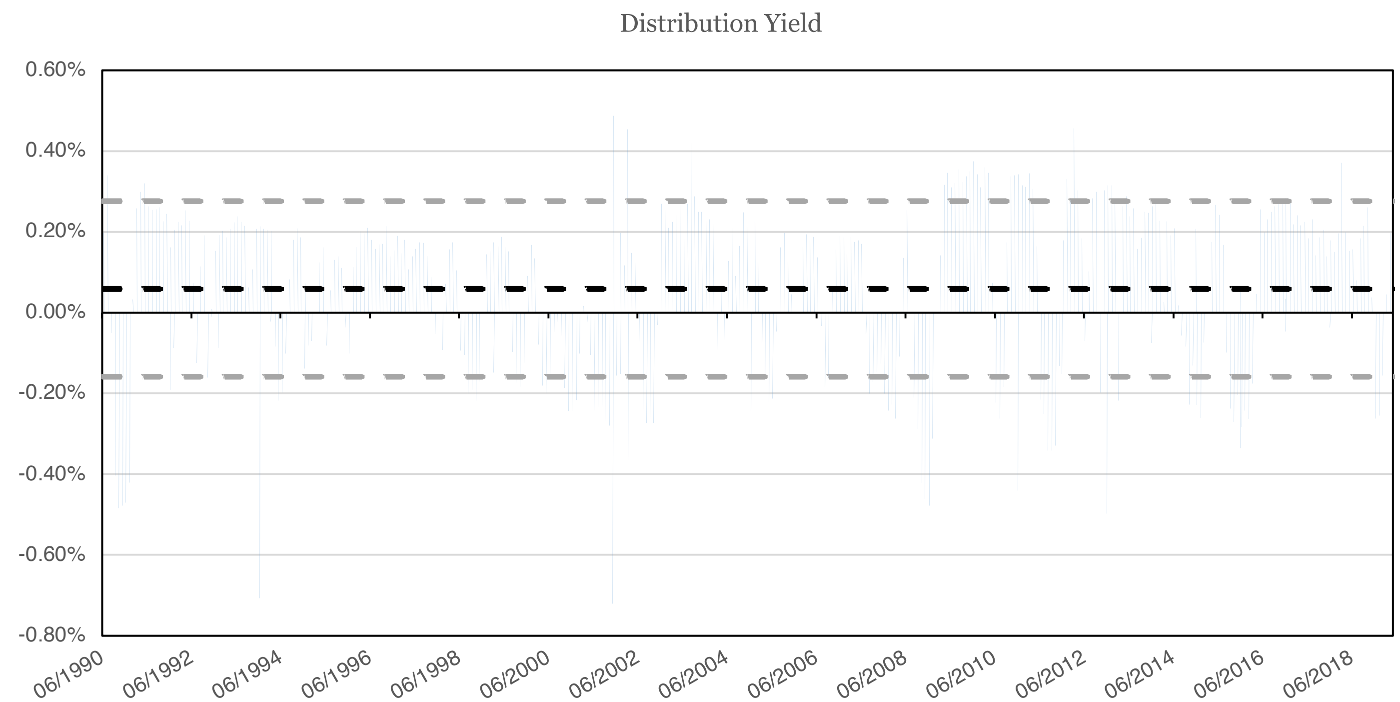
Source: St. Louis Federal Reserve. Calculations by Newfound Research.

Capturing price returns due to changes in credit spreads are not responsible for all of the strategy's returns, however.

Below we explicitly calculate the yield generated by the long/short strategy over time. As high yield corporates tend to offer higher yields, when the strategy is net long high yield, the strategy's yield is positive. On the other hand, when the strategy is net short high yield, the strategy's yield is negative.

This is consistent with our initial view about why these sorts of tactical strategies can be so difficult. During the latter stages of the 2008 crisis, the long/short strategy had a net negative yield of close to -0.5% per month.[32] Thus, the cost of carrying this tactical position is rather expensive and places a larger burden on the strategy accurately timing price return.

[32] Note that this does not account for actual borrowing costs. However, as these tactical strategies are implemented as long only, our "short" is implicit, not explicit, and therefore borrowing costs are not relevant.



Source: Tiingo. Calculations by Newfound Research. Tactical Credit strategy returns are hypothetical and backtested. Returns gross of all management fees and taxes, but net of underlying fund fees.

From this graph, we believe there are two interesting things worth calling out:

- The long-run average yield is positive, representing the strategy's ability to capture carry differences between high yield and core bonds.
- In the post-crisis environments, the strategy generates yields in excess of one standard deviation of the full-period sample, indicating that the strategy may have benefited from allocating to high yield when yields were abnormally large.

To better determine whether capturing changes in credit spreads or carry differences had a larger impact on strategy returns, we can explicitly calculate the –price– and –total return– indices of the ensemble strategy.

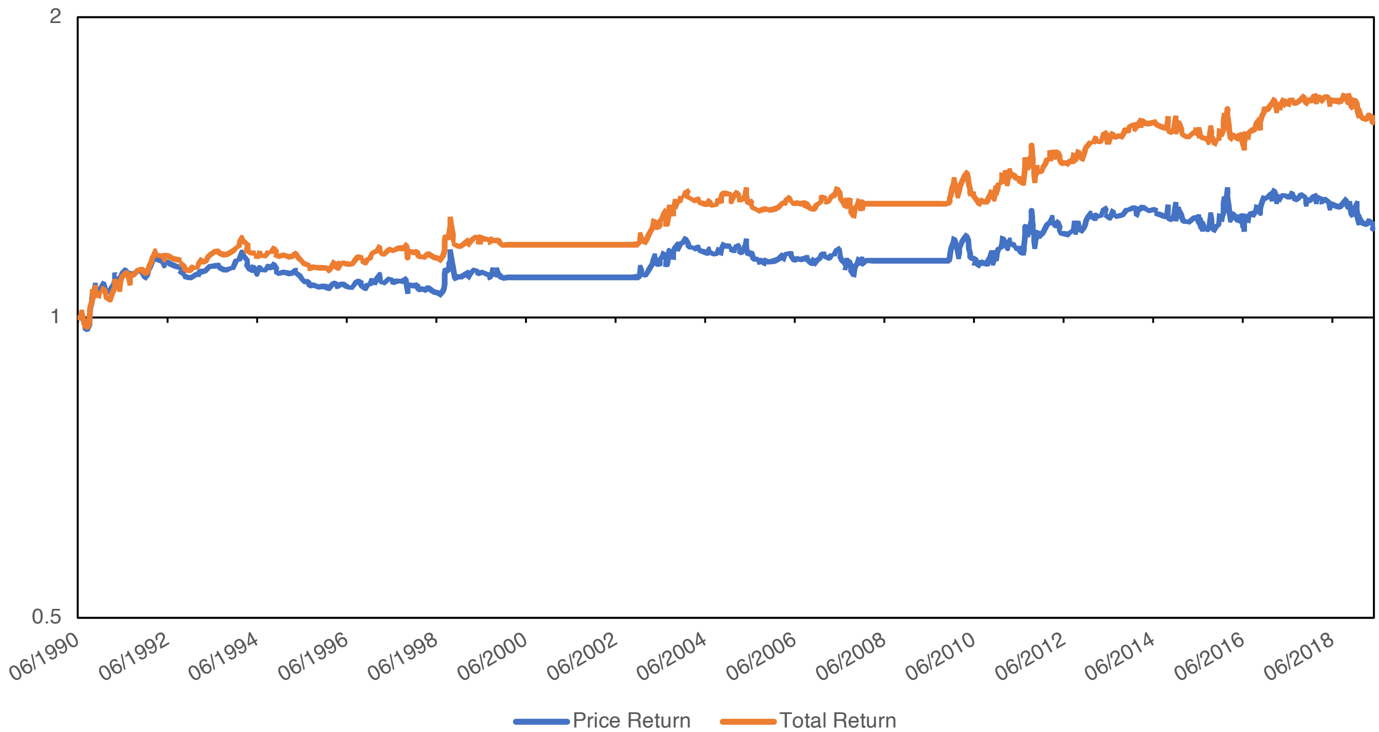
Source: St. Louis Federal Reserve. Calculations by Newfound Research. Tactical Credit strategy returns are hypothetical and backtested. Returns gross of all management fees and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

The –price return– and –total return– series return 2.1% and 2.9% annualized respectively, implying that capturing price return effects account for approximately 75% of the strategy's total return.

This is potentially concerning, because we have seen that the majority of the price return comes from a single regime: when credit spreads are above their long-term median and expanding. As we further saw, simply removing the 2000-2003 and 2008-2009 periods significantly reduced the strategy's ability to harvest these credit spread changes.

While the strategy may appear to be supported by nearly 30-years of empirical evidence, in reality we have a situation where the vast majority of the strategy's returns were generated in just two regimes.

If we remove 2000-2003 and 2008-2009 from the return series, however, we can see that the total return of the strategy only falls to 0.7% and. 1.6% annualized for –price return– and –total return– respectively. While this may appear to be a precipitous decline, it indicates that there may be potential to capture both changes in credit spread and net carry differences even in normal market environments so long as implementation costs are kept low enough.



Source: St. Louis Federal Reserve. Calculations by Newfound Research. Tactical Credit strategy returns are hypothetical and backtested. Returns gross of all management fees and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

Conclusion

In this commentary, we explored a tactical credit strategy that switched between high yield corporate bonds and core fixed income. We decompose these strategies into a 50% high yield / 50% core fixed income portfolio that is overlaid with 50% exposure to a dollar-neutral long/short strategy that captures the tactical tilts. We focus our exploration on the dollar-neutral long/short portfolio, as it isolates the active bets of the strategy.

Using cross-sectional momentum, we found that short-term signals with formation periods ranging from 3-to-5 months were statistically significant, so long as the holding period was sufficiently short.

We used this information to construct an ensemble strategy made out of more than 1,800 underlying strategy specifications. Consistent with past research, we found that the ensemble closely tracked the median annualized return of the underlying strategies, but had significantly lower volatility and maximum drawdown, leading to a higher information ratio.

We then attempted to deconstruct where the strategy generated its returns from. We found that a significant proportion of total returns were achieved during periods when credit spreads were above their long-term median and expanding. This is consistent with periods of economic volatility such as 2000-2003 and 2008-2009.

The strategy also benefited from harvesting net carry differences between high yield and core fixed income. Explicitly calculating strategy price and total return, we find that this carry component accounts for approximately 25% of strategy returns.

The impact of the 2000-2003 and 2008-2009 periods on strategy returns should not be understated. Removing these time periods reduced strategy returns from 2.9% to 1.6% annualized. Interestingly, however, the proportion of total return explained by net carry only increased from 25% to 50%, potentially indicating that the strategy was still able to harvest some opportunities in changing credit spreads.

For investors evaluating these types of strategies, cost will be an important component. While environments like 2008 may lead to opportunities for significant out-performance, without them the strategy may offer anemic returns. This is especially true when we recall that a long-only implementation only has 50% implicit exposure to the long/short strategy we evaluated in this piece.

Thus, the 2.9% annualized return is really closer to a 1.5% annualized excess return above the 50/50 portfolio. For the ex-crisis periods, the number is closer to 0.8% annualized. When we consider that this analysis was done without explicit consideration for management costs or trading costs and we have yet to apply an appropriate expectation haircut given the fact that this analysis was all backtested, there may not be sufficient juice to squeeze.

That said, we only evaluated a single signal in this piece. Combining momentum with valuation, carry, or even macro-economic signals may lead to significantly better performance. Further, high yield corporates is a space where empirical evidence suggests that security selection can make a large difference. Careful selection of funds may lead to meaningfully better performance than just broad asset class exposure.

QUANTITATIVE STYLES AND MULTI-SECTOR BONDS

June 10, 2019

SUMMARY

- In this commentary we explore the application of several quantitative signals to a broad set of fixed income exposures.
- Specifically, we explore value, momentum, carry, long-term reversals, and volatility signals.
- We find that value, 3-month momentum, carry, and 3-year reversals all create attractive quantile profiles, potentially providing clues for how investors might consider pursuing higher returns or lower risk.
- This study is by no means comprehensive and only intended to invite further research and conversation around the application of quantitative styles across fixed income exposures.

In *Navigating Municipal Bonds with Factors*, we employed momentum, value, carry, and low-volatility signals to generate a sector-based approach to navigating municipal bonds.

In this article, we will introduce an initial data dive into applying quantitative signals to a broader set of fixed income exposures. Specifically, we will incorporate 17 different fixed income sectors, spanning duration, credit, and geographic exposure.

- **U.S. Treasuries**: Near (3-Month), short (1-3 Year), mid (3-5 Year) intermediate (7-10 Year), and long (20+ Year).
- **Investment-Grade Corporates**: Short-term, intermediate-term, and Floating Rate corporate bonds.
- **High Yield**: Short- and intermediate-term high yield.
- **International Government Bonds**: Currency hedged and un-hedged government bonds.
- **Emerging Market**: Local and US dollar denominated.
- **TIPs**: Short- and intermediate-term TIPs.
- **Mortgage-Backed**: Investment grade mortgage-backed bonds.

In this study, each exposure is represented by a corresponding ETF. We extend our research prior to ETF launch by employing underlying index data the ETF seeks to track.

The quantitative styles we will explore are:

- **Momentum:** Buy recent winners and sell recent losers.
- **Value:** Buy cheap and sell expensive.
- **Carry:** Buy high carry and sell low carry.
- **Reversal:** Buy long-term losers and sell long-term winners.

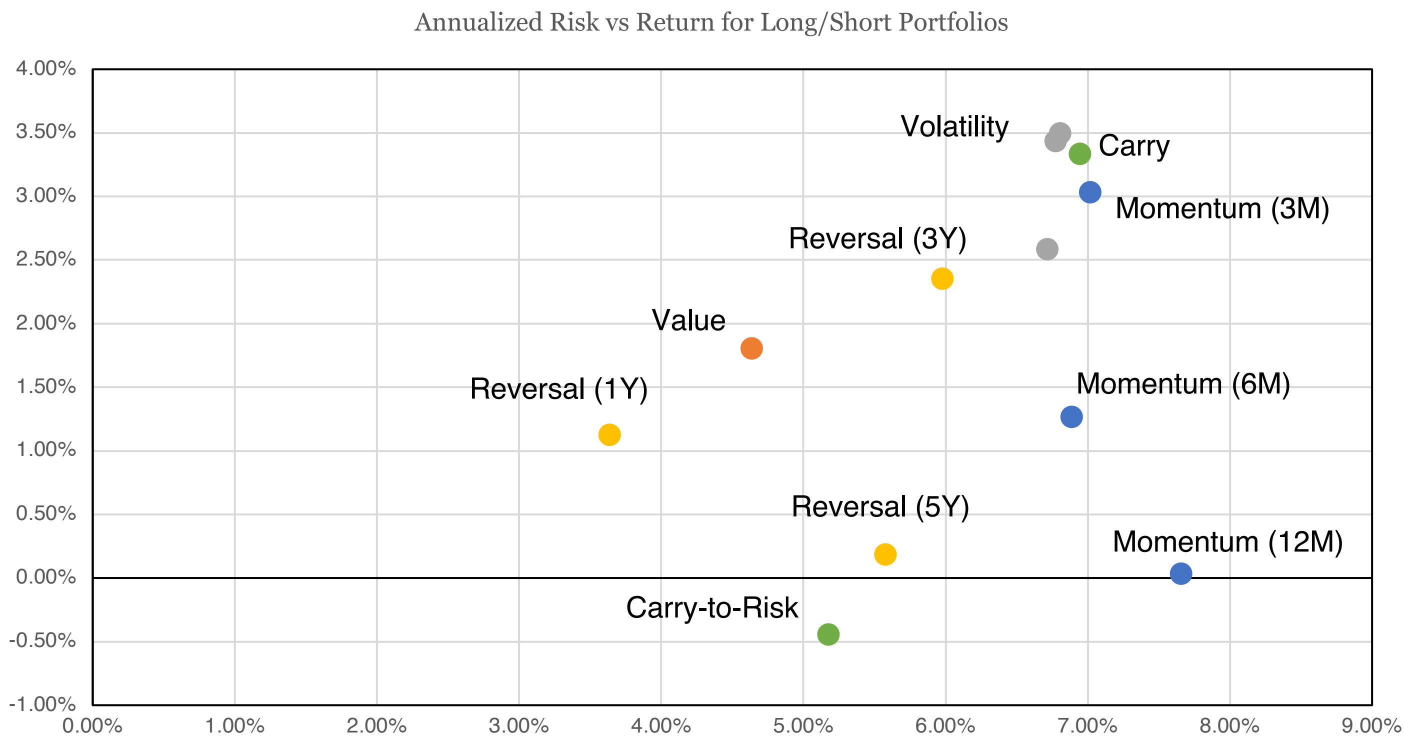
- **Volatility:** Buy high volatility and sell low volatility.[33]

The details of each style are explained in greater depth in each section below.

Note that the analysis herein is by no means meant to be prescriptive in any manner, nor is it a comprehensive review. Rather, it is meant as a launching point for further commentaries we expect to write.

At the risk of spoiling the conclusion, below we plot the annualized returns and volatility profiles of dollar-neutral long-short portfolios.[34] We can see that short-term Momentum, Value, Carry, and Volatility signals generate positive excess returns over the testing period.

Curiously, longer-term Momentum does not seem to be a profitable strategy, despite evidence of this approach being rather successful for many other asset classes.



Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

---

[33] Note that this is the opposite of typical betting-against-beta and low-volatility portfolio construction. In this construction, we are basically confirming "higher risk implies higher reward." More on this later.

[34] Exposures are rank-weighted in the same fashion long/short portfolios were constructed in *Navigating Municipal Bonds with Factors*.

However, these results are not achievable by most investors who may be constrained to a long-only implementation. Even when interpreted as over- and under-weight signals, the allocations in the underlying long/short portfolios differ so greatly from benchmark exposures, they would be nearly impossible to implement.

For a long-only investor, then, what is more relevant is how these signals forecast performance of different rank orderings of portfolios. For example, how does a portfolio of the best-ranking 3-month momentum exposures compare to a portfolio of the worst-ranking?

In the remainder of this commentary, we explore the return and risk profiles of quintile portfolios formed on each signal. To construct these portfolios, we rank order our exposures based on the given quantitative signal and equally-weight the exposures falling within each quintile.

Momentum

We generate momentum signals by computing 12-, 6- and 3- month prior total returns to reflect slow, intermediate, and fast momentum signals. Low-ranking exposures are those with the lowest prior total returns, while high ranking exposures have the highest total returns.

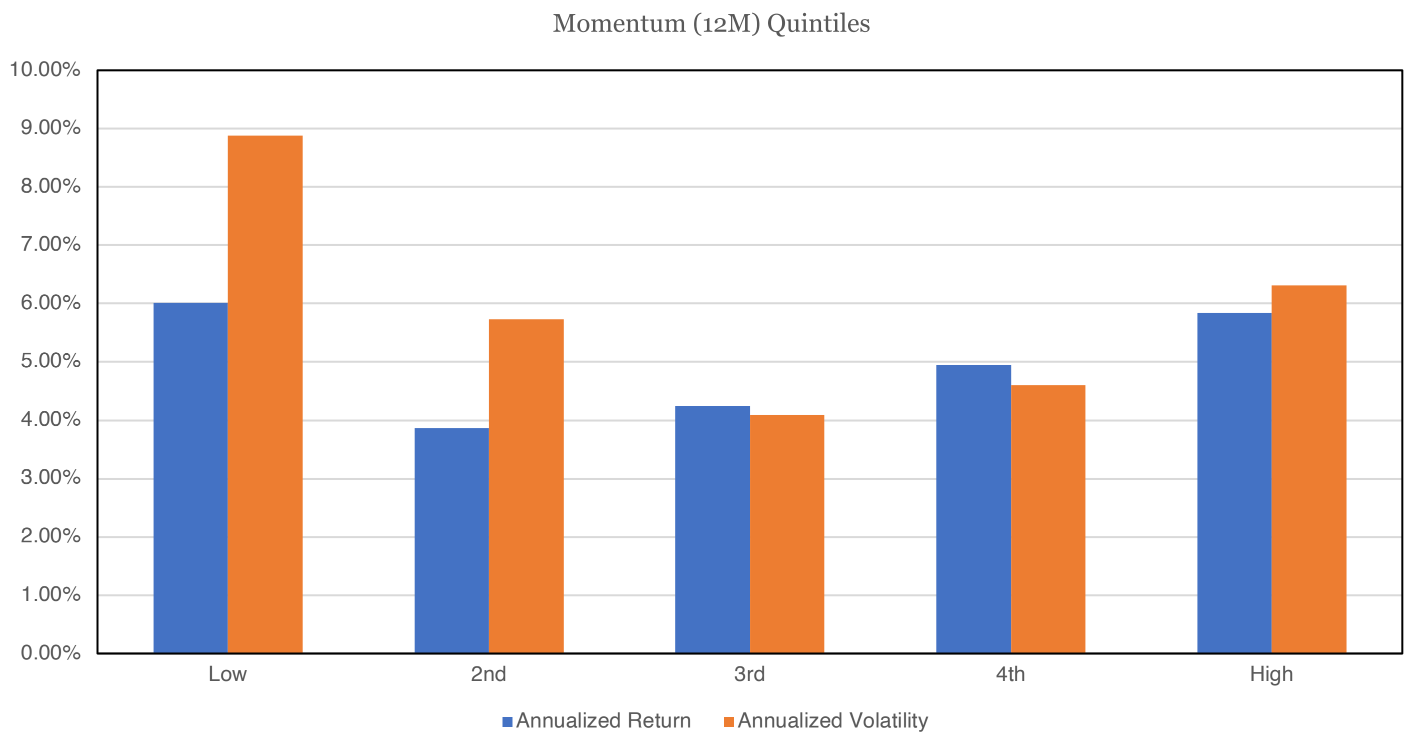
The portfolios assume a 1-month holding period for momentum signals. To avoid timing luck, four sub-indexes are used, each rebalancing on a different week of the month.

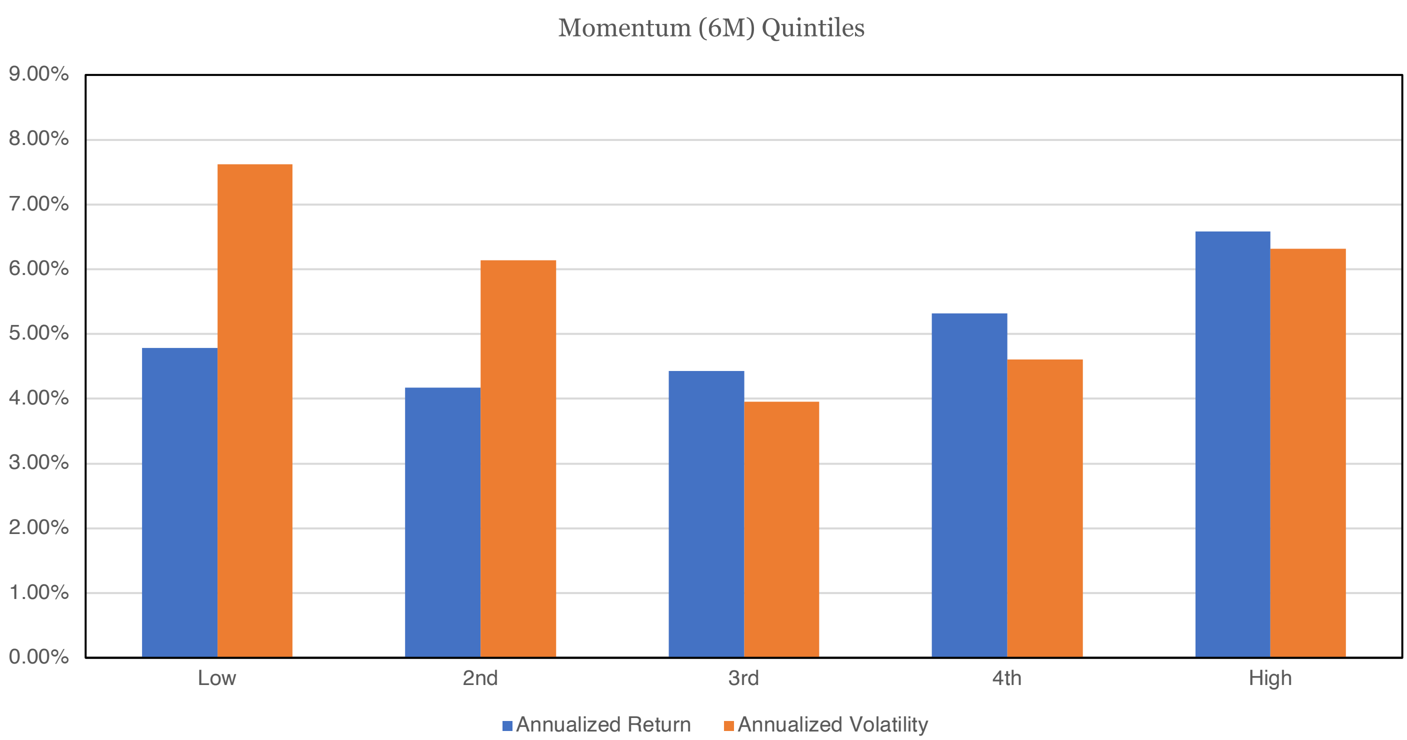
Annualized return and volatility numbers for the quintiles are plotted below.

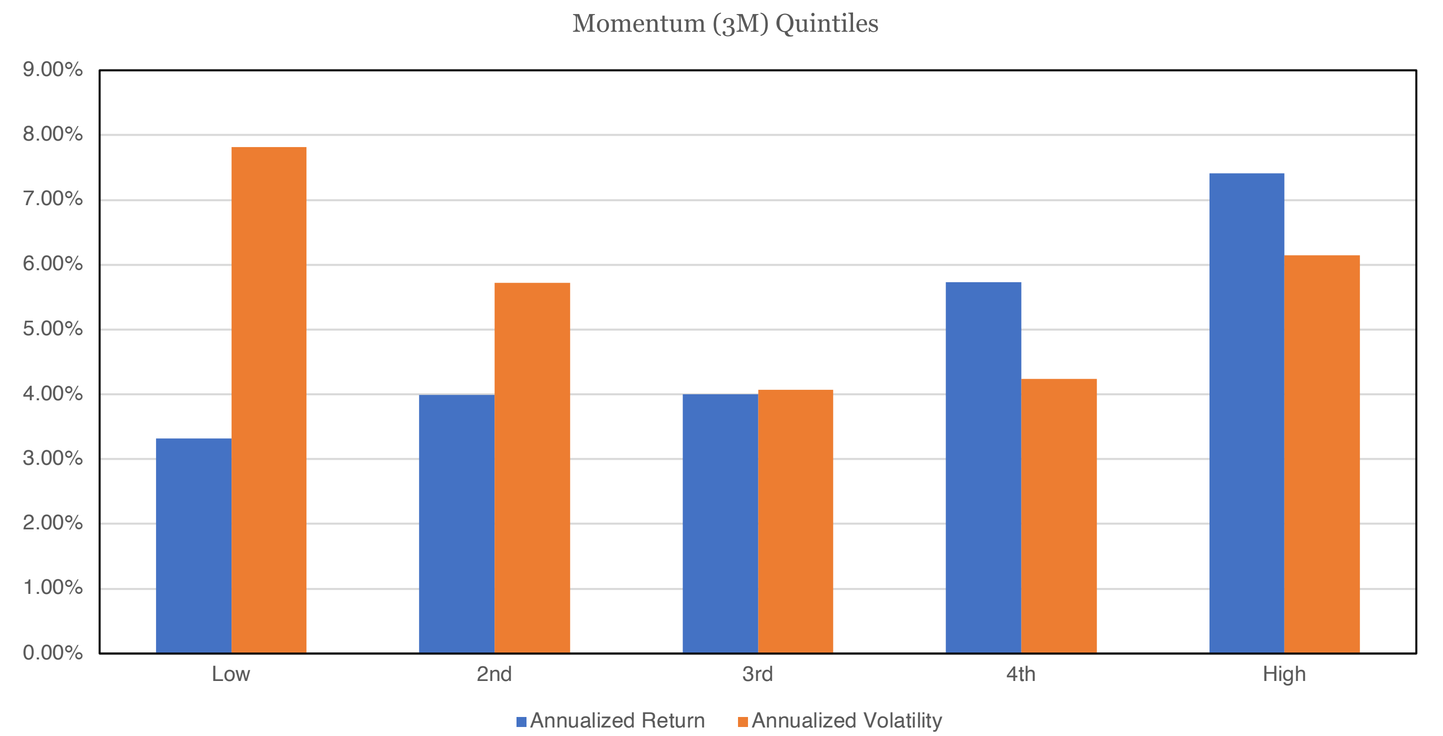
A few interesting data-points stand out:

- For 12-month prior return, the lowest quintile actually had the highest total return. However, it has a dramatically lower Sharpe ratio than the highest quintile, which only slightly underperforms it.

- Total returns among the highest quintile increase by 150 basis points ("bps") from 12-month to 3-month signals, and 3-month rankings create a more consistent profile of increasing total return *and* Sharpe ratio. This may imply that short-term signals are more effective for fixed income.


Momentum (12M) Quintiles
10.00%
9.00%
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
Low
2nd
3rd
4th
High
Annualized Return
Annualized Volatility


Momentum (6M) Quintiles
9.00%
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
Low
2nd
3rd
4th
High
Annualized Return
Annualized Volatility



Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

Carry

Carry is the expected excess return of an asset assuming price does not change. For our fixed income universe, we proxy carry using yield-to-worst minus the risk-free rate. For non-Treasury holdings, we adjust this figure for expected defaults and recovery.

For reasonably efficient markets, we would expect higher carry to imply higher return, but not necessarily higher *risk-adjusted* returns. In other words, we earn higher carry as a reward for bearing more risk.

Therefore, we also calculate an alternate measure of carry: carry-to-risk. Carry-to-risk is calculated by taking our carry measure and dividing it by recent realized volatility levels. One way of interpreting this figure is as forecast of Sharpe ratio. Our expectation is that this signal may be able to identify periods when carry is episodically cheap or rich relative to prevailing market risk.

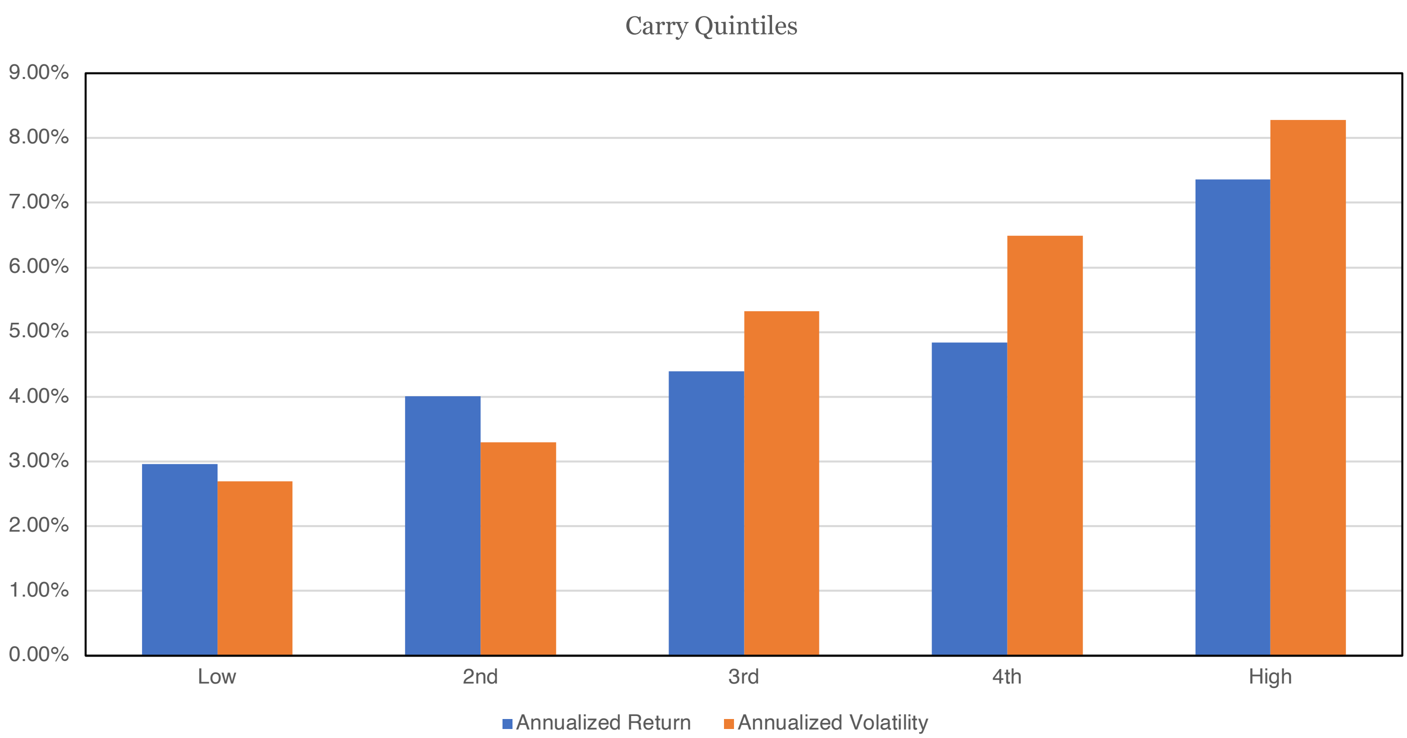
The portfolios assume a 12-month holding period for carry signals. To avoid timing luck, 52 sub-indexes are used, each rebalancing on a different week of the year.

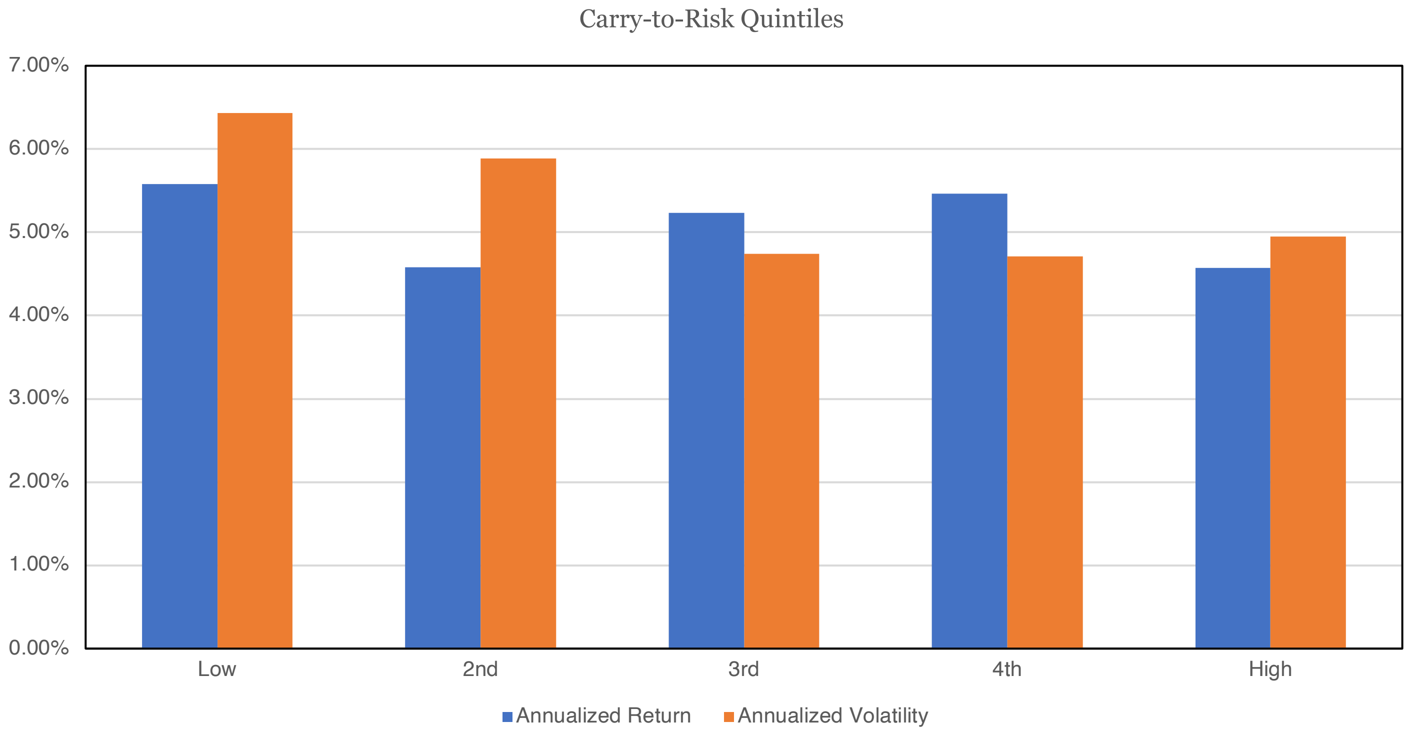
We see:

- Higher carry implies a higher return as well as a higher volatility. As expected, no free lunch here.

- Carry-to-risk does not seem to provide a meaningful signal. In fact, low carry-to-risk outperforms high carry-to-risk by 100bps annualized.

- Volatility meaningfully declines for carry-to-risk quintiles, potentially indicating that this integrated carry/volatility signal is being too heavily driven by volatility.





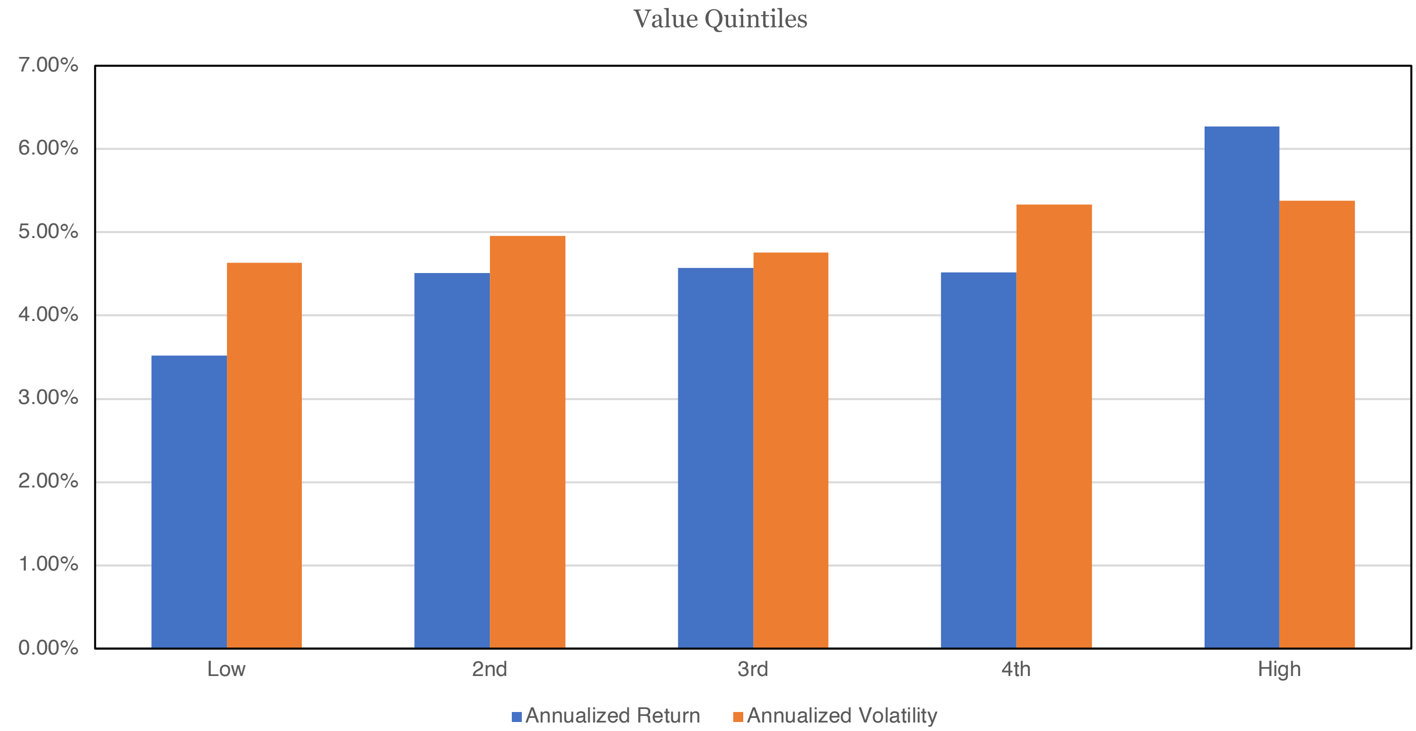
Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

Value

In past commentaries, we have used real yield as our value proxy in fixed income. In this commentary, we deviate from that methodology slightly and use a time-series z-score of carry as our value of measure. Historically high carry levels are considered to be cheap while historically low carry levels are considered to be expensive.

The portfolios assume a 12-month holding period for value signals. To avoid timing luck, 52 sub-indexes are used, each rebalancing on a different week of the year.

We see not only a significant increase in total return in buying cheap versus expensive holdings, but also an increase in risk-adjusted returns.



Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

Reversal

Reversal signals are the opposite of momentum: we expect past losers to outperform and past winners to underperform. Empirically, reversals tend to occur over very short time horizons (e.g. 1 month) and longer-term time horizons (e.g. 3- to 5-years). In many ways, long-term reversals can be thought of as a naive proxy for value, though there may be other behavioral and structural reasons for the historical efficacy of reversal signals.

We must be careful implementing reversal signals, however, as exposures in our universe have varying return dynamics (e.g. expected return and volatility levels).

To illustrate this problem, consider the simple two-asset example of equities and cash. A 3-year reversal signal would sell the asset that has had the best performance over the prior 3-years and buy the asset that has performed the worst. The problem is that we expect stocks to outperform cash due to the equity risk premium. Naively ranking on prior returns alone would have us out of equities during most bull markets.

Therefore, we must be careful in ranking assets with meaningfully different return dynamics.

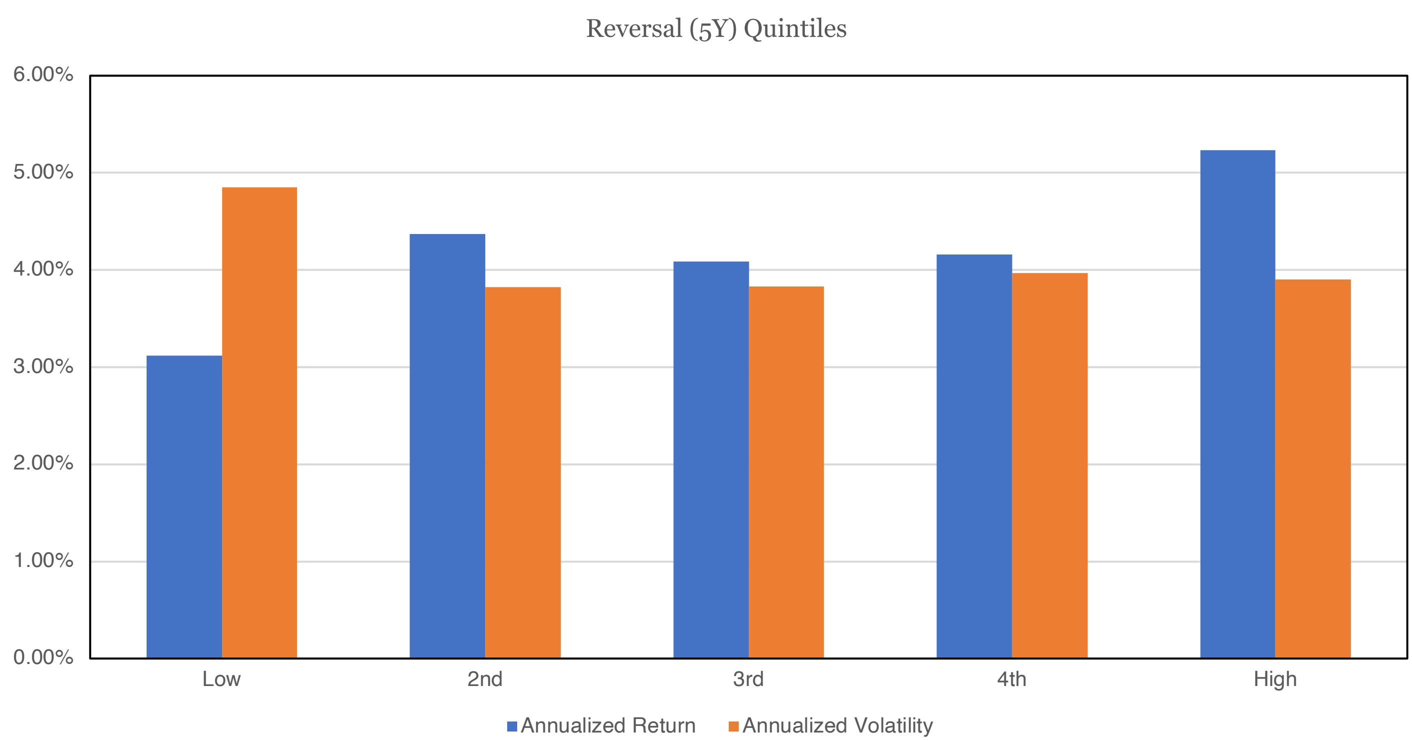
(Why, then, can we do it for momentum? In a sense, momentum is explicitly trying to exploit the relative time-series properties over a short-term horizon. Furthermore, in a universe that contains low-risk, low-return assets, cross-sectional momentum can be thought of as an integrated process between time-series momentum and cross-sectional momentum, as the low-risk asset will bubble to the top when absolute returns are negative.)

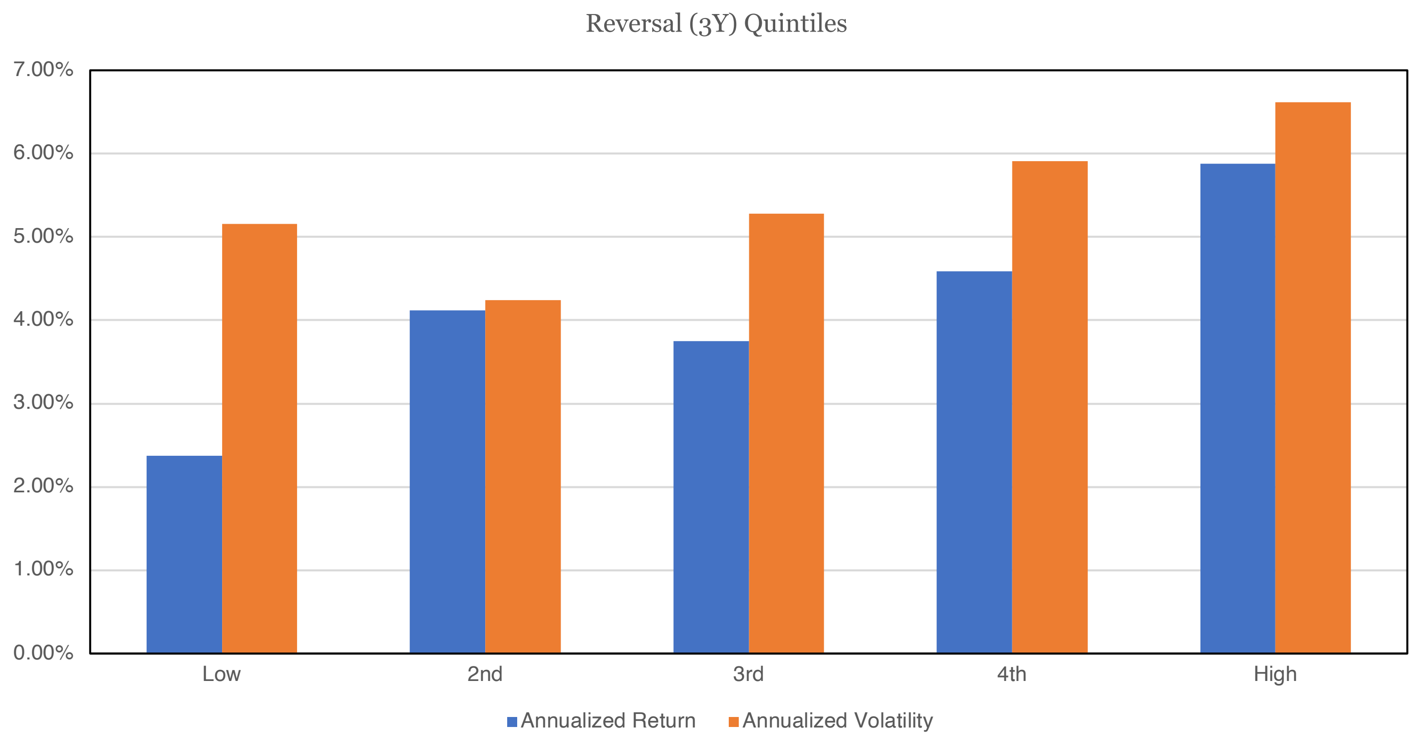
To account for this, we use a time-series z-score of prior returns to create a reversal signal. For example, at each point in time we calculate the current 3-year return and z-score it against all prior rolling 3-year periods.

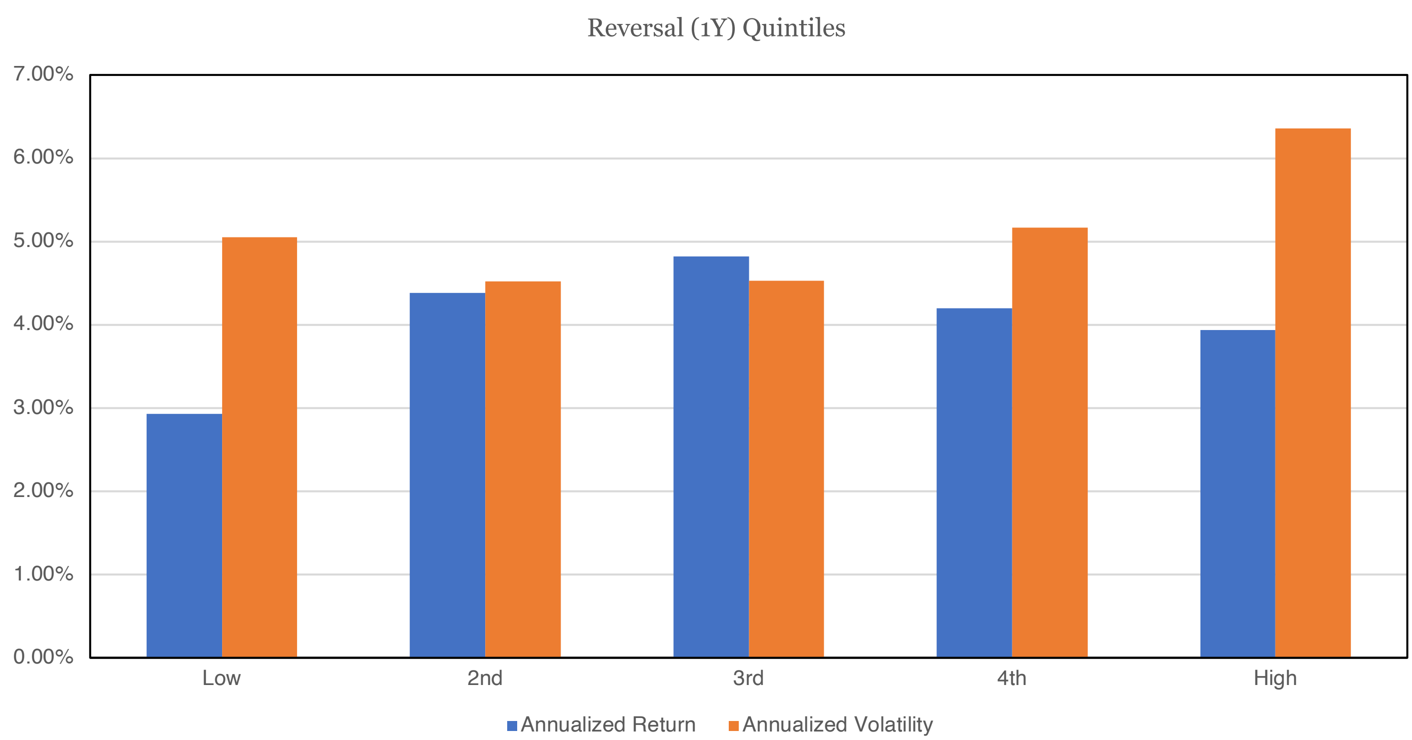
Note that in this construction, high z-scores will reflect higher-than-normal 3-year numbers and low z-scores will reflect lower-than-normal 3-year returns. Therefore, we negate the z-score to generate our signal such that low-ranked exposures reflect those we want to sell and high-ranked exposures reflect those we want to buy.

The portfolios assume a 12-month holding period for value signals. To avoid timing luck, 52 sub-indexes are used, each rebalancing on a different week of the year.

Plotting the results below for 1-, 3-, and 5-year reversal signals, we see that 3- and 5-year signals see a meaningful increase in both total return and Sharpe ratio between the lowest quintile.




Reversal (3Y) Quintiles
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
Low
2nd
3rd
4th
High
Annualized Return
Annualized Volatility


Reversal (1Y) Quintiles
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
Low
2nd
3rd
4th
High
Annualized Return
Annualized Volatility

Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

Volatility

Volatility signals are trivial to generate: we simply sort assets based on prior realized volatility. Unfortunately, exploiting the low-volatility anomaly is difficult without leverage, as the empirically higher risk-adjusted return exhibited by low-volatility assets typically coincides with lower total returns.

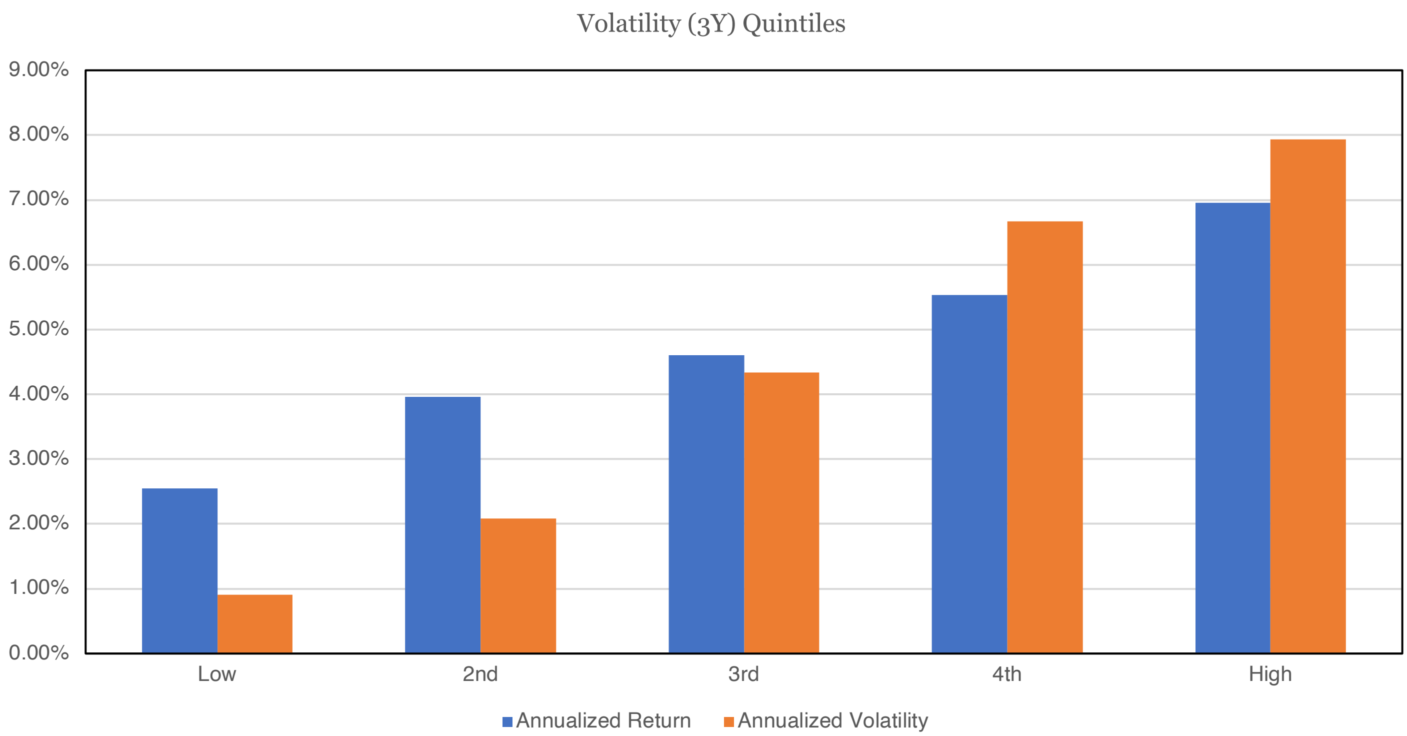
For example, in the tests below the low quintile is mostly comprised of short-term Treasuries and floating rate corporates. The top quintile is allocated across local currency emerging market debt, long-dated Treasuries, high yield bonds, and unhedged international government bonds.

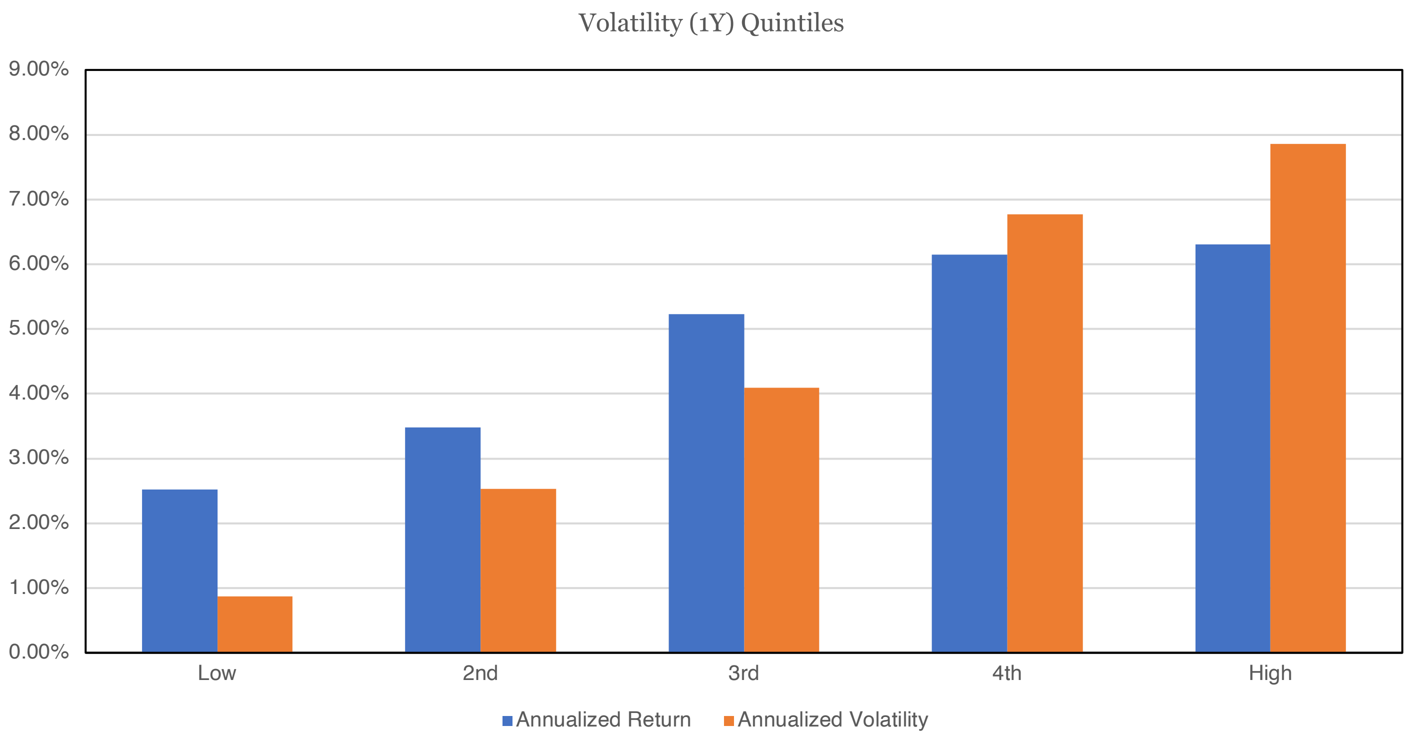
As a side note, for the same reason we z-scored reversal signals, we also hypothesized that z-scoring may work on volatility. Beyond these two sentences, the results were nothing worth writing about.

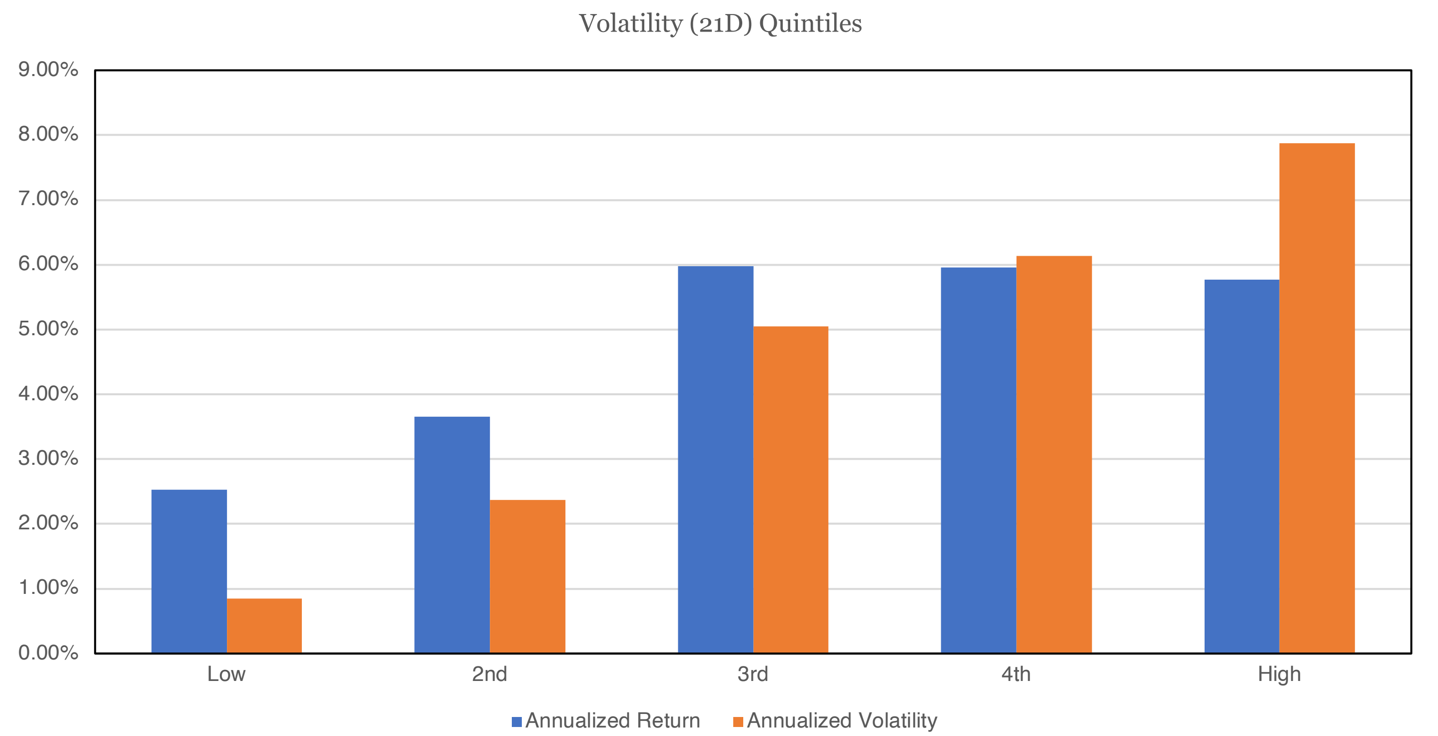
Nevertheless, we can still attempt to confirm the existence of the low-volatility anomaly in our investable universe by ranking assets on their past volatility.

The portfolios assume a 1-month holding period for momentum signals. To avoid timing luck, four sub-indexes are used, each rebalancing on a different week of the month.

Indeed, in plotting results we see that the lowest volatility quintiles have significantly higher realized Sharpe ratios.


Volatility (3Y) Quintiles
9.00%
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
Low
2nd
3rd
4th
High
Annualized Return
Annualized Volatility


Volatility (1Y) Quintiles
9.00%
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
Low
2nd
3rd
4th
High
Annualized Return
Annualized Volatility



Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

Of the results plotted above, our eyes might be drawn to the results in the short-term volatility measure. It would appear that the top quintile has both a lower total return and much higher volatility than the 3rd and 4th quintiles. This might suggest that we could improve our portfolios risk-adjusted returns without sacrificing total return by avoiding those top-ranked assets.

Unfortunately, this is not so clear cut. Unlike the other signals where the portfolios had meaningful turnover, these quintiles are largely stable. This means that the results are driven more by the composition of the portfolios than the underlying signals. For example, the 3rd and 4th quintiles combine both Treasuries and credit exposure, which allows the portfolio to realize lower volatility due to correlation. The highest volatility quintile, on the other hand, holds both local currency emerging market debt and un-hedged international government bonds, introducing (potentially uncompensated) currency risk into the portfolio.

Thus, the takeaway may be more strategic than tactical: diversification is good and currency exposure is going to increase your volatility.

Oh – and allocating to zero-to-negatively yielding foreign bonds isn't going to do much for your return unless currency changes bail you out.

Conclusion

In this study, we explored the application of value, momentum, carry, reversal, and volatility signals across fixed income exposures. We found that value, 3-month momentum, carry, and 3-year reversal signals may all provide meaningful information about forward expected returns and risk.

Our confidence in this analysis, however, is potentially crippled by several points:

- The time horizon covered is, at best, two decades, and several economic variables are constant throughout it.
    - The inflation regime over the time period was largely uniform.
    - A significant proportion of the period covered had near-zero short-term Treasury yields and negative yields in foreign government debt.
- Reversal signals require a significant amount of formation data. For example, the 3-year reversal signal requires 6 years (i.e. 3-years of rolling 3-year returns) of data before a signal can be generated. This represents nearly 1/3rd of the data set.
- The dispersion in return dynamics (e.g. volatility and correlation) of the underlying assets can lead to the emergence of unintended artifacts in the data that may speak more to portfolio composition than the value-add from the quantitative signal.
- We did not test whether certain exposures or certain time periods had an outsized impact upon results.
- We did not thoroughly test stability regions for different signals.
- We did not test the impact of our holding period assumptions.
- Holdings within quantile portfolios were assumed to be equally weighted.

Some of these points can be addressed simply. Stability concerns, for example, can be addressed by testing the impact of varying signal parameterization.

Others are a bit trickier and require more creative thinking or more computational horsepower.

Testing for the outsized impact of a given exposure or a given time period, for example, can be done through sub-sampling and cross-validation techniques. We can think of this as the application of randomness to efficiently cover our search space.

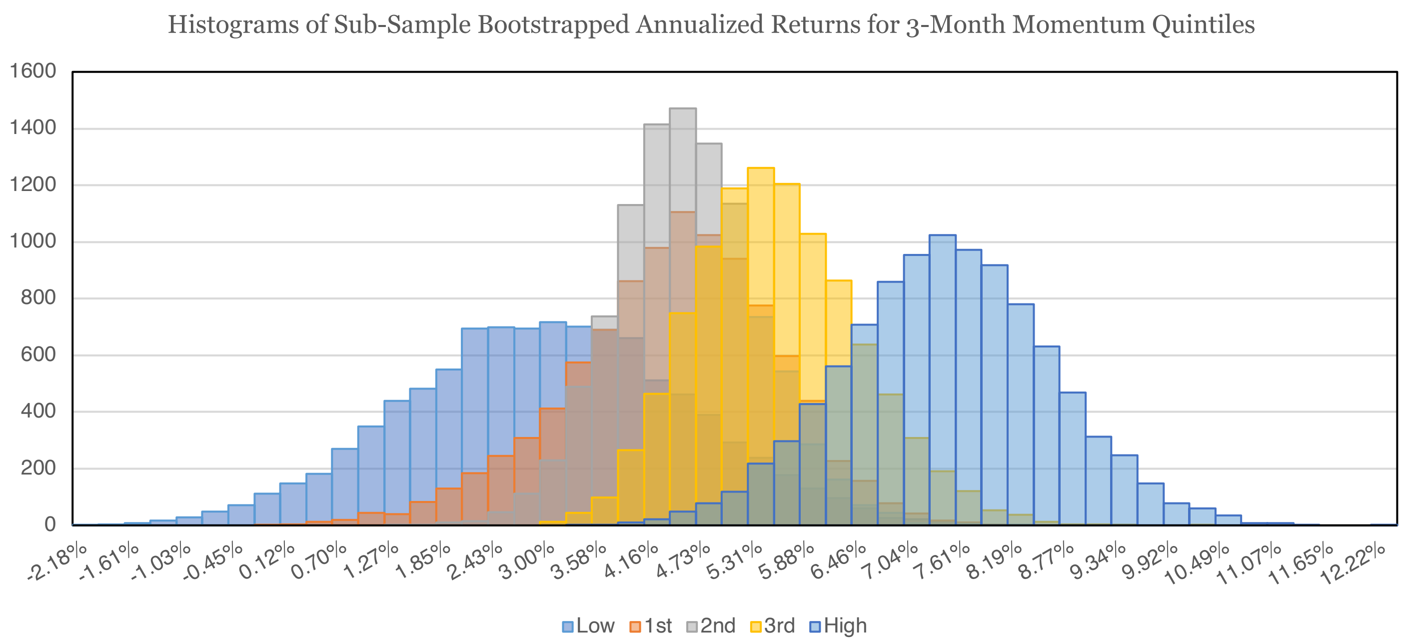
For example, below we re-create our 3-month momentum quintiles, but do so by randomly selecting only 10 of the exposures and 75% of the return period to test. We repeat this resampling 10,000 times for each quintile and plot the distribution of annualized returns below.

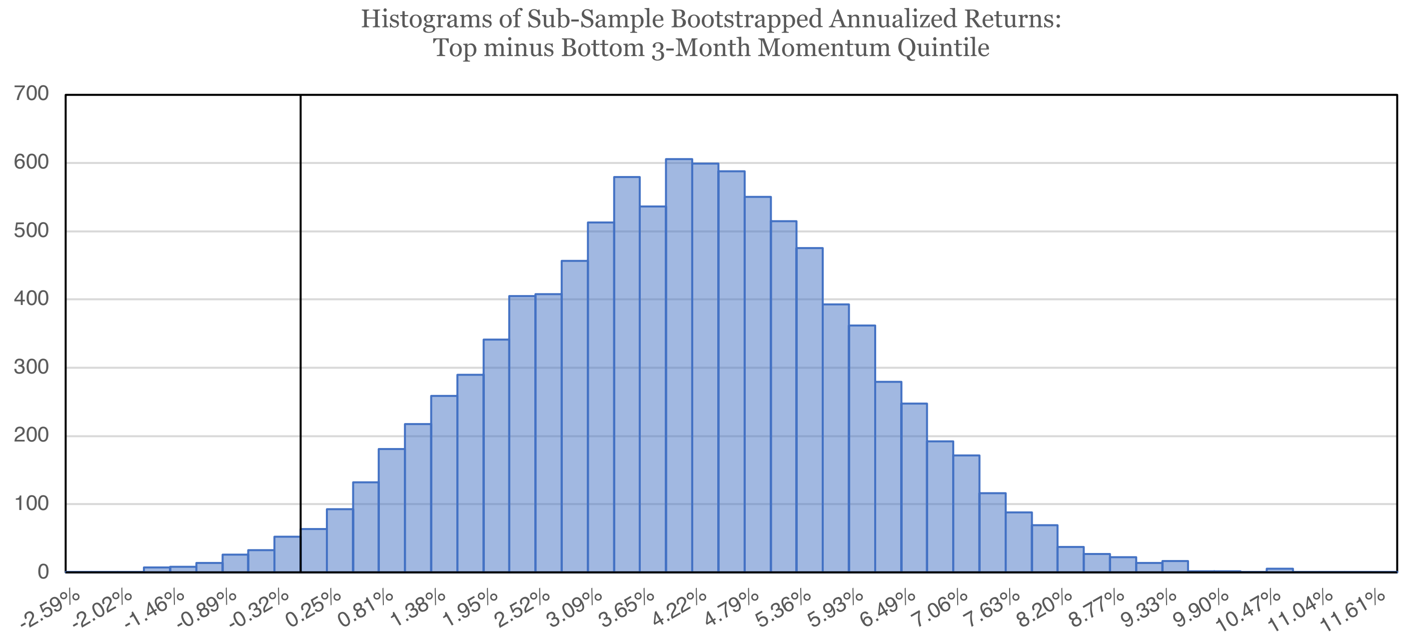
Even without performing an official difference-in-means test, the separation between the low and high quintile annualized return distributions provides a clue that the performance difference between these two is more likely to be a pervasive effect rather than due to an outlier holding or outlier time period.

We can make this test more explicit by using this subset resampling technique to bootstrap a distribution of annualized returns for a top-minus-bottom quintile long/short portfolio. Specifically, we randomly select a subset of assets and generate our 3-month momentum signals. We construct a dollar-neutral long/short portfolio by going long assets falling in the top quintile and short assets falling in the bottom quintile. We then select a random sub-period and calculate the annualized return.

Only 207 of the 10,000 samples fall below 0%, indicating a high statistical likelihood that the outperformance of recent winners over recent losers is not an effect dominated by a specific subset of assets or time-periods.

While this commentary provides a first step towards analyzing quantitative style signals across fixed income exposures, more tests need to be run to develop greater confidence in their efficacy.





Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

TIME-SERIES SIGNALS AND MULTI-SECTOR BONDS

June 17, 2019

SUMMARY

- We expand last week's commentary to explore momentum, carry, value, and long-term reversal signals in a time-series context.
- Using these signals, we generate long/short portfolios for each asset class. We use a sub-sampling methodology to bootstrap and annualized return distribution.
- We find that the signals are only selectively significant, and rarely consistent.
- We believe this initial study supports the idea that the application of these signals for the purpose of generating excess returns may not be supported. It is still possible, however, that these signals can meaningfully manipulate return distributions in other ways (e.g. reduce drawdowns) that investors may value.

This week's commentary aims to extend last week's commentary (*Quantitative Styles and Multi-Sector Bonds*) by evaluating the same quantitative signals in a time-series, rather than cross-sectional, context.

With cross-sectional signals, we are making a relative comparison and asking, "which of these securities do we prefer?" With time-series signals, we are making an absolute comparison and are asking a different question, "do I want to hold this security at all?"

For example, consider a simple momentum signal. In last week's commentary, we used prior returns to rank securities in quintiles and compared the performance of those quintiles, attempting to determine if we could find an edge by selecting securities that had recently outperformed their peers.

This week, we will employ the same total return calculation, but simply measure whether it is positive or negative. When it is positive, we will invest and when it is negative, we will go short.

In contrast to last week's analysis where we built portfolios, this week we will evaluate each security individually. Specifically, we will construct dollar-neutral long/short portfolios for each security, going long the asset and short a 1-3 Year U.S. Treasuries index when signals are positive and short the asset and long a 1-3 Year U.S. Treasury index when signals are negative.

Here we should pause and address the fact that the assets in this universe can have significant borrowing costs. For example, going short a high yield bond ETF such as HYG could cost you well in excess of 0.5% annualized in borrowing costs. Thus, this analysis may not be sufficient for investors actually considering explicitly shorting. However, this analysis may still be highly relevant for investors looking to construct long/flat portfolios, where short positions are implicitly achieved through the reduction of position size.

In an effort to perform more robust analysis, we employ a sub-sampling approach. For each asset, we calculate long/short portfolio returns and then randomly drop 25% of periods, using the remaining 75% to estimate the annualized return. We repeat this 1,000 times to generate a distribution of annualized returns. Our goal in taking this approach is to attempt to determine if the results are highly regime dependent or due to an outlier time-period. If either of these are the case, we would expect to see a wide dispersion in annualized returns.

For each asset, we plot the intra-quartile range of the data as well as whiskers that cover approximately 99% of the data.

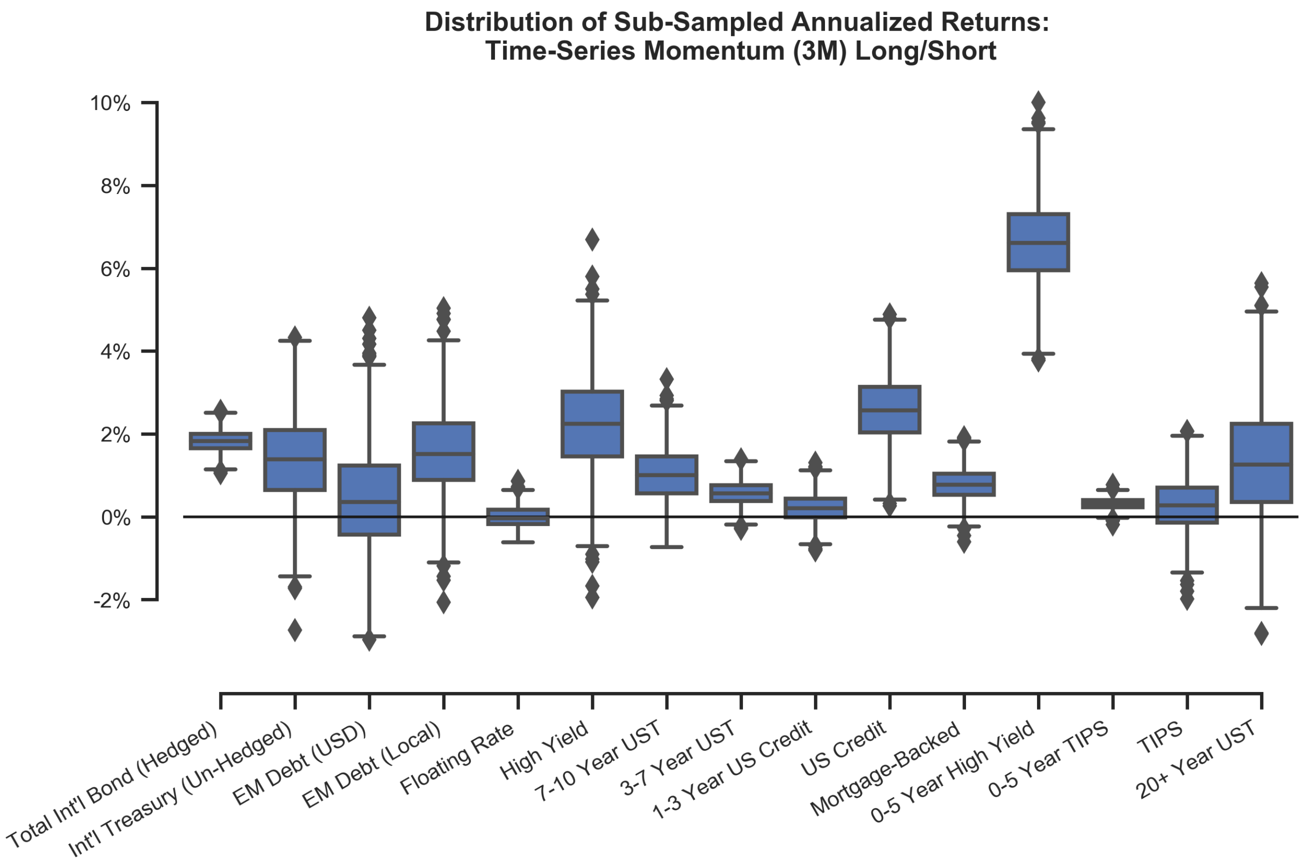
In the remainder of this rather brief commentary, we will review four separate sigals: momentum, carry, value, and reversal.

Momentum

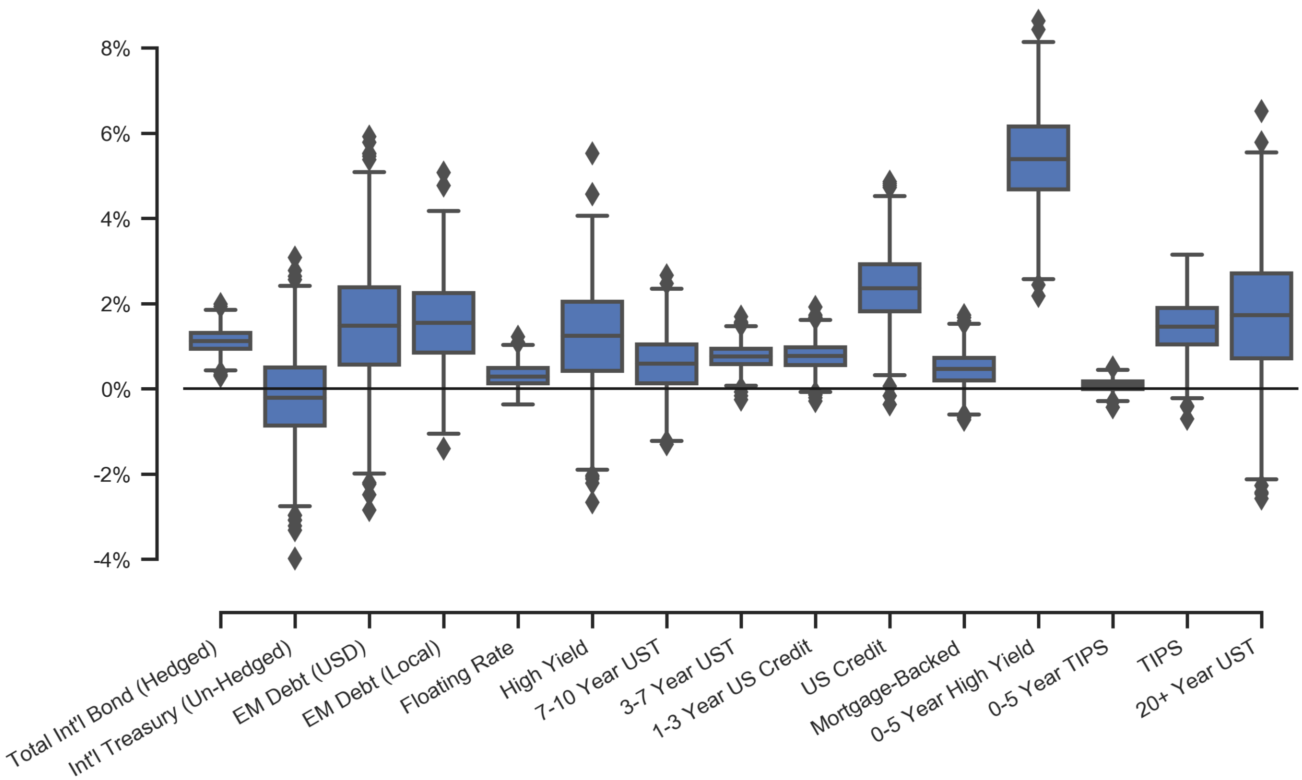
We generate momentum signals by computing 12-, 6- and 3- month prior total returns to reflect slow, intermediate, and fast momentum signals. Portfolios are long when prior returns are positive and short when prior returns are negative. The portfolios assume a 1-month holding period for momentum signals. To avoid timing luck, four sub-indexes are used, each rebalancing on a different week of the month.

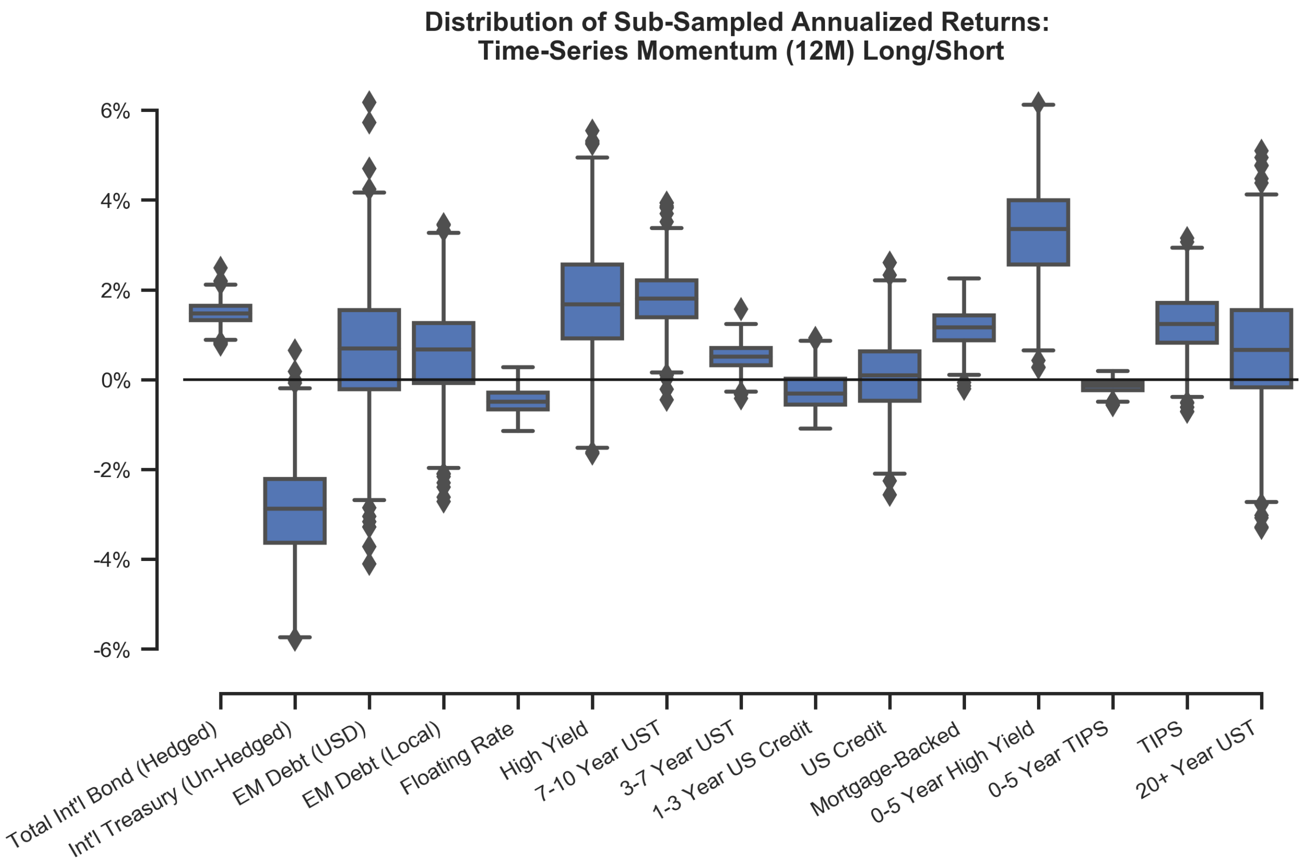
A few interesting data-points stand out:

- Results are highly inconsistent across 3-, 6-, and 12-month measures.
- Only three of the assets are significant for the 3- and 6-month momentum signals at the 1% level: hedged international bonds, U.S. credit, and short-term high yield.
- The significant dispersion in returns is reminiscent of the highly regime-driven results we identified when attempting to apply trend following to high yield in our commentary Tactical Credit.


Distribution of Sub-Sampled Annualized Returns:
Time-Series Momentum (3M) Long/Short
10%
8%
6%
4%
2%
0%
-2%
Total Int'l Bond (Hedged)
Int'l Treasury (Un-Hedged)
EM Debt (USD)
EM Debt (Local)
Floating Rate
High Yield
7-10 Year UST
3-7 Year UST
1-3 Year US Credit
US Credit
Mortgage-Backed
0-5 Year High Yield
0-5 Year TIPS
TIPS
20+ Year UST

Distribution of Sub-Sampled Annualized Returns:
Time-Series Momentum (6M) Long/Short





Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

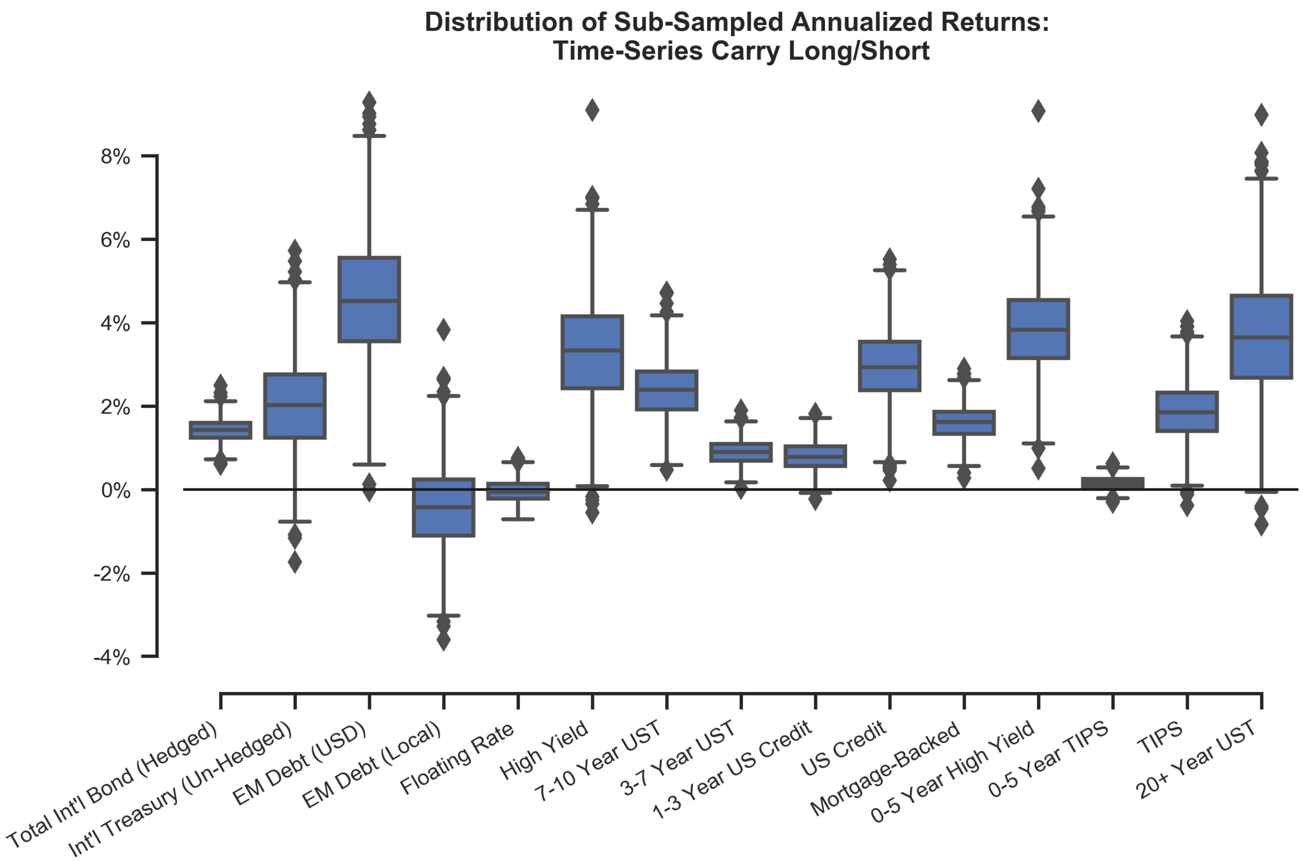
Carry

Carry is the expected excess return of an asset assuming price does not change. For our fixed income universe, we proxy carry using yield-to-worst minus the risk-free rate. For non-Treasury holdings, we adjust this figure for expected defaults and recovery.

Strategies are long when carry is positive and short when it is negative. The portfolios assume a 12-month holding period for carry signals. To avoid timing luck, 52 sub-indexes are used, each rebalancing on a different week of the year.

- Carry generates statistically positive results for 11 of the assets at a 5% level, and 9 of the assets at a 1% level.
- Two stand-outs are unhedged international government bonds and local currency emerging market debt, whose returns are highly influenced by currency fluctuations.

- It is worth acknowledging that carry was largely positive for most assets over the period, and this graph may largely represent just a buy-and-hold result.



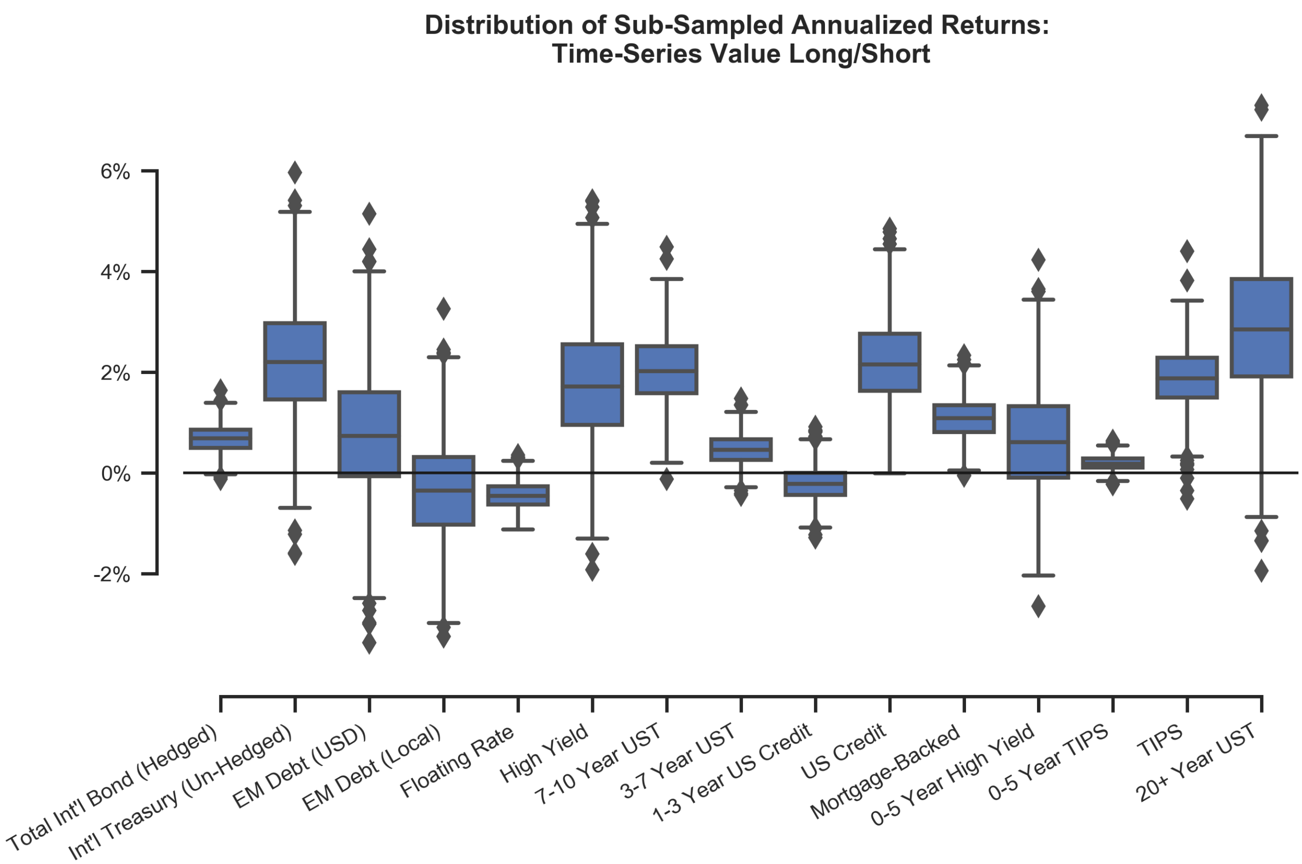
Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

Value

In past commentaries, we have used real yield as our value proxy in fixed income. In this commentary, we deviate from that methodology slightly and use a time-series z-score of carry as our value of measure. Historically high carry levels are considered to be cheap while historically low carry levels are considered to be expensive.

Strategies are long when z-scores are positive and short when they are negative. The portfolios assume a 12-month holding period for value signals. To avoid timing luck, 52 sub-indexes are used, each rebalancing on a different week of the year.

- 7-10 year U.S. Treasuries, mortgage-backed securities, and TIPS are all significant at a 1% level. Hedged international bonds and US credit are significant at a 5% level.



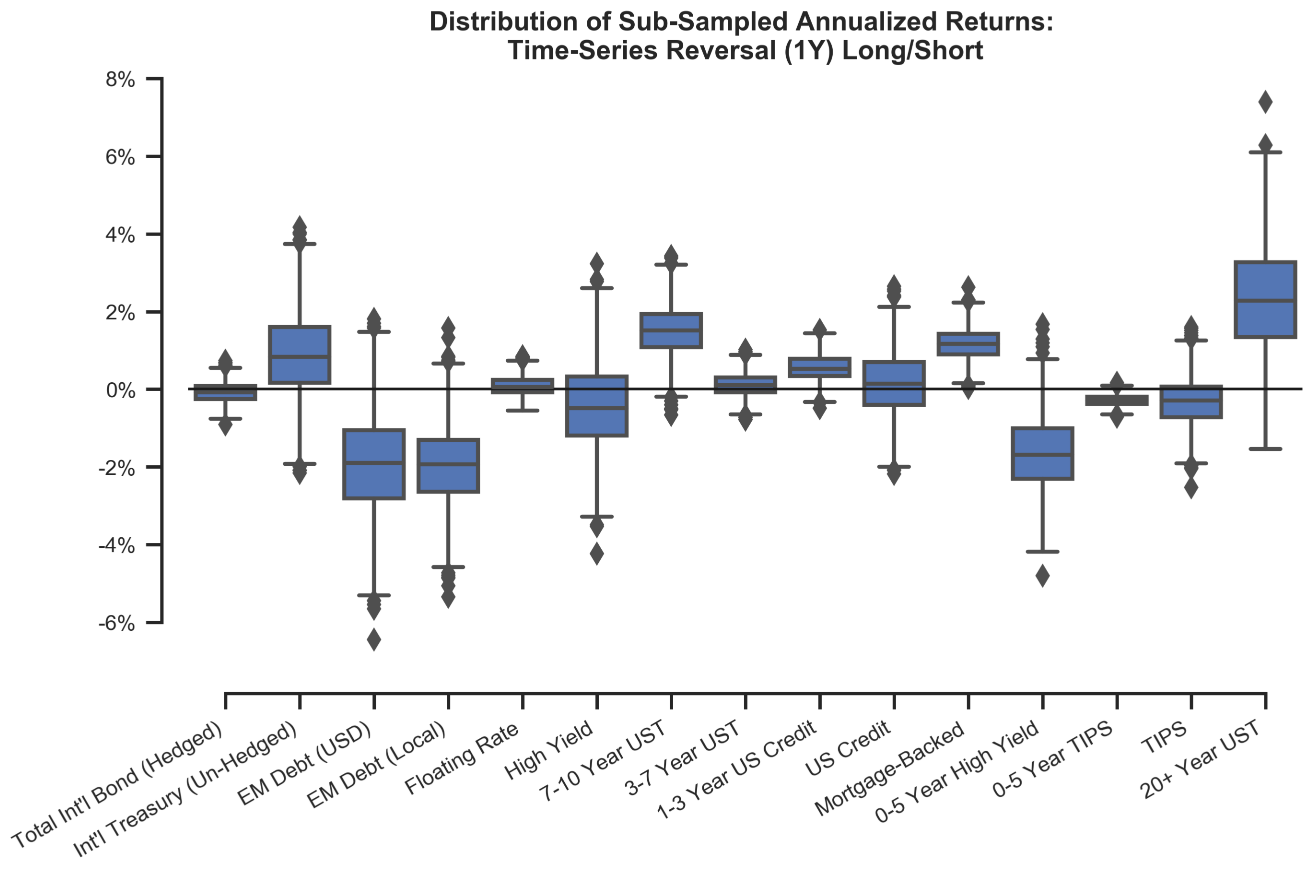
Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

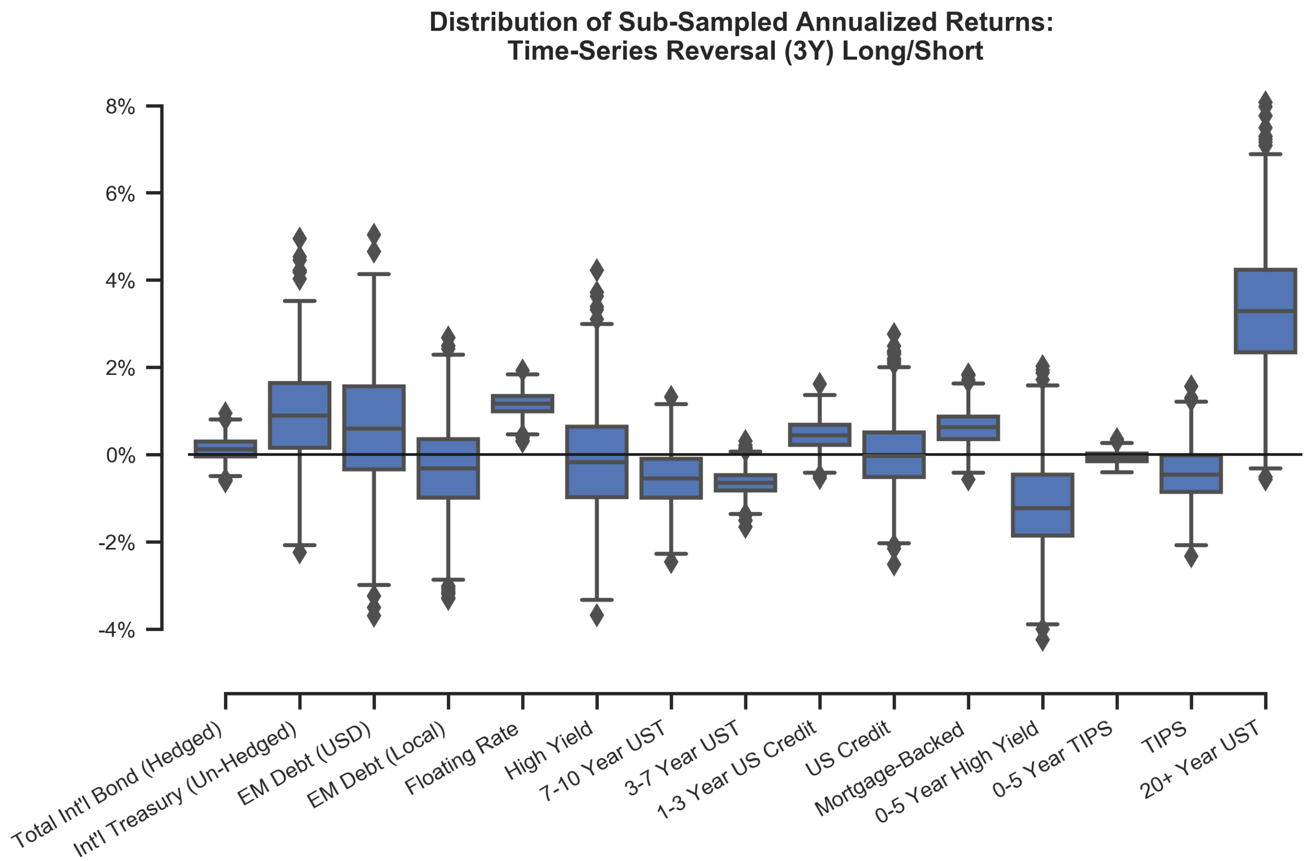
Reversal

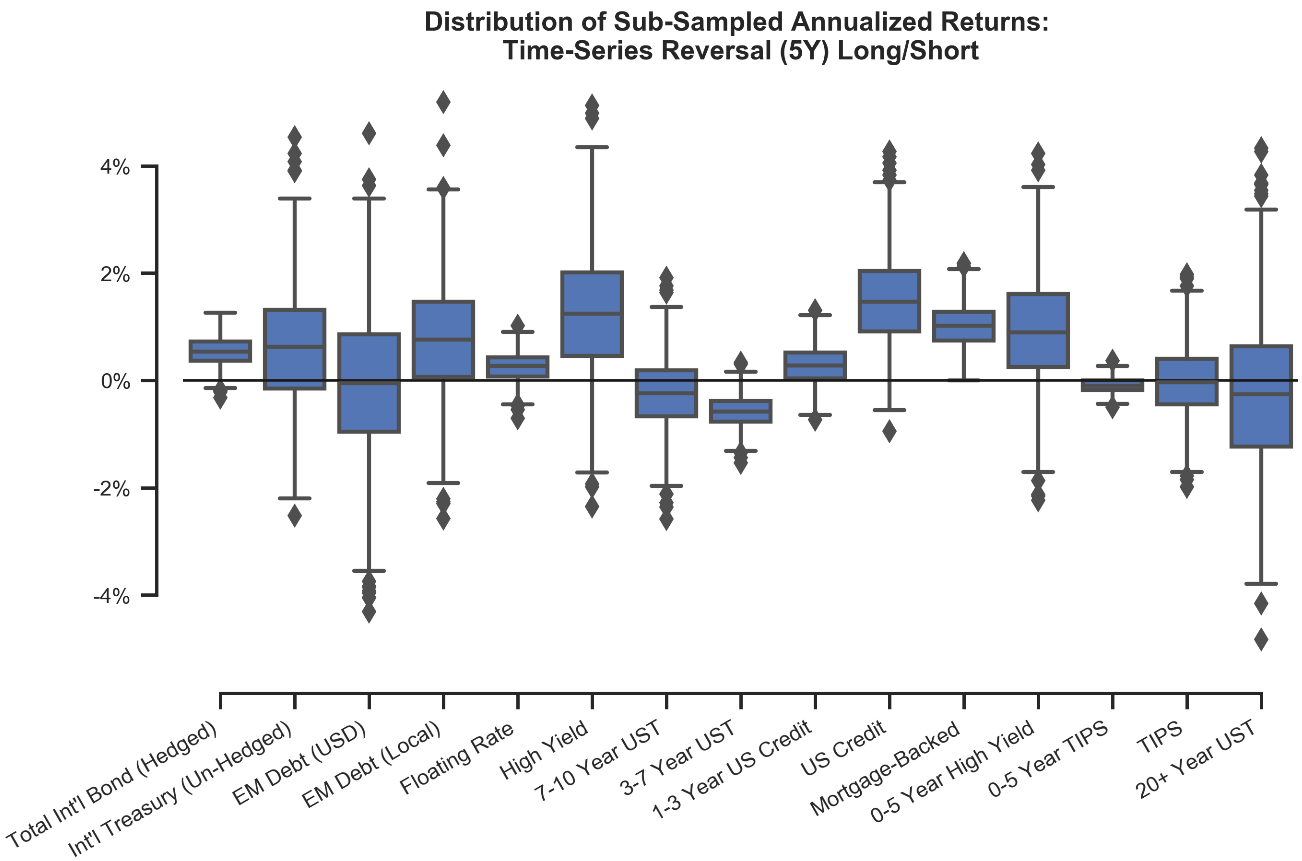
Reversal signals are the opposite of momentum: we expect past losers to outperform and past winners to underperform. Empirically, reversals tend to occur over very short time horizons (e.g. 1 month) and longer-term time horizons (e.g. 3- to 5-years). In many ways, long-term reversals can be thought of as a naive proxy for value, though there may be other behavioral and structural reasons for the historical efficacy of reversal signals.

As we did in our prior commentary, we employ a z-score on prior rolling returns. Strategies short when z-scores are positive (return dynamics above average) and go long when z-scores are negative (return dynamics below average). The portfolios assume a 12-month holding period for value signals. To avoid timing luck, 52 sub-indexes are used, each rebalancing on a different week of the year.

We can see that reversal signals are largely unsuccessful, telling us that simply buying something because it has done poorly in the past, or selling it because it has done well, has not been a successful approach.




Distribution of Sub-Sampled Annualized Returns:
Time-Series Reversal (3Y) Long/Short
8%
6%
4%
2%
0%
-2%
-4%
Total Int'l Bond (Hedged)
Int'l Treasury (Un-Hedged)
EM Debt (USD)
EM Debt (Local)
Floating Rate
High Yield
7-10 Year UST
3-7 Year UST
1-3 Year US Credit
US Credit
Mortgage-Backed
0-5 Year High Yield
0-5 Year TIPS
TIPS
20+ Year UST



Source: Bloomberg; Tiingo. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

Conclusion

In this study, we extend last week's commentary and use the same quantitative signals – momentum, carry, value, and reversals – in a time-series manner.

In almost all cases, we find little statistical significance.

Time-series momentum – i.e. trend following – is not robust across formation periods or asset classes. In fact, it only appears to work for a small subset of asset classes (hedged international bonds, broad credit, and short-term high yield) in short-term formation periods.

Carry signals appear more significant, but likely only due to the fact that they remained largely positive over the entire testing period, leading to largely buy-and-hold portfolios.

The design of our value signals partially addresses the carry signal's flaws through z-scoring. Unfortunately, we see – as with many other contexts – that value timing on its own rarely works.

Finally, we see that reversal signals offer almost no potential.

Taken together, we see little evidence herein supporting an *excess-return* driven motive for pursuing these signals in a time-series fashion. It is entirely possible, however, that these signals may manipulate return distributions in other meaningful ways (e.g. cut drawdown risk) and are still worth further exploration.

VALUE AND THE CREDIT CURVE

July 1, 2019

SUMMARY

- We continue our exploration of quantitative signals in fixed income.
- We use a measure of credit curve steepness as a valuation signal for timing exposure between corporate bonds and U.S. Treasuries.
- The value signal generates a 0.84% annualized return from 1950 to 2019 but is highly regime dependent with meaningful drawdowns.
- Introducing a naïve momentum strategy significantly improves the realized Sharpe ratio and drawdown profile, but does not reduce the regime-based nature of the returns.
- With a combined return of just 1.0% annualized, this strategy may not prove effective after appropriate discounting for hindsight bias, costs, and manager fees. The signal itself, however, may be useful in other contexts.

In the last several weeks, we have been exploring the application of quantitative signals to fixed income.

- In *Tactical Credit* we explored trend-following strategies with high yield bonds.
- In *Quantitative Styles and Multi-Sector Bonds* we built off a prior piece (*Navigating Municipal Bonds with Factors)* and explored the cross-sectional application of momentum, value, carry, reversal, and volatility signals in a broad fixed income universe.
- In *Time Series Signals and Multi-Sector Bonds* we explored the same momentum, value, carry, and reversal signals as market timing signals.

Recent cross-sectional studies also build off of further research we've done in the past on applying trend, value, carry, and explicit measures of the bond risk premium as duration timing mechanisms (see *Duration Timing with Style Premia; Timing Bonds with Value, Momentum, and Carry; and A Carry-Trend-Hedge Approach to Duration Timing*).

Broadly, our studies have found:

- Value (measured as deviation from real yield), momentum (prior 12-month returns), and carry (yield-to-worst) were all profitable factors in cross-section municipal bond sector long/short portfolios.

- Value (measured as deviation from real yield), trend (measured as prior return), and carry (measured as term spread + roll yield) have historically been effective timing signals for U.S. duration exposure.

- Prior short-term equity returns proved to be an effective signal for near-term returns in U.S. Treasuries (related to the "flight-to-safety premium").

- Short-term trend proved effective for high yield bond timing, but the results were vastly determined by performance in 2000-2003 and 2008-2009. While the strategy appeared to still be able to harvest relative carry between high-yield bonds and core fixed income in other environments, a significant proportion of returns came from avoiding large drawdowns in high yield.

- Short-term cross-section momentum (prior total returns), value (z-score of loss-adjusted yield-to-worst), carry (loss-adjusted yield-to-worst), and 3-year reversals all appeared to offer robust signals for relative selection in fixed income sectors. The time period covered in the study, however, was limited and mostly within a low-inflation regime.

- Application of momentum, value, carry, and reversal as timing signals proved largely ineffective for generating excess returns.

In this week's commentary, we want to further contribute to research by introducing a value timing signal for credit.

Finding Value in Credit

Identifying a value signal requires some measure or proxy of an asset's "fair" value. What can make identifying value in credit so difficult is that there are a number of moving pieces.

Conceptually, credit spreads should be proportional to default rates, recovery rates, and aggregate risk appetite, making determining whether spreads are cheap or expensive rather complicated. Prior literature typically tackles the problem with one of three major categories of models:

- *Econometric*: "Fair value" of credit spreads is modeled through a regression that typically explicitly accounts for default and recovery rates. Inputs are often related to economic and market variables, such as equity market returns, 10-year minus 2-year spreads, corporate leverage, and corporate profitability. Bottom-up analysis may use metrics such as credit quality, maturity, supply, and liquidity.

- *Merton Model*: Based upon the idea the bond holders have sold a put on a company's asset value. Therefore, options pricing models can be used to calculate a credit spread. Inputs include the total asset value, asset volatility, and leverage of the firm under analysis.

- *Spread Signal*: A simple statistical model derived from credit spread themselves. For example, a rolling z-score of option-adjusted spreads or deviations from real yield. Other models (e.g. Haghani and Dewey (2016)) have used spread plus real yield versus a long-run constant (e.g. "150 basis points").

The first method requires a significant amount of economic modeling. The second approach requires a significant amount of extrapolation from market data. The third method, while computationally (and intellectually) less intensive, requires a meaningful historical sample that realistically needs to cover at least one full market cycle.

While attractive for its simplicity, there are a number of factors that complicate the third approach.

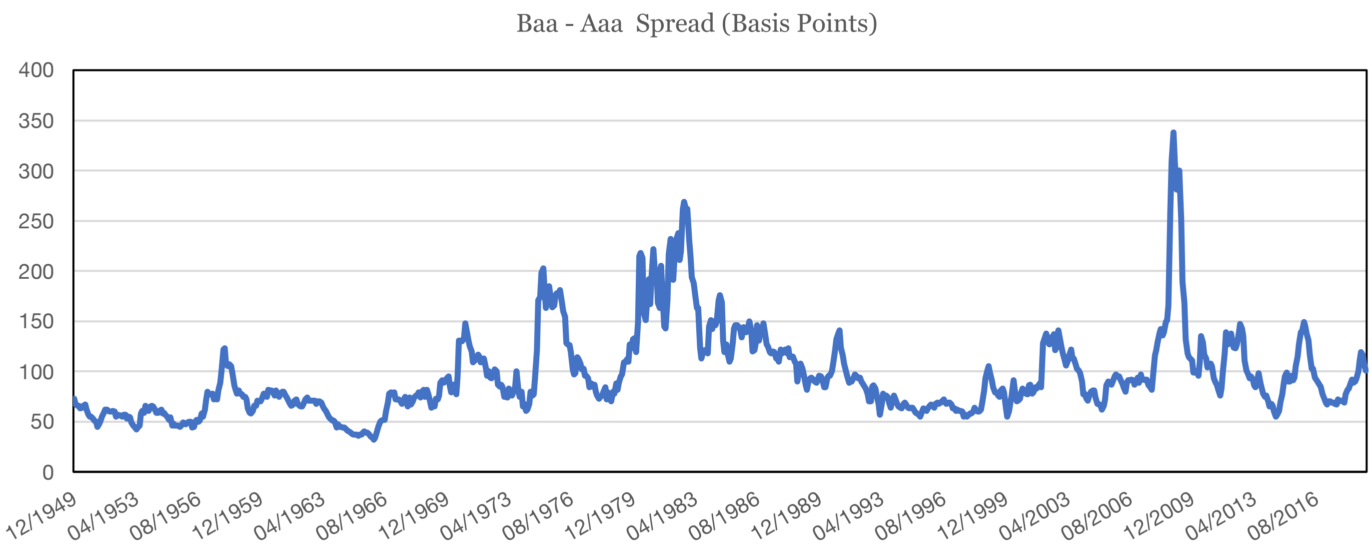
First, if spreads are measured against U.S. Treasuries, the metric may be polluted by information related to Treasuries due to their idiosyncratic behavior (e.g. scarcity effects and flight-to-safety premiums). Structural shifts in default rates, recovery rates, and risk appetites may also cause a problem, as spreads may appear unduly thin or wide compared to past regimes.

In light of this, in this piece we will explore a similarly simple-to-calculate spread signal, but one that hopefully addresses some of these short-comings.

Baa vs. Aaa Yields

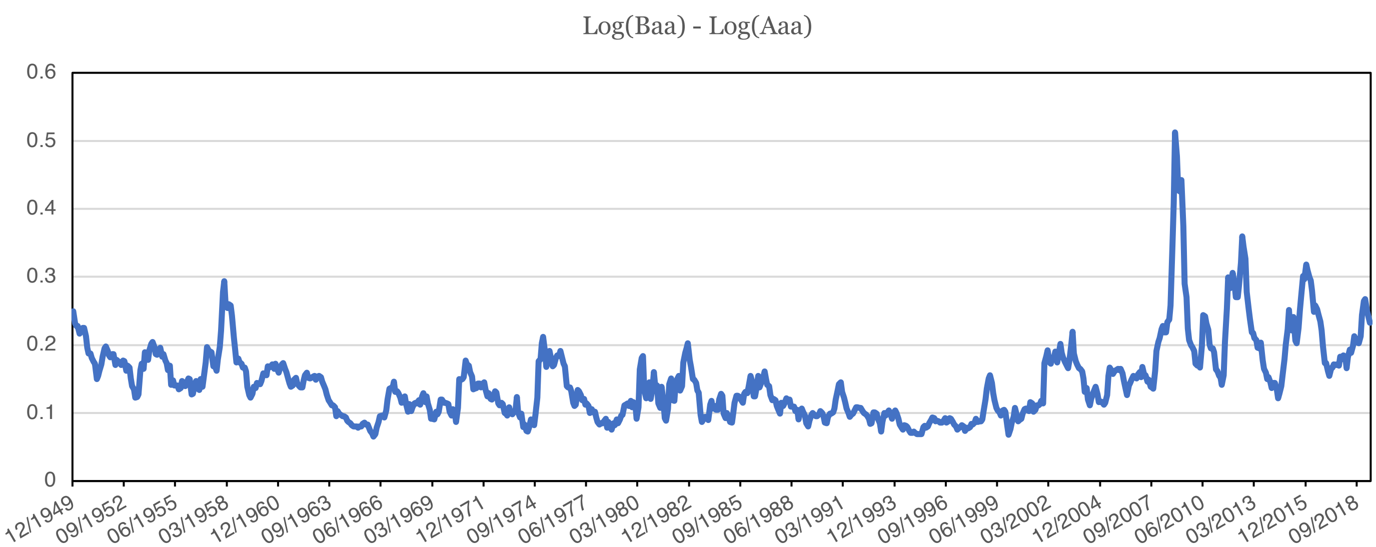
In order to adjust for these problems, we propose looking at the steepness of the credit curve itself by comparing prime / high-grade yield versus lower-medium grade yields. For example, we could compare Moody's Season Aaa Corporate Bond Yield and Moody's Season Baa Corporate Bond Yield. In fact, we will use these yields for the remainder of this study.

We may be initially inclined to measure the steepness of the credit curve by taking the difference in yield spreads, which we plot below.



Source: Federal Reserve of St. Louis. Calculations by Newfound Research.

We can find a stronger mean-reverting signal, however, if we calculate the log-difference in yields.



Source: Federal Reserve of St. Louis. Calculations by Newfound Research.

We believe this transformation is appropriate for two reasons. First, the log transformation helps control for the highly heteroskedastic and skewed nature of credit spreads.

Second, it helps capture both the steepness *and* the level of the credit curve simultaneously. For example, a 50-basis-point premium when Aaa yield is 1,000 basis points is very different than when Aaa yield is 100 basis points. In the former case, investors may not feel any pressure to bear excess risk to achieve their return objectives, and therefore a 50-basis-point spread may be quite thin. In the latter case, 50 basis points may represent a significant step-up in relative return level in an environment where investors have either low default expectations, high recovery expectations, high risk appetite, or some combination thereof.

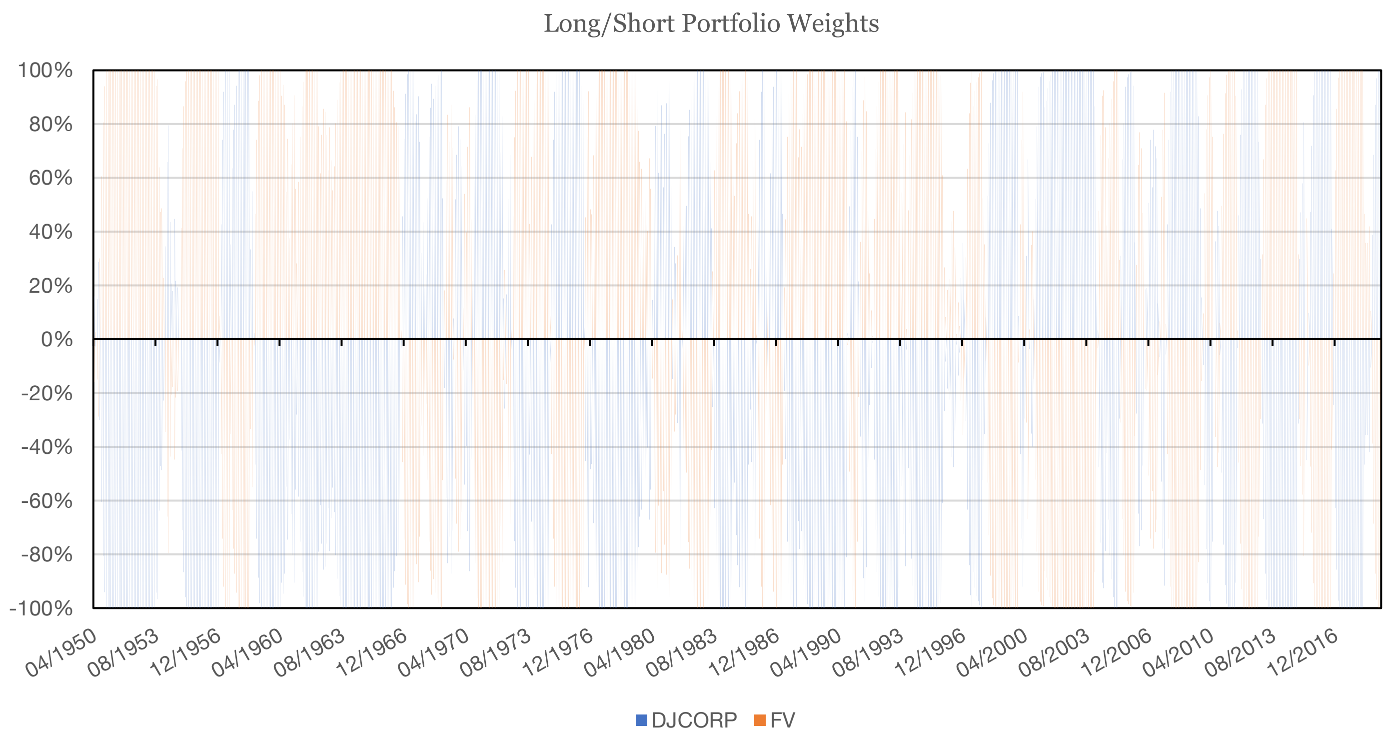
Another way of interpreting our signal is that it informs us about the relative decisions investors must make about their expected dispersion in terminal wealth.

Constructing the Value Strategy

With our signal in hand, we can now attempt to time credit exposure. When our measure signals that the credit curve is historically steep, we will take credit risk. When our signal indicates that the curve is historically flat we will avoid it.

Specifically, we will construct a dollar-neutral long/short portfolio using the Dow Jones Corporate Bond Index ("DJCORP") and a constant maturity 5-year U.S. Treasury index ("FV"). We will calculate a rolling z-score of our steepness measure and go long DJCORP and short FV when the z-score is positive and place the opposite trade when the z-score is negative.

In line with prior studies, we will apply an ensemble approach. Portfolios are reformed monthly using formation ranging from 3-to-6 years with holding periods ranging from 1-to-6 months. Portfolio weights for the resulting strategy are plotted below.



Source: Federal Reserve of St. Louis and Global Financial Data. Calculations by Newfound Research.

We should address the fact that while both corporate bond yield and index data is available back to the 1930s, we have truncated our study to ignore dates prior to 12/1949 to normalize for a post-war period. It should be further acknowledged that the Dow Jones Corporate Bond index used in this study did not technically exist until 2002. Prior to that date, the index return tracks a Dow Jones Bond Aggregate, which was based upon four sub-indices: high-grade rails, second-grade rails, public utilities, and industries. This average existed from 1915 to 1976, when it was replaced with a new average at that point when the number of railway bonds was no longer sufficient to maintain the average.

Below we plot the returns of our long/short strategy.

Source: Federal Reserve of St. Louis and Global Financial Data. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

The strategy has an annualized return of 0.84% with a volatility of 3.89%, generating a Sharpe ratio of 0.22. Of course, long-term return statistics belie investor and manager experience, with this strategy exhibiting at least two periods of decade-plus-long drawdowns. In fact, the strategy really has just four major return regimes: 1950 to 1970 (-0.24% annualized), 1970 to 1987 (2.59% annualized), 1987 to 2002 (-0.33%), and 2002 to 2019 (1.49% annualized).

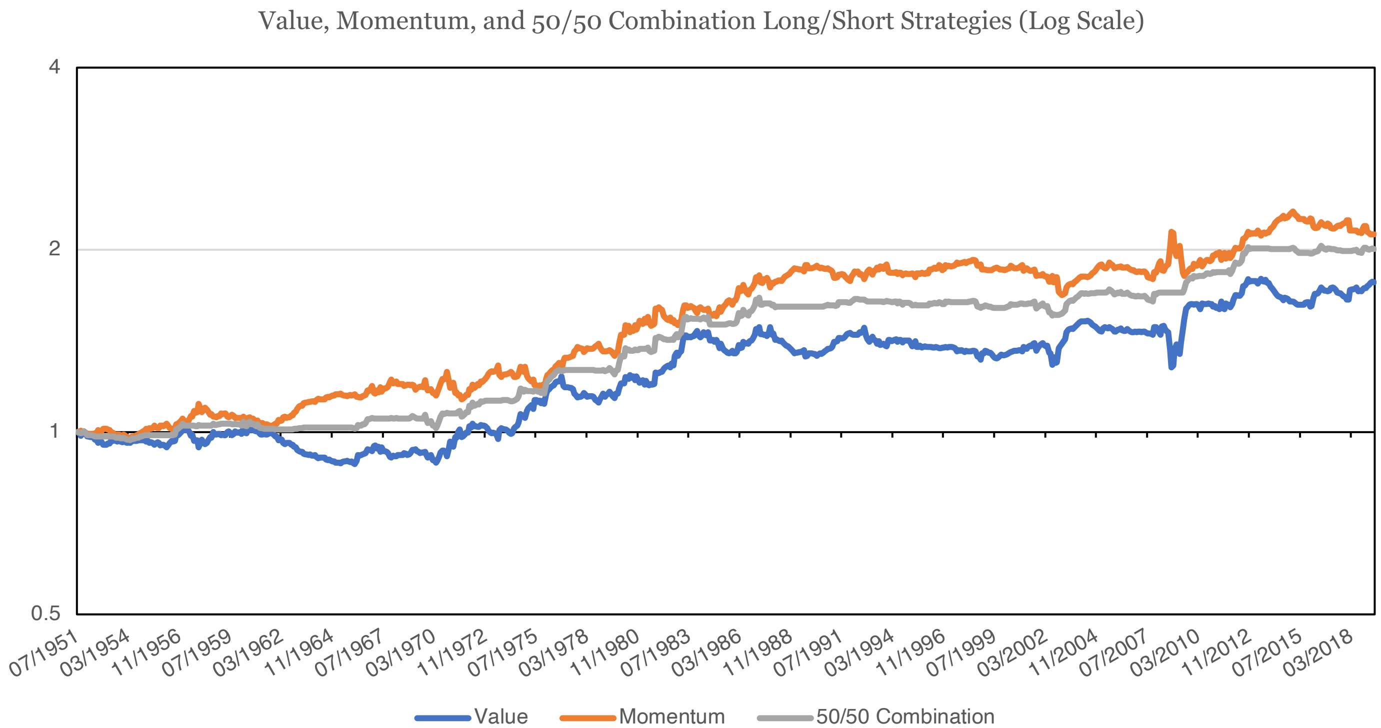
Try the strategy out in the wrong environment and we might be in for a lot of pain.

Momentum to the Rescue?

It is no secret that value and momentum go together like peanut butter and jelly. Instead of tweaking our strategy to death in order to improve it, we may just find opportunity in combining it with a negatively correlated signal.

Using an ensemble model, we construct a dollar-neutral long/short momentum strategy that compares prior total returns of DJCORP and FV. Rebalanced monthly, the portfolios use formation periods ranging from 9-to-15 months and holding periods ranging from 1-to-6 months.

Below we plot the growth of $1 in our value strategy, our momentum strategy, and a 50/50 combination of the two strategies that is rebalanced monthly.



Source: Federal Reserve of St. Louis and Global Financial Data. Calculations by Newfound Research. Returns are hypothetical and backtested. Returns are gross of all management fees, transaction fees, and taxes, but net of underlying fund fees. Total return series assumes the reinvestment of all distributions.

The first thing we note is – even without calculating any statistics – the meaningful negative correlation we see in the equity curves of the value and momentum strategies. This should give us confidence that there is the potential for significant improvement through diversification.

The momentum strategy returns 1.11% annualized with a volatility of 3.92%, generating a Sharpe ratio of 0.29. The 50/50 combination strategy, however, returns 1.03% annualized with a volatility of just 2.16% annualized, resulting in a Sharpe ratio of 0.48.

While we still see significant regime-driven behavior, the negative regimes now come at a far lower cost.

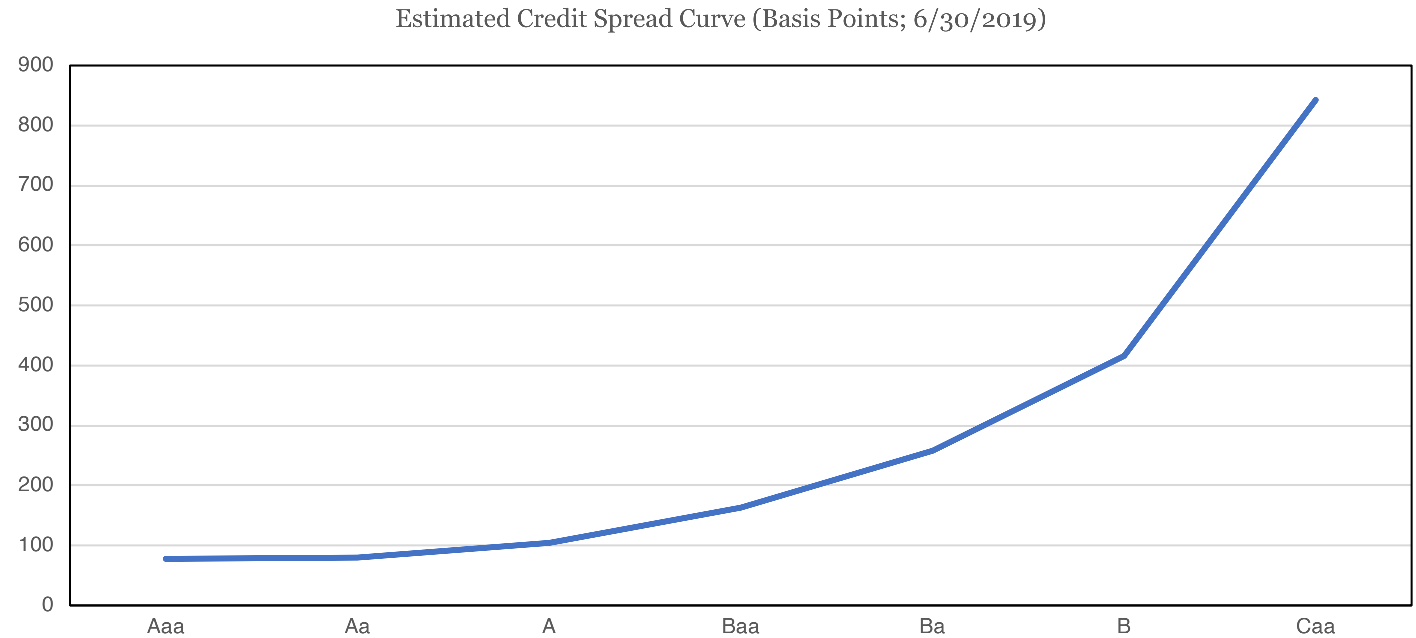
DECOMPOSING THE CREDIT CURVE

July 8, 2019

SUMMARY

- In this research note, we continue our exploration of credit.
- Rather than test a quantitative signal, we explore credit changes through the lens of statistical decomposition.
- As with the Treasury yield curve, we find that changes in the credit spread curve can be largely explained by Level, Slope, and Curvature (so long as we adjust for relative volatility levels).
- We construct stylized portfolios to reflect these factors, adjusting position weights such that they contribute an equal amount of credit risk. We then neutralize interest rate exposure such that the return of these portfolios represents credit-specific information.
- We find that the Level trade suggests little-to-no realized credit premium over the last 25 years, and Slope suggests no realized premium of junk-minus-quality within credit either. However, results may be largely affected by idiosyncratic events (e.g. LTCM in 1998) or unhedged risks (e.g. sector differences in credit indices).

In this week's research note, we continue our exploration of credit with a statistical decomposition of the credit spread curve. Just as the U.S. Treasury yield curve plots yields versus maturity, the credit spread curve plots *excess* yield versus credit quality, providing us insight into how much extra return we demand for the risks of declining credit quality.



Source: Federal Reserve of St. Louis; Bloomberg. Calculations by Newfound Research.

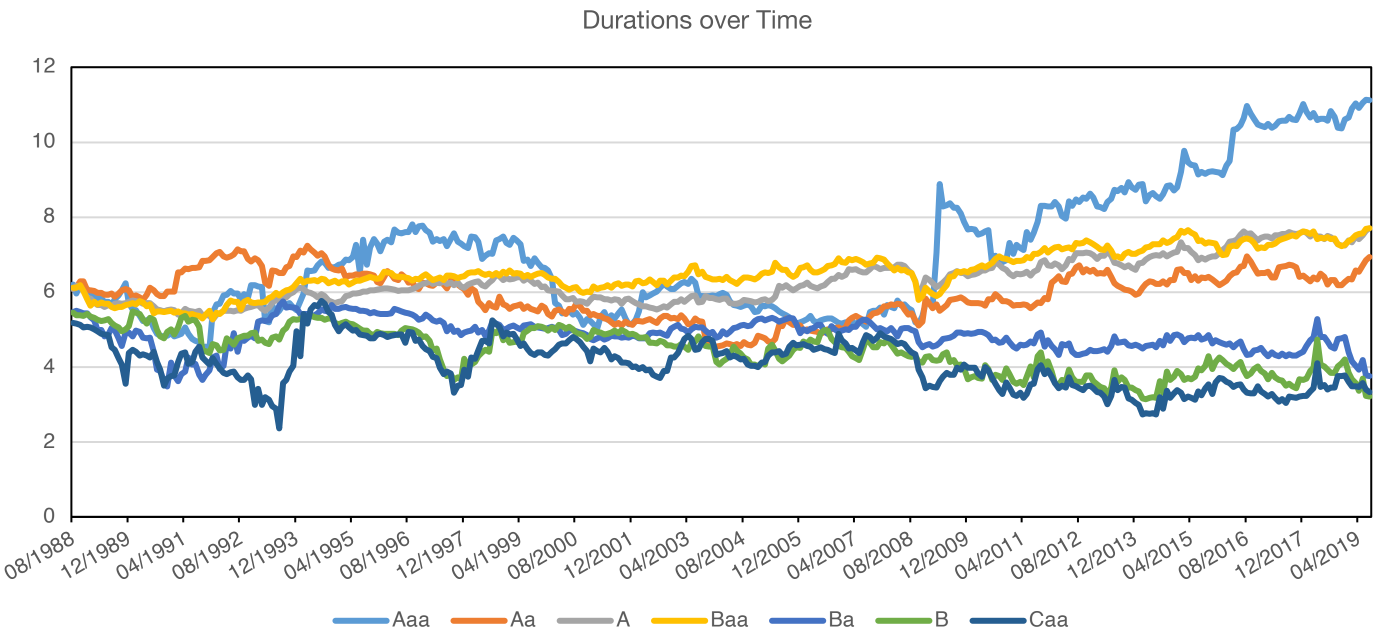
Our goal in analyzing the credit spread curve is to gain a deeper understanding of the principal drivers behind its changes. In doing so, we hope to potentially gain intuition and ideas for trading signals between low- and high-quality credit.

To begin our, we must first construct our credit spread curve. We will use the following index data to represent our different credit qualities.

- **Aaa:** Bloomberg U.S. Corporate Aaa Index (LCA3TRUU)
- **Aa:** Bloomberg U.S. Corporate Aa Index (LCA2TRUU)
- **A:** Bloomberg U.S. Corporate A Index (LCA1TRUU)
- **Baa:** Bloomberg U.S. Corporate Baa Index (LCB1TRUU)
- **Ba:** Bloomberg U.S. Corporate HY Ba Index (BCBATRUU)
- **B:** Bloomberg U.S. Corporate HY B Index (BCBHTRUU)
- **Caa:** Bloomberg U.S. Corporate HY Caa Index (BCAUTRUU)

Unfortunately, we cannot simply plot the yield-to-worst for each index, as spread captures the *excess* yield. Which raises the question: *excess to what*? As we want to isolate the credit component of the yield, we need to remove the duration-equivalent Treasury rate.

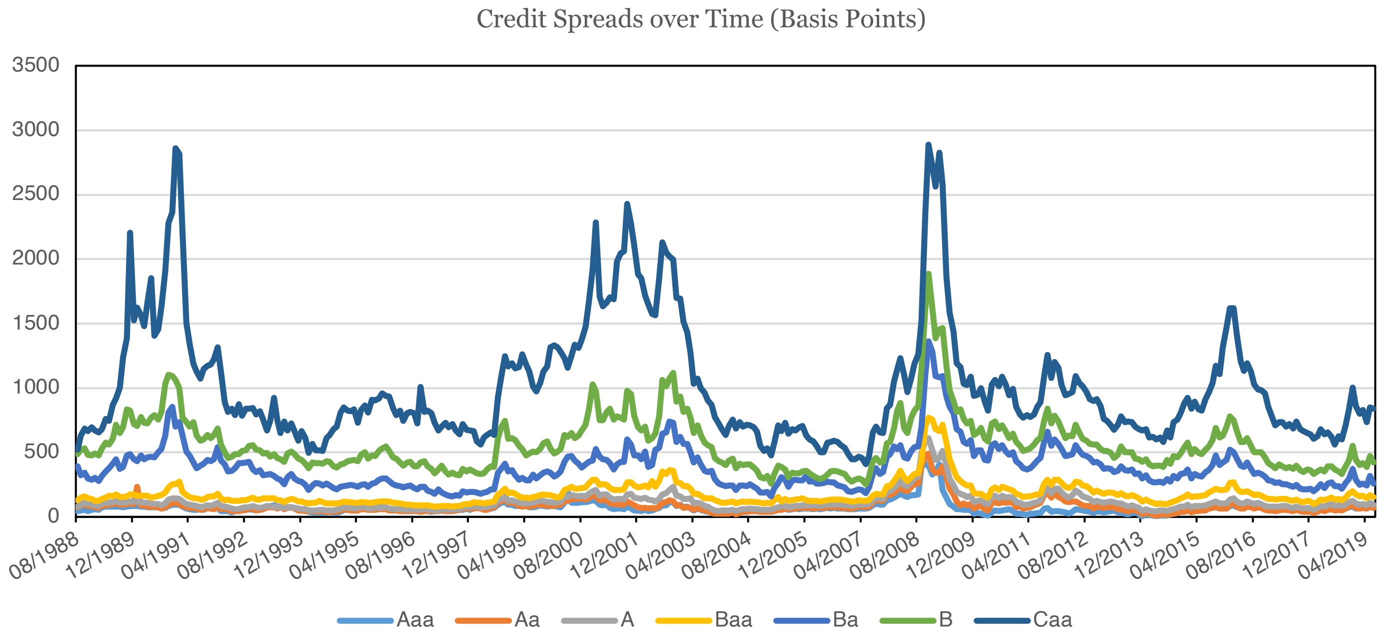
Plotting the duration of each credit index over time, we can immediately see why incorporating this duration data will be important. Not only do durations vary meaningfully over time (e.g. Aaa durations varying between 4.95 and 11.13), but they also deviate across quality (e.g. Caa durations currently sit near 3.3 while Aaa durations are north of 11.1).



Source: Bloomberg.

To calculate our credit spread curve, we must first calculate the duration-equivalent Treasury bond yield for each index at each point in time. For each credit index at each point in time, we use the historical Treasury yield curve to numerically solve for the Treausry maturity that matches the credit index's duration. We then subtract that matching rate from the credit index's reported yield-to-worst to estimate the credit spread.

We plot the spreads over time below.



Source: Federal Reserve of St. Louis; Bloomberg. Calculations by Newfound Research.

Statistical Decomposition: Eigen Portfolios

With our credit spreads in hand, we can now attempt to extract the statistical drivers of change within the curve. One method of achieving this is to:

- Calculate month-to-month differences in the curve.
- Calculate the correlation matrix of the differences.
- Calculate an eigenvalue decomposition of the correlation matrix.

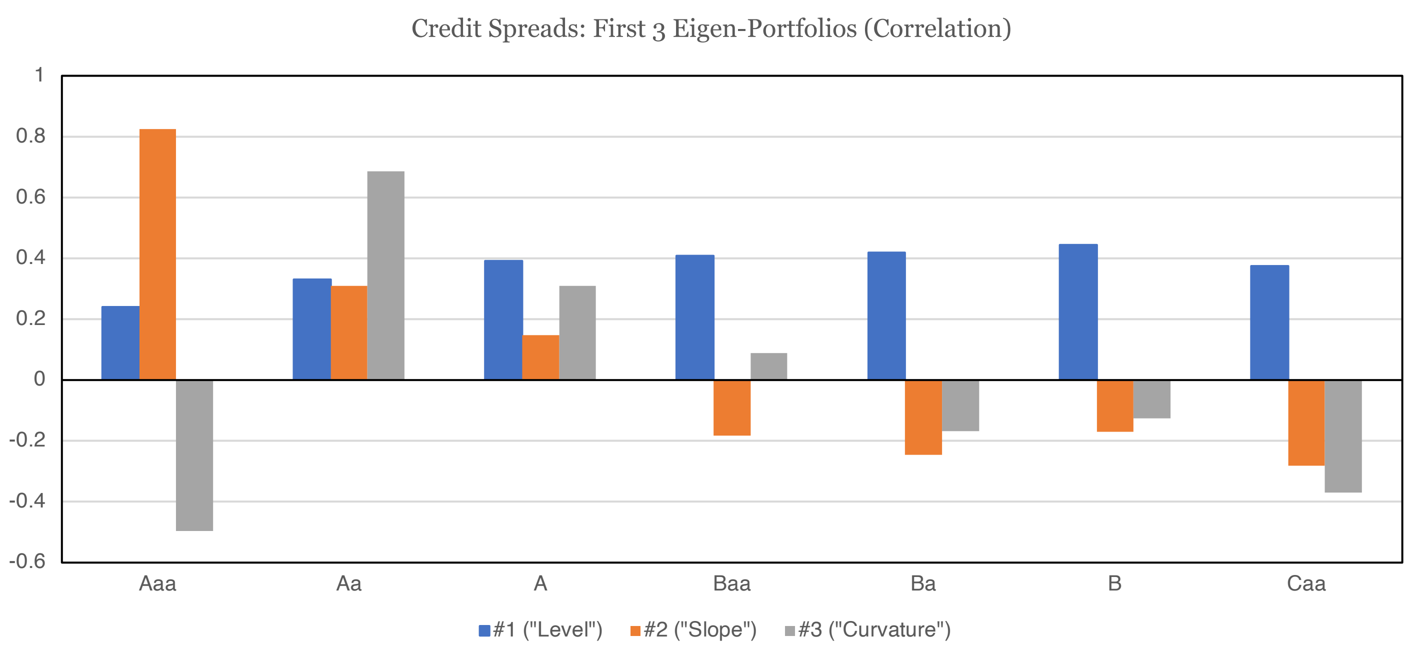
Stopping after just the first two steps, we can begin to see some interesting visual patterns emerge in the correlation matrix.

- There is not a monotonic decline in correlation between credit qualities. For example, Aaa is not more highly correlated to Aa than Ba and A is more correlated to B than it is Aa.
- Aaa appears to behave rather uniquely.
- Baa, Ba, B, and to a lesser extent Caa, appear to visually cluster in behavior.
- Ba, B, and Caa do appear to have more intuitive correlation behavior, with correlations increasing as credit qualities get closer.

	Aaa	Aa	A	Baa	Ba	B	Caa
Aaa	1.00	0.32	0.37	0.28	0.30	0.35	0.28
Aa	0.32	1.00	0.55	0.48	0.43	0.50	0.37
A	0.37	0.55	1.00	0.61	0.58	0.64	0.46
Baa	0.28	0.48	0.61	1.00	0.67	0.70	0.57
Ba	0.30	0.43	0.58	0.67	1.00	0.80	0.63
B	0.35	0.50	0.64	0.70	0.80	1.00	0.69
Caa	0.28	0.37	0.46	0.57	0.63	0.69	1.00

Step 3 might seem foreign for those unfamiliar with the technique, but in this context eigenvalue decomposition has an easy interpretation. The process will take our universe of credit indices and return a universe of statistically independent factor portfolios, where each portfolio is made up of a combination of credit indices.

As our eigenvalue decomposition was applied to the correlation matrix of credit spread changes, the factors will explain the principal vectors of variance in credit spread changes. We plot the weights of the first three factors below.



Source: Federal Reserve of St. Louis; Bloomberg. Calculations by Newfound Research.

For anyone who has performed an eigenvalue decomposition on the yield curve before, three familiar components emerge.

We can see that Factor #1 applies nearly equal-weights across all the credit indices. Therefore, we label this factor "level" as it represents a level shift across the entire curve.

Factor #2 declines in weight from Aaa through Caa. Therefore, we label this factor "slope," as it controls steepening and flattening of the credit curve.

Factor #3 appears as a barbell: negative weights in the wings and positive weights in the belly. Therefore, we call this factor "curvature," as it will capture convexity changes in the curve.

Together, these three factors explain 80% of the variance in credit spread changes. Interestingly, the 4th factor – which brings variance explained up to 87.5% – also looks very much like a curvature trade, but places zero weight on Aaa and barbells Aa/Caa against A/Baa. We believe this serves as further evidence as to the unique behavior of Aaa credit.

Tracking Credit Eigen Portfolios

As we mentioned, each factor is constructed as a combination of exposure to our Aaa-Caa credit universe; in other words, they are portfolios! This means we can track their performance over time and see how these different trades behave in different market regimes.

To avoid overfitting and estimation risk, we decided to simplify the factor portfolios into more stylized trades, whose weights are plotted below (though ignore, for a moment, the actual weights, as they are meant only to represent relative weighting within the portfolio and not absolute level). Note that the Level trade has a cumulative positive weight while the Slope and Curvature trades sum to zero.

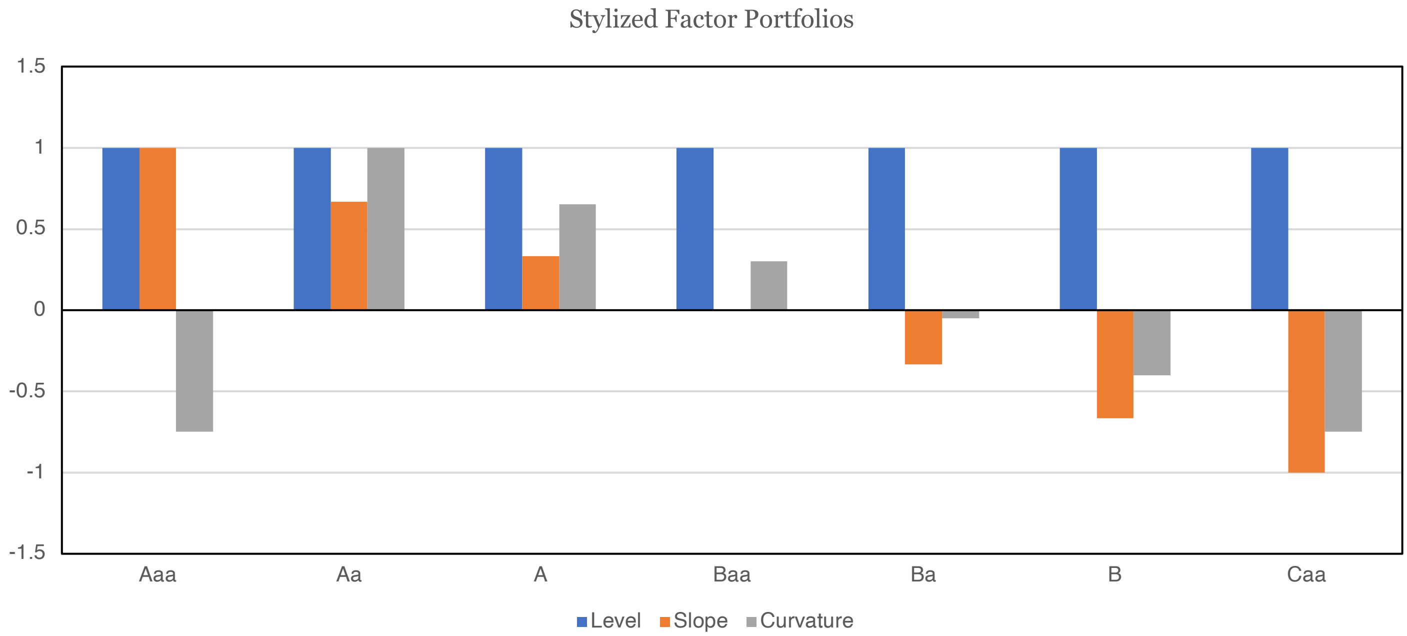
To actually implement these trades, we need to account for the fact that each credit index will have a different level of *credit duration*.

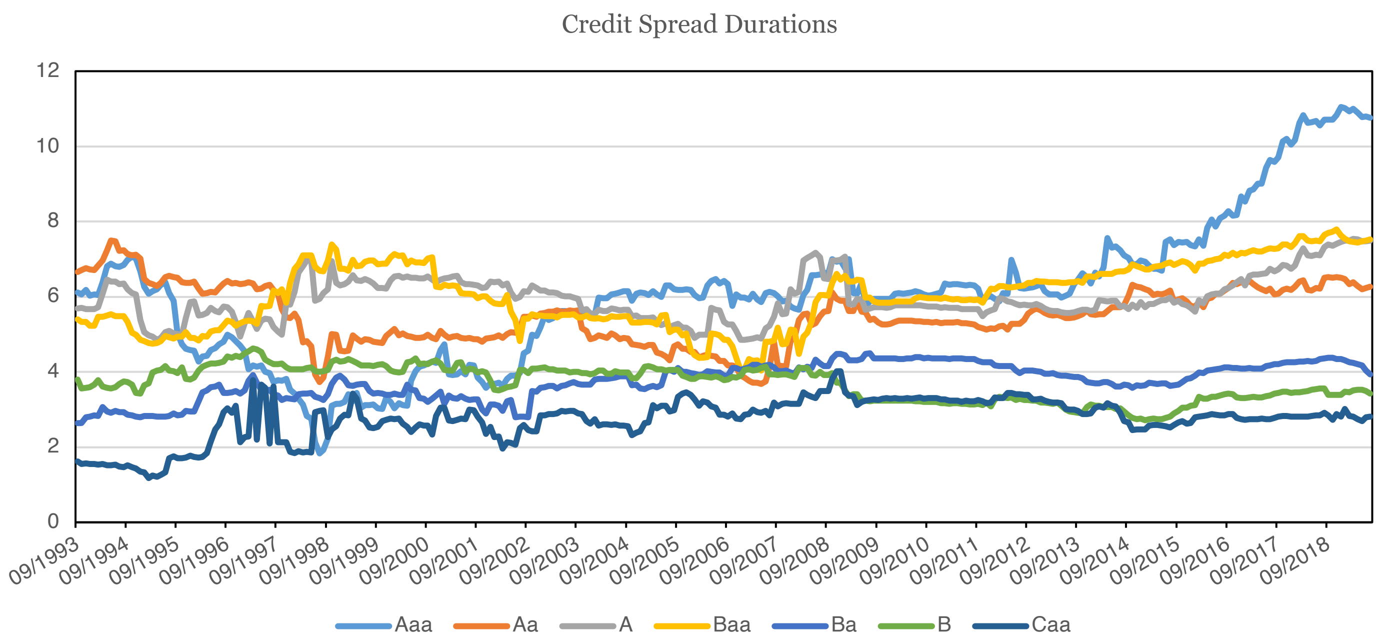
Akin to duration, which measure's a bond's sensitivity to interest rate changes, credit duration measures a bond's sensitivity to changes in its credit spread. As with Treasuries, we need to adjust the weights of our trades to account for this difference in credit durations across our indices.

For example, if we want to place a trade that profits in a steepening of the Treasury yield curve, we might sell 10-year US Treasuries and buy 2-year US Treasuries. However, we would not buy and sell the same notional amount, as that would leave us with a significantly negative duration position. Rather, we would scale each leg such that their durations offset. In the end, this causes us to buy significantly more 2s than we sell 10s.

To continue, therefore, we must calculate credit spread durations.

Without this data on hand, we employ a statistical approach. Specifically, we take monthly total return data and subtract yield return and impact from interest rate changes (employing the duration-matched rates we calculated above). What is left over is an estimate of return due to changes in credit spreads. We then regress these returns against changes in credit spreads to calculate credit spread durations, which we plot below.





Source: Federal Reserve of St. Louis; Bloomberg. Calculations by Newfound Research.

The results are a bit of a head scratcher. Unlike duration in the credit curve which typically increases monotonically across maturities, we get a very different effect here. Aaa credit spread duration is 10.7 today while Caa credit spread duration is 2.8. How is that possible? Why is lower-quality credit not *more sensitive* to credit changes than higher quality credit?

Here we run into a very interesting empirical result in credit spreads: spread change is proportional to spread level. Thus, a true "level shift" rarely occurs in the credit space; e.g. a 1bp change in the front-end of the credit spread curve may actually manifest as a 10bp change in the back end. Therefore, the lower credit spread duration of the back end of the curve is offset by larger changes.

There is some common-sense intuition to this effect. Credit has a highly non-linear return component: defaults. If we enter an economic environment where we expect an increase in default rates, it tends to happen in a non-linear fashion across the curve. To offset the larger increase in defaults in lower quality credit, investors will demand larger corresponding credit spreads.

(Side note: this is why we saw that the Baa–Aaa spread did not appear to mean-revert as cleanly as the log-difference of spreads did in last week's commentary, *Value and the Credit Spread*.)

While our credit spread durations may be correct, we still face a problem: weighting such that each index contributes equal credit spread duration will create an outsized weight to the Caa index.

DTS Scaling

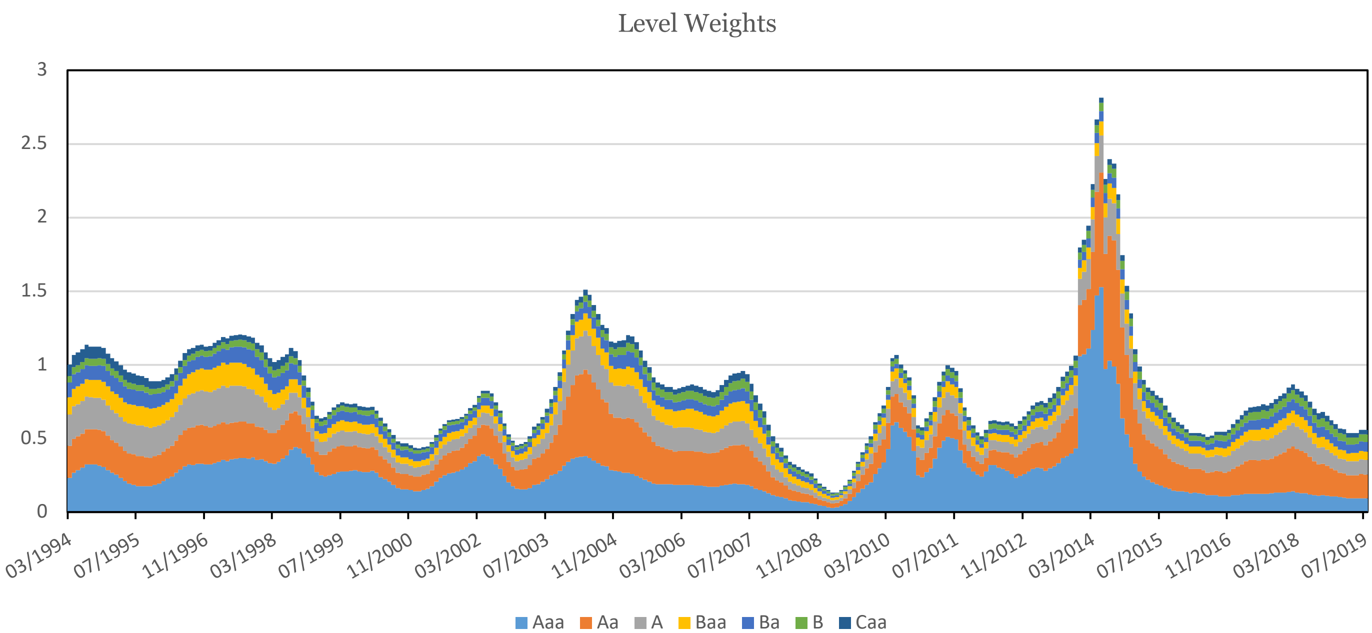
Fortunately, some very smart folks thought about this problem many years ago. Recognizing the stability of relative spread changes, Dor, Dynkin, Hyman, Houweling, van Leeuwen, and Penninga (2007) recommend the measure of duration times spread ("DTS") for credit risk.

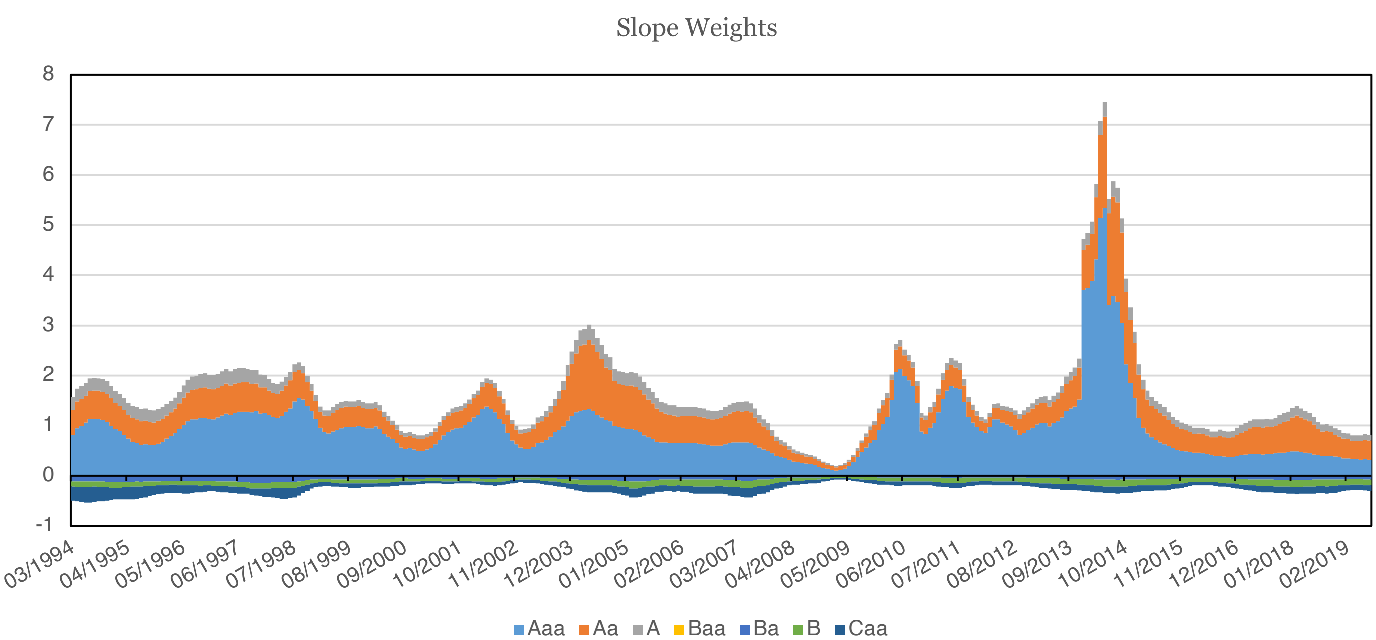
With a more appropriate measure of credit sensitivity, we can now scale our stylized factor portfolio weights such that each position contributes an equal level of DTS. This will have two effects: (1) the relative weights in the portfolios will change over time, and (2) the notional size of the portfolios will change over time.

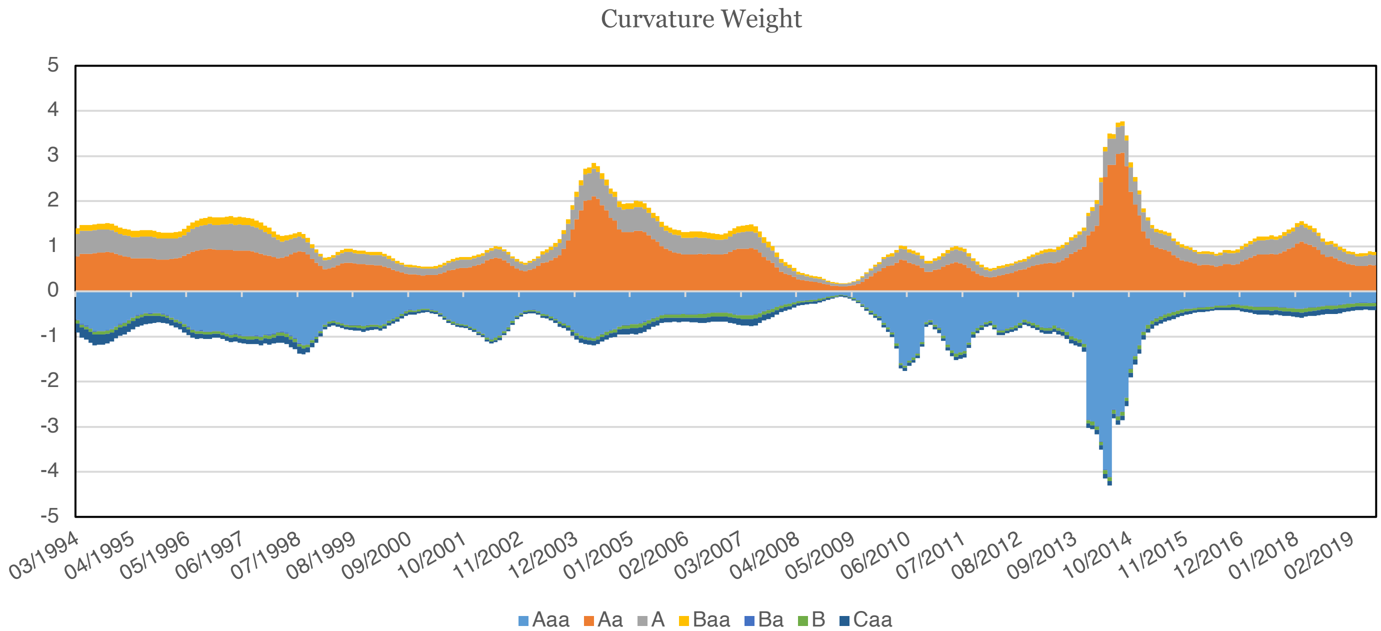
We scale each position such that (1) they contribute an equal level of DTS to the portfolio and (2) each leg of the portfolio has a total DTS of 500bps. The Level trade, therefore, represents a constant 500bps of DTS risk over time, while the Slope and Curvature trades represent 0bps, as the longs and short legs net out.

One problem still remains: interest rate risk. As we plotted earlier in this piece, the credit indices have time-varying – and sometimes substantial – interest rate exposure. This creates an unintended bet within our portfolios.

Fortunately, unlike the credit curve, true level shift does empirically apply in the Treasury yield curve. Therefore, to simplify matters, we construct a 5-year zero-coupon bond, which provides us with a constant duration instrument. At each point in time, we calculate the net duration of our credit trades and use the 5-year ZCB to neutralize the interest rate risk. For example, if the Level portfolio has a duration of 1, we would take a -20% notional position in the 5-year ZCB.


Level Weights
3
2.5
2
1.5
1
0.5
0
03/1994
07/1995
11/1996
03/1998
07/1999
11/2000
03/2002
07/2003
11/2004
03/2006
07/2007
11/2008
03/2010
07/2011
11/2012
03/2014
07/2015
11/2016
03/2018
07/2019
Aaa Aa A Baa Ba B Caa


Slope Weights
8
7
6
5
4
3
2
1
0
-1
03/1994
04/1995
05/1996
06/1997
07/1998
08/1999
09/2000
10/2001
11/2002
12/2003
01/2005
02/2006
03/2007
04/2008
05/2009
06/2010
07/2011
08/2012
09/2013
10/2014
11/2015
12/2016
01/2018
02/2019
Aaa Aa A Baa Ba B Caa



Source: Federal Reserve of St. Louis; Bloomberg. Calculations by Newfound Research.

Some things we note when evaluating the portfolios over time:

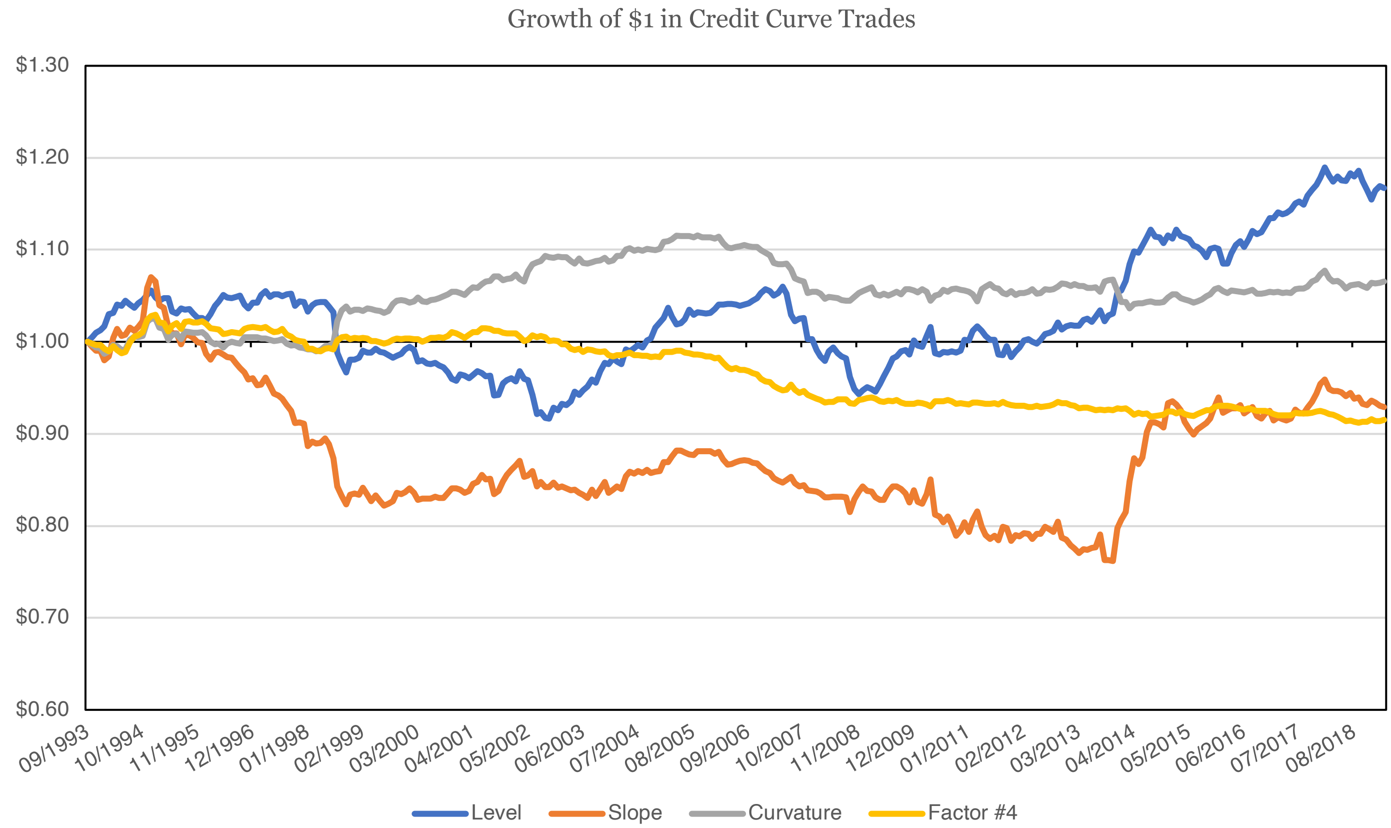
- In all three portfolios, notional exposure to higher credit qualities is substantially larger than lower credit qualities. This captures the meaningfully higher exposure that lower credit quality indices have to credit risk than higher quality indices.

- The total notional exposure of each portfolio varies dramatically over time as market regimes change. In tight spread environments, DTS is low, and therefore notional exposures increase. In wide spread environments – like 2008 – DTS levels expand dramatically and therefore only a little exposure is necessary to achieve the same risk target.

- 2014 highlights a potential problem with our approach: as Aaa spreads reached just 5bps, DTS dipped as low as 41bps, causing a significant swing in notional exposure to maintain the same DTS contribution.

Conclusion

The fruit of our all our labor is the graph plotted below, which shows the growth of $1 in our constant DTS, stylized credit factor portfolios.



What can we see?

First and foremost, constant credit exposure has not provided much in the last 25 years until recently. It would appear that investors did not demand a high enough premium for the risks that were realized over the period, which include the 1998 LTCM blow-up, the burst of the dot-com bubble, and the 2008 recession.

From 12/31/2008 lows through Q1 2019, however, a constant 500bps DTS exposure generated a 2.0% annualized return with 2.4% annualized volatility, reflecting a nice annual premium for investors willing to bear the credit risk.

Slope captures the high-versus-low-quality trade. We can see that junk meaningfully out-performed quality in the 1990s, after which there really did not appear to be a meaningful difference in performance until 2013 when oil prices plummeted and high yield bond prices collapsed. This result does highlight a potential problem in our analysis: the difference in sector composition of the underlying indices. High yield bonds had an outsized reaction compared to higher quality investment grade credit due to a more substantial exposure to the energy sector, leading to a lop-sided reaction.

What is also interesting about the Slope trade is that the market did not seem to price a meaningful premium for holding low quality credit over high quality credit.

Finally, we can see that Curvature ("barbell versus belly") – trade was rather profitable for the first decade, before deflating pre-2008 and going on a mostly-random walk ever since. However, as mentioned when the curvature trade was initially introduced, the 4th factor in our decomposition also appeared to reflect a similar trade but shorts Aa and Caa versus a long position in A and Baa. This trade has been a fairly consistent money-loser since the early 2000s, indicating that a barbell of high quality (just not Aaa) and junk might do better than the belly of the curve.

It is worth pointing out that these trades represent a significant amount of compounding estimation – from duration-matching Treasury rates to credit spread durations – which also means a significant risk of compounding estimation error. Nevertheless, we believe there are a few takeaways worth exploring further:

- The Level trade appears highly regime dependent (in positive and negative economic environments), suggesting a potential opportunity for on/off credit trades.

- The 4th factor is a consistent loser, suggesting a potential structural tilt that can be made by investors by holding quality and junk (e.g. QLTA + HYG) rather than the belly of the curve (LQD). Implementing this in a long-only fashion would require more substantial analysis of duration trade-offs, as well as a better intuition as to *why* the returns are emerging as they are.

- Finally, a recognition that maintaining a constant credit risk level requires reducing notional exposure as rates go up, as rate changes are proportional to rate levels. This is an important consideration for strategic asset allocation.

DYNAMIC SPENDING IN RETIREMENT MONTE CARLO

July 15, 2019

SUMMARY

- Many retirement planning analyses rely on Monte Carlo simulations with static assumptions for withdrawals.
- Incorporating dynamic spending rules can more closely align the simulations with how investors would likely behave during times when the plan looked like it was on a path to failure.
- Even a modest reduction in withdrawals (e.g. 10%) can have a meaningful impact on reducing failure rates, nearly cutting it in half in a sample simulation.
- Combining dynamic spending rules with other marginal improvements, such as supplemental income and active risk management, can lead to more robust retirement plans and give investors a better understanding of the variables that are within their realm of control.

Monte Carlo simulations are a prevalent tool in financial planning, especially pertaining to retirement success calculations.

Under a typical framework of normally distributed portfolio returns and constant inflation-adjusted withdrawals, calculating the success of a given retirement portfolio is straightforward. But as with most tools in finance, the art lies both in the assumptions that go into the calculation and in the proper interpretation of the result.

If a client is told they have a 10% chance of running out of money over their projected retirement horizon, what does that mean for them?

They cannot make 9 copies of themselves to live out separate lives, with one copy (hopefully not the original) unfortunately burning through the account prematurely.

They also cannot create 9 parallel universes and ensure they do not choose whichever one does not work out.

We wrote previously how investors follow a single path (You Are Not a Monte-Carlo Simulation). If that path hits zero, the other hypothetical simulation paths don't mean a thing.

A simulation path is only as valuable as the assumptions that go into creating it, and fortunately, we can make our simulations align more closely with investor behavior.

The best way to interpret the 10% failure rate is to think of it as a 10% chance of *having* to make an adjustment before it hits zero. Rarely would an investor stand by while their account went to zero. There are circumstances that are entirely out of investor control, but to the extent that there was something they could do to prevent that event, they would most likely do it.

Derek Tharp, on Michael Kitces' blog, wrote a post a few years ago weighing the relative benefit of implementing small but permanent adjustments vs. large but temporary adjustments to retirement withdrawals and found that making small adjustments and leaving them in place led to greater likelihoods of success over retirement horizons (Dynamic Retirement Spending Adjustments: Small-But-Permanent Vs Large-But-Temporary).

In this week's commentary, we want to dig a little deeper into some simple path dependent modifications that we can make to retirement Monte-Carlo simulations with the hope of creating a more robust toolset for financial planning.

The Initial Plan

Suppose an investor is 65 and holds a moderate portfolio of 60% U.S. stocks and 40% U.S. Treasuries. From 1871 until mid-2019, this portfolio would have returned an inflation-adjusted 5.1% per year with 10.6% volatility according to Global Financial Data.

Sticking with the rule-of-thumb 4% annual withdrawal of the initial portfolio balance and assuming a 30-year retirement horizon, this yields a predicted failure rate of 8% (plus or minus about 50 bps).

The financial plan is complete.

If you start with $1,000,000, simply withdraw $3,333/month and you should be fine 92% of the time.

But what if the portfolio drops 5% in the first month? (It almost did that in October 2018).

The projected failure rate over the next 29 years and 11 months has gone up to 11%. That violates a 10% threshold that may have been a target in the planning process.

Or what if it drops 30% in the first 6 months, like it would have in the second half of 1931?

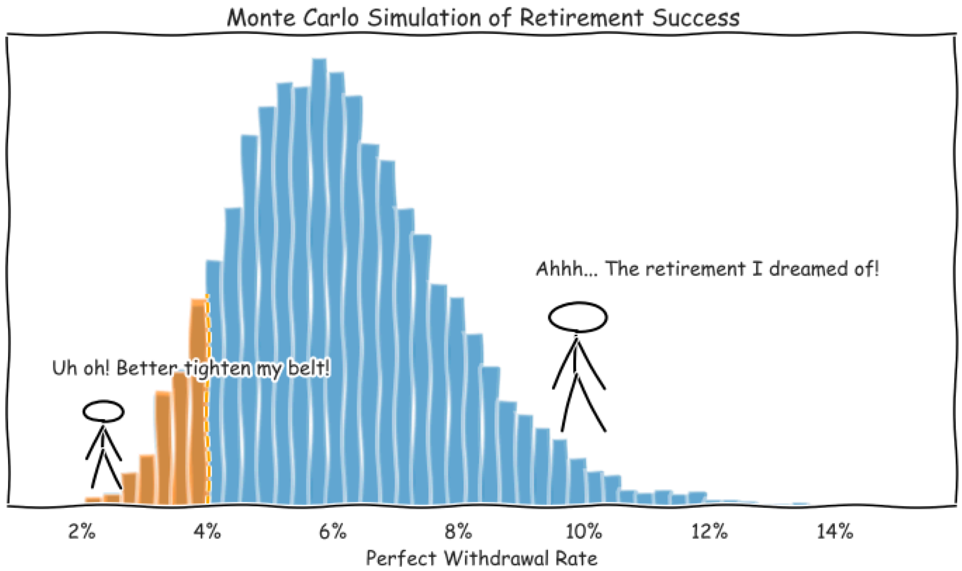
Now the project failure rate is a staggering 46%. Retirement success has been reduced to a coin flip.

Admittedly, these are trying scenarios, but these numbers are a key driver for financial planning. If we can better understand the risks and spell out a course of action beforehand, then the risk of making a rash emotion-driven decision can be mitigated.

Aligning the Plan with Reality

When the market environment is challenging, investors can benefit by being flexible. The initial financial plan does not have to be jettisoned; agreed upon actions within it are implemented.

One of the simplest – and most impactful – modifications to make is an adjustment to spending. For instance, an investor might decide at the outset to scale back spending by a set amount when the probably of failure crosses a threshold.

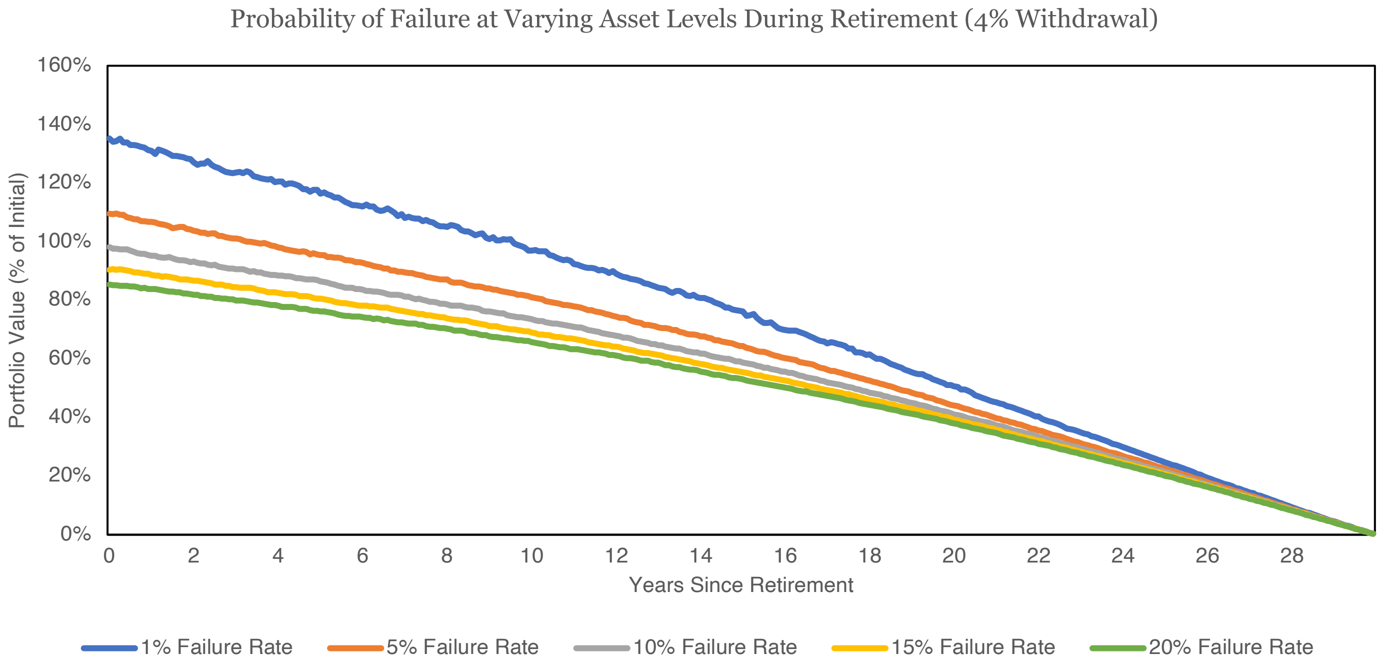


Source: Global Financial Data. Calculations by Newfound.

This reduction in spending would increase the probability of success going forward through the remainder of the retirement horizon.

And if we knew that this spending cut would likely happen if it was necessary, then we can quantify it as a rule in the initial Monte Carlo simulation used for financial planning.

Graphically, we can visualize this process by looking at the probabilities of failure for varying asset levels over time. For example, at 10 years after retirement, the orange line indicates that a portfolio value ~80% of the initial value would have about a 5% failure rate.



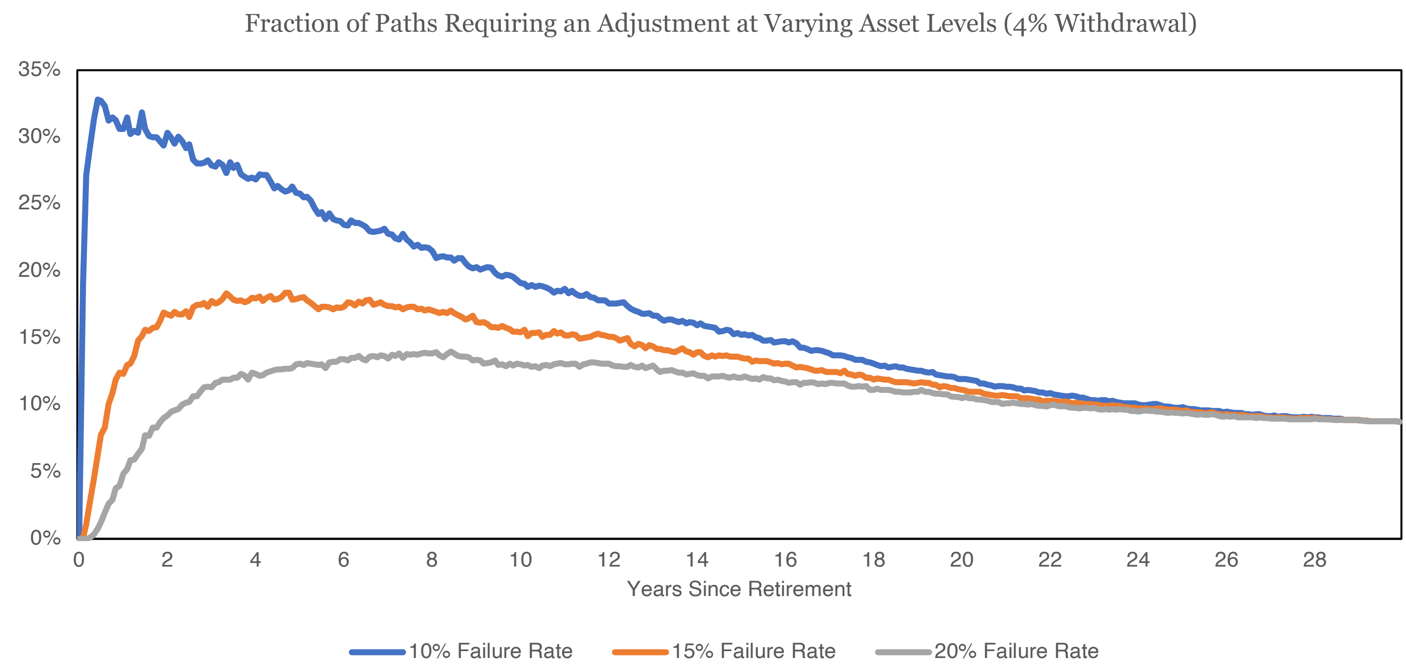
Source: Global Financial Data. Calculations by Newfound.

As long as the portfolio value remains above a given line, no adjustment would be needed based on a standard Monte Carlo analysis. Once a line is crossed, the probability of success is below that threshold.

This chart presents a good illustration of sequence risk: the lines are flatter initially after retirement and the slope progressively steepens as the time progresses. A large drawdown initially puts the portfolio below the threshold for making and adjustment.

For instance, at 5 years, the portfolio has more than a 10% failure rate if the value is below 86%. Assuming zero real returns, withdrawals alone would have reduced the value to 80%. Positive returns over this short time period would be necessary to feel secure in the plan.

Looking under the hood along the individual paths used for the Monte Carlo simulation, at 5 years, a quarter of them would be in a state requiring an adjustment to spending at this 10% failure level.



Source: Global Financial Data. Calculations by Newfound.

This belies the fact that some of the paths that would have crossed this 10% failure threshold prior to the 5-year mark improved before the 5-year mark was hit. 75% of the paths were below this 10% failure rate at some point prior to the 5-year mark. Without more appropriate expectations of a what these simulations mean, under this model, most investors would have felt like their plan's failure rate was uncomfortable at some point in the first 5 years after retirement!

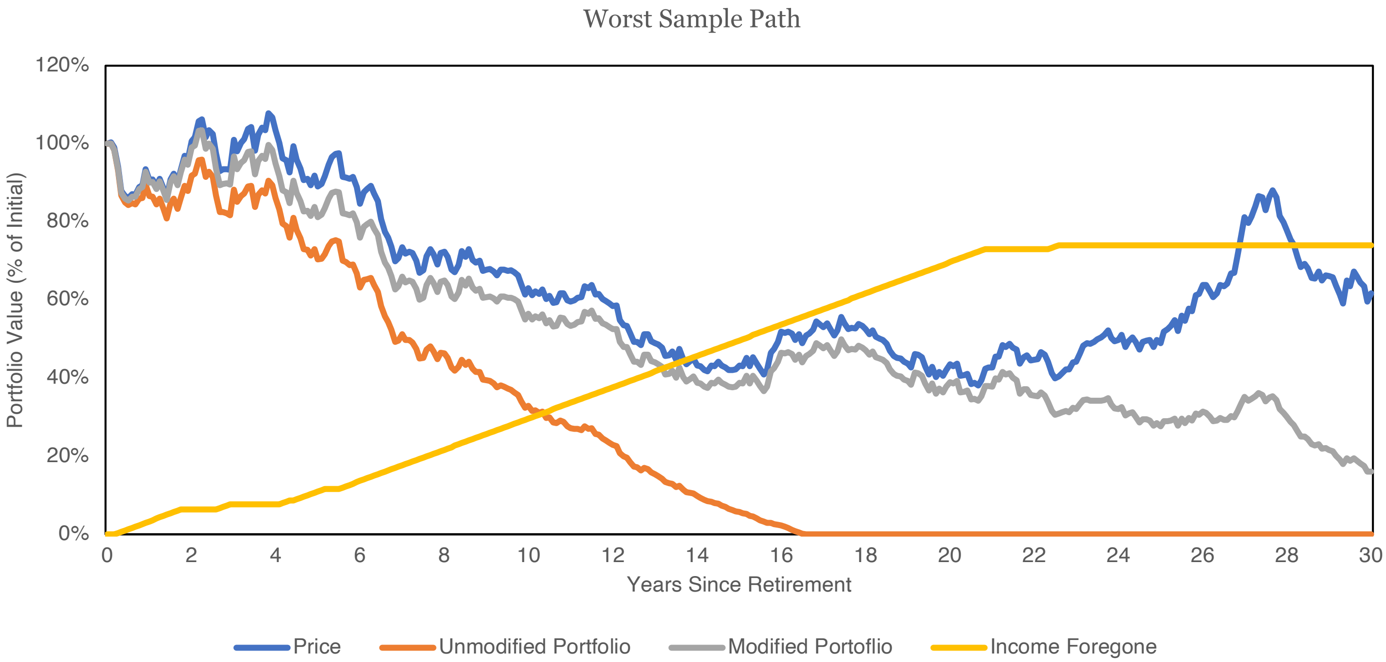
Dynamic Spending Rules

If the goal is ultimately not to run out of funds in retirement, the first spending adjustment case can substantially improve those chances (aside from a large negative return in the final periods prior to the last withdrawals).

Each month, we will compare the portfolio value to the 90% success value. If the portfolio is below that cutoff, we will size the withdrawal to hit improve the odds of success back to that level, if possible.

The benefit of this approach is greatly improved success along the different paths. The cost is forgone income.

But this can mean forgoing a lot of income over the life of the portfolio in a particularly bad state of the world. The worst case in terms of this total forgone income is shown below.



Source: Global Financial Data. Calculations by Newfound.

The portfolio gives up withdrawals totaling 74%, nearly *19 years' worth*. Most of this is given up in consecutive periods during the prolonged drawdown that occurs shortly after retirement.

This is an extreme case that illustrates how large of income adjustments could be required to ensure success under a Monte Carlo framework.

The median case foregoes 9 months of total income over the portfolio horizon, and the worst 5% of cases all give up 30% (7.5 years) of income based off the initial portfolio value.

That is still a bit extreme in terms of potential cutbacks.

As a more realistic scenario that is easier on the pocketbook, we will limit the total annual cutback to 30% of the withdrawal in the following manner:

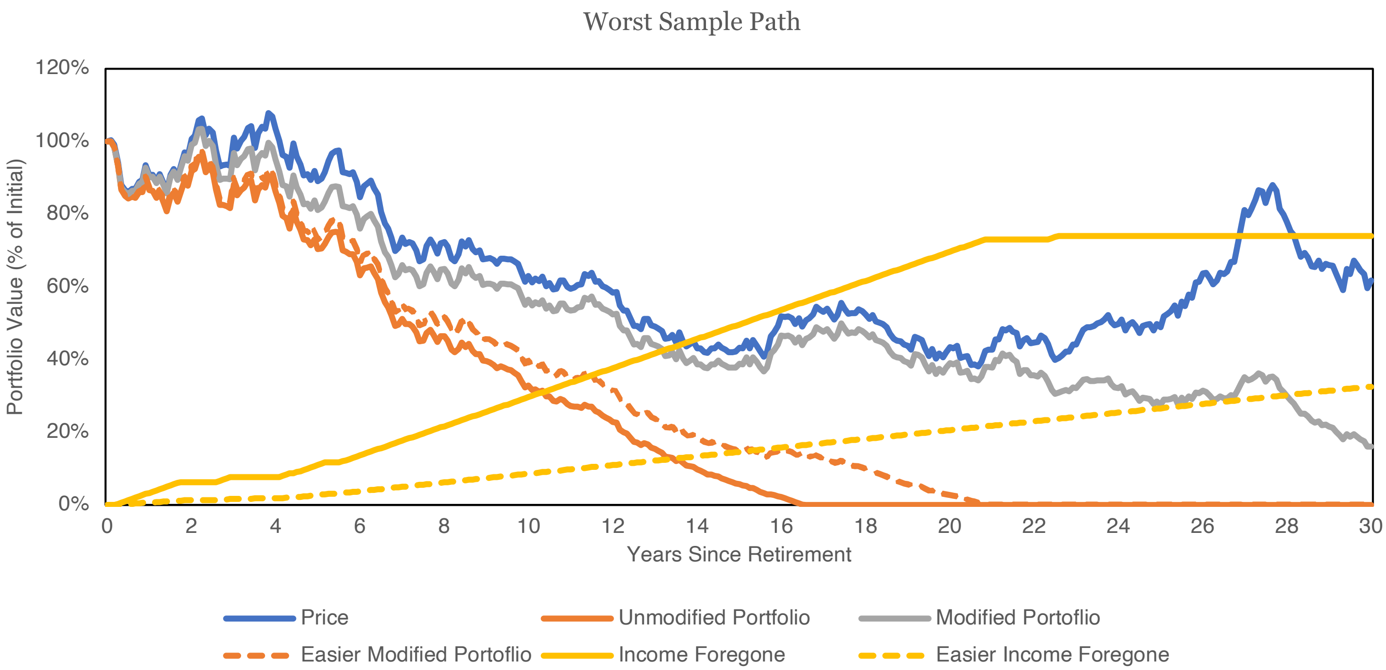
1) If the current chance of failure is greater than 20%, cut spending by 30%. This equates to reducing the annual withdrawal by $12,000 assuming a $1,000,000 initial balance.

2) If the current chance of failure is between 15% and 20%, cut spending by 20%. This equates to reducing the annual withdrawal by $8,000 assuming a $1,000,000 initial balance.

3) If the current chance of failure is between 10% and 15%, cut spending by 10%. This equates to reducing the annual withdrawal by $4,000 assuming a $1,000,000 initial balance.

These rules still increase the success rate to 99% but substantially reduce the amount of reductions in income.

Looking again at the worst-case scenario, we see that this case still "fails" (even though it lasts another 4.5 years) but that its reduction in come is now less than half of what it was in the extreme cutback case. This pattern is in line with the "lower for longer" reductions that Derek had looked at in the blog post.



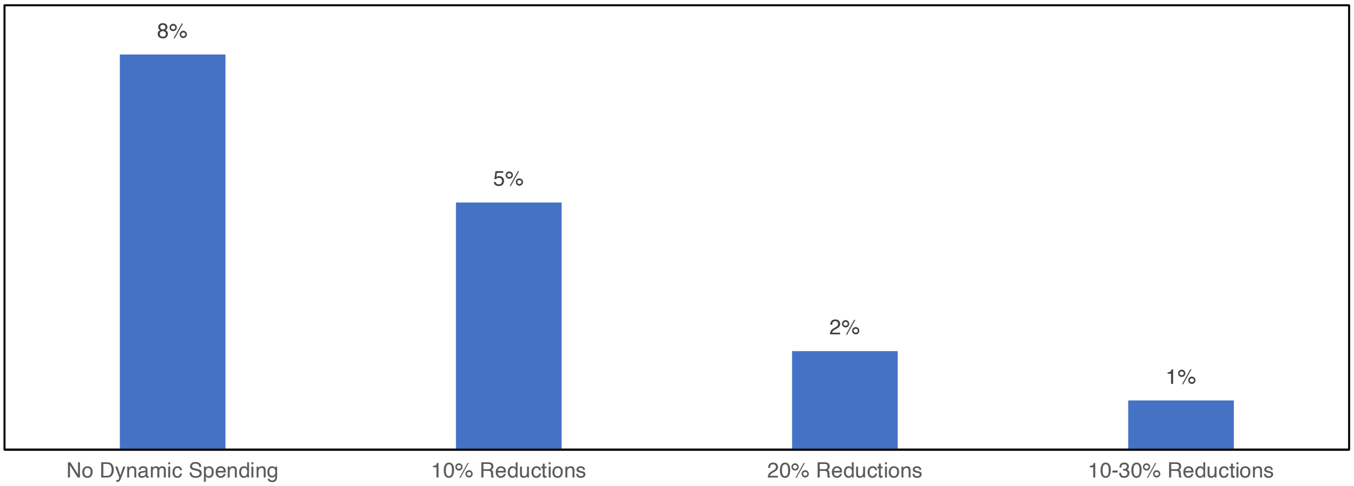
Source: Global Financial Data. Calculations by Newfound.

On the 66% of sample paths where there was a cut in spending at some point, the average total cut amounted to 5% of the portfolio (a little over a year of withdrawals spread over the life of the portfolio).

Even moving to an even less extreme reduction regime where only 10% cuts are ever made if the probability of failure increases above 10%, the average reduction in the 66% of cases that required cuts was about 9 months of withdrawals over the 30-year period.

In these scenarios, the failure rate is reduced to 5% (from 8% with no dynamic spending rules).

Failure Rates for Different Spending Reductions



Source: Global Financial Data. Calculations by Newfound.

Conclusion

Retirement simulations can be a powerful planning tool, but they are only as good as their inputs and assumptions. Making them align as closes with reality as possible can be a way to quantify the impact of dynamic spending rules in retirement.

While the magnitude of spending reductions necessary to guarantee success of a retirement plan in all potential states of the world is prohibitive. However, small modifications to spending can have a large impact on success.

For example, reducing withdrawal by 10% when the forecasted failure rate increases above 10% nearly cut the failure rate of the entire plan in half.

But dynamic spending rules do not exist in a vacuum; they can be paired with other marginal improvements to boost the likelihood of success:

- Seek out higher returns – small increases in portfolio returns can have a significant impact over the 30 -ear planning horizon.
- Supplement income – having supplements to income, even small ones, can offset spending during any market environment, improving the success rate of the financial plan.
- Actively manage risk – managing risk, especially early in retirement is a key factor to now having to reduce withdrawals in retirement.

- Plan for more flexibility – having the ability to reduce spending when necessary reduces the need to rely on the portfolio balance when the previous factors are not working.

While failure is certainly possible for investors, a “too big to fail” mentality is much more in line with the reality of retirement.

Even if absolute failure is unlikely, adjustments will likely be a requirement. These can be built into the retirement planning process and can shed light on stress testing scenarios and sensitivity.

From a retirement planning perspective, flexibility is simply another form of risk management.

ENSEMBLE MULTI-ASSET MOMENTUM

July 22, 2019

SUMMARY

- We explore a representative multi-asset momentum model that is similar to many bank-based indexes behind structured products and market-linked CDs.

- With a monthly rebalance cycle, we find substantial timing luck risk.

- Using the same basic framework, we build a simple ensemble approach, diversifying both process and rebalance timing risk.

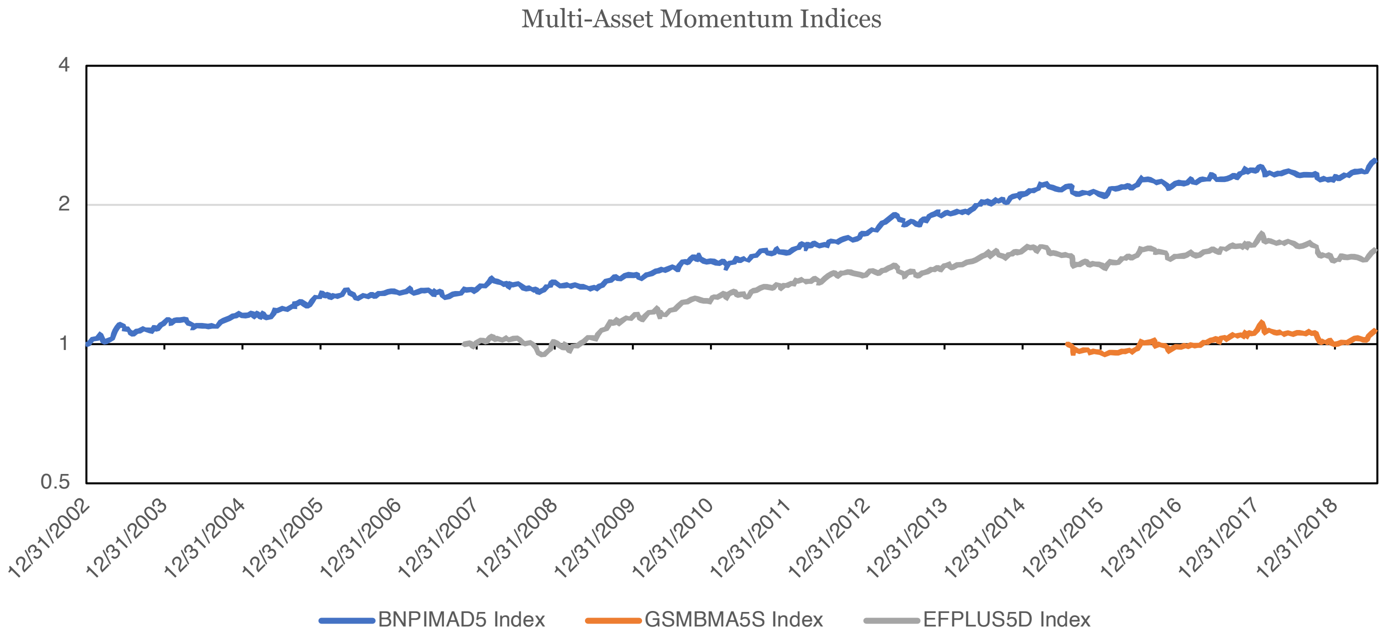
- We find that the virtual strategy-of-strategies is able to harvest diversification benefits, realizing a top-quartile Sharpe ratio with a bottom-quartile maximum drawdown.

Early in the 2010s, a suite of index-linked products came to market that raised billions of dollars. These products – offered by just about every major bank – sought to simultaneously exploit the diversification benefits of modern portfolio theory and the potential for excess returns from the momentum anomaly.

While each index has its own bells and whistles, they generally follow the same approach:

- A global, multi-asset universe covering equities, fixed income, and commodities.
- Implemented using highly liquid ETFs.
- Asset class and position-level allocation limits.
- A monthly rebalance schedule.
- A portfolio optimization that seeks to maximize weighted *prior* returns (e.g. prior 6 month returns) while limiting portfolio volatility to some maximum threshold (e.g. 5%).

And despite their differences, we can see in plotting their returns below that these indices generally share a common return pattern, indicating a common, driving style.



Source: Bloomberg.

Frequent readers will know that "monthly rebalance" is an immediate red flag for us here at Newfound: an indicator that timing luck is likely lurking nearby.

Replicating Multi-Asset Momentum

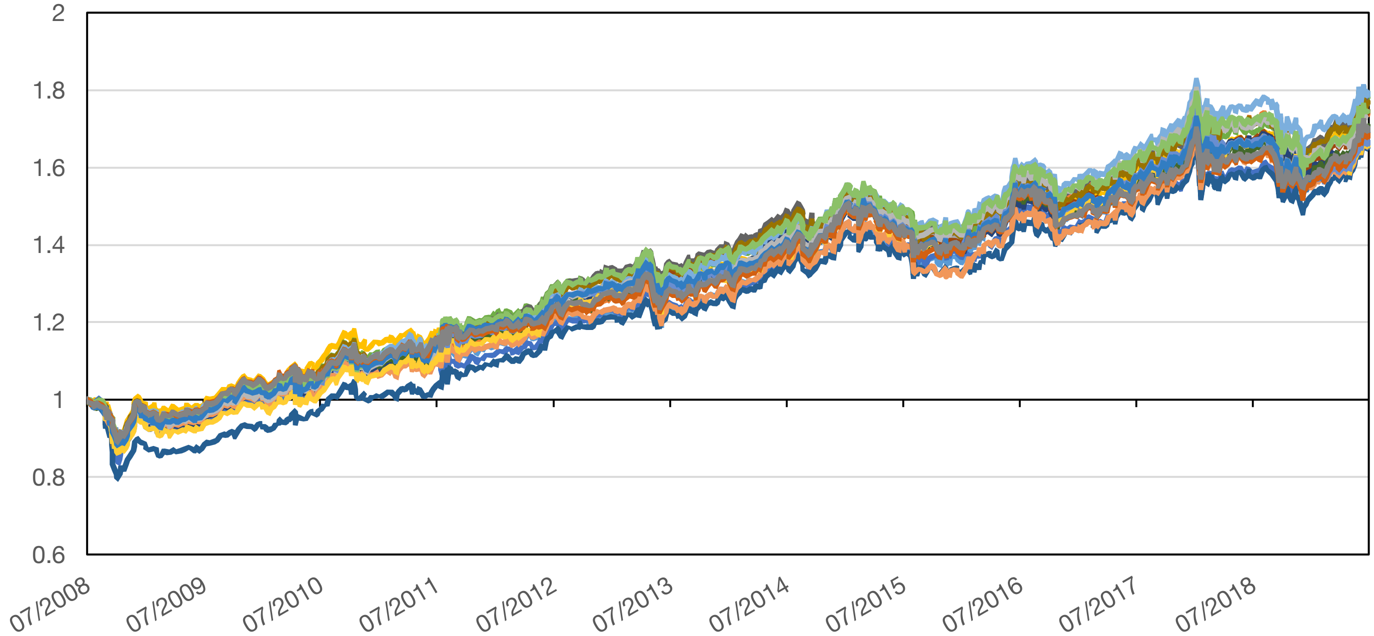
To test the impact of timing luck, we replicate a simple multi-asset momentum strategy based upon available index descriptions.

We rebalance the portfolio at the end of each month. Our optimization process seeks to identify the portfolio with a realized volatility less than 5% that would have maximized returns over the prior six months, subject to a number of position and asset-level limits. If the 5% volatility target is not achievable, the target is increased by 1% until a portfolio can be constructed that satisfies our constraints.

We use the following ETFs and asset class limits:

	Sector Cap	Basket Constituent	Asset Cap
1	Developed Equities 50%	SPDR® S&P 500® ETF Trust	20%
2		iShares® Russell 2000 Index Fund	10%
3		iShares® MSCI EAFE Index Fund	20%
4	Bonds 50%	iShares® Barclays 20+ Year Treasury Bond Fund	20%
5		iShares® iBOXX Investment Grade Corporate Bond Fund	20%
6		iShares® iBOXX High Yield Corporate Bond Fund	20%
7	Emerging Markets 25%	iShares® MSCI Emerging Markets Index Fund	20%
8		iShares® Emerging Markets Bond Fund	20%
9	Alternative Investments 25%	iShares® Dow Jones Real Estate Index Fund	20%
10		iShares® S&P GSCI™ Commodity-Indexed Trust	10%
11		SPDR® Gold Trust	10%
12	Inflation Protected Bonds and Cash 50%	iShares® Barclays TIPS Bond Fund	50%
13		JPMorgan Cash Index USD 3 Month*	50%

As a naïve test for timing luck, rather than assuming the index rebalances at the end of each month, we will simply assume the index rebalances every 21 trading days. In doing so, we can construct 21 different variations of the index, each representing the results from selecting a different rebalance date.



Source: CSI Analytics; Calculations by Newfound Research. Results are backtested and hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes, with the exception of underlying ETF expense ratios. Past performance is not an indicator of future results.

As expected, the choice of rebalance date has a meaningful impact. Annualized returns range from 4.7% to 5.5%, Sharpe ratios range from 0.6 to 0.9, and maximum drawdowns range from 9.9% to 20.8%.

On a year-by-year basis, the only thing that is consistent is the large spread between the worst and best-performing rebalance date. On average, the yearly spread exceeds 400 basis points.

	Min	**Max**
2008*	-9.91%	0.85%
2009	2.36%	4.59%
2010	6.46%	9.65%
2011	3.31%	10.15%
2012	6.76%	10.83%
2013	3.42%	6.13%
2014	5.98%	10.60%
2015	-5.93%	-2.51%
2016	4.18%	8.45%
2017	9.60%	11.62%

2018	-6.00%	-2.53%
2019 YTD	5.93%	10.01%

** Partial year starting 7/22/2018*

We've said it in the past and we'll say it again: timing luck can be the difference between hired and fired. And while we'd rather be on the side of good luck, the lack of control means we'd rather just avoid this risk all together.

If it isn't nailed down for a reason, diversify it

The choice of when to rebalance is certainly not the only free variable of our multi-asset momentum strategy. Without an explicit view as to why a choice is made, our preference is always to diversify so as to avoid specification risk.

We will leave the constraints (e.g. volatility target and weight constraints) well enough alone in this example, but we should consider the process by which we're measuring past returns as well as the horizon over which we're measuring it. There is plenty of historical efficacy to using prior 6-month total returns for momentum, but no *lack* of evidence supporting other lookback horizons or measurements.

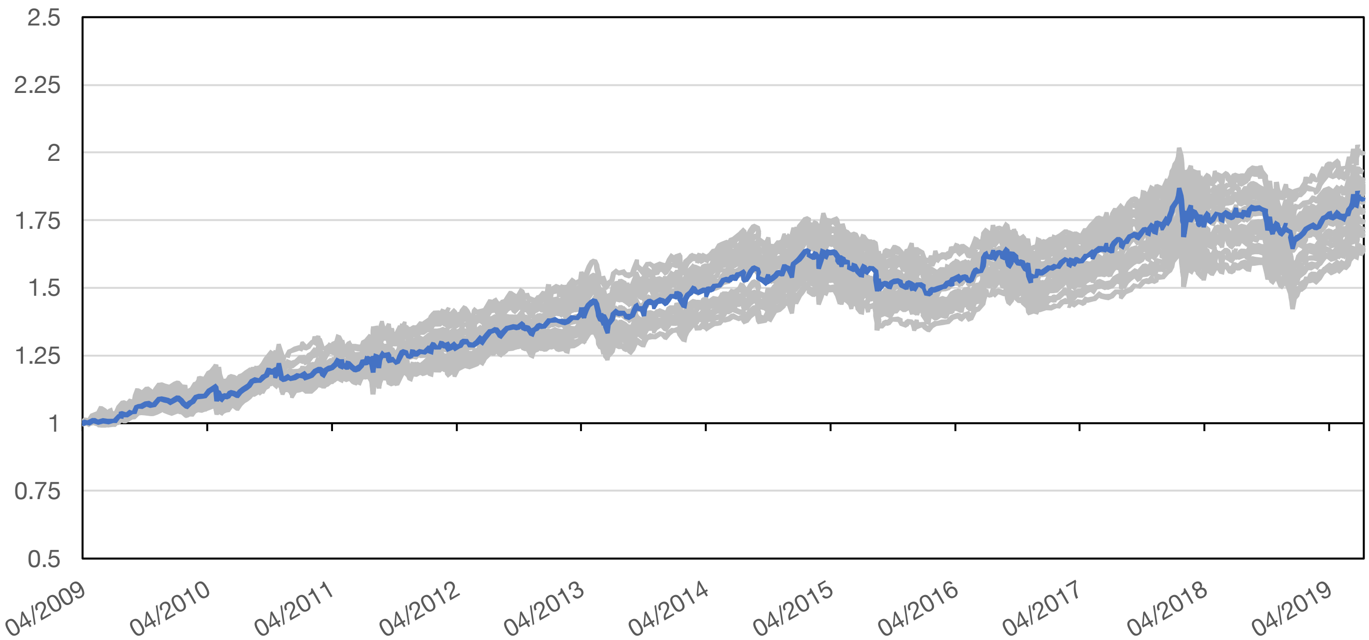
Therefore, we will use three models of momentum: prior total return, the distance of price from its moving average, and the distance of a short-term moving average from a longer-term moving average. We will vary the parameterization of these signals to cover horizons ranging from 3- to 15-months in length.

We will also vary which day of the month the portfolio rebalances on.

By varying the signal, the lookback horizon, and the rebalance date, we can generate hundreds of different portfolios, all supported by the same theoretical evidence but having slightly different realized results due to their particular specification.

Our *robust* portfolio emerges by calculating the weights for all these different variations and averaging them together, in many ways creating a virtual strategy-of-strategies.

Below we plot the result of this –ensemble approach– as compared to a –random sample of the underlying specifications–. We can see that while there are specifications that do much better, there are also those that do much worse. By employing an ensemble approach, we forgo the opportunity for good luck and avoid the risk of bad luck. Along the way, though, we may pick up some diversification benefits: the Sharpe ratio of the ensemble approach fell in the top quartile of specifications and its maximum drawdown was in the bottom quartile (i.e. lower drawdown).



Source: CSI Analytics; Calculations by Newfound Research. Results are backtested and hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes, with the exception of underlying ETF expense ratios. Past performance is not an indicator of future results.

Conclusion

In this commentary, we again demonstrate the potential risk of needless specification and the potential power of diversification.

Using a popular multi-asset momentum model as our example, we again find a significant amount of timing luck lurking in a monthly rebalance specification. By building a virtual strategy-of-strategies, we are able to manage this risk by partially rebalancing our portfolio on different days.

We go a step further, acknowledging that *process* represents another axis of risk. Specifically, we vary both how we measure momentum and the horizon over which it is measured. Through the variation of rebalance days, model specifications, and lookback horizons, we generate over 500 different strategy specifications and combine them into a virtual strategy-of-strategies to generate our robust multi-asset momentum model.

As with prior commentaries, we find that the robust model is able to effectively reduce the risk of both specification and timing luck. But perhaps most importantly, it was able to harvest the benefits of diversification, realizing a Sharpe ratio in the top quartile of specifications and a maximum drawdown in the lowest quartile.

TIMING LUCK AND SYSTEMATIC VALUE

July 29, 2019

SUMMARY

- We have shown many times that timing luck – when a portfolio chooses to rebalance – can have a large impact on the performance of tactical strategies.

- However, fundamental strategies like value portfolios are susceptible to timing luck, as well.

- Once the rebalance frequency of a strategy is set, we can mitigate the risk of choosing a poor rebalance date by diversifying across all potential variations.

- In many cases, this mitigates the risk of realizing poor performance from an unfortunate choice of rebalance date while achieving a risk profile similar to the top tier of potential strategy variations.

- By utilizing strategies that manage timing luck, the investors can more accurately assess performance differences arising from luck and skill.

On August 7th, 2013 we wrote a short blog post titled *The Luck of Rebalance Timing.* That means we have been prattling on about the impact of timing luck for over six years now (with apologies to our compliance department…).

(For those still unfamiliar with the idea of timing luck, we will point you to a recent publication from Spring Valley Asset Management that provides a very approachable introduction to the topic.[35])

While most of our earliest studies related to the impact of timing luck in tactical strategies, over time we realized that timing luck could have a profound impact on just about any strategy that rebalances on a fixed frequency. We found that even a simple fixed-mix allocation of stocks and bonds could see annual performance spreads exceeding 700bp due only to the choice of when they rebalanced in a given year.

In seeking to generalize the concept, we derived a formula that would estimate how much timing luck a strategy might have. The details of the derivation can be found in our paper recently published in the Journal of Index Investing, but the basic formula is:

$$L = \left(\frac{T}{2F}\right)S$$

[35] We're biased towards any publication that cites our work.

Here *T* is strategy turnover, *F* is how many times per year the strategy rebalances, and *S* is the volatility of a long/short portfolio capturing the difference between what the strategy is currently invested in versus what it could be invested in.

We're biased, but we think the intuition here works out fairly nicely:

- The higher a strategy's turnover, the greater the impact of our choice of rebalance dates. For example, if we have a value strategy that has 50% turnover per year, an implementation that rebalances in January versus one that rebalances in July might end up holding very different securities. On the other hand, if the strategy has just 1% turnover per year, we don't expect the differences in holdings to be very large and therefore timing luck impact would be minimal.

- The more frequently we rebalance, the lower the timing luck. Again, this makes sense as more frequent rebalancing limits the potential difference in holdings of different implementation dates. Again, consider a value strategy with 50% turnover. If our portfolio rebalances every other month, there are two potential implementations: one that rebalances January, March, May, etc. and one that rebalances February, April, June, etc. We would expect the difference in portfolio holdings to be much more limited than in the case where we rebalance only annually.[36]

- The last term, *S*, is most easily explained with an example. If we have a portfolio that can hold either the Russell 1000 or the S&P 500, we do not expect there to be a large amount of performance dispersion regardless of when we rebalance or how frequently we do so. The volatility of a portfolio that is long the Russell 1000 and short the S&P 500 is so small, it drives timing luck near zero. On the other hand, if a portfolio can hold the Russell 1000 or be *short* the S&P 500, differences in holdings due to different rebalance dates can lead to massive performance dispersion. Generally speaking, *S* is larger for more highly concentrated strategies with large performance dispersion in their investable universe.

Timing Luck in Smart Beta

To date, we have not meaningfully tested timing luck in the realm of systematic equity strategies.[37] In this commentary, we aim to provide a concrete example of the potential impact.

A few weeks ago, however, we introduced our Systematic Value portfolio, which seeks to deliver concentrated exposure to the value style while avoiding unintended process and timing luck bets.

[36] It should be noted that turnover and rebalance frequency are not independent variables. Increasing rebalance frequency often increases turnover, as signals are noisy.
[37] Though this space was previously explored by Blitz, van der Grient, and van Vliet in the 2010 paper *Fundamental Indexing: Rebalancing Assumptions and Performance.*

To achieve this, we implement an overlapping portfolio process. Each month we construct a concentrated deep value portfolio, selecting just 50 stocks from the S&P 500. However, because we believe the evidence suggests that value is a slow-moving signal, we aim for a holding period between 3-to-5 years. To achieve this, our capital is divided across the prior 60 months of portfolios.[38]

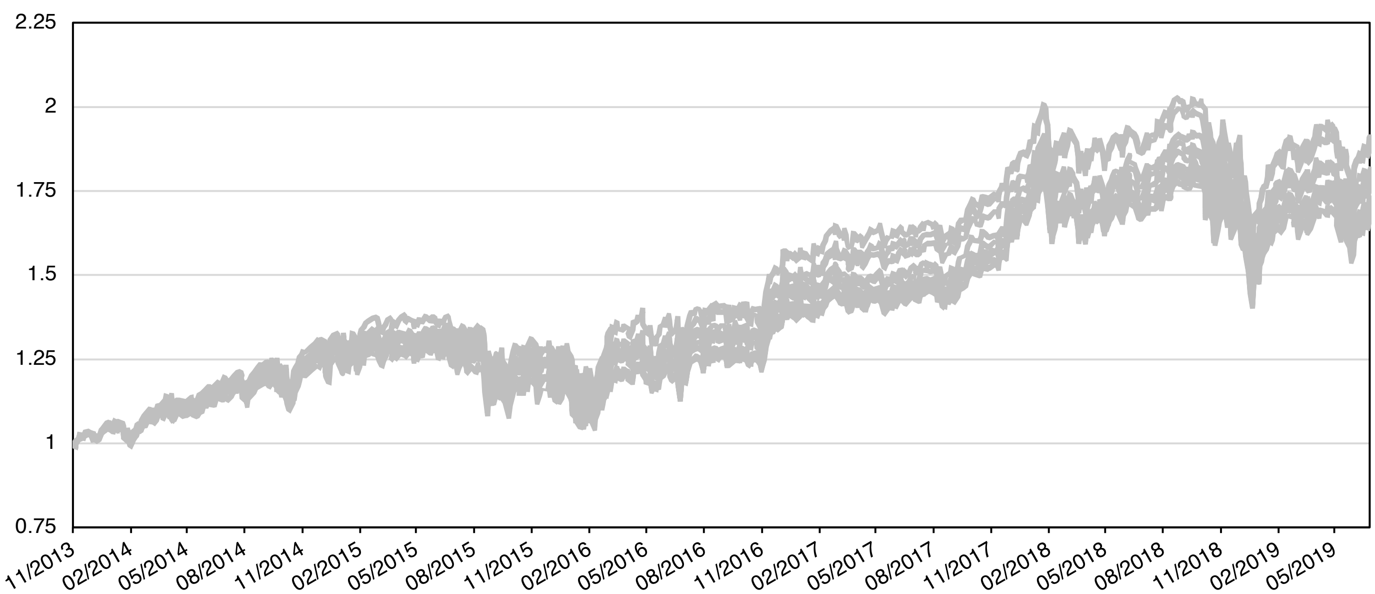
Which all means that we have monthly snapshots of deep value[39] portfolios going back to November 2012, providing us data to construct all sorts of rebalance variations.

The Luck of Annual Rebalancing

Given our portfolio snapshots, we will create annually rebalanced portfolios. With monthly portfolios, there are twelve variations we can construct: a portfolio that reconstitutes each January; one that reconstitutes each February; a portfolio that reconstitutes each March; et cetera.

Below we plot the equity curves for these twelve variations.

Annual Rebalance Variations



[38] With generally increasing weights towards the forward months due to selling securities previously purchased that have re-valued positively. See our commentary on the process for more details.

[39] Based upon our specific process, it should be noted.

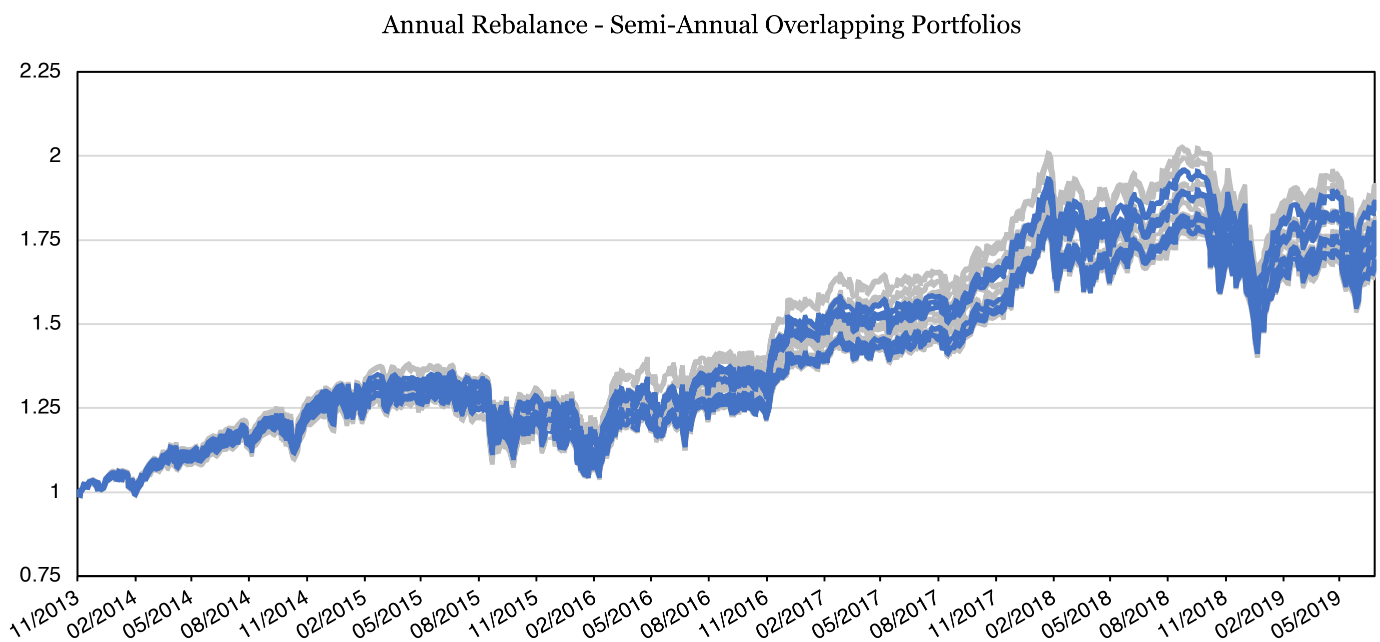
Source: CSI Analytics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

We cannot stress enough that these portfolios are all implemented using a *completely identical* process. The only difference is *when* they run that process. The annualized returns range from 9.6% to 12.2%. And those two portfolios with the largest disparity rebalanced just a month apart: January and February.

To avoid timing luck, we want to diversify when we rebalance. The simplest way of achieving this goal is through overlapping portfolios. For example, we can build portfolios that rebalance annually, but allocate to two different dates. One portfolio could place 50% of its capital in the January rebalance index and 50% in the July rebalance index.

Another variation could place 50% of its capital in the February index and 50% in the August index.[40] There are six possible variations, which we plot below.

The best performing variation (January and July) returned 11.7% annualized, while the worst (February and August) returned 9.7%. While the spread has narrowed, it would be dangerous to confuse 200bp annualized for alpha instead of rebalancing luck.



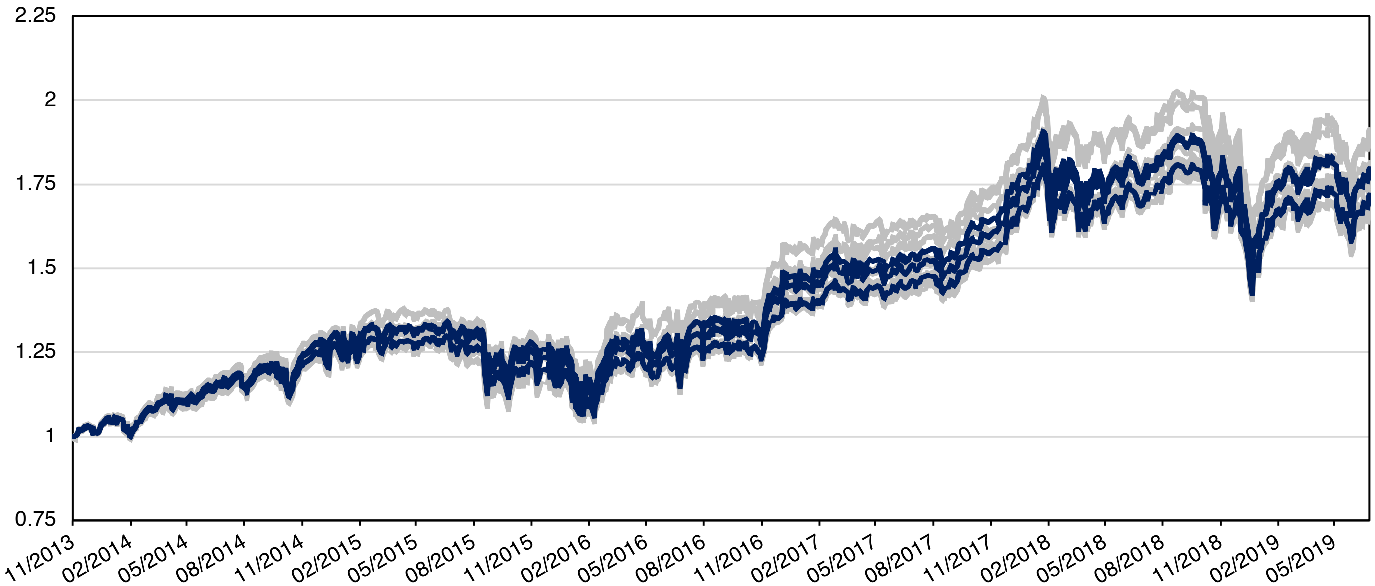
Source: CSI Analytics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

[40] We will leave out the math, but generally speaking you want to maximize the distance between overlapping portfolios.

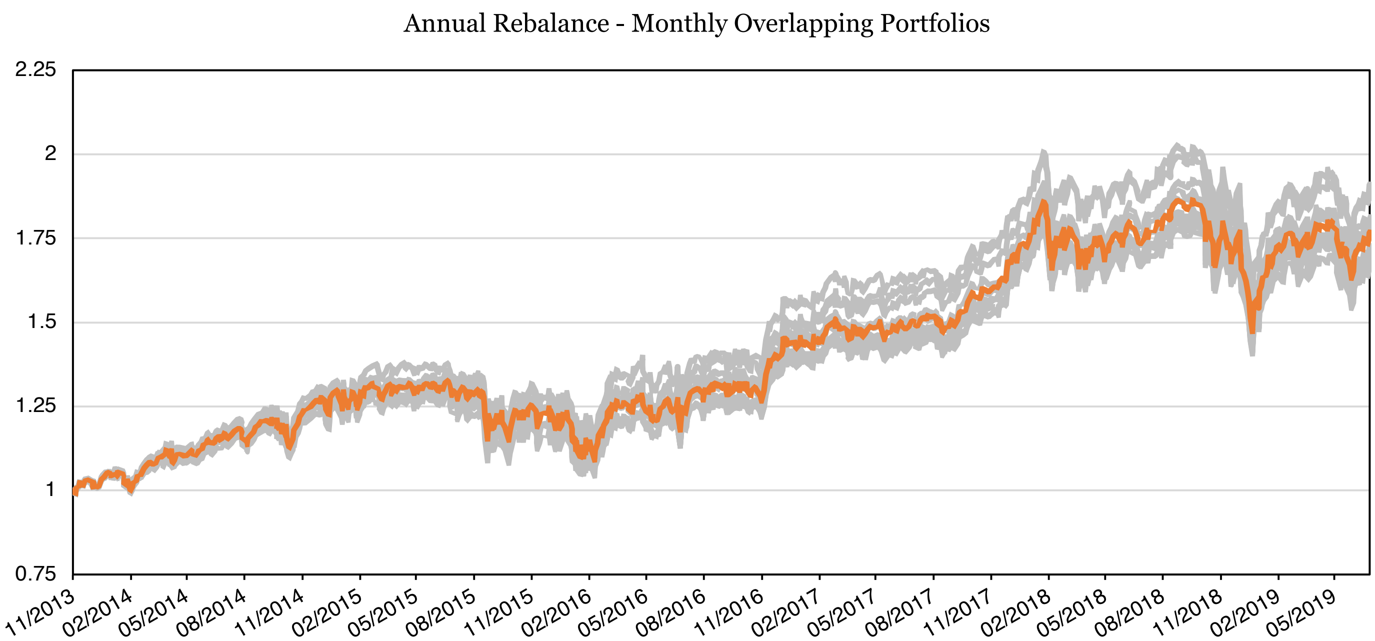
We can go beyond just two overlapping portfolios, though. Below we plot the three variations that contain four overlapping portfolios (January-April-July-October, February-May-August-November, and March-June-September-December). The best variation now returns 10.9% annualized while the worst returns 10.1% annualized. We can see how overlapping portfolios are shrinking the variation in returns.

Finally, we can plot the variation that employs 12 overlapping portfolios. This variation returns 10.6% annualized; almost perfectly in line with the average annualized return of the underlying 12 variations. No surprise: diversification has neutralized timing luck.

Annual Rebalance - Quarterly Overlapping Portfolios



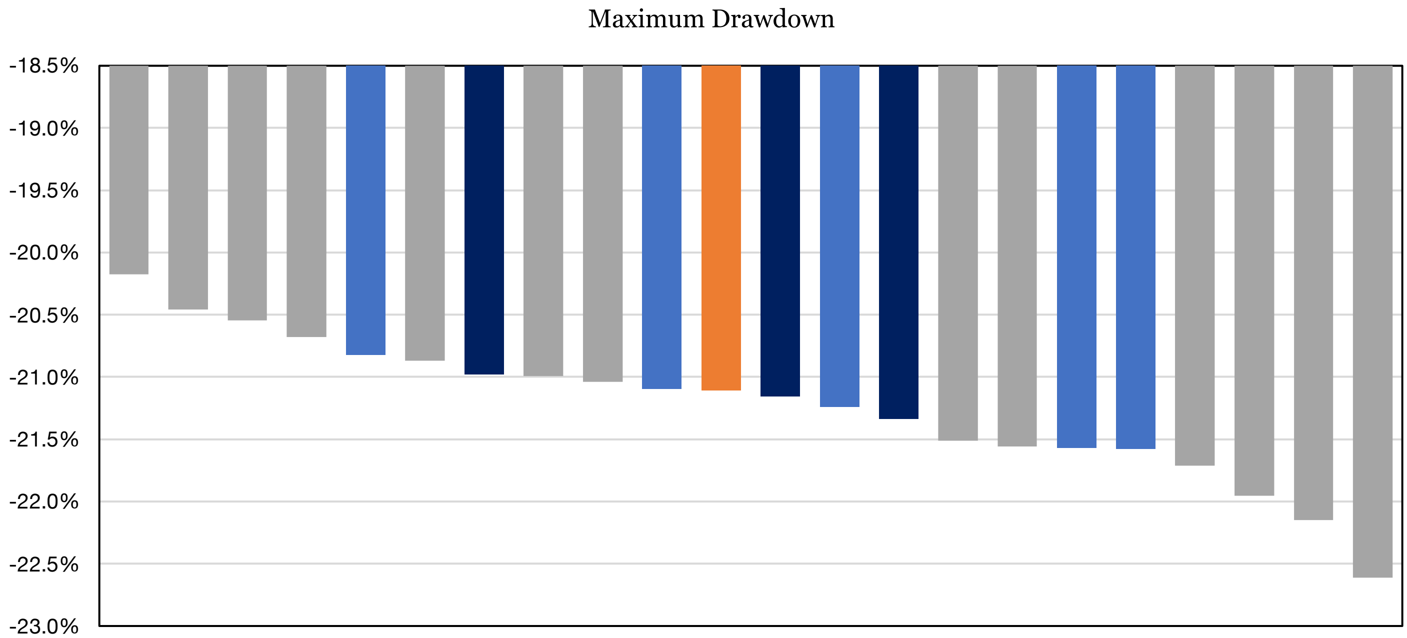
Source: CSI Analytics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.



Source: CSI Analytics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

But besides being "average by design," how can we measure the benefits of diversification?

As with most ensemble approaches, we see a reduction in realized risk metrics. For example, below we plot the maximum realized drawdown for **annual variations**, **semi-annual variations**, **quarterly variations**, and the **monthly variation**. While the dispersion is limited to just a few hundred basis points, we can see that the diversification embedded in the **monthly variation** is able to reduce the bad luck of choosing an unfortunate rebalance date.



Source: CSI Analytics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

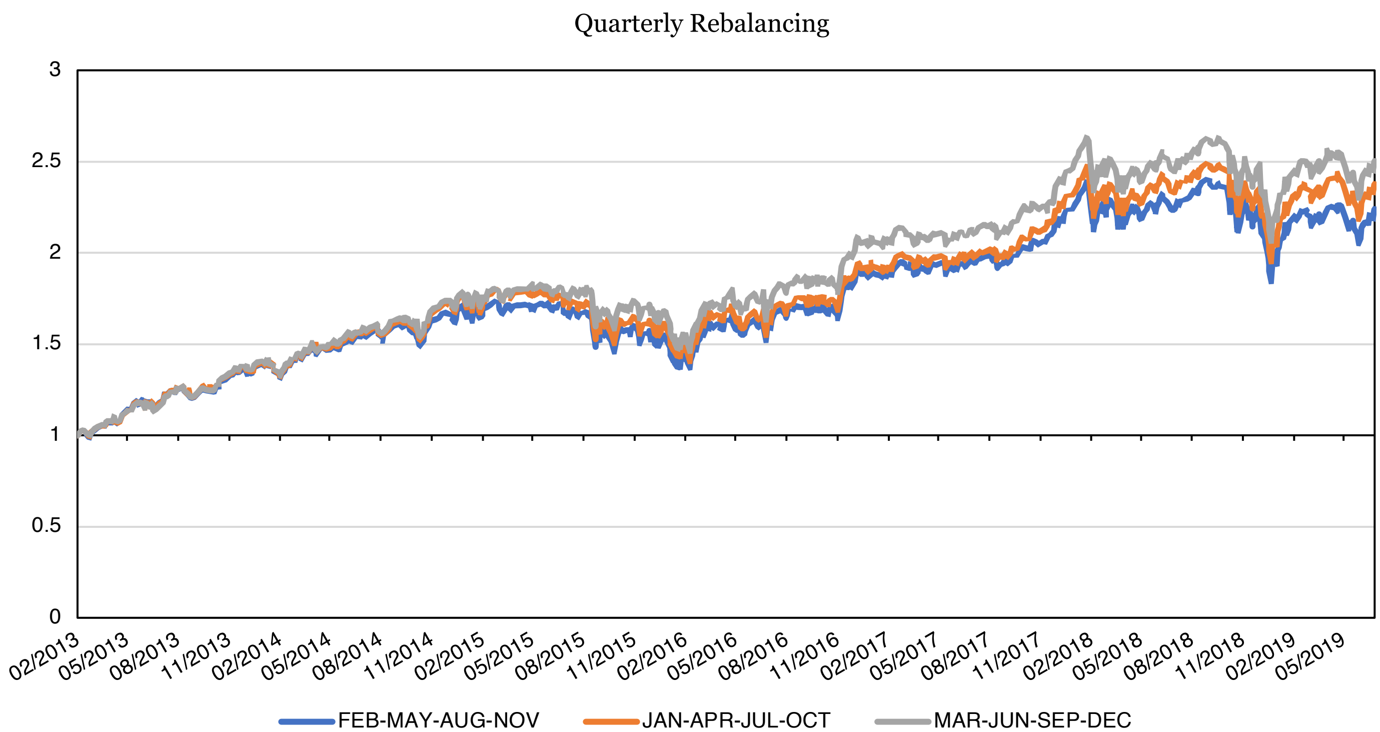
Just Rebalance more Frequently?

One of the major levers in the timing luck equation is how frequently the portfolio is rebalanced. However, we firmly believe that while rebalancing frequency impacts timing luck, timing luck should not be a driving factor in our choice of rebalance frequency.

Rather, rebalance frequency choices should be a function of the speed at which our signal decays (e.g. fast-changing signals such as momentum versus slow-changing signals like value) versus implementation costs (e.g. explicit trading costs, market impact, and taxes). Only after this choice is made should we seek to limit timing luck.

Nevertheless, we can ask the question, "how does rebalancing more frequently impact timing luck in this case?"

To answer this question, we will evaluate quarterly-rebalanced portfolios. The distinction here from the quarterly overlapping portfolios above is that the entire portfolio is rebalanced each quarter rather than only a quarter of the portfolio. Below, we plot the equity curves for the three possible variations.



Source: CSI Analytics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

The best performing variation returns 11.7% annualized while the worst returns 9.7% annualized, for a spread of 200 basis points. This is actually *larger* than the spread we saw with the three quarterly overlapping portfolio variations, and likely due to the fact that turnover within the portfolios increased meaningfully.

While we can see that increasing the frequency of rebalancing can help, in our opinion the choice of rebalance frequency should be distinct from the choice of managing timing luck.

Conclusion

In our opinion, there are at least two meaningful conclusions here:

The first is for product manufacturers (e.g. index issuers) and is rather simple: if you're going to have a fixed rebalance schedule, please implement overlapping portfolios. It isn't hard. It is literally just averaging. We're all better off for it.

The second is for product users: realize that performance dispersion between similarly-described systematic strategies can be heavily influenced by when they rebalance. The excess return may really just be a phantom of luck, not skill.

The solution to this problem, in our opinion, is to either: (1) pick an approach and just stick to it regardless of perceived dispersion, accepting the impact of timing luck; (2) hold multiple approaches that rebalance on different days; or (3) implement an approach that accounts for timing luck.

We believe the first approach is easier said than done. And without a framework for distinguishing between timing luck and alpha, we're largely making arbitrary choices.

The second approach is certainly feasible but has the potential downside of requiring more holdings as well as potentially forcing an investor to purchase an approach they are less comfortable with. For example, blending IWD (Russell 1000 Value), RPV (S&P 500 Pure Value), VLUE (MSCI U.S. Enhanced Value), and QVAL (Alpha Architect U.S. Quantitative Value) may create a portfolio that rebalances on many different dates (annual in May; annual in December; semi-annual in May and November; and quarterly, respectively), it also introduces significant process differences. Though research suggests that investors may benefit from further manager/process diversification.

For investors with conviction in a single strategy implementation, the last approach is certainly the best. Unfortunately, as far as we are aware, there are only a few firms who actively implement overlapping portfolios (including Newfound Research, O'Shaughnessy Asset Management, AQR, and Research Affiliates). Until more firms adopt this approach, timing luck will continue to loom large.

HARVESTING THE BOND RISK PREMIUM

August 5, 2019

SUMMARY

- The bond risk premium is the return that investors earn by investing in longer duration bonds.
- While the most common way that investors can access this return stream is through investing in bond portfolios, bonds often significantly de-risk portfolios and scale back returns.
- Investors who desire more equity-like risk can tap into the bond risk premium by overlaying bond exposure on top of equities.
- Through the use of a leveraged ETP strategy, we construct a long-only bond risk premium factor and investigate its characteristics in terms of rebalance frequency and timing luck.
- By balancing the costs of trading with the risk of equity overexposure, investors can incorporate the bond risk premium as a complementary factor exposure to equities without sacrificing return potential from scaling back the overall risk level unnecessarily.

The discussion surrounding factor investing generally pertains to either equity portfolios or bond portfolios in isolation. We can calculate value, momentum, carry, and quality factors for each asset class and invest in the securities that exhibit the best characteristics of each factor or a combination of factors.

There are also ways to use these factors to shift allocations between stocks and bonds (e.g. trend and standardizing based on historical levels). However, we do not typically discuss bonds as their own standalone factor.

The bond risk premium – or term premium – can be thought of as the premium investors earn from holding longer duration bonds as opposed to cash. In a sense, it is a measure of carry. Its theoretical basis is generally seen to be related to macroeconomic factors such as inflation and growth expectations. [41]

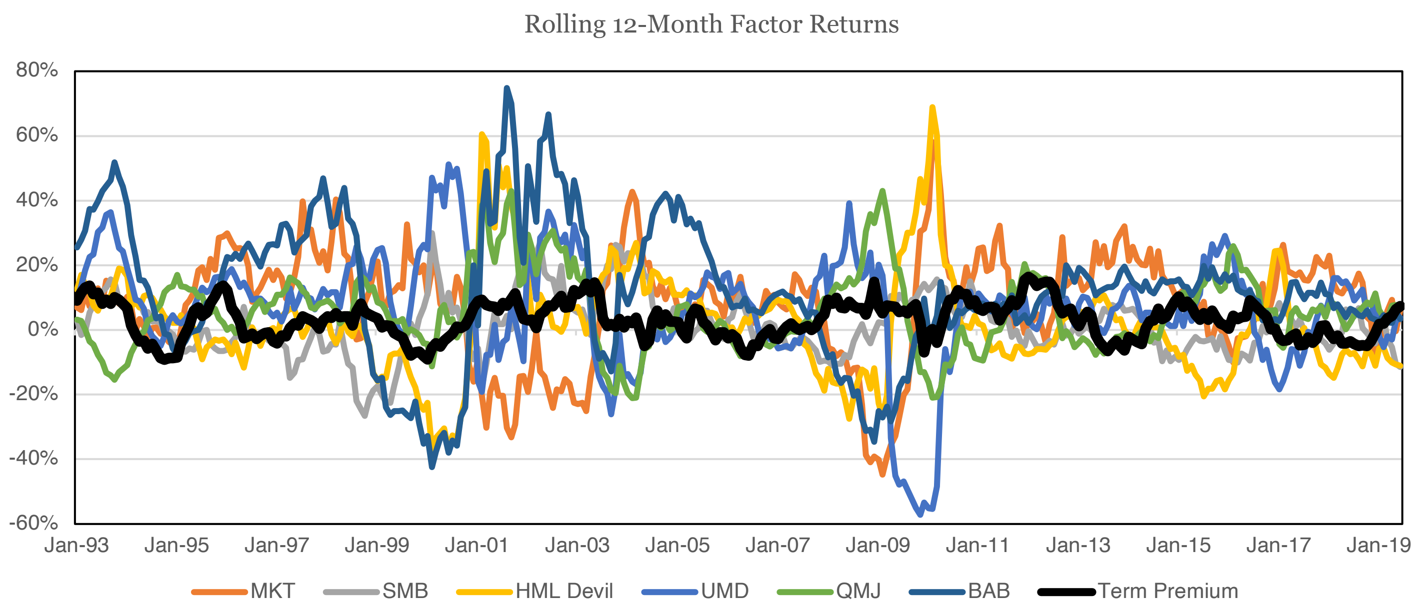
While timing the term premium using factors within bond duration buckets is definitely a possibility, this commentary will focus on the term premium in the context of an equity investor who wants long-term exposure to the factor.

[41] Kopp, E. and Williams, P. (2018). IMF Working Paper. A Macroeconomic Approach to the Term Premium.

The Term Premium as a Factor

For the term premium, we can take the usual approach and construct a self-financing long/short portfolio of 100% intermediate (7-10 year) U.S. Treasuries that borrows the entire portfolio value at the risk-free rate.

This factor, shown in bold in the chart below, has exhibited a much tamer return profile than common equity factors.



Source: CSI Analytics, AQR, and Bloomberg. Calculations by Newfound Research. Data from 1/31/1992 to 6/28/2019. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

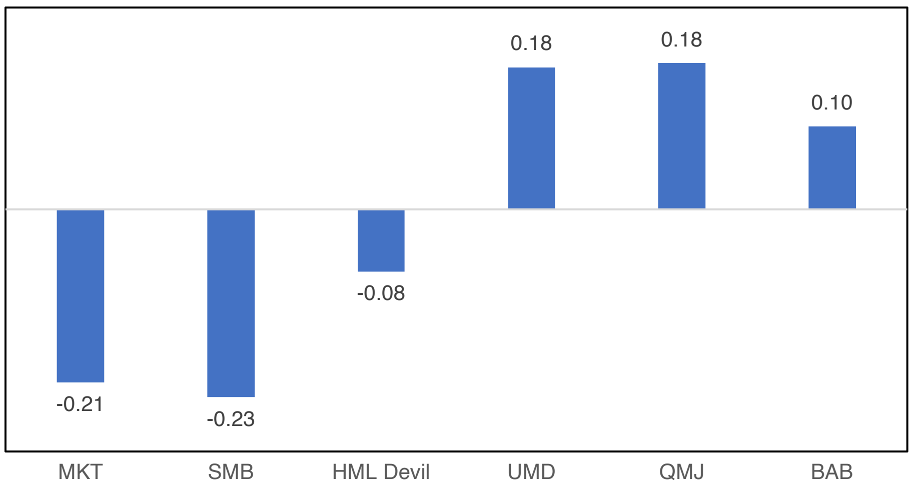
	Term Premium	MKT	SMB	HML Devil	UMD	QMJ	BAB
Annualized Return	3.0%	6.9%	0.6%	1.0%	5.1%	5.0%	9.9%
Annualized Volatility	5.8%	15.0%	9.2%	12.8%	17.3%	9.1%	13.5%
Sharpe Ratio	0.52	0.46	0.06	0.08	0.30	0.55	0.74
Maximum Drawdown	10.8%	52.7%	36.5%	50.6%	58.4%	28.5%	54.9%

Source: CSI Analytics, AQR, and Bloomberg. Calculations by Newfound Research. Data from 1/31/1992 to 6/28/2019. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

But over the entire time period, its returns have been higher than those of both the Size and Value factors. Its maximum drawdown has been less than 40% of that of the next best factor (Quality), and it is worth acknowledging that its volatility – which is generally correlated to drawdown for highly liquid assets with non-linear payoffs – has also been substantially lower.

The term premium also has exhibited very low correlation with the other equity factors.

Correlation of the Term Premium Factor to Other Equity Factors



Source: CSI Analytics, AQR, and Bloomberg. Calculations by Newfound Research. Data from 1/31/1992 to 6/28/2019. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

A Little Free Lunch

Whether we are treating bonds as factor or not, they are generally the primary way investors seek to diversify equity portfolios.

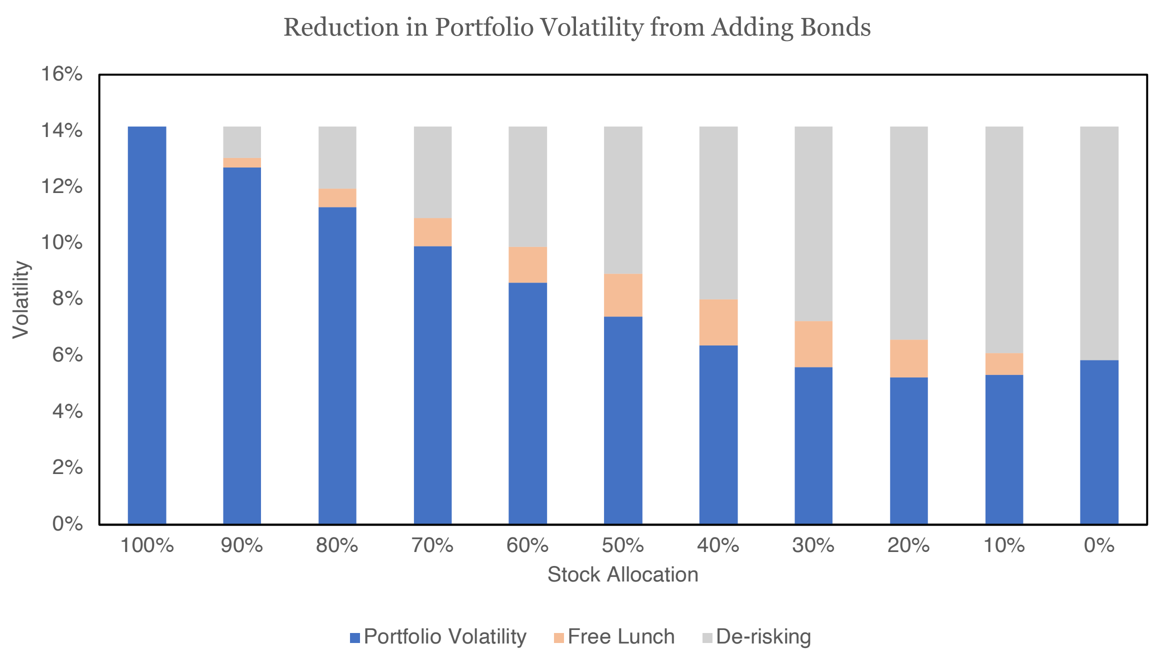
The problem is that they are also a great way to reduce returns during most market environments through their inherently lower risk.

Anytime that an asset with lower volatility is added to a portfolio, the risk will be reduced. Unless the asset class also has a particularly high Sharpe ratio, maintaining the same level of return is virtually impossible even if risk-adjusted returns are improved.

In a 2016 paper[42], Salient broke down this reduction in risk into two components: de-risking and the "free lunch" affect.

The reduction in risk form the free lunch effect is desirable, but the risk reduction from de-risking may or may not be desirable, depending on the investor's target risk profile.

The following chart shows the volatility breakdown of a range of portfolios of the S&P 500 (IVV) and 7-10 Year U.S. Treasuries (IEF).



Source: CSI Analytics and Bloomberg. Calculations by Newfound Research. Data from 1/31/1992 to 6/28/2019. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

Moving from an all equity portfolio to a 50/50 equity reduces the volatility from 14.2% to 7.4%. But only 150 bps of this reduction is from the free lunch effect that stems from the lower correlation between the two assets (-0.18). The remaining 530 bps of volatility reduction is simply due to lower risk.

In this case, annualized returns were dampened from 9.6% to 7.8%. While the Sharpe ratio climbed from 0.49 to 0.70, an investor seeking higher risk would not benefit without the use of leverage.

Despite the strong performance of the term premium factor, risk-seeking investors (e.g. those early in their careers) are generally reluctant to tap into this factor too much because of the de-risking effect.

[42] Croce, R., Guinn, R., and Robinson, T. (2016) The Free Lunch Effect: The Value of Decoupling Diversification and Risk

How do investors who want to bear risk commensurate with equities tap into the bond risk premium without de-risking their portfolio?

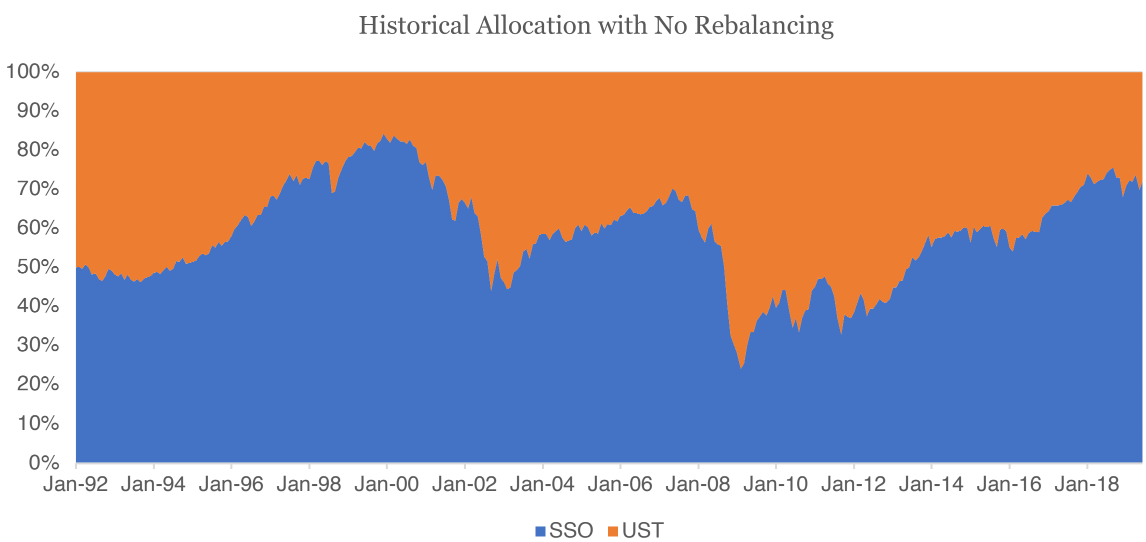
One solution is using leveraged ETPs.

Long-Only Term Premium

By taking a 50/50 portfolio of the 2x Levered S&P 500 ETF (SSO) and the 2x Levered 7-10 Year U.S. Treasury ETF (UST), we can construct a portfolio that has 100% equity exposure and 100% of the term premium factor.[43]

But managing this portfolio takes some care.

Left alone to drift, the allocations can get very far away from their target 50/50, spanning the range from 85/15 to 25/75. Periodic rebalancing is a must.



Source: CSI Analytics and Bloomberg. Calculations by Newfound Research. Data from 1/31/1992 to 6/28/2019. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

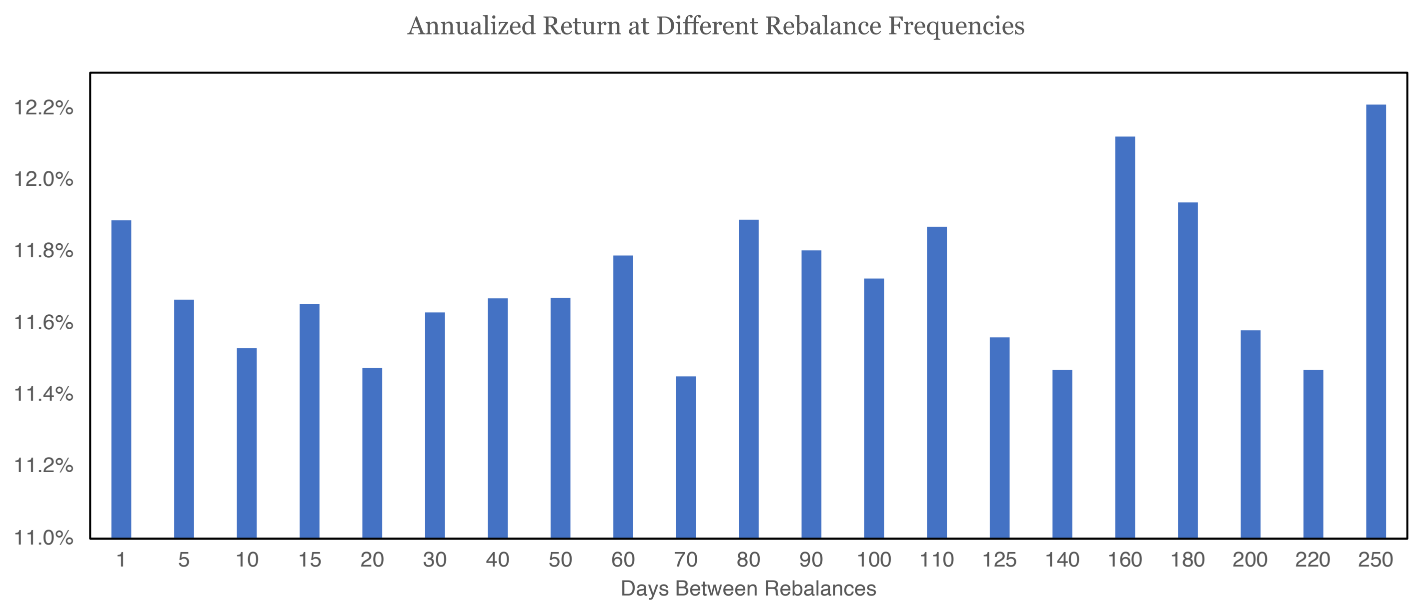
Of course, now the question is, "How frequently should we rebalance the portfolio?"

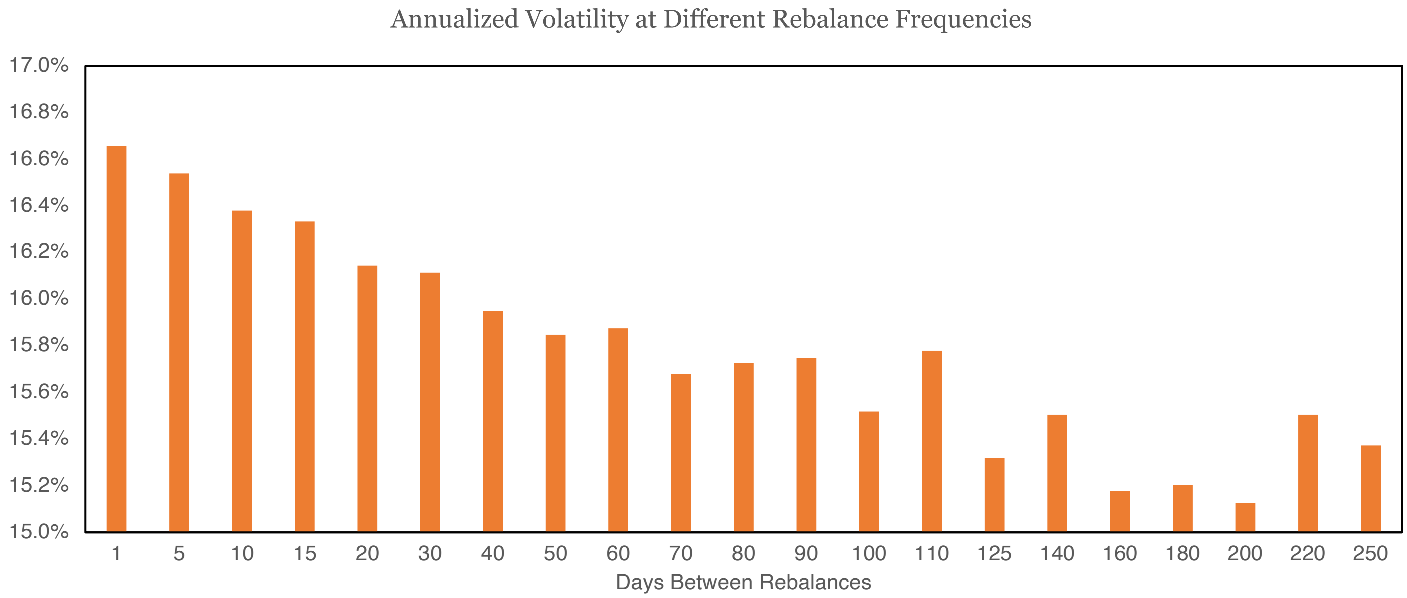
[43] Index data for the S&P 500 and the 7-10 Year U.S. Treasury index were used prior to inception of the ETPs. The historical fees on the ETPs were assumed to be their present values, and the 3-month LIBOR rate was used for borrowing. The cost of the total return swap was calculated by calibrating the model to the live data for each security.

This boils down to a balancing act between performance and costs (e.g. ticket charges, tax impacts, operational burden, etc.).

On one hand, we would like to remain as close to the 50/50 allocation as possible to maintain the desired exposure to each asset class. However, this could require a prohibitive amount of trading.

From a performance standpoint, we see improved results with longer holding periods (take note of the y-axes in the following charts; they were scaled to highlight the differences).





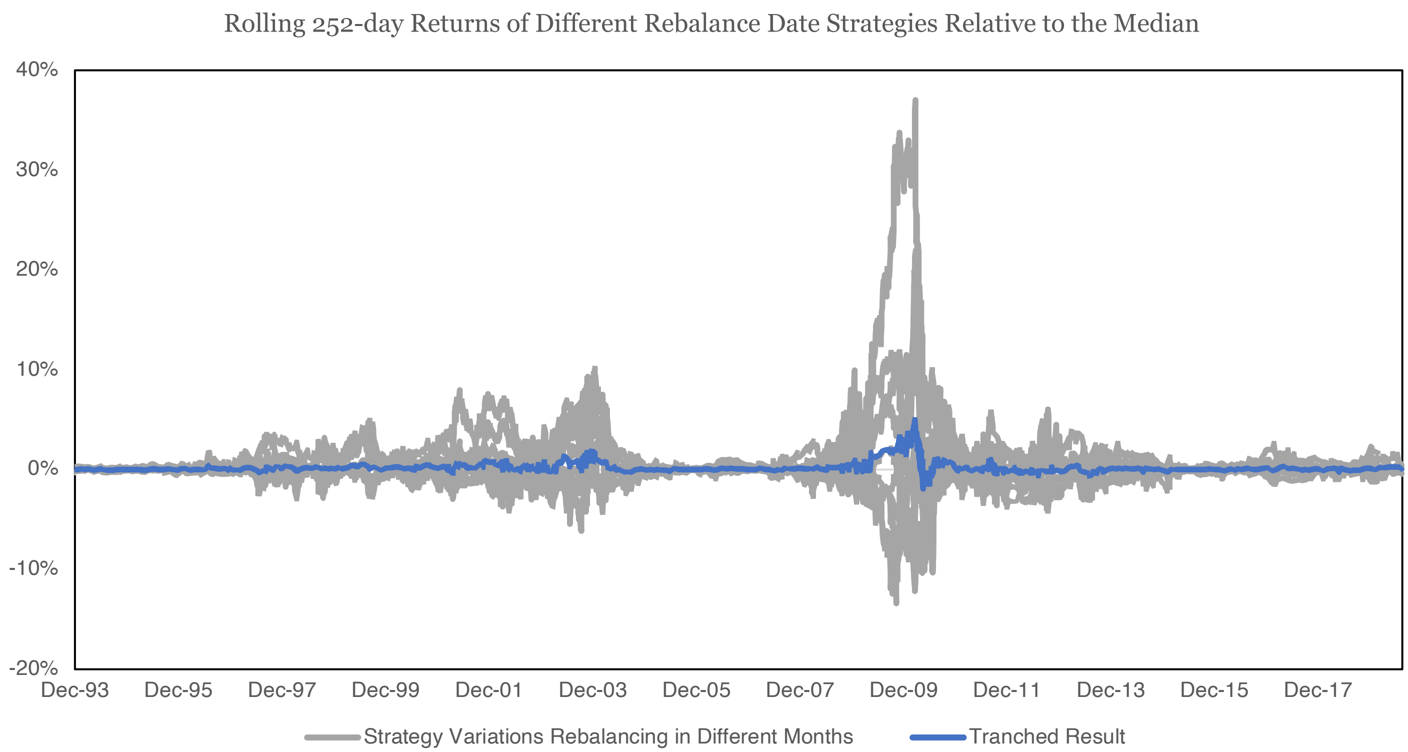
Source: CSI Analytics and Bloomberg. Calculations by Newfound Research. Data from 1/31/1992 to 6/28/2019. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

The returns do not show a definitive pattern based on rebalance frequency, but the volatility decreases with increasing time between rebalances. This seems like it would point to waiting longer between rebalances, which would be corroborated by a consideration of trading costs.

The issues with waiting longer between the rebalance are twofold:

1. Waiting longer is essentially a momentum trade. The better performing asset class garners a larger allocation as time progresses. This can be a good thing – especially in hindsight with how well equities have done – but it allows the portfolio to become overexposed to factors that we are not necessarily intending to exploit.

2. Longer rebalances are more exposed to timing luck. For example, a yearly rebalance may have done well from a performance perspective, but the short-term performance could vary by as much as 50,000 bps between the best performing rebalance month and the worst! The chart below shows the performance of each iteration relative to the median performance of the 12 different monthly rebalance strategies.



Source: CSI Analytics and Bloomberg. Calculations by Newfound Research. Data from 1/31/1992 to 6/28/2019. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

As the chart also shows, tranching can help mitigate timing luck. Tranching also gives the returns of the strategies over the range of rebalance frequencies a more discernable pattern, with longer rebalance period strategies exhibiting slightly higher returns due to their higher average equity allocations.

Under the assumption that we can tranche any strategy that we choose, we can now compare only the tranched strategies at different rebalance frequencies to address our concern with taking bets on momentum.

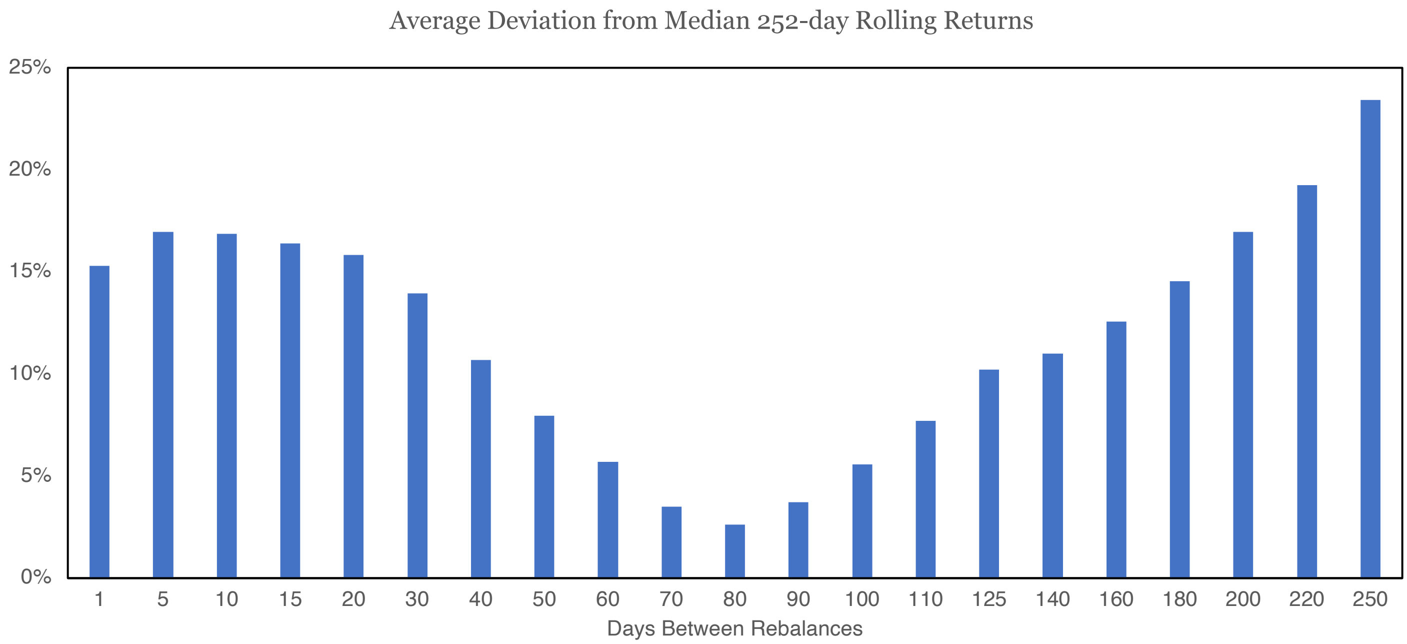
Pausing for a minute, we should be clear that we do not actually *know* what the true factor construction should be; it is a moving target. We are more concerned with robustness than simply trying to achieve outperformance. So we will compare the strategies to the median performance of the previously monthly offset annual rebalance strategies.

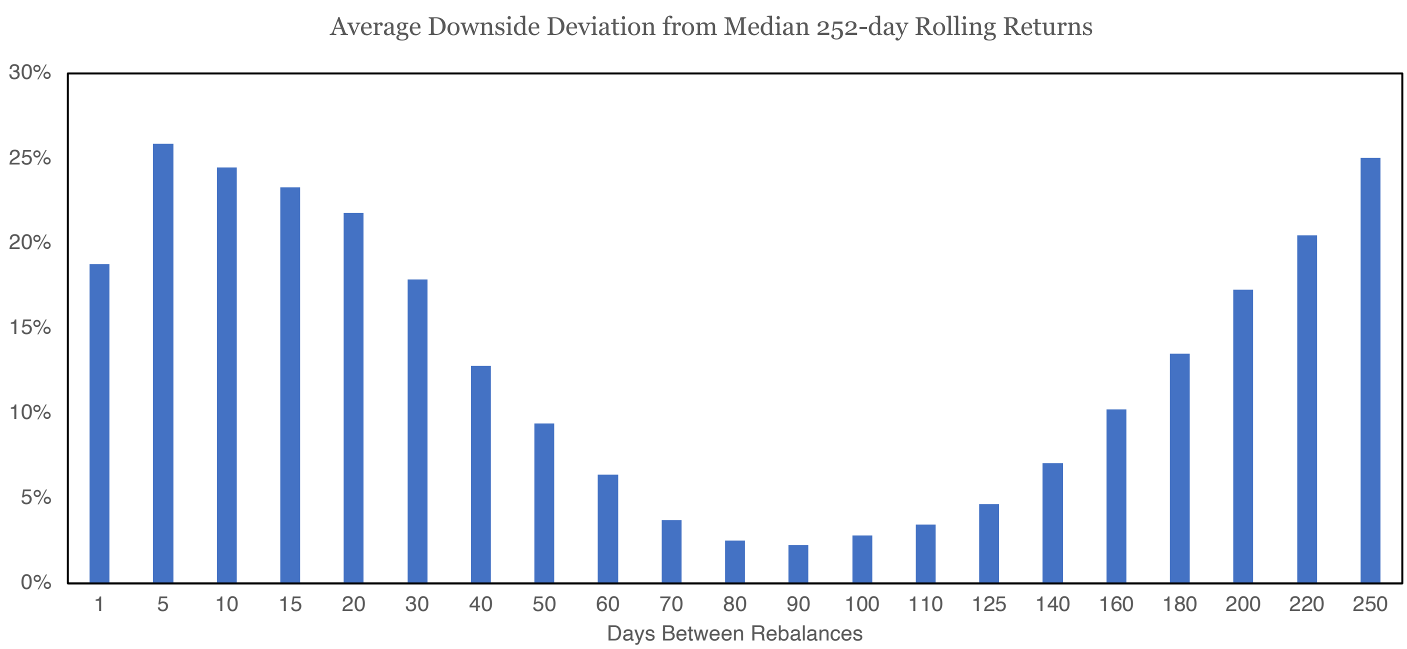
The following charts shows the aggregate risk of short-term performance deviations from this benchmark.

The first one shows the aggregate deviations, both positive and negative, and the second focuses on only the downside deviation (i.e. performance that is worse than the median).[44]

[44] We are using root-mean square errors here.

Both charts support a choice of rebalance frequency somewhere in the range of 3-6 months.





Source: CSI Analytics and Bloomberg. Calculations by Newfound Research. Data from 1/31/1992 to 6/28/2019. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

With the rebalance frequency set based on the construction of the factor, the last part is a consideration of costs.

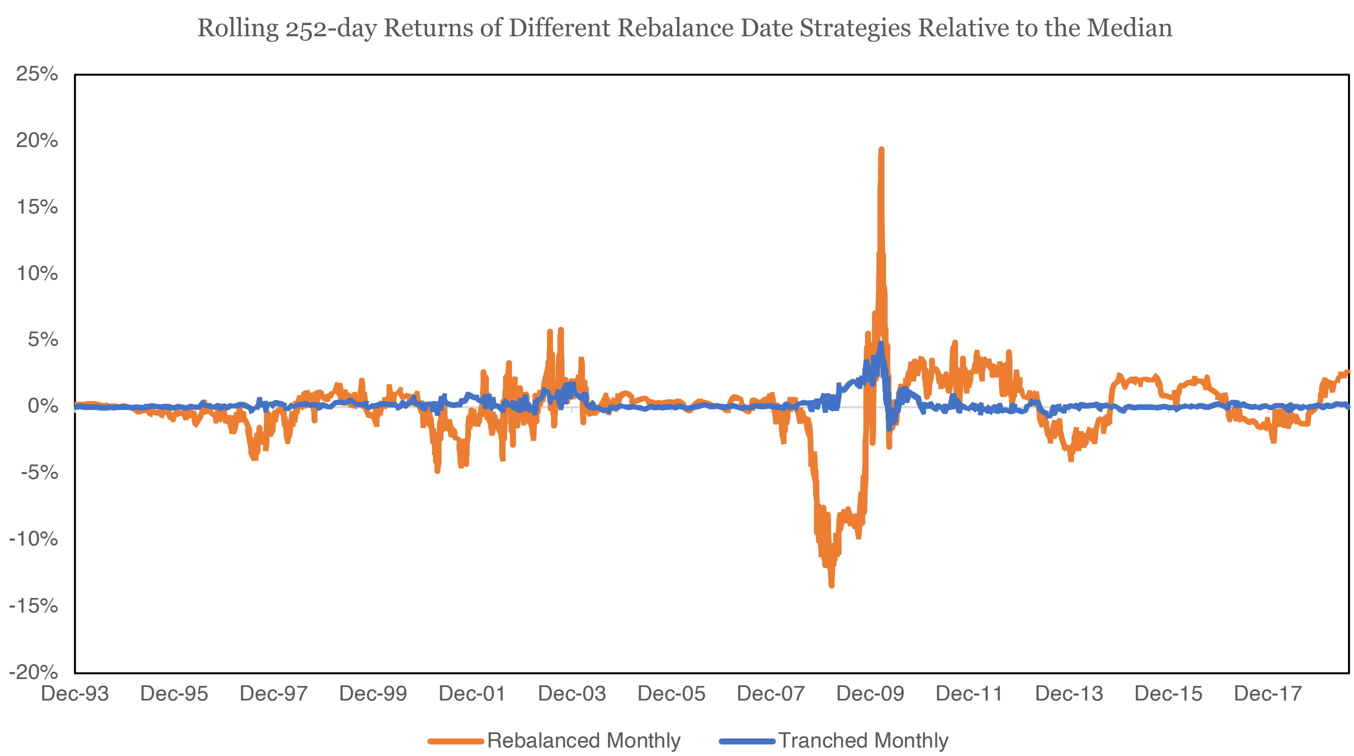
Unfortunately, this is more situation-specific (e.g. what commissions does your platform charge for trades?).

From an asset manager point-of-view, where we can trade with costs proportional to the size of the trade, execute efficiently, and automate much of the operational burden, tranching is our preferred approach.

We also prefer this approach over simply rebalancing back to the static 50/50 allocation more frequently.

In our previous commentary on constructing value portfolios to mitigate timing luck, we described how tranching monthly is a different decision than rebalancing monthly and that tranching frequency and rebalance frequency are distinct decisions.

We see the same effect here where we plot the monthly tranched annually rebalanced strategy (blue line) and the strategy rebalanced back to 50/50 every month (orange line).



Source: CSI Analytics and Bloomberg. Calculations by Newfound Research. Data from 1/31/1992 to 6/28/2019. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

Tranching wins out.

However, since the target for the term premium factor is a 50/50 static allocation, running a simple allocation filter to keep the portfolio weights within a certain tolerance can be a way to implement a more dynamic rebalancing model while reducing costs.

For example, rebalancing when the allocations for SSO and UST we outside a 5% band (i.e. the portfolio was beyond a 55/45 or 45/55) achieved better performance metrics than the monthly rebalanced version with an average of only 3 rebalances per year.

Conclusion

The bond term premium does not have to be reserved for risk-averse investors. Investors desiring portfolios tilted heavily toward equities can also tap into this diversifying return stream as a factor within their portfolio.

Utilizing leveraged ETPs is one way to maintaining exposure to equities while capturing a significant portion of the bond risk premium. However, it requires more oversight than investing in other factors such as value, momentum, and quality, which are typically packaged in easy-to-access ETFs.

If a fixed frequency rebalance approach is used, tranching is an effective way to reduce timing risk, especially when markets are volatile. Aside from tranching, we find that, historically, holding periods between 3 and 6 months yield results close in line with the median rolling short-term performance of the individual strategies. Implementing a methodology like this can reduce the risk of poor luck in choosing the rebalance frequency or starting the strategy at an unfortunate time.

If frequent rebalances – like those seen with tranching – are infeasible, a dynamic schedule based on a drift in allocations is also a possibility.

Leveraged ETPs are often seen as risk trading instruments that are not fit for retail investors who are more focused on buy-and-hold systems. However, given the right risk management, these investment vehicles can be a way for investors to access the bond term premium, getting a larger free lunch, and avoiding undesired de-risking along the way.

YOUR STYLE-AGE MAY VARY

August 12, 2019

SUMMARY

- New research from Axioma suggests that tilting less – through lower target tracking error – can actually create more academically pure factor implementation in long-only portfolios.
- This research highlights an important question: how should long-only investors think about factor exposure in their portfolios? Is measuring against an academically-constructed long/short portfolio really appropriate?
- We return to the question of style versus specification, plotting year-to-date excess returns for long-only factor ETFs. While the general style serves as an anchor, we find significant specification-driven performance dispersion.
- We believe that the "right answer" to this dispersion problem largely depends upon the investor.

When quants speak about factor and style returns, we often do so with some sweeping generalizations. Typically, we're talking about some long/short specification, but precisely how that portfolio is formed can vary.

For example, one firm might look at deciles while another looks at quartiles. One shop might equal-weight the holdings while another value-weights them. Some might include mid- and small-caps, while others may work on a more realistic liquidity-screened universe.

More often than not, the precision does not matter a great deal (with the exception of liquidity-screening) because the general conclusion is the same.

But for investors who are actually realizing these returns, the precision matters quite a bit. This is particularly true for long-only investors, who have adopted smart-beta ETFs to tap into the factor research.

As we have discussed in the past, any active portfolio can be decomposed into its benchmark plus a dollar-neutral long/short portfolio that encapsulates the active bets. The active bets, then, can actually approach the true long/short implementation.

To a point, at least. The "shorts" will ultimately be constrained by the amount the portfolio can under-weight a given security.

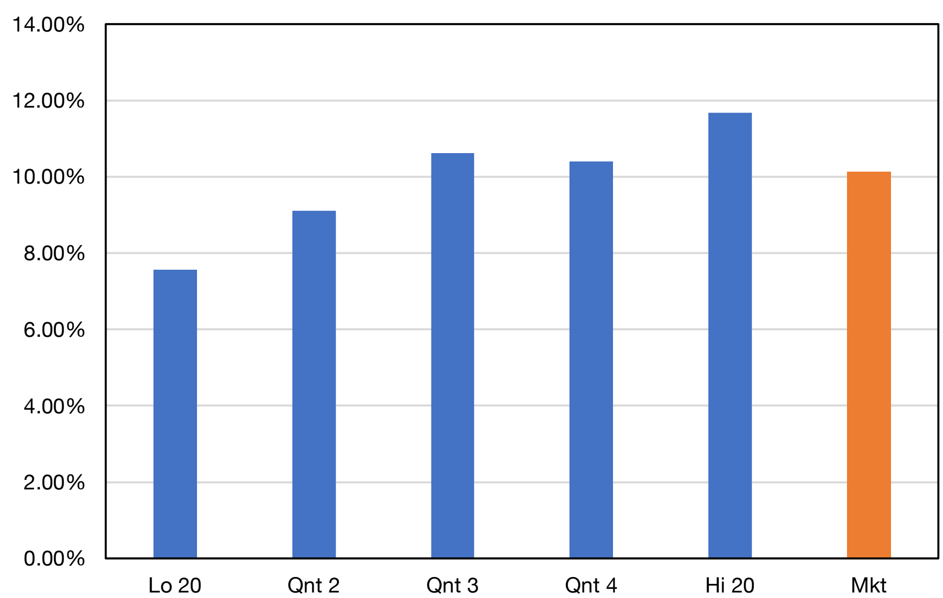
For long-only portfolios, increasing active share often means having to lean more heavily into the highest quintile or decile holdings. This is not a problem in an idealized world where factor scores have a monotonically increasing relationship with excess returns. In this perfect world, increasing our allocation to high-ranking stocks creates just as much excess return as shorting low-ranking stocks does.

Unfortunately, we do not live in a perfect world and for some factors the premium found in long/short portfolios is mostly found on the short side.[45] For example, consider the Profitability Factor. The annualized spread between the top- and bottom-quintile portfolios is 410 basis points. The difference between the top quintile portfolio and the market, though, is just 154 basis points. Nothing to scoff at, but when appropriately discounted for data-mining risk, transaction costs, and management costs, there is not necessarily a whole lot left over.

Which leads to some interesting results for portfolio construction, at least according to a recent study by Axioma.[46] For factors where the majority of the premium arises from the short side, tilting *less* might mean achieving more.

For example, Axioma found that a portfolio optimized maximize exposure to the profitability factor while targeting a tracking error to the market of just 10 basis points had a meaningfully higher correlation than the excess returns of a long-only portfolio that simply bought the top quintile. In fact, the excess returns of the top quintile portfolio had *zero correlation* to the long/short factor returns. Let's repeat that: the active returns of the top quintile portfolio had zero correlation to the returns of the profitability factor. Makes us sort of wonder what we're actually buying…

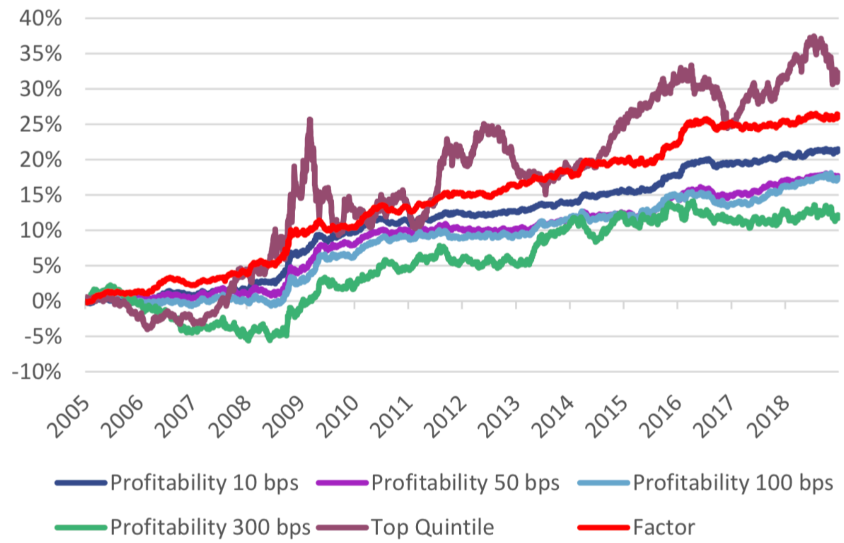
Annualized Return of Profitability-Sorted Quintiles



Source: Kenneth French Data Library; Calculations by Newfound Research.

[45] Some may even argue that the factor premia themselves really just specters in the data, existing only due to limits of arbitrage (e.g. shorting costs) that prohibit investors from actually pursuing the returns.
[46] What is a Factor? Part 2: The Impact of the Long-Only Constraint (https://go.axioma.com/rs/240-ASI-005/images/WhatIsAFactorPt2final1.pdf)

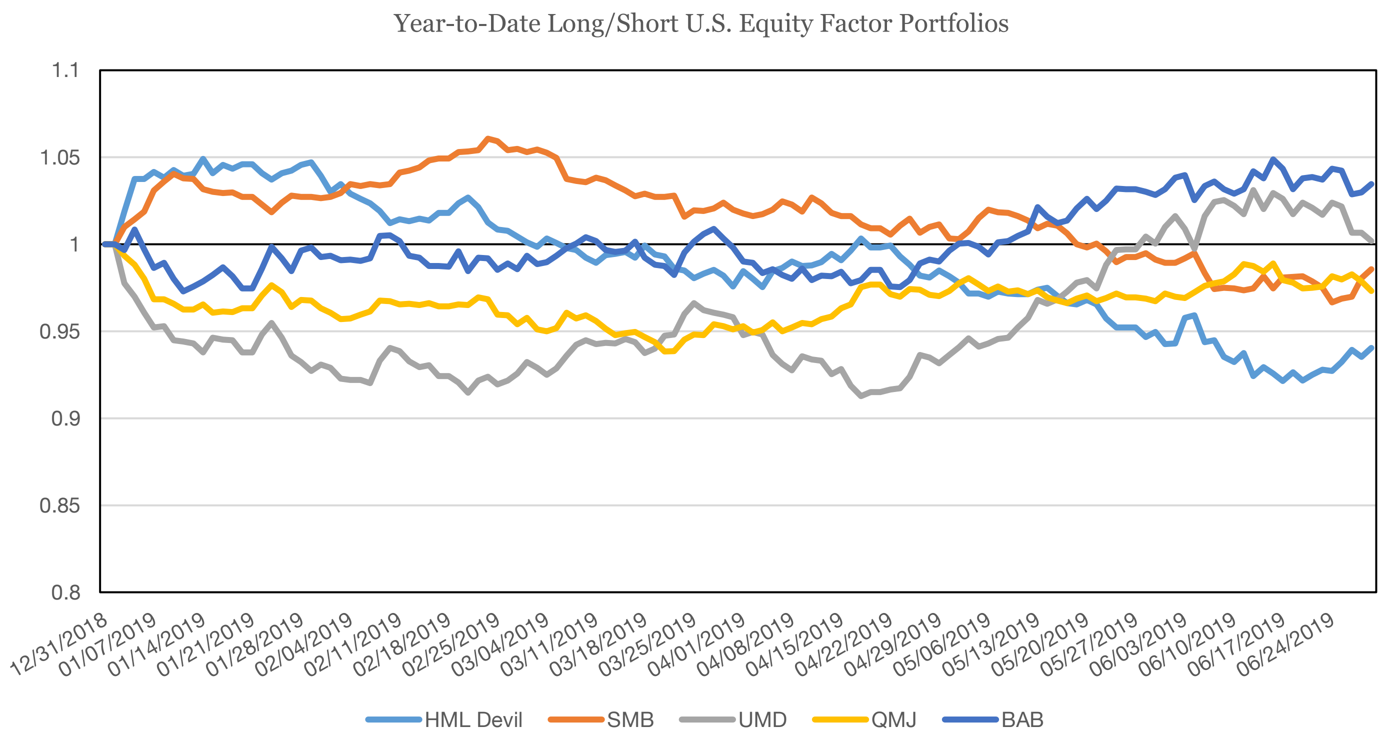
Cumulative Active Returns of Long-Only Portfolios



Note: Returns are scaled to a Profitability exposure of 1
Source: Axioma

So, what does it actually mean for long-only investors when we plot long/short equity factor returns? When we see that the Betting-Against-Beta ("BAB") factor is up 3% on the year, what does that imply for our low-volatility factor ETF? Momentum ("UMD") was down nearly 10% earlier this year; were long-only momentum ETFs really under-performing by that much?

And what does this all mean for the results in those fancy factor decomposition reports the nice consultants from the big asset management firms have been running for me over the last couple of years?

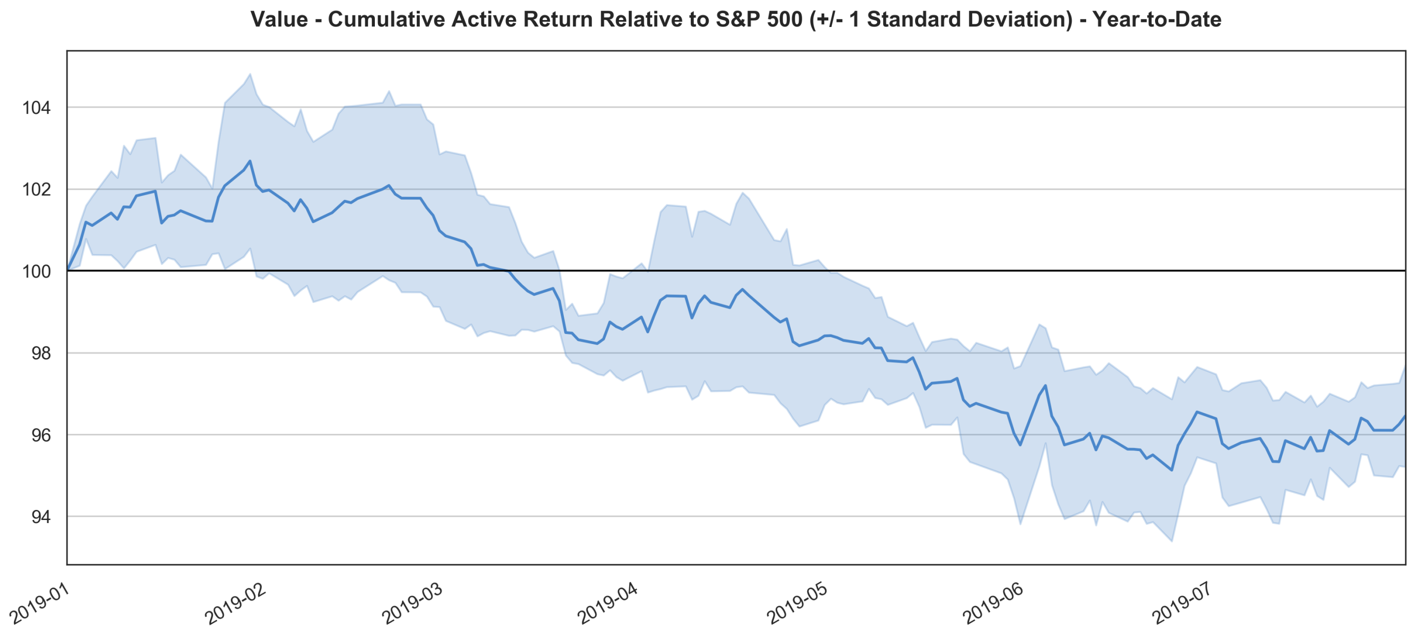


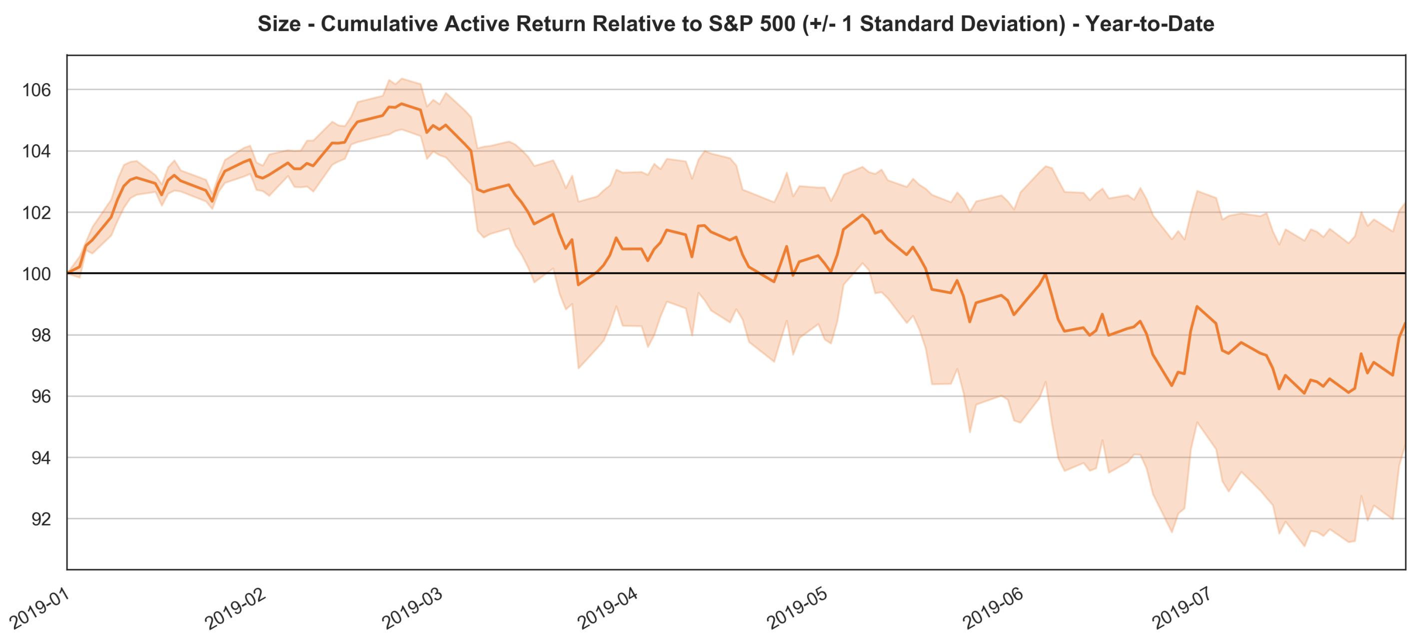
Source: AQR. Calculations by Newfound Research.

We find ourselves back to a theme we've circled many times over the last few years: *style versus specification*. Choices such as how characteristics are measured, portfolio concentration, the existence or absence of position- and industry/sector-level constraints, weighting methodology, and rebalance frequency (and even date!) can have a profound impact on realized results. The little details compound to matter quite a bit.

To highlight this disparity, below we have plotted the excess return of an equally-weighted portfolio of long-only style ETFs versus the S&P 500 as well as a standard deviation cone for individual style ETF performance.

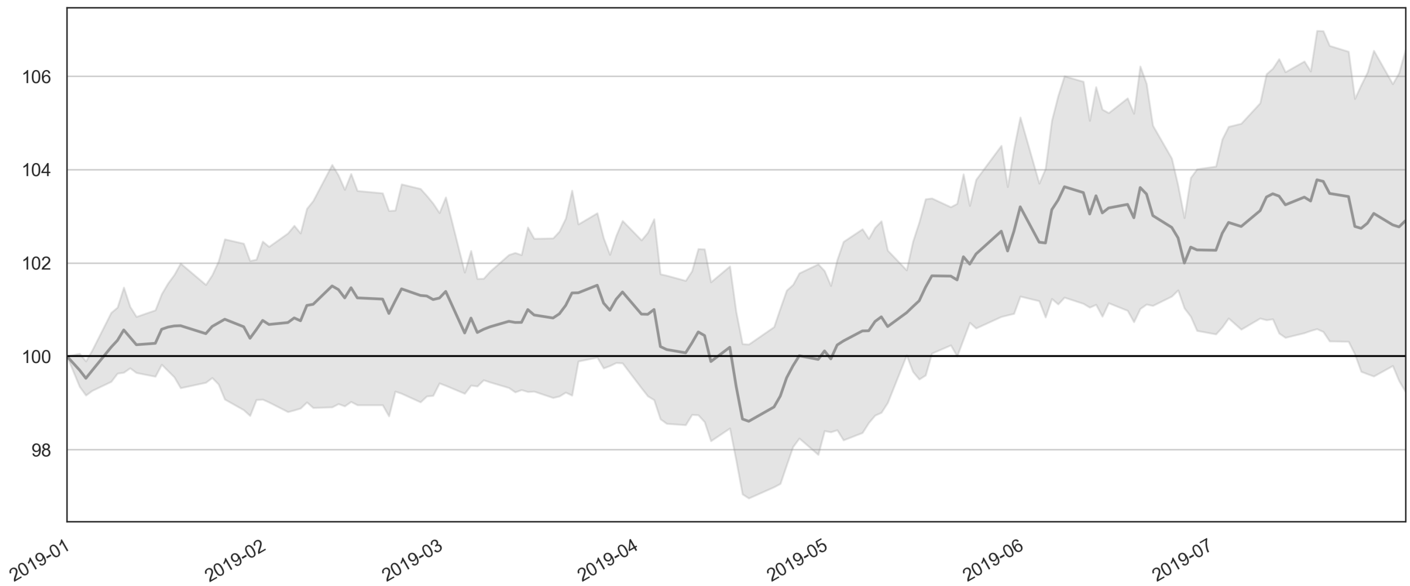
While most of the ETFs are ultimately anchored to their style, we can see that short-term performance can meaningfully deviate.


Value - Cumulative Active Return Relative to S&P 500 (+/- 1 Standard Deviation) - Year-to-Date
104
102
100
98
96
94
2019-01
2019-02
2019-03
2019-04
2019-05
2019-06
2019-07

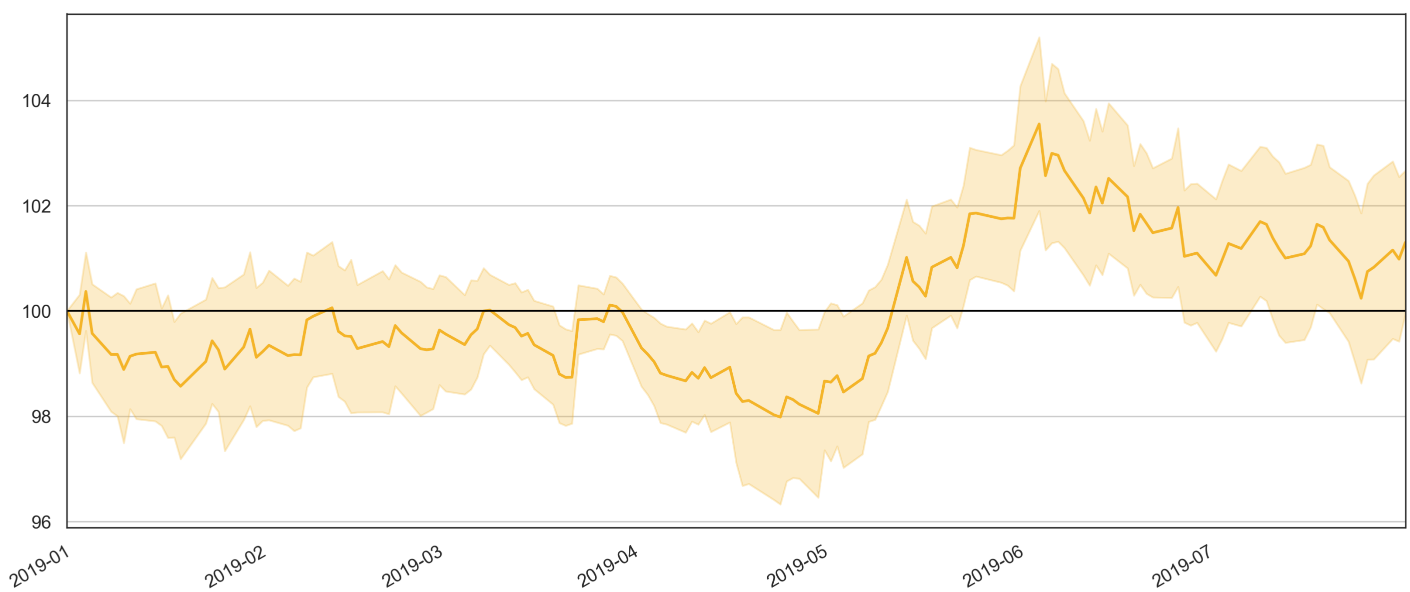

Size - Cumulative Active Return Relative to S&P 500 (+/- 1 Standard Deviation) - Year-to-Date
106
104
102
100
98
96
94
92
2019-01
2019-02
2019-03
2019-04
2019-05
2019-06
2019-07

December 30th, 2019

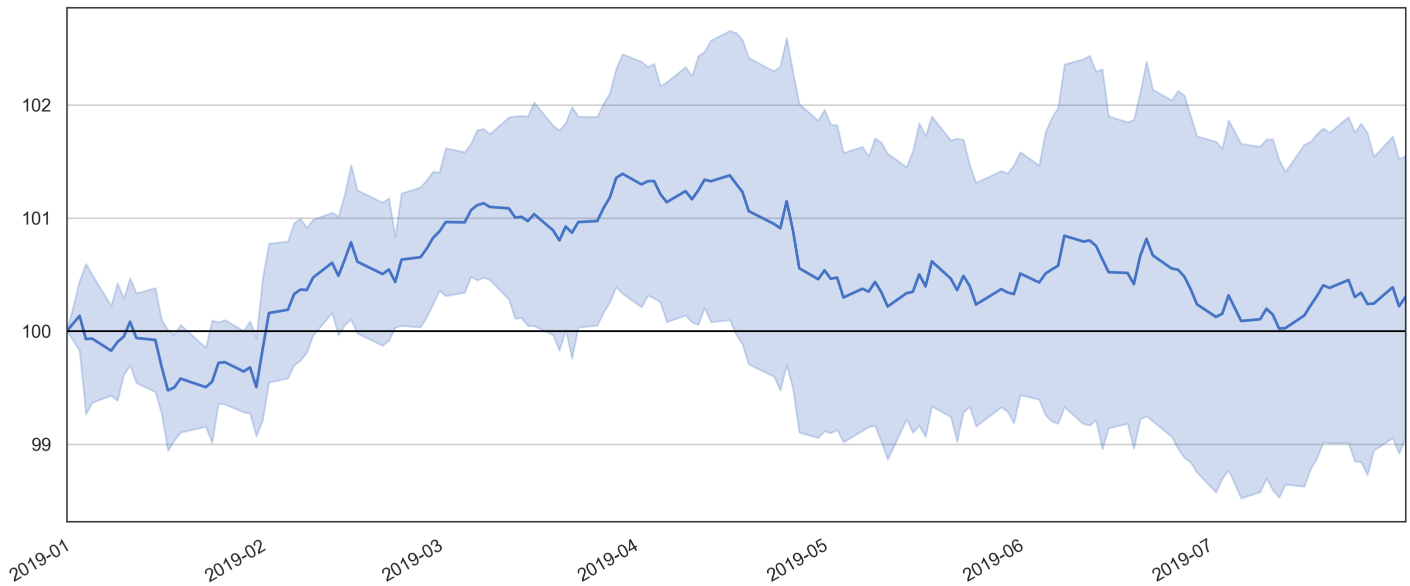
Momentum - Cumulative Active Return Relative to S&P 500 (+/- 1 Standard Deviation) - Year-to-Date



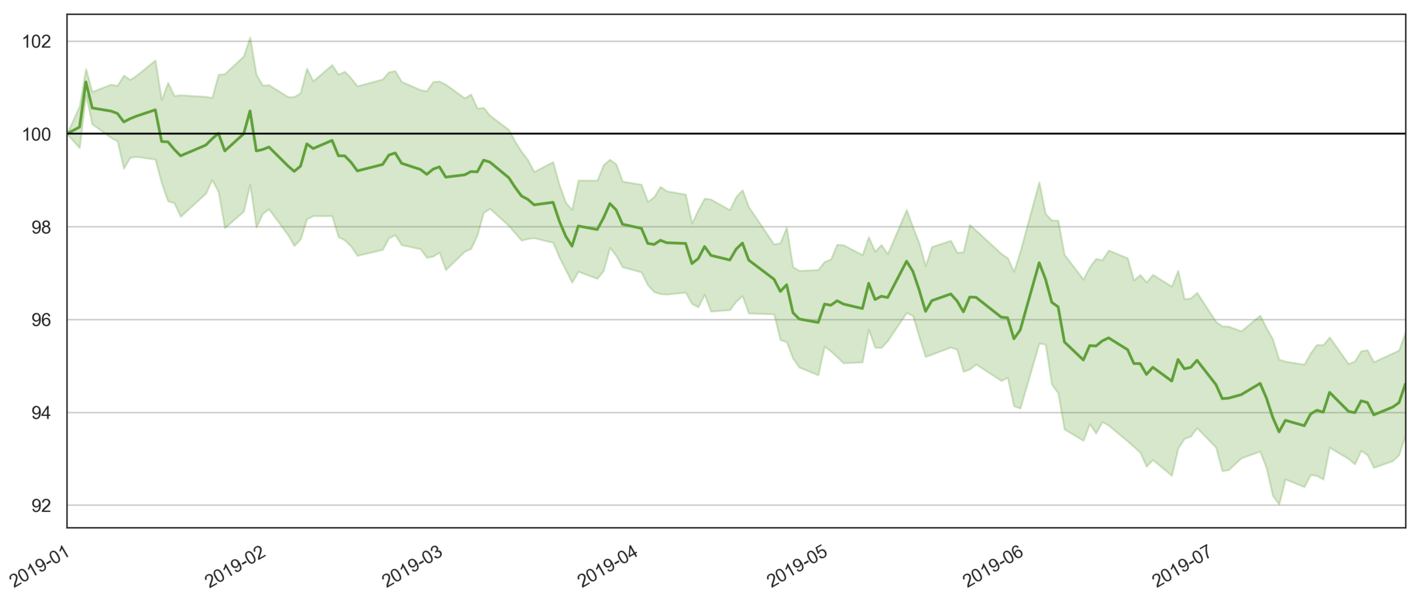
Low Volatility - Cumulative Active Return Relative to S&P 500 (+/- 1 Standard Deviation) - Year-to-Date



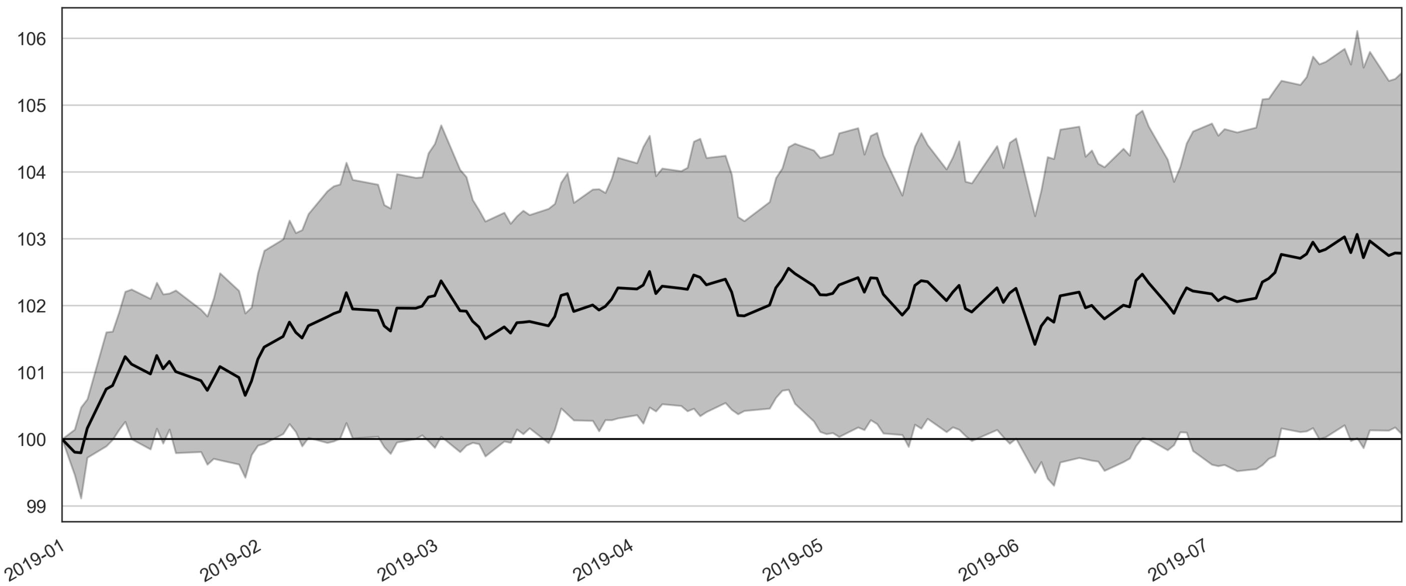
Quality - Cumulative Active Return Relative to S&P 500 (+/- 1 Standard Deviation) - Year-to-Date



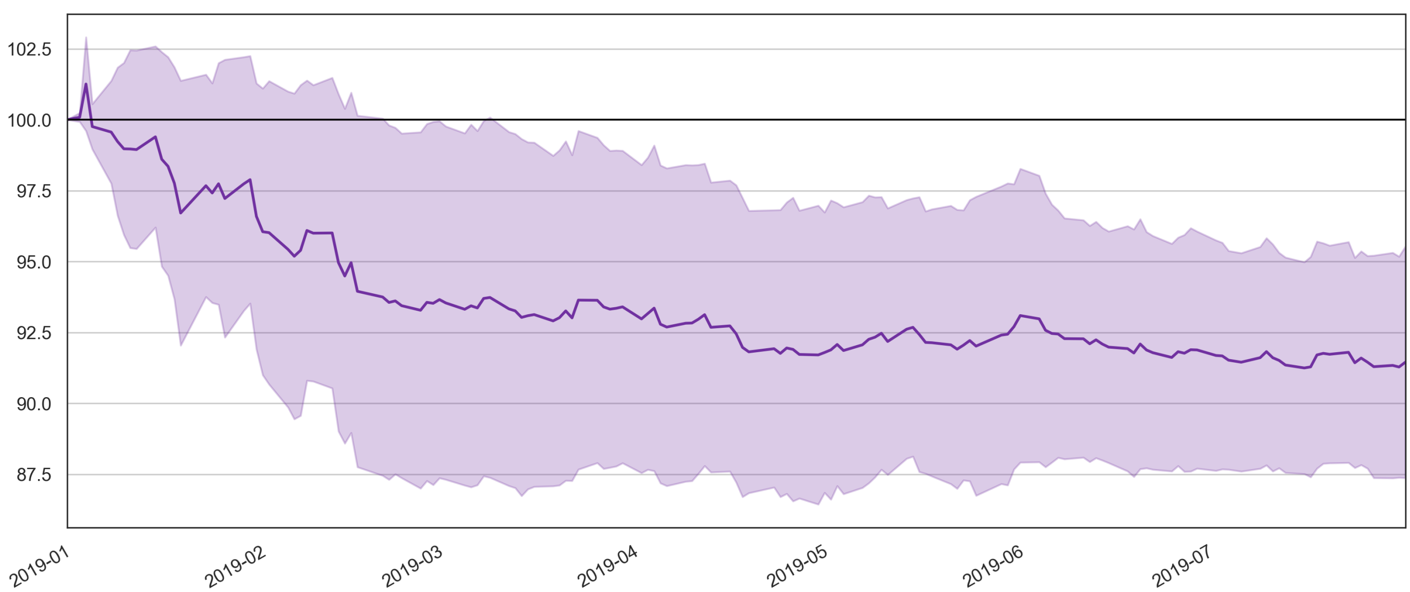
Yield - Cumulative Active Return Relative to S&P 500 (+/- 1 Standard Deviation) - Year-to-Date



Growth - Cumulative Active Return Relative to S&P 500 (+/- 1 Standard Deviation) - Year-to-Date



Trend - Cumulative Active Return Relative to S&P 500 (+/- 1 Standard Deviation) - Year-to-Date



Source: CSI Analytics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes, with the exception of underlying ETF expense ratios. Past performance is not an indicator of future results. Year-to-Date returns are computed by assuming an equal-weight allocation to representative long-only ETFs for each style. Returns are net of underlying ETF expense ratios. Returns are calculated in excess of the SPDR&P 500 ETF ("SPY"). The ETFs *used for each style are (in alphabetical order): Value: FVAL, IWD, JVAL, OVLU, QVAL, RPV, VLU, VLUE; Size: IJR, IWM, OSIZ; Momentum: FDMO, JMOM, MMTM, MTUM, OMOM, QMOM, SPMO; Low Volatility: FDLO, JMIN, LGLV, OVOL, SPLV, SPMV, USLB, USMV; Quality; FQAL, JQUA, OQAL, QUAL, SPHQ; Yield: DVY, FDVV, JDIV, OYLD, SYLD, VYM; Growth: CACG, IWF, QGRO, RPG, SCHG, SPGP, SPYG; Trend: BEMO, FVC, LFEQ, PTLC. Newfound may hold positions in any of the above securities.*

Conclusion

In our opinion, the research and data outlined in this commentary suggests a few potential courses of action for investors.

- For certain styles, we might consider embracing smaller tilts for purer factor exposure.
- To avoid specification risk, we might embrace the potential benefits of multi-manager diversification.
- Or, if there is a particular approach we prefer, simply acknowledge that it may not behave anything like the academic long/short definition – or even other long-only implementations – in the short-term.

Academically, we might be able to argue for one approach over another. Practically, the appropriate solution is whatever is most suitable for the investor and the approach that they will be able to stick with.

If a client measures their active returns with respect to academic factors, then understanding how portfolio construction choices deviate from the factor definitions will be critical.

An advisor trying to access a style but not wanting to risk choosing the wrong ETF might consider asking themselves, "why choose?" Buying a basket of a few ETFs will do wonders to reduce specification risk.

On the other hand, if an investor is simply trying to maximize their compound annualized return and nothing else, then a concentrated approach may very well be warranted.

Whatever the approach taken, it is important to remember that results between two strategies that claim to implement the same style can and will deviate significantly, especially in the short run.

Using PMI to Trade Cyclicals vs Defensives

August 19, 2019

Summary

- After stumbling across a set of old research notes from 2009 and 2012, we attempt to implement a Cyclicals versus Defensives sector trade out-of-sample.

- Post-2012 returns prove unconvincing and we find little evidence supporting the notion that PMI changes can be used for constructing this trade.

- Using data from the Kenneth French website, we extend the study to 1948, and similarly find that changes in PMI (regardless of lookback period) are not an effective signal for trading Cyclical versus Defensive sectors.

I love coming across old research because it allows for truly out-of-sample testing.

Earlier this week, I stumbled across a research note from 2009 and a follow-up note from 2012, both exploring the use of macro-based signals for constructing dollar-neutral long/short sector trades. Specifically, the pieces focused on using manufacturing Purchasing Manager Indices (PMIs) as a predictor for Cyclical versus Defensive sectors.[47]

The strategy outlined is simple: when the prior month change in manufacturing PMI is positive, the strategy is long Cyclicals and short Defensives; when the change is negative, the strategy is long Defensives and short Cyclicals. The intuition behind this signal is that PMIs provide a guide to hard economic activity.

The sample period for the initial test is from 1998 to 2009, a period over which the strategy performed quite well on a global basis and even better when using the more forward-looking ratio of new orders to inventory.

Red flags start to go up, however, when we read the second note from 2012. "It appears that the new orders-to-inventory ratio has lost its ability to forecast the output index." "In addition, the optimal lookback period … has shifted from one to two months."

At this point, we can believe one of a few things:

- The initial strategy works, has simply hit a rough patch in the three years after publishing, and will work again in the future.
- The initial strategy worked but has broken since publishing.

[47] Where Cyclicals were defined as an equal-weighted basket of Industrials, Materials, IT and Consumer Discretionary equities and Defensives were defined as an equal-weight basket of Utilities, Telecom, Consumer Staples, and Health Care equities, with Financials and Energy being neither Cyclicals nor Defensives.

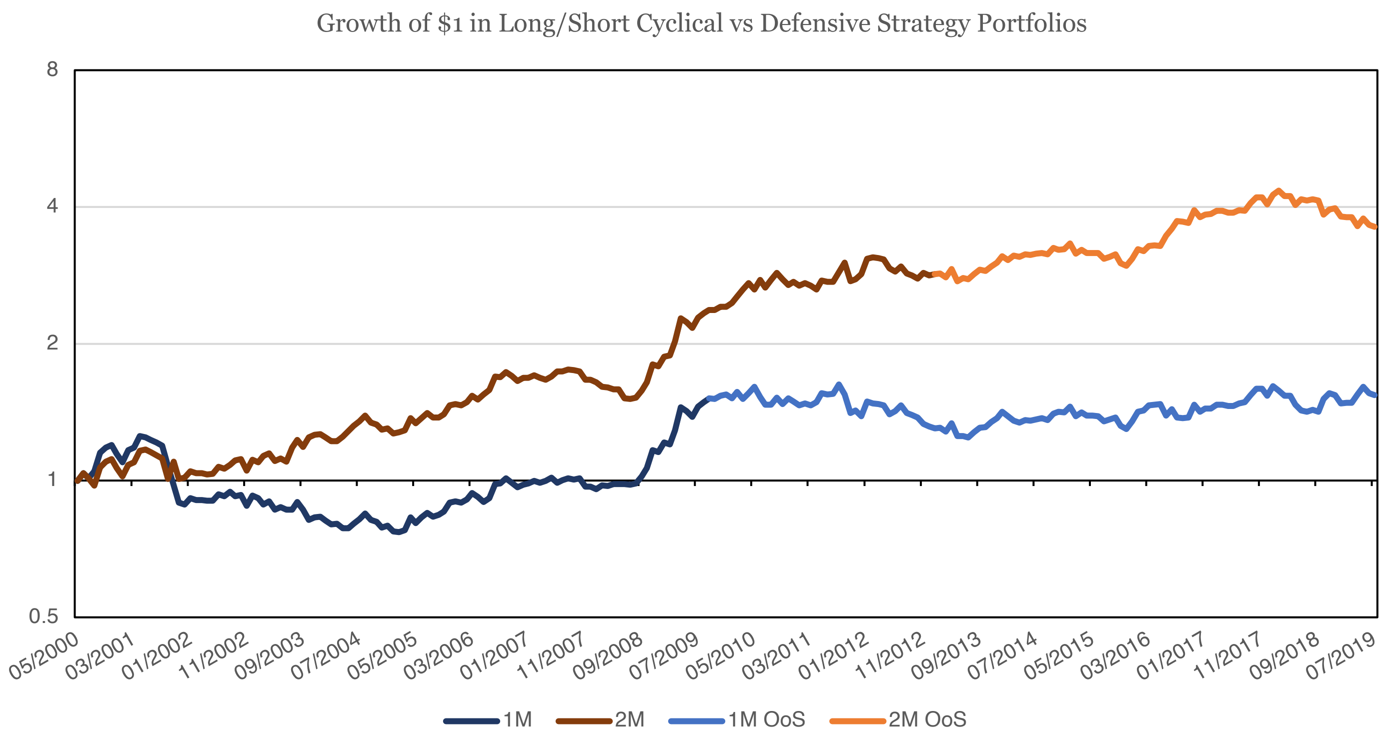
- The initial strategy never worked and was an artifact of datamining.

I won't even bother addressing the whole "one-month versus two-month" comment. Long-time readers know where we come down on ensembles versus parameter specification…

Fortunately, we do not have to pass qualitative judgement: we can let the numbers speak for themselves.

While the initial notes focused on global implementation, we can rebuild the strategy using U.S. equity sectors and US manufacturing PMI as the driving signal. This will serve both as an out-of-sample test for assets, as well as provide approximately 7 more years of out-of-sample time to evaluate returns.

Below we plot the results of this strategy for both 1-month and 2-month lookback periods, highlighting the in-sample and out-of-sample periods for each specification based upon the date the original research notes were published. We use the State Street SPDR Sector Select ETFs as our implementation vehicles, with the exception of the iShares Dow Jones US Telecom ETF.



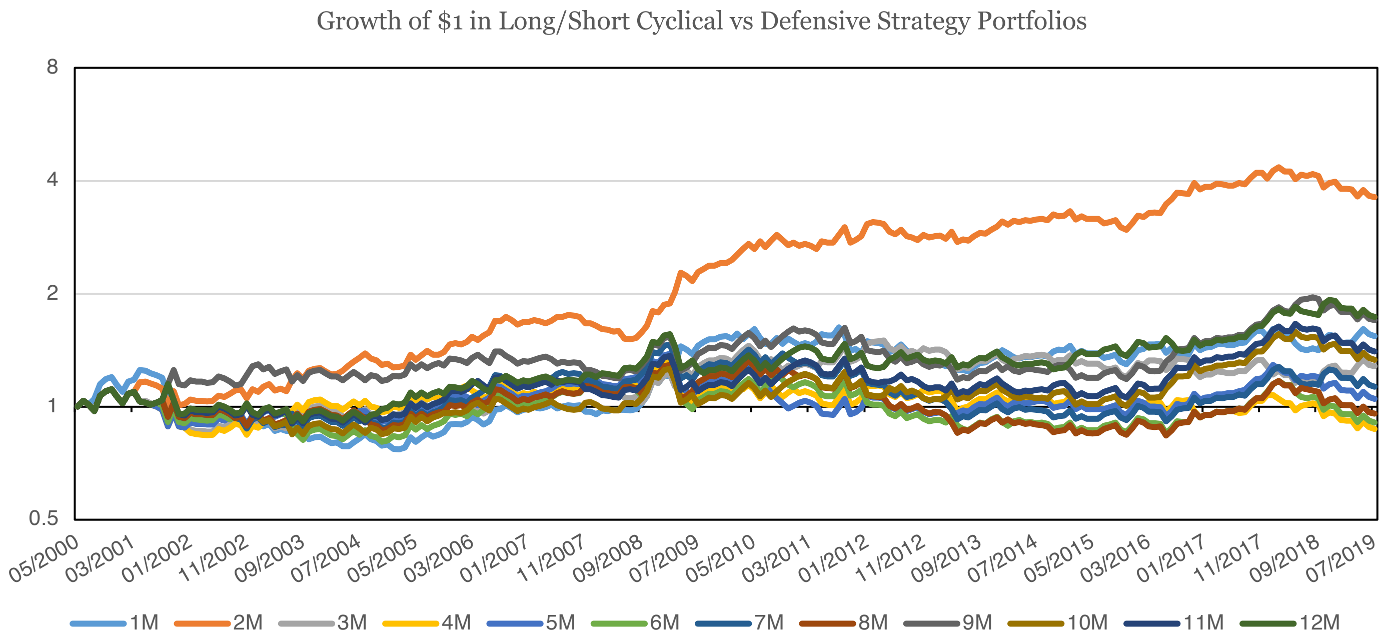
Source: CSI Data; Quandl. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

The first thing we notice is that the original 1-month implementation – which appeared to work on a global scale – does not seem particularly robust when implemented with U.S. sectors. Post publish date results do not fare much better.

The 2-month specification, however, does appear to work reasonably well both in- and out-of-sample.

But is there something inherently magical about that two-month specification? We are hard-pressed to find a narrative explanation.

If we plot lookback specifications from 3- to 12-months, we see that the 2-month specification proves to be a significant outlier. Given the high correlation between all the other specifications, it is more likely that the 2-month lookback was the beneficiary of luck rather than capturing a special particular edge.



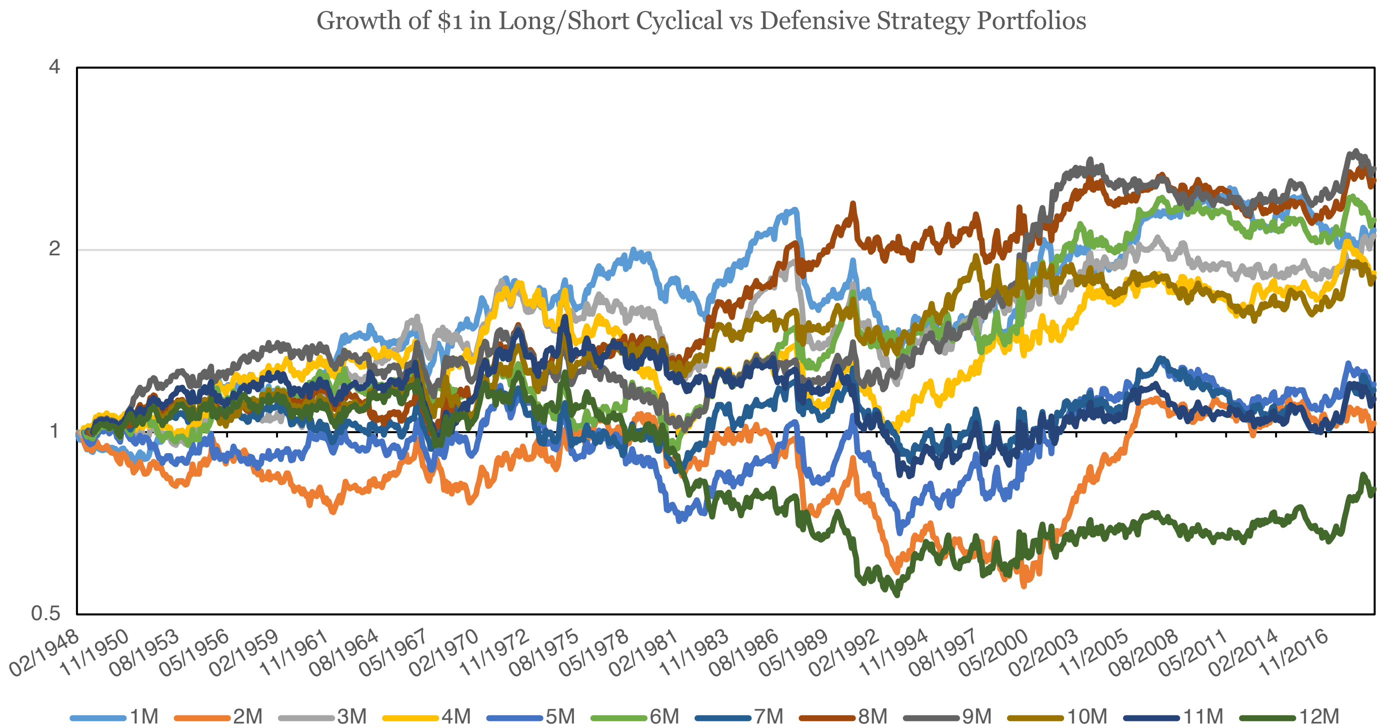
Source: CSI Data; Quandl. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

Perhaps we're not giving this idea enough breathing room. After all, were we to evaluate most value strategies in the most recent decades, we'd likely declare them insignificant as well.

With manufacturing PMI data extending back to the 1948, we can use sector index data from the Kenneth French website to reconstruct this strategy.

Unfortunately, the Kenneth French definitions do not match GICs perfectly, so we have to change the definition of Cyclicals and Defensives slightly. Using the Kenneth French data, we will define Cyclicals to be Manufacturing, Non-Durables, Technology, and Shops. Defensives are defined to be Durables, Telecom, Health Care, and Utilities.

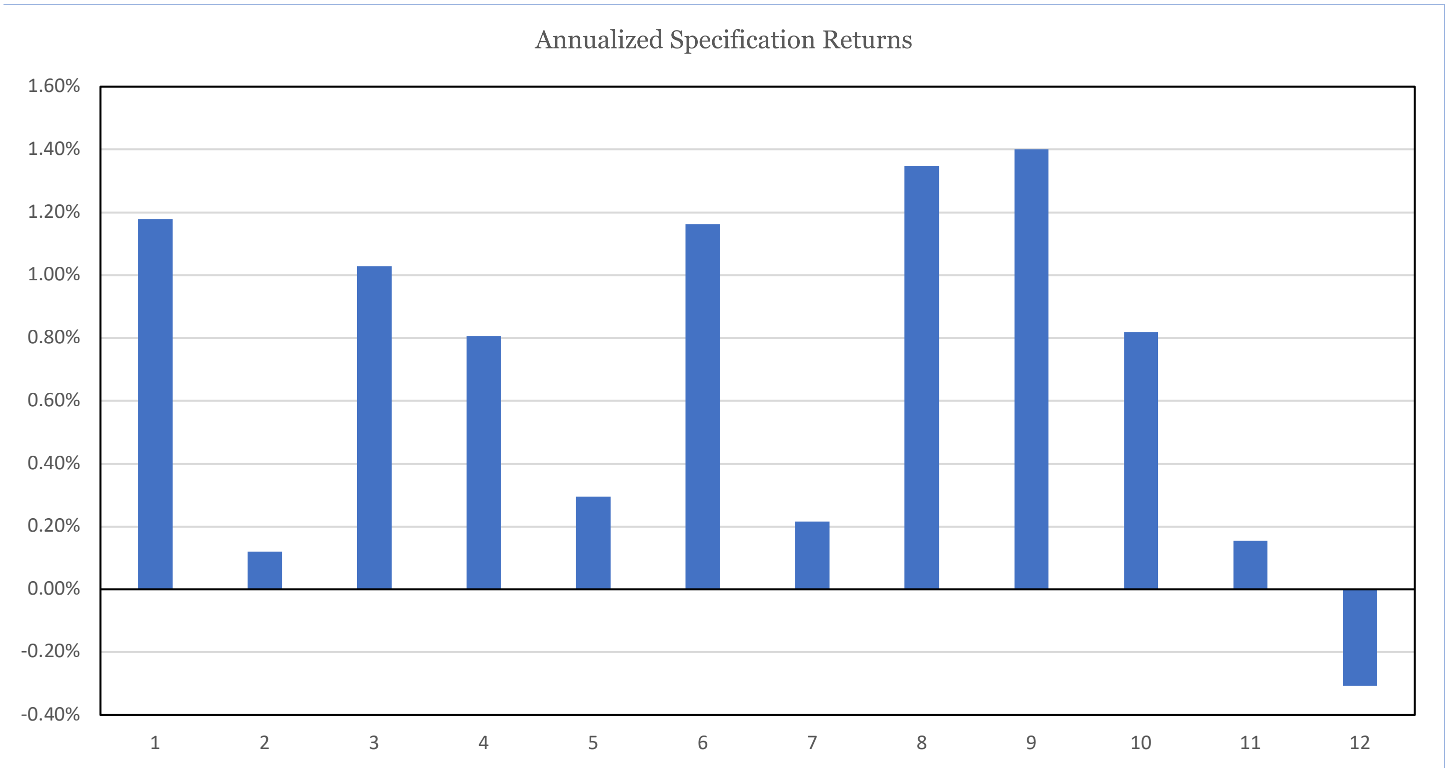
We use the same strategy as before, going long Cyclicals and short Defensives when changes in PMI are positive, and short Cyclicals and long Defensives when changes to PMI are negative. We again vary the lookback period from 1- to 12-months.



Source: Kenneth French Data Library; Quandl. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

The results are less than convincing. Not only do we see significant dispersion across implementations, but there is also no consistency in those implementations that do well versus those that do not.

Perhaps worse, the best performing variation only returned a paltry 1.40% annualized *gross* of any implementation costs. Once we start accounting for transaction costs, slippage, and management fees, this figure deflates towards zero rather quickly.



Source: Kenneth French Data Library; Quandl. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

Conclusion

There is no shortage of quantitative research in the market and the research can be particularly compelling when it seems to fit a pre-existing narrative.

Cyclicals versus Defensives are a perfect example. Their very names imply the regimes during which they are supposed to add value, but actually translating this notion into a robust strategy proves to be less than easy.

I would make the philosophical argument that it quite simply cannot be easy. Consider the two pieces of information we need to believe for this strategy to work:

- Cyclicals outperform Defensives in an economic expansion and Defensives outperform Cyclicals in an economic contraction.
- We can forecast economic expansions and contractions *before* it is priced into the market.

If we have very high confidence in both statements, it effectively implies an arbitrage.

Therefore, if we have very high confidence in the truth of the first statement, then for markets to be reasonably efficient, we must have little confidence in the second statement.

Similarly, if we have high confidence in the trust of the second statement, then for markets to be reasonably efficient, we must have little confidence in the first statement.

Thus, a more reasonable expectation might be that Cyclicals *tend* to outperform Defensives during an expansion, and Defensives *tend* to outperform Cyclicals in a contraction, but there may be meaningful exceptions depending upon the particular cycle.

Furthermore, we may believe we have an edge in forecasting expansions and contractions (perhaps not with just PMI, though), but there will be many false positives and false negatives along the way.

Taken together, we might believe we can construct such a strategy, but errors in both assumptions will lead to periods of frustration. However, we should recognize that for such an "open secret" strategy to work in the long run, there have to be troughs of sorrow deep enough to avoid permanent crowding.

In this case, we believe there is little evidence to suggest that level changes in PMI provide particular insight into Cyclicals versus Defensives, but that does not mean there are no macro signals that might.

ES-CAPE VELOCITY: VALUE-DRIVEN SECTOR ROTATION

August 26, 2019

SUMMARY

- Systematic value strategies have struggled in the post-2008 environment, so one that has performed well catches our eye.

- The Barclays Shiller CAPE sector rotation strategy – a value-based sector rotation strategy – has out-performed the S&P 500 by 267 basis points annualized since it launched in 2012.

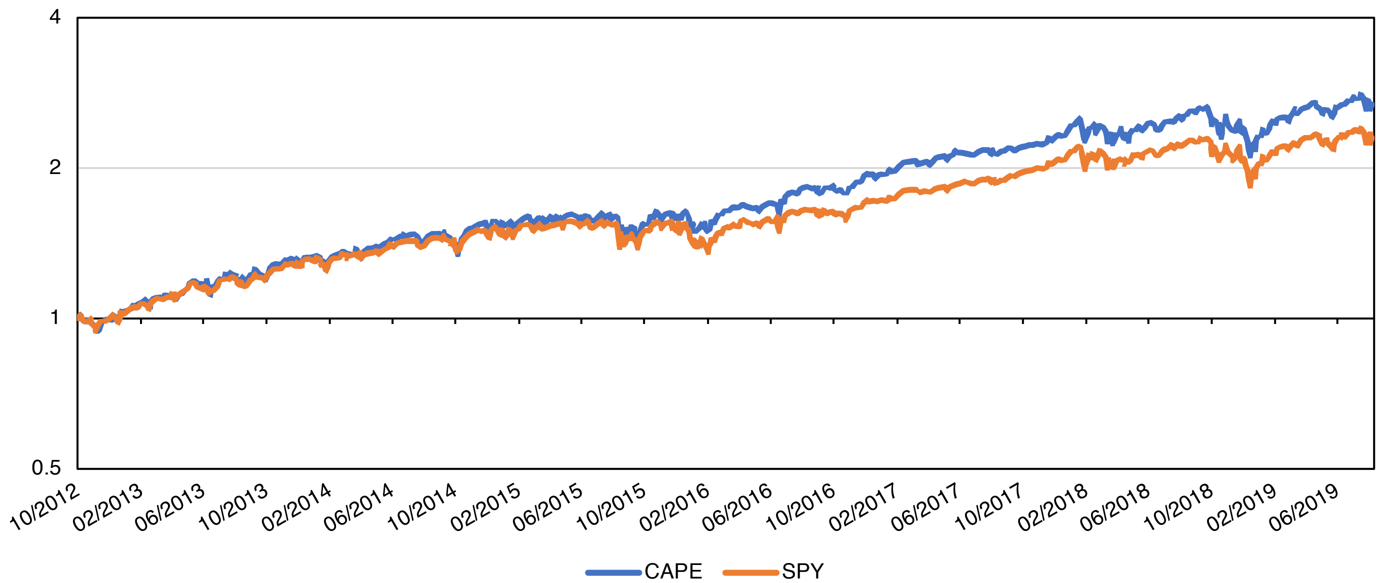
- The strategy applies a unique Relative CAPE metric to account for structural differences in sector valuations as well as a momentum filter that seeks to avoid "value traps."

- In an effort to derive the source of out-performance, we explore various other valuation metrics and model specifications.

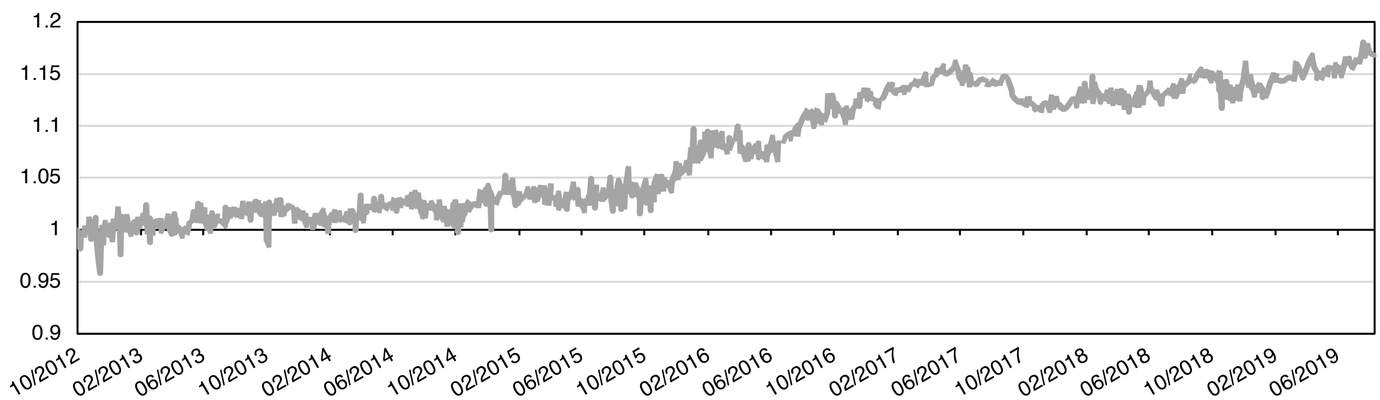
- We find that what has actually driven performance in the past may have little to do with value at all.

It is no secret that systematic value investing of all sorts has struggled as of late. With the curious exception, that is, of the Barclays Shiller CAPE sector rotation strategy, a strategy explored by Bunn, Staal, Zhuang, Lazanas, Ural and Shiller in their 2014 paper *Es-cape-ing from Overvalued Sectors: Sector Selection Based on the Cyclically Adjusted Price-Earnings (CAPE) Ratio.* Initial performance suggests that the idea has performed quite well out-of-sample, which stands out among many "smart-beta" strategies which have failed to live up to their backtests.

Barclays Shiller CAPE Index ETN ("CAPE") and SSgA S&P 500 ETF ("SPY")



Ratio of CAPE / SPY



Source: CSI Data. Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

Why is this strategy finding success where other value strategies have not? That is what we aim to explore in this commentary.

On a monthly basis, the Shiller CAPE sector rotation portfolio is rebalanced into an equal-weight allocation across four of the ten primary GICS sectors. The four are selected first by ranking the 10 primary sectors based upon their Relative

CAPE ratios and choosing the cheapest five sectors. Of those cheapest five sectors, the sector with the worst trailing 12-month return ("momentum") is removed.

The CAPE ratio – standing for Cyclically-Adjusted Price-to-Earnings ratio – is the current price divided by the 10-year moving average of inflation-adjusted earnings. The purpose of this smoothing is to reduce the impact of business cycle fluctuations.

The potential problem with using the raw CAPE value for each sector is that certain sectors have structurally higher and lower CAPE ratios than their peers. High growth sectors – e.g. Technology – tend to have higher CAPE ratios because they reinvest a substantial portion of their earnings while more stable sectors – e.g. Utilities – tend to have much lower CAPE ratios. Were we to simply sort sectors based upon their current CAPE ratio, we would tend to create structural over- and under-weights towards certain sectors.

To adjust for this structural difference, the strategy uses the idea of a Relative CAPE ratio, which is calculated by taking the current CAPE ratio and dividing it by a rolling 20-year average CAPE ratio[48] for that sector. The thesis behind this step is that dividing by a long-term mean normalizes the sectors and allows for better comparison. Relative CAPE values above 1 mean that the sector is more expensive than it has historically been, while values less than 1 mean it is cheaper.

It is important to note here that the actual selection is still performed on a cross-sector basis. It is entirely possible that all the sectors appear cheap or expensive on a historical basis at the same time. The portfolio will simply pick the cheapest sectors available.

Poking and Prodding the Parameters

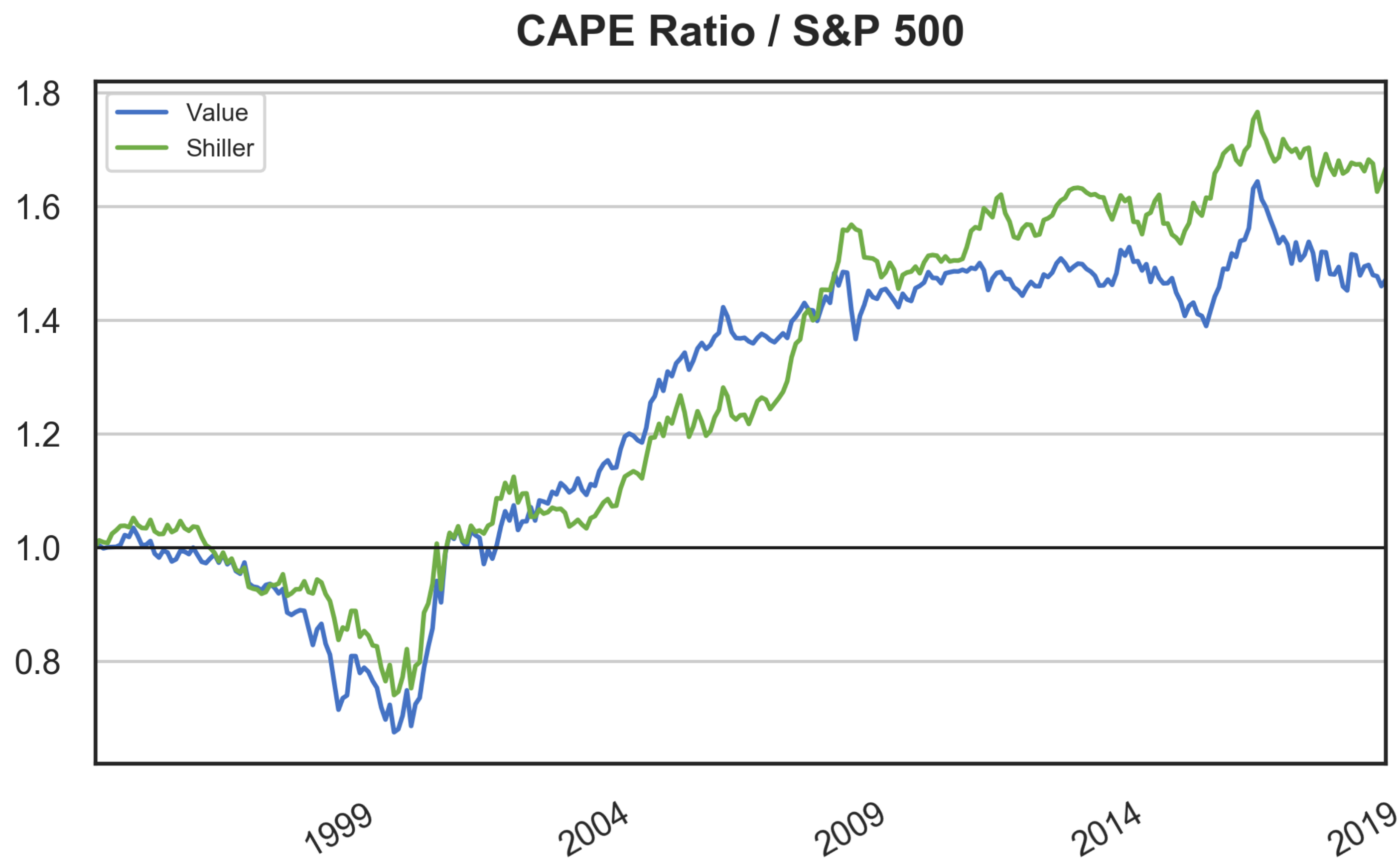
With an understanding of the rules, our first step is to poke and prod a bit to figure out what is really driving the strategy.

We begin by first exploring the impact of using the Relative CAPE ratio versus just the CAPE ratio.

For each of these ratios, we'll plot two strategies. The first is a naïve Value strategy, which will equally-weight the four cheapest sectors. The second is the Shiller strategy, which chooses the top five cheapest sectors and drops the one with the worst momentum. This should provide a baseline for comparing the impact of the momentum filter.

Strategy returns are plotted relative to the S&P 500.

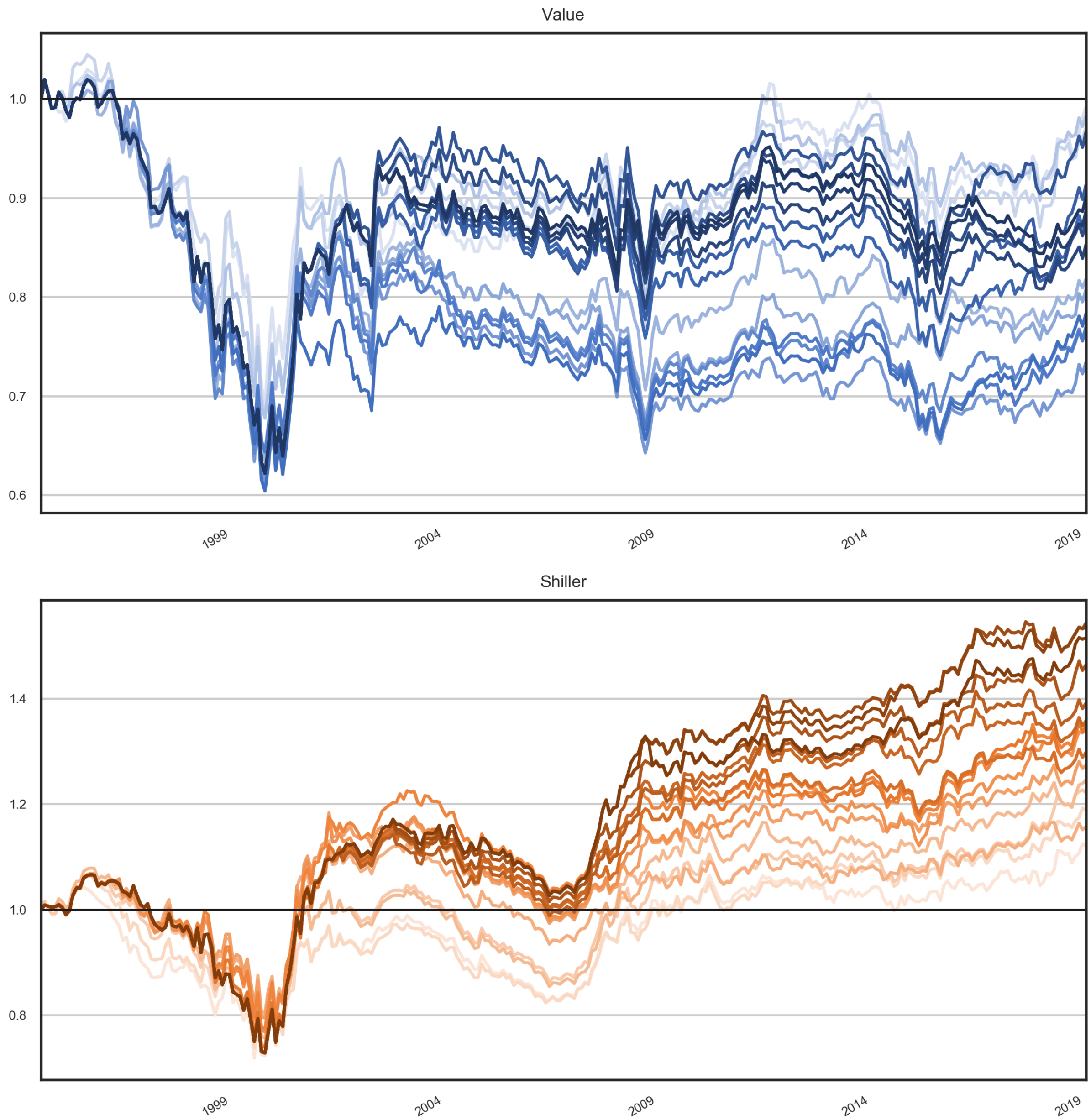
[48] Winsorized at the 5th percentile



Source: Siblis Research; Morningstar; CS Data. Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

For the Relative CAPE ratio, we also vary the lookback period for calculating the rolling average CAPE from 5- to 20-years.

Relative CAPE Ratio / S&P 500



Source: Siblis Research; Morningstar; CSI Data. Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

A few things immediately stand out:

- Interestingly, standard CAPE actually appears to perform better than Relative CAPE for both the traditional value and Shiller implementations.

- The Relative CAPE approach fared much more poorly from 2004-2007 than the simple CAPE approach.

- There is little difference in performance for the Value and Shiller strategy for standard CAPE, but a meaningful difference for Relative CAPE.

- While standard CAPE value has stagnant relative performance since 2007, Relative CAPE appears to continue to work for the Shiller approach.

- A naïve value implementation seems to perform quite poorly for Relative CAPE, while the Shiller strategy appears to perform rather well.

- There is meaningful performance dispersion based upon the lookback period, with longer-dated lookbacks (darker shades) appearing to perform better than shorter-period lookbacks (lighter shades) for the Relative CAPE variation.

The second-to-last point is particularly curious, as it implies that using momentum to "avoid the value trap" creates significant value (no pun intended; okay, pun intended) for the strategy.

Varying the Value Metric (in Vain)

To gain more insight, we next test the impact of the choice of the CAPE ratio. Below we plot the relative returns of different Shiller-based strategies (again varying lookbacks from 5- to 20-years), but use price-to-book, trailing 12-month price-to-earnings, and trailing 12-month EV/EBITDA as our value metrics.

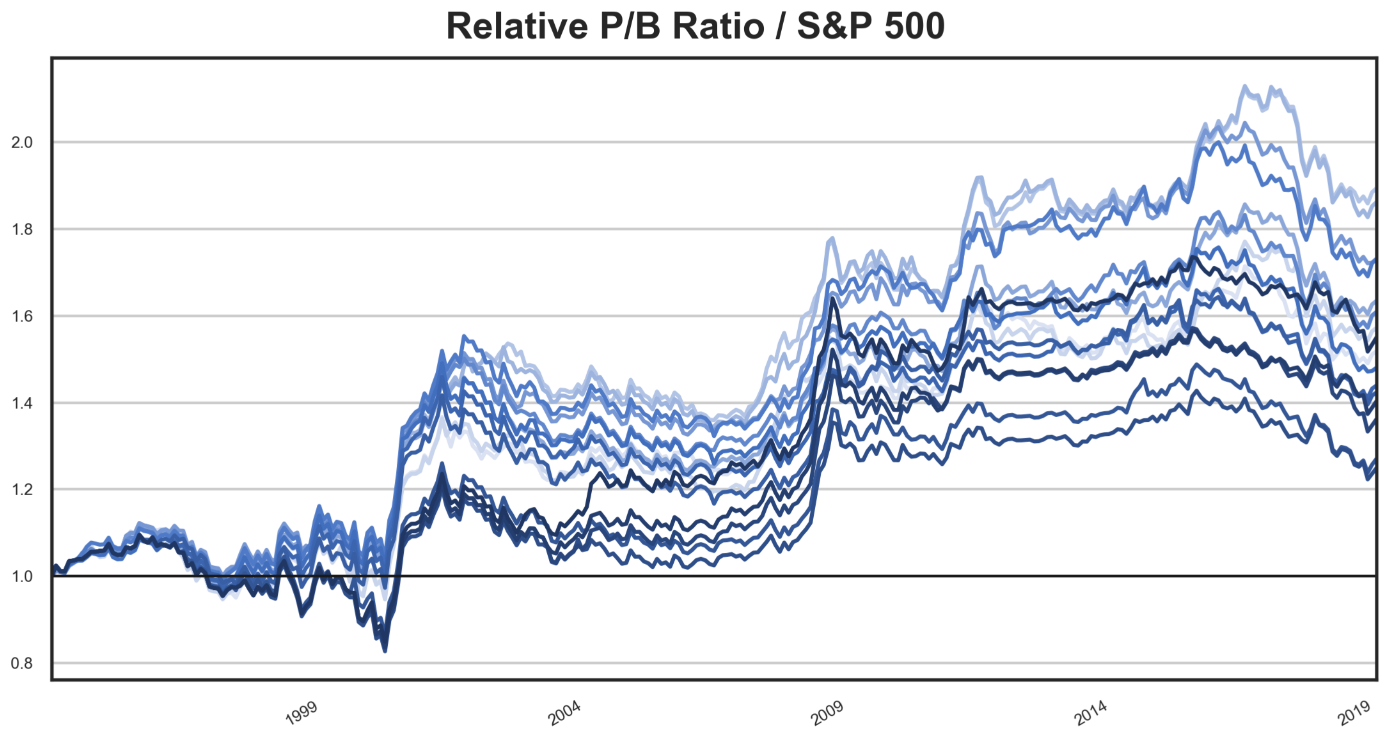
A few things stand out:

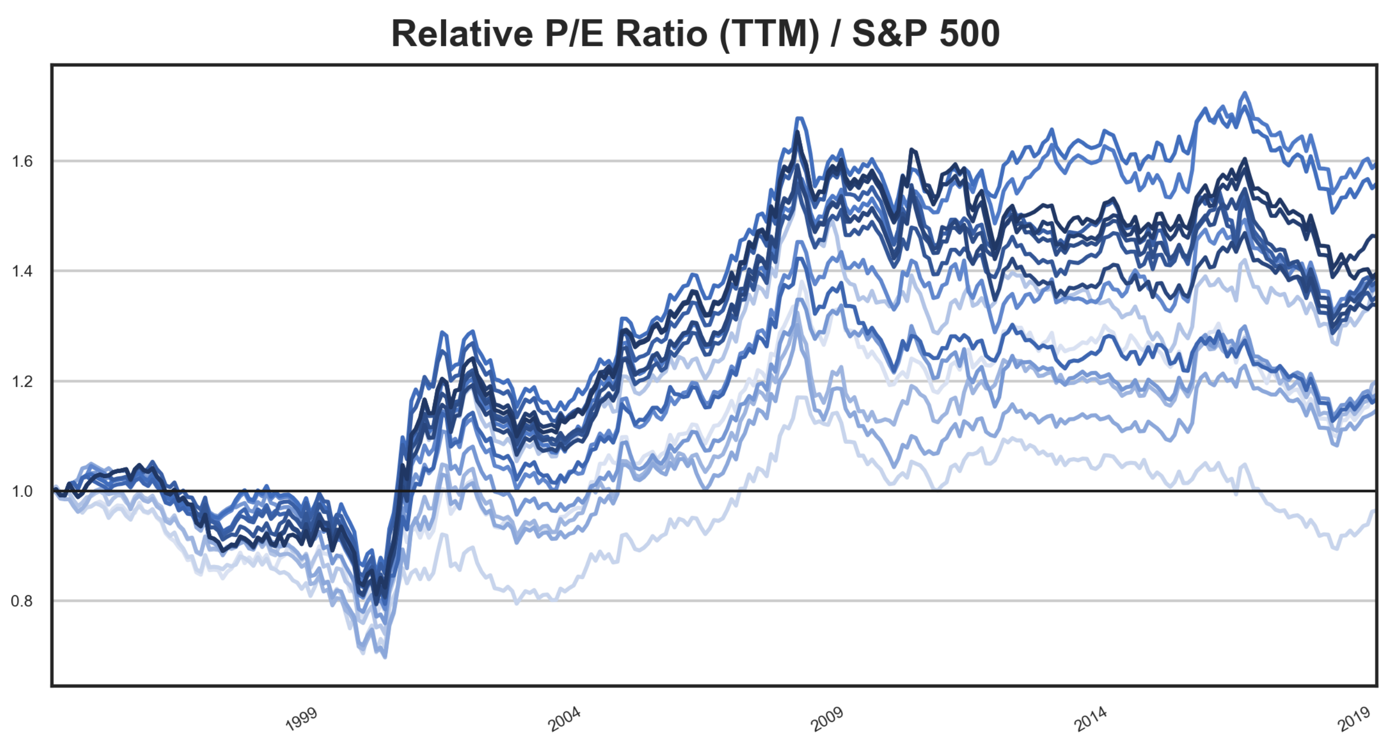
- Value-based sector rotation seems to have "worked" from 2000 to 2009, regardless of our metric of choice.

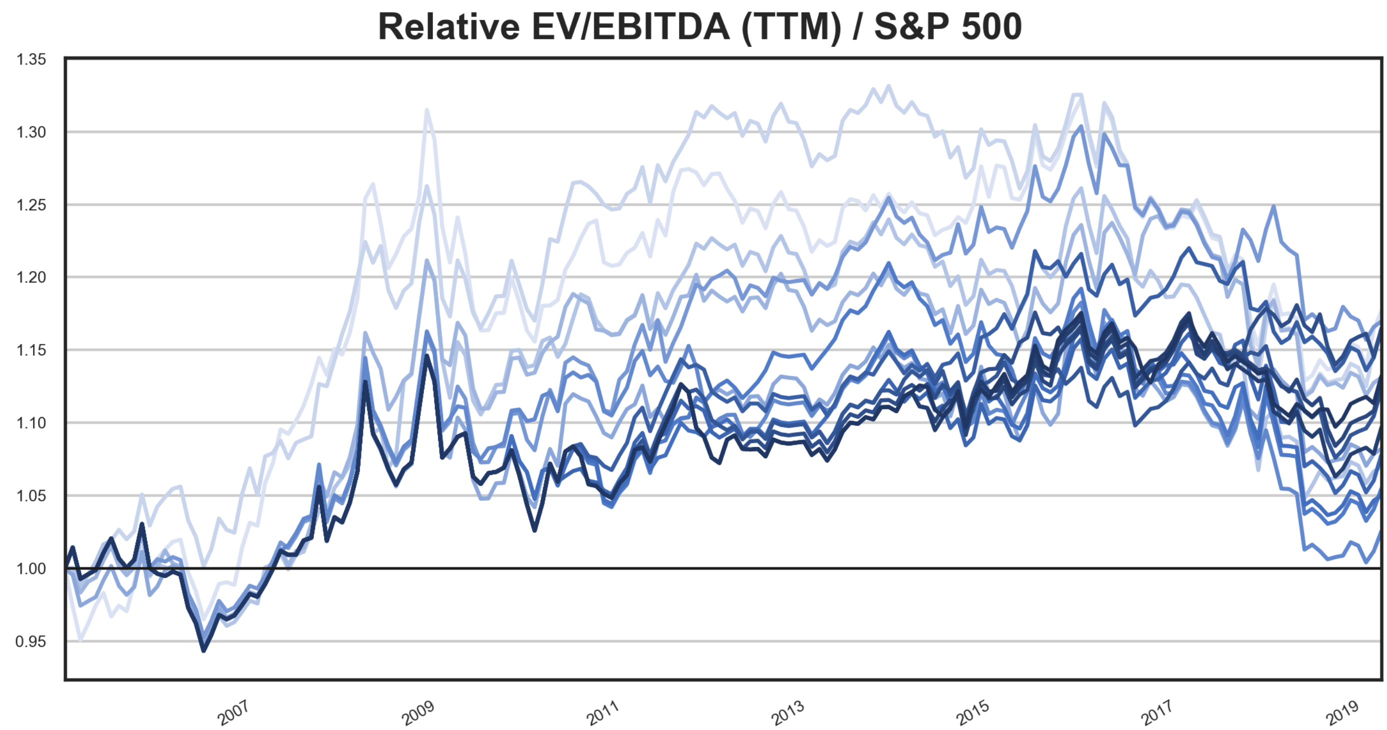
- Almost all value-based strategies appear to exhibit significant relative out-performance during the dot-com and 2008 recessions.

- After 2009, most value strategies appear to exhibit random relative performance versus the S&P 500.

- All three approaches appear to suffer since 2016.


Relative P/B Ratio / S&P 500
2.0
1.8
1.6
1.4
1.2
1.0
0.8
1999
2004
2009
2014
2019


Relative P/E Ratio (TTM) / S&P 500
1.6
1.4
1.2
1.0
0.8
1999
2004
2009
2014
2019



Source: Siblis Research; Morningstar; CSI Data. Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

At this point, we have to ask: is there something special about the Relative CAPE that makes it inherently superior to other metrics?

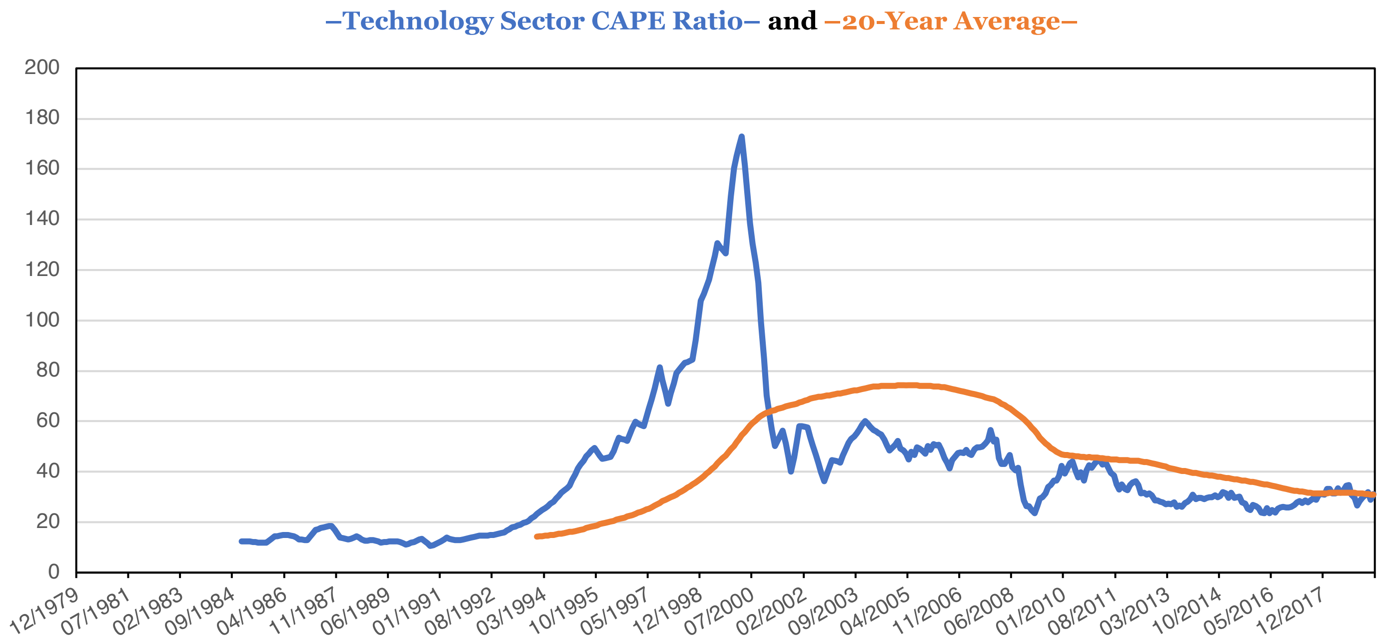
A Big Bubble-Based Bet?

If we take a step back for a moment, it is worth asking ourselves a simple question: what would it take for a sector rotation strategy to out-perform the S&P 500 over the last decade?

With the benefit of hindsight, we know Consumer Discretionary and Technology have led the pack, while traditionally stodgy sectors like Consumer Staples and Utilities have lagged behind (though not nearly as poorly as Energy).

As we mentioned earlier, a naïve rank on the CAPE ratio would almost certainly prefer Utilities and Staples over Technology and Discretionary. Thus, for us to outperform the market, we must somehow construct a value metric that identifies the two most chronically expensive sectors (ignoring back-dated valuations for the new Communication Services sector) as being among the cheapest.

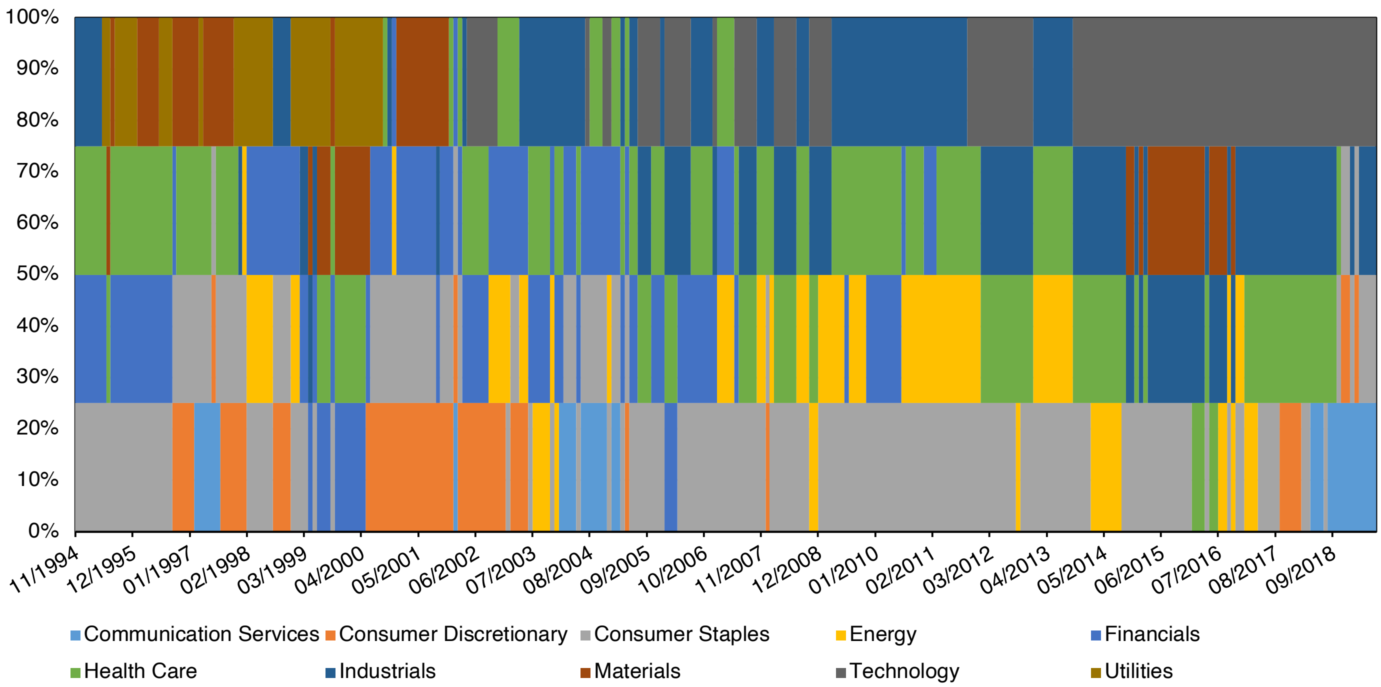
This is where dividing by the rolling 20-year average comes into play. In spirit, it makes a certain degree of sense. In practice, however, this plays out perfectly for Technology, which went through such an enormous bubble in the late 1990s that the 20-year average was meaningfully skewed upward by an outlier event. Thus, for almost the entire 20-year period *after* the dot-com bubble, Technology appears to be relatively cheap by comparison. After all, you can buy for 30x earnings today what you used to be able to buy for 180x!



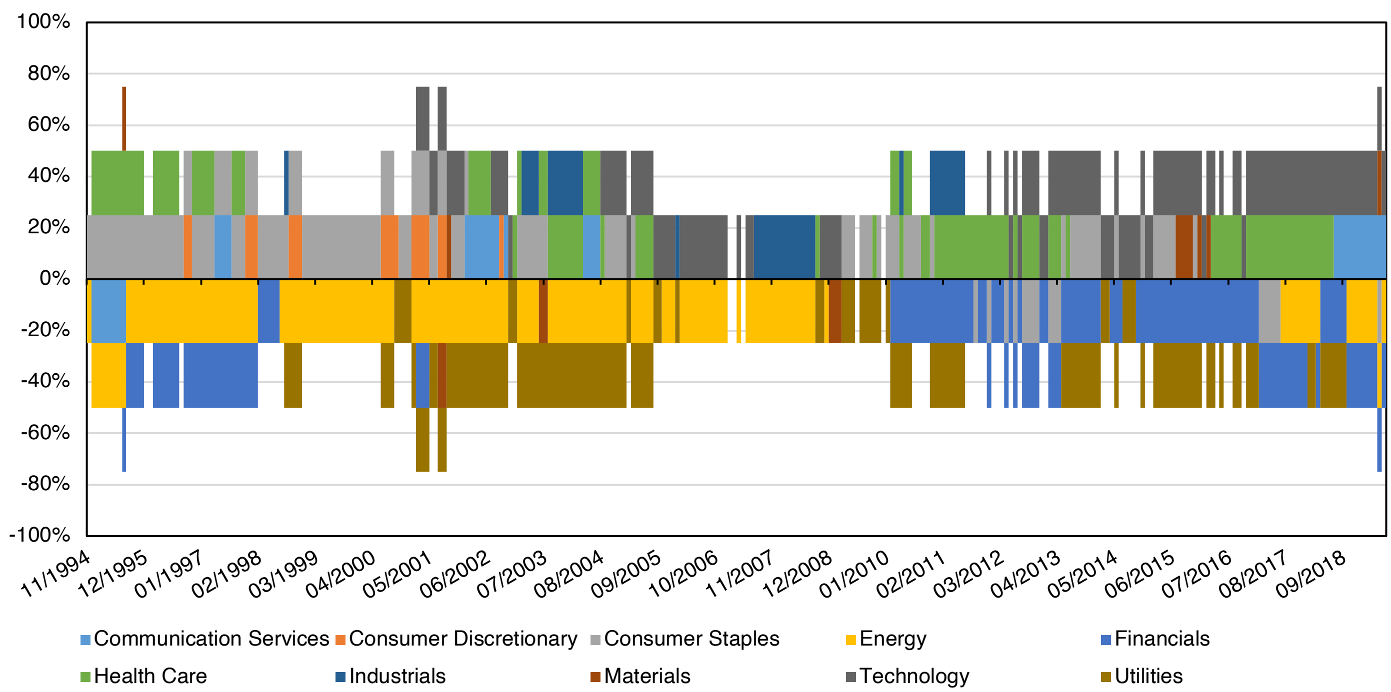
The result is a significant – and near-permanent tilt – towards Technology since the beginning of 2012, which can be seen in the graph of strategy weights below.

One way to explore the impact of this choice is calculate the weight differences between a top-4 CAPE strategy and a top-4 Relative CAPE strategy, which we also plot below. We can see that after early 2012, the Relative CAPE strategy is structurally overweight Technology and underweight Financials and Utilities. Prior to 2008, we can see that it is structurally underweight Energy and overweight Consumer Staples.

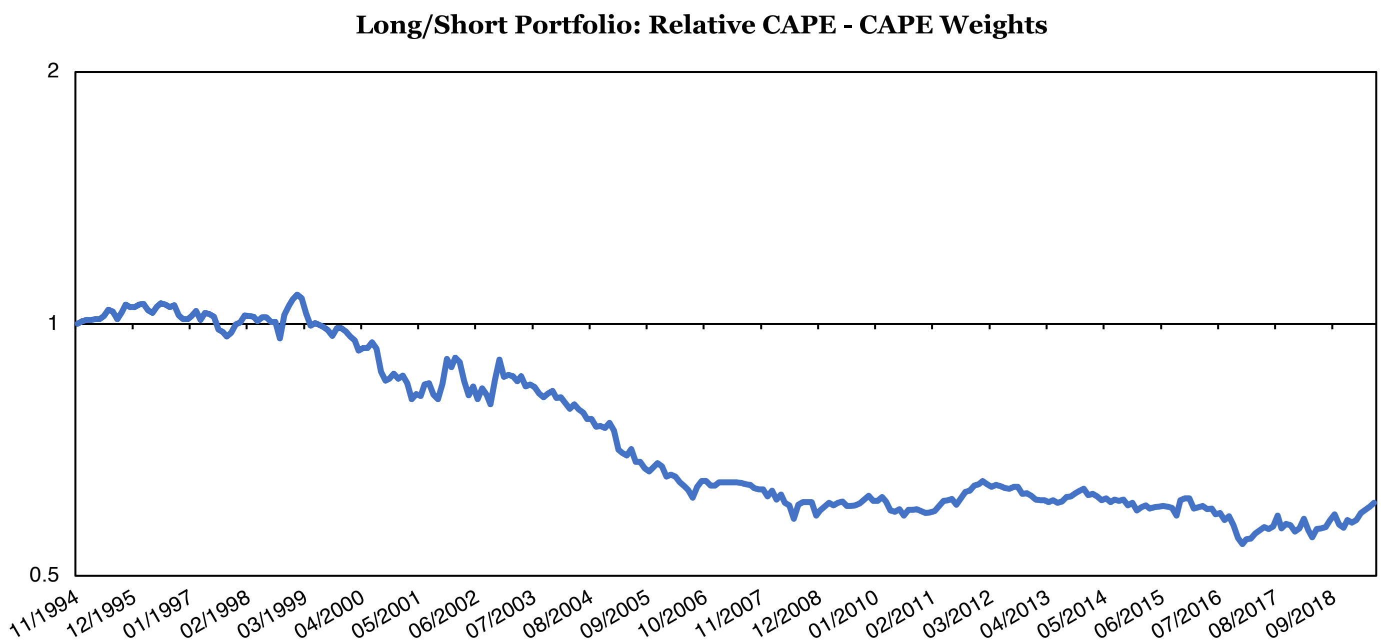
Shiller CAPE Sector Rotation Strategy Weights



Long/Short Portfolio Weights



If we take these weights and use them to construct a return stream, we can isolate the return impact the choice of using Relative CAPE versus CAPE has. Interestingly, the long Technology / short Financials & Utilities trade did not appear to generate meaningful out-performance in the post-2012 era, suggesting that something else is responsible for post-2012 performance.



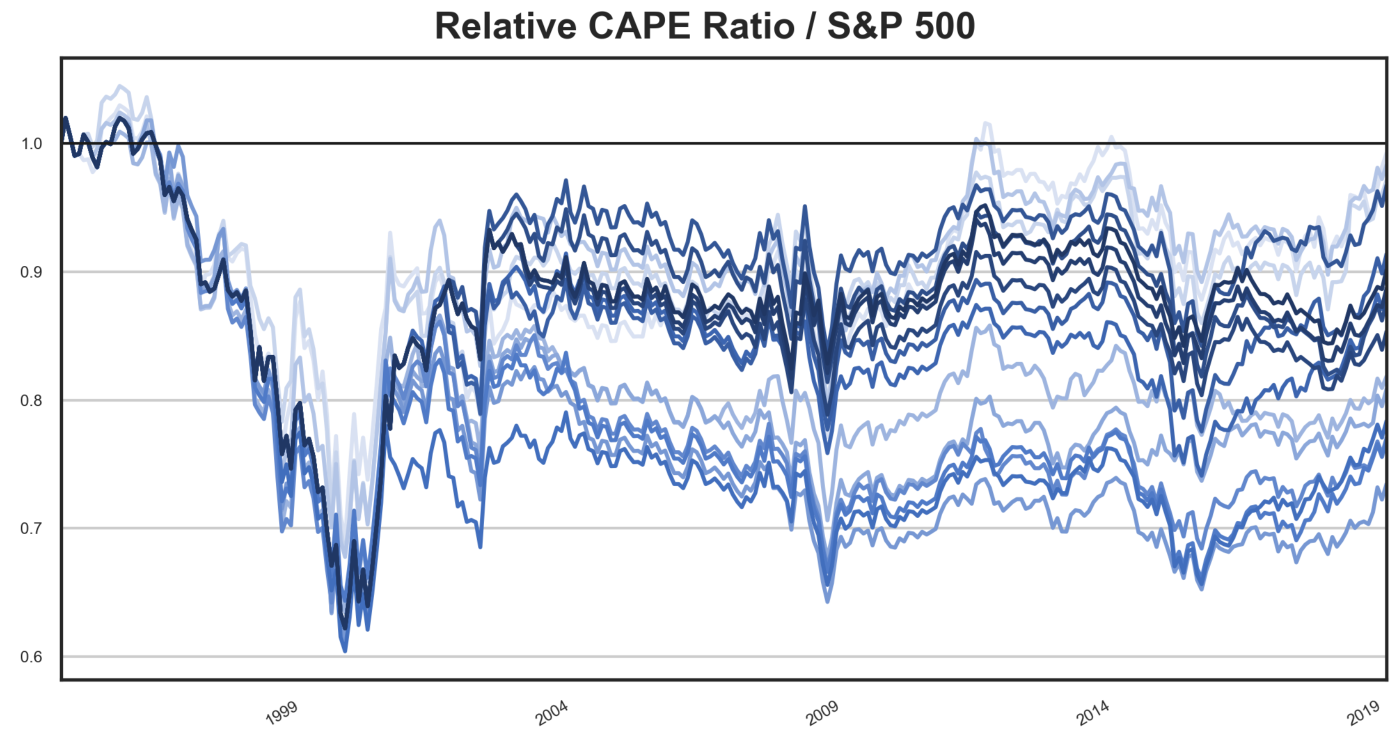
Source: Siblis Research; Morningstar; CSI Data. Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

The Miraculous Mojo of Momentum

This is where the 12-month momentum filter plays a crucial role. Narratively, it is to avoid value traps. Practically, it helps the strategy deftly dodge Financials in 2008, avoiding a significant melt-down in one of the S&P 500's largest sectors.

Now, you might think that valuations alone should have allowed the strategy to avoid Technology in the dot-com fallout. As it turns out, the Technology CAPE fell so precipitously that in using the Relative CAPE metric the Technology sector was still ranked as one of the top five cheapest sectors from 3/2001 to 11/2002. The only way the strategy was able to avoid it? The momentum filter.

Removing this filter makes the relative results a lot less attractive. Below we re-plot the relative performance of a simple "top 4" Relative CAPE strategy.



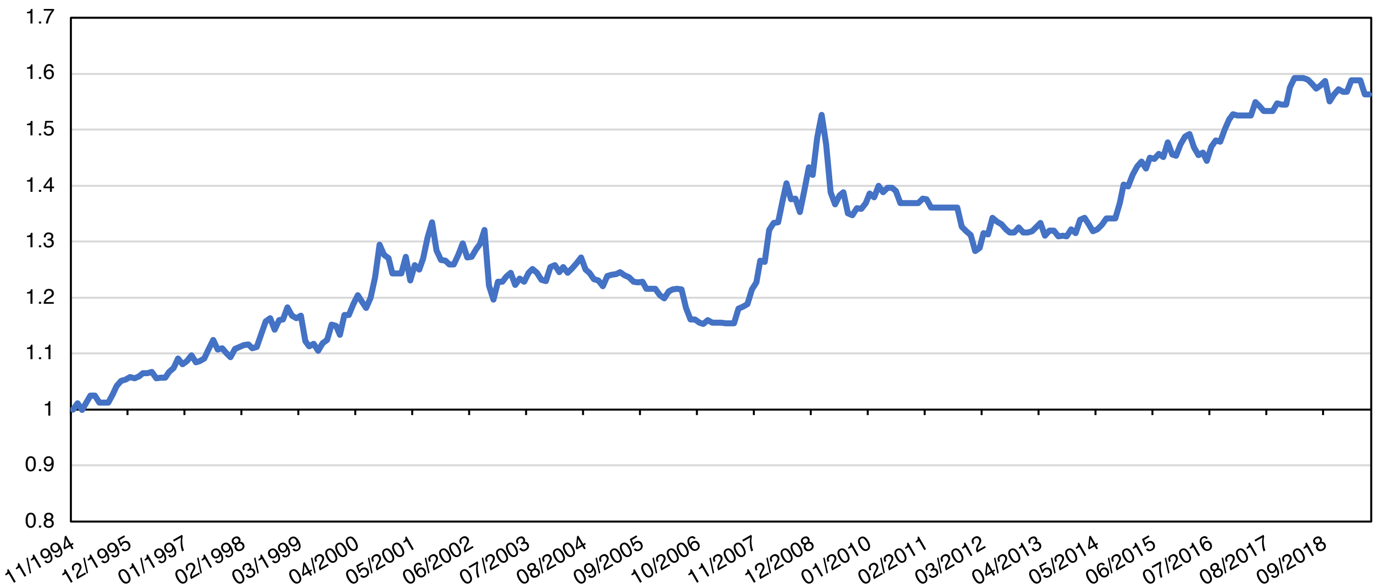
Source: Siblis Research; Morningstar; CSI Data. Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

Just how much impact does the momentum filter have? We can isolate the effect by taking the weights of the Shiller strategy and subtracting the weights of the Value strategy to construct a long/short index that isolates the effect. Below we plot the returns of this index.

It should be noted that the legs of the long/short portfolio only have a notional exposure of 25%, as that is the most the Value and Shiller strategies can deviate by. Nevertheless, even with this relatively small weight, when isolated the filter generates an annualized return of 1.8% per year with an annualized volatility of 4.8% and a maximum drawdown of 11.6%.

Scaled to a long/short with 100% notional per leg, annualized returns jump to 6.0%. Though volatility and maximum drawdown both climb to 20.4% and 52.6% respectively.

Growth of $1 in Isolated Momentum Filter



Source: Siblis Research; Morningstar; CSI Data. Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

Conclusion

Few, if any, systematic value strategies have performed well as of late. When one does – as with the Shiller CAPE sector rotation strategy – it is worth further review.

As a brief summary of our findings:

- Despite potential structural flaws in measuring cross-sectional sector value, CAPE outperformed Relative CAPE for a naïve rank-based value strategy.

- There is significant dispersion in results using the Relative CAPE metric depending upon which lookback parameterization is selected. Initial tests suggest that the longer lookbacks appear to have been more effective.

- Using valuation metrics other than CAPE – e.g. P/B, P/E (TTM), and EV/EBITDA (TTM) – do not appear as effective in recent years.

- Longer lookbacks allow the Relative CAPE methodology to create a structural overweight to the Technology sector over the last 15 years.

- The momentum filter plays a crucial role in avoiding the Technology sector in 2001-2002 and the Financial sector in 2008.

Taken all together, it is hard to not question whether these results are unintentionally datamined. Unfortunately, we just do not have enough data to extend the tests further back in time for truly out-of-sample analysis.

What we can say, however, is that the backtested and live performance hinges almost entirely a few key trades:

- Avoiding Technology in 2001-2002 due to the momentum filter.
- Avoiding Financials in 2008 due to the momentum filter.
- Avoiding a Technology underweight in recent years due to an inflated "average" historical CAPE due to the dot-com bubble.
- Avoiding Energy in 2014-2016 due to the momentum filter.

Three of these four trades are driven by the momentum filter. When we further consider that the Shiller strategy is in effect the returns of the pure value implementation – which suffered in the dot-com run-up and was a mostly random walk thereafter – and the returns of the isolated momentum filter, it becomes rather difficult to call this a value strategy at all.

As of the date of this commentary, neither Newfound Research nor Corey Hoffstein holds a position in the securities discussed in this article and do not have any plans to trade in such securities. Newfound Research and Corey Hoffstein do not take a position as to whether this security should be recommended for any particular investor.

SECTOR MOMENTUM

September 3, 2019

SUMMARY

- We explore "top N" sector rotation strategies based upon momentum signals.
- We find that too much concentration (i.e. N is too small) leads to poor performance, whereas performance does not appear to materially degrade for larger N.
- We find that short- to long-term signals all appear to generate higher total returns than the S&P 500 and there may be room to benefit from diversification by using multiple signals.
- However, in attempting to use momentum information in an optimization, we struggle to generate any value.
- We find that the majority of "top N" returns actually come from the tilt towards an equal-weight sector approach, with momentum adding little-to-no beneficial information.
- Of 1,000 randomly generated "top 4" sector strategies, 97% out-performed the S&P 500 from 2000-2012, highlighting the importance of decomposing strategy returns into the contribution of each step of the portfolio's construction. In this case, by more appropriately measuring the impact of momentum information after accounting for the return generated by equal-weight tilts, we find no benefit since the turn of the century.

In last week's commentary ("Es-CAPE Velocity: Value-Driven Sector Rotation," August 26th, 2019) we explored value-based sector rotation in the 20th century. Specifically, we deconstructed the Shiller CAPE US Sector Rotation strategy and ultimately found that the largest driver of performance was *not* value, but a momentum-based filter.

Given the success of the momentum-based filter and the fact that momentum often exhibits negative correlation to value, in this week's research note we wanted to explore the application of momentum-based signals in US sector rotation.

To perform this analysis, we will apply momentum-based signals to portfolios build from the SSgA Sector Select ETFs. After REITs were spun out of the Financial sector in 2016, a Vanguard REIT ETF was included in the investible universe. With the introduction of the Communication Services sector and a significant re-classification of Consumer Discretionary and Technology companies in September 2018, we utilized hypothetical indices that more accurately reflect the re-categorization (when historically applied) to generate momentum signals.

How Many Sectors?

To begin our analysis, we will start with a very traditional sector rotation model: a "top N" system. In this system, sectors are ranked on a quantitative signal and then the top N sectors with the strongest signals are equally weighted.

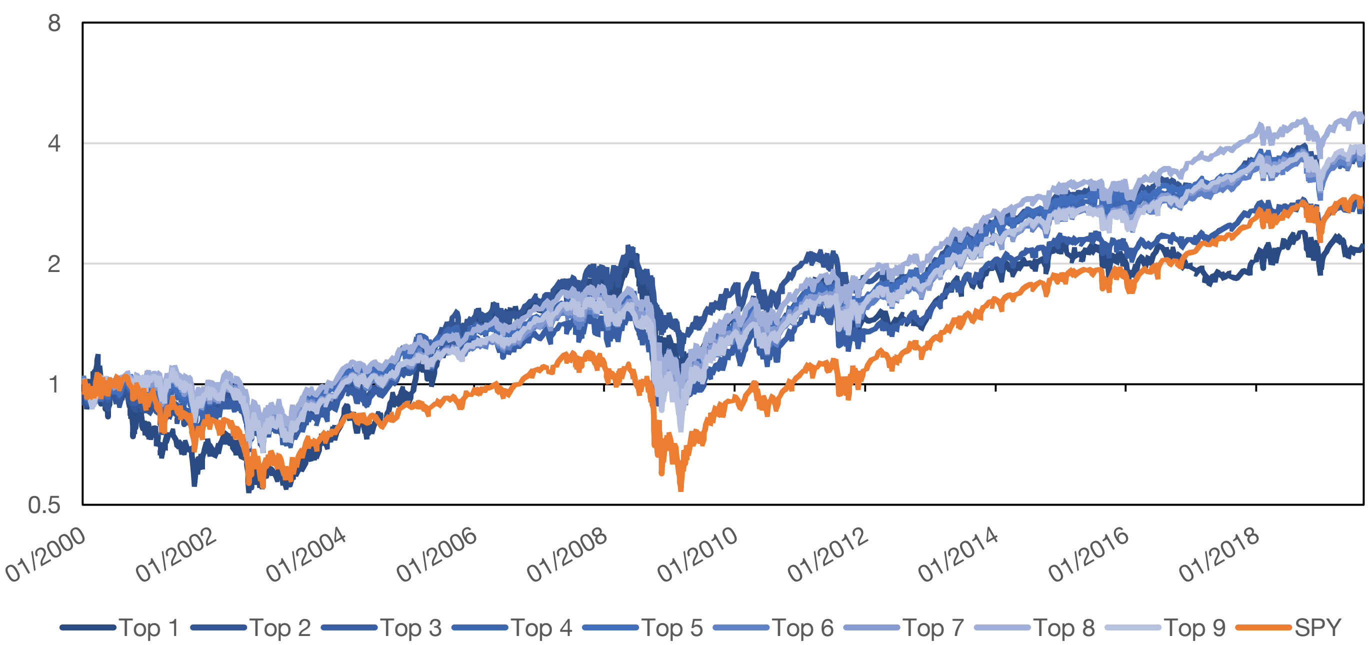
For example, in a top 4 12-month momentum system, sectors would be ranked based upon their prior 12-month total return and the strategy would allocate 25% of the portfolio to the top 4 ranking sectors for the next month.

The choice of N largely depends upon our expectations of signal strength, accuracy, and relative performance between ranks. For example, if the signal demonstrates a monotonic improvement in return (i.e. higher ranks imply higher returns) with a strong degree of accuracy, we might choose a small N to maximize our returns. The stronger the improvement, the more concentration risk we may be willing to bear.

On the other hand, if the signal is largely flat, with a drop-off in performance for lower ranking sectors, than we might treat it more as a screen, equally weighting a large number of sectors and avoiding just a few.

Below we plot the equity curves for different 12-month top N momentum strategies, where N is varied from 1 to 9. We have shaded the graph from darker-to-lighter blues in effort to determine if there are any parameterization-based patterns that emerge visually.

We can quickly see that the more concentrated strategies – for example, Top 1 and Top 3 – have the lowest total returns. However, among more diversified strategies, there appears to be a large degree of consistency in their terminal wealth. This might suggest that momentum, when applied in the last century, served better in avoiding the worst sectors rather than picking the best.



Source: CSI Data; S&P Dow Jones; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

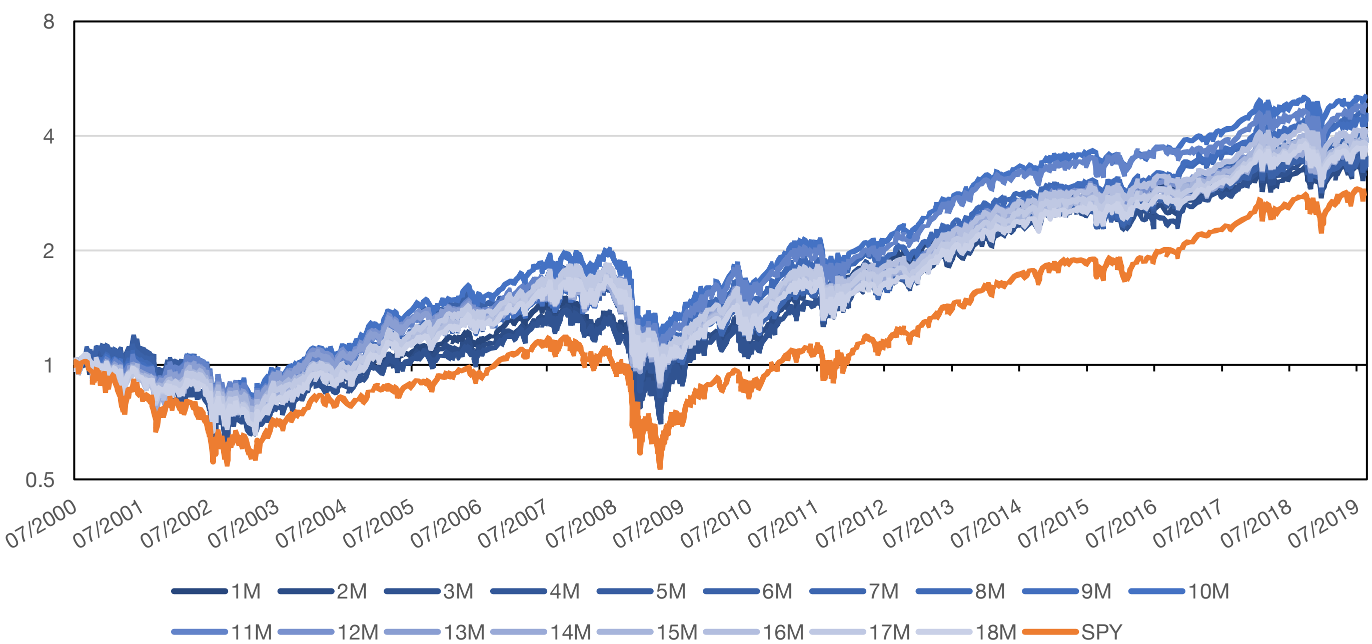
Long- versus Short-Term Momentum

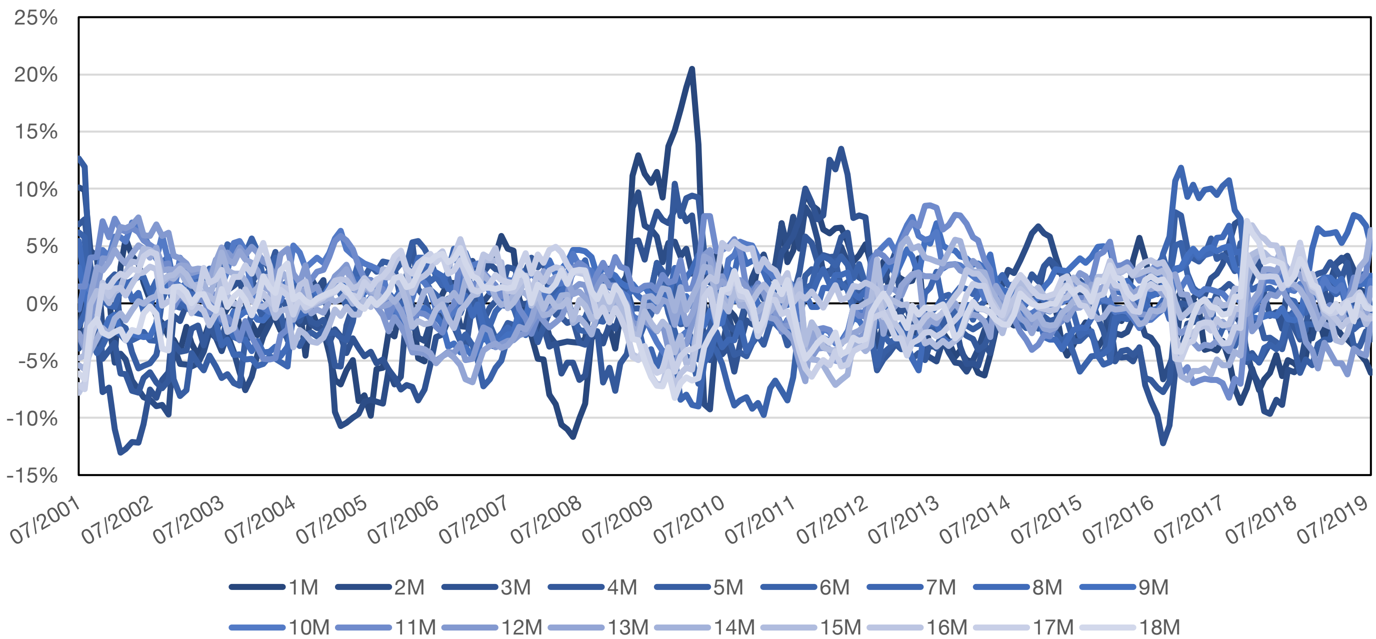
Below we plot all the top 4 systems with momentum lookbacks ranging from 1 to 18 months. Again, we vary the shades from darker-to-lighter blues in an effort to determine if any visual patterns emerge.

Interestingly, we can see that both short- (1-to-3 months) and long-term (16-to-18 months) momentum appear to perform the worst on a total return basis, while more intermediate-term measures (10- and 11-month momentum) appear to perform the best.

Of course, in the short-term, results can vary dramatically and even a 20-year period is not sufficient to determine if a cluster of parameterizations is superior. If we plot rolling 1-year returns, we can see dramatic dispersion in results versus the average return of the different parameterizations. For example, from June 2007 to June 2008, the 1M variation under-performed the average result by over 1100 basis points, but from February 2009 to February 2010, it out-performed by over 2000 basis points.

With no statistical difference in Sharpe ratios over the period, we would suggest that – once again – different parameterizations represent an opportunity for diversification.





Source: CSI Data; S&P Dow Jones; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

Tracking-Error Based

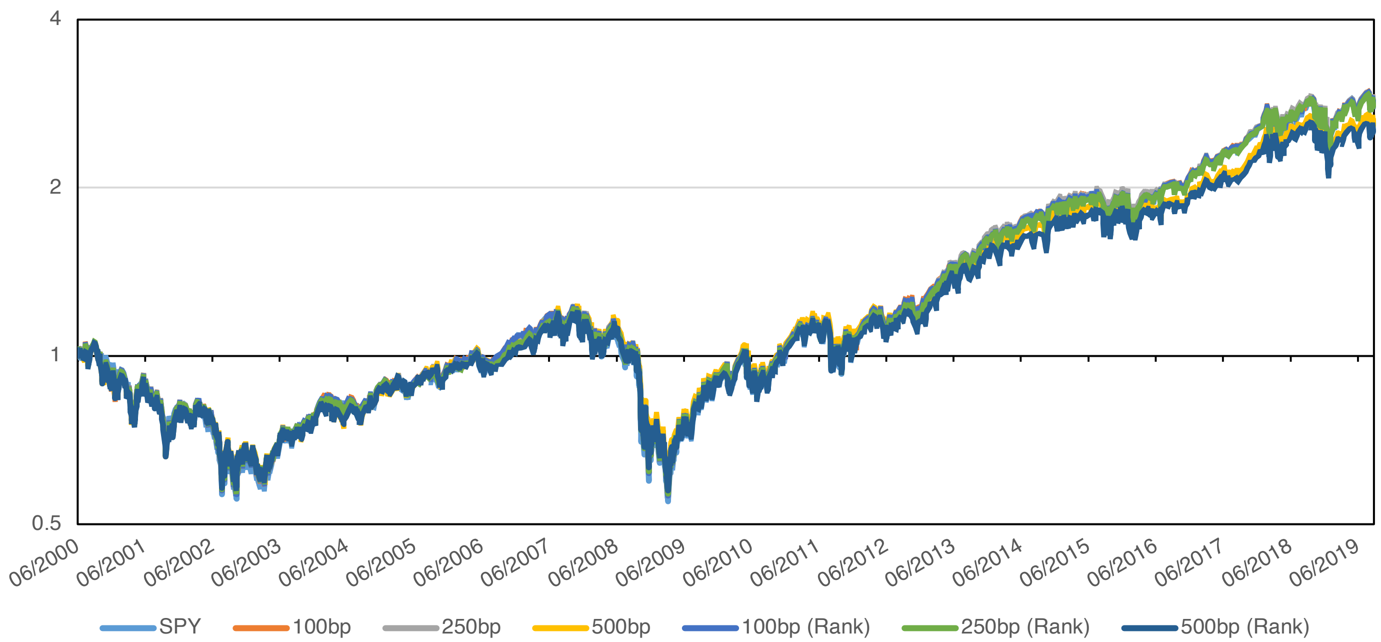
The above analysis suggests that sector-based momentum may be most effective if we simply try to avoid the worst sectors and apply an ensemble of parameterizations. One potential way of improving our model, then, is via optimization. Specifically, we can seek to maximize our momentum exposure subject to tracking-error constraints with respect to the S&P 500.

To implement this idea, we:

1. Generate N-month returns for each sector
2. Calculate a covariance matrix from sector returns (exponentially weighted over the prior 252 days).
3. Given current S&P 500 sector weights, select portfolio that maximizes weighted N-month returns subject to a given tracking error constraint.

The goal of this process is for the optimization process – despite the added computational complexity and embedded estimation risk – to more intelligently apply our active risk budget.

We also construct a variation where instead of maximizing the weighted N-month returns, we will transform the N-month returns to ranks and maximize the weighted rank. This can be thought of as a regularization step, sacrificing potential non-linear information but reducing the impact of outlier returns.



Source: CSI Data; S&P Dow Jones; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

The results here are a bit of a head scratcher, as we have almost completely lost any out-performance generated in the top N models.

Which suggests that, perhaps, the out-performance had nothing to do with momentum-based information, but rather something else entirely.

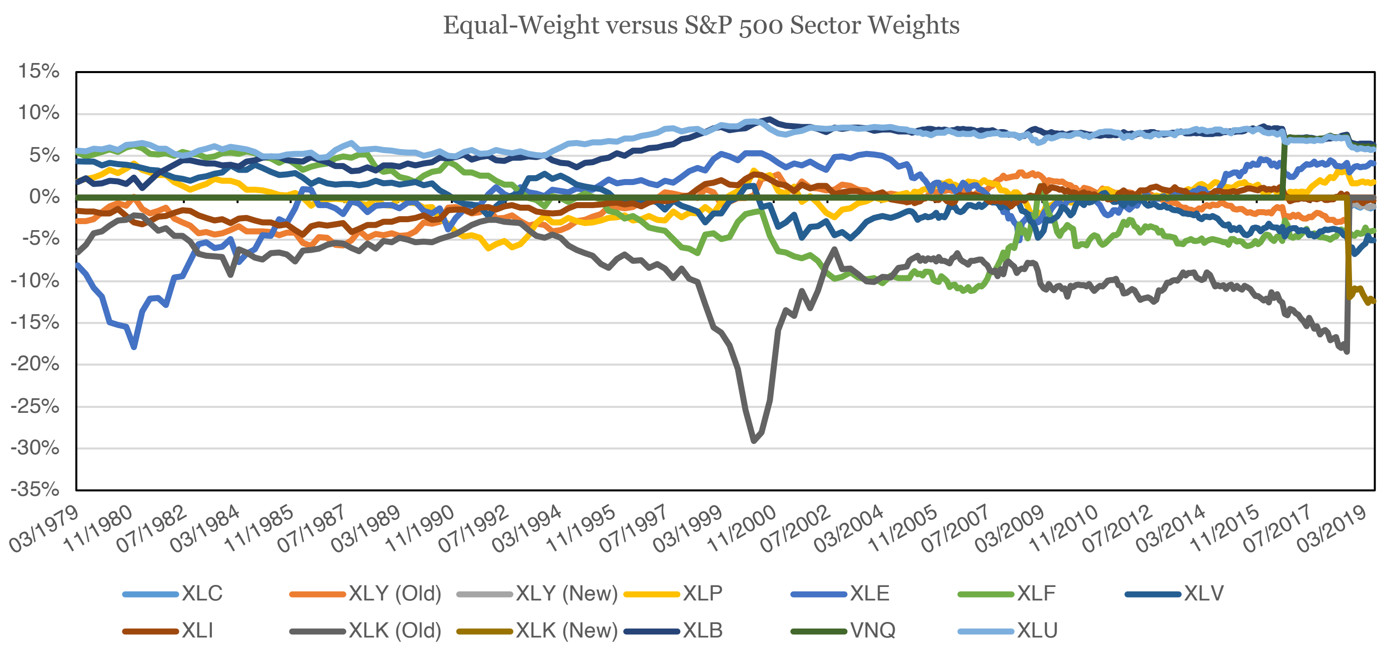
It Was Always You, Equal Weight

If we return to our foundational framework that any active strategy is really just the market-capitalization-weighted portfolio plus a dollar-neutral long/short portfolio overlaid on top, then we can really think of our top N portfolios in a two-step active process:

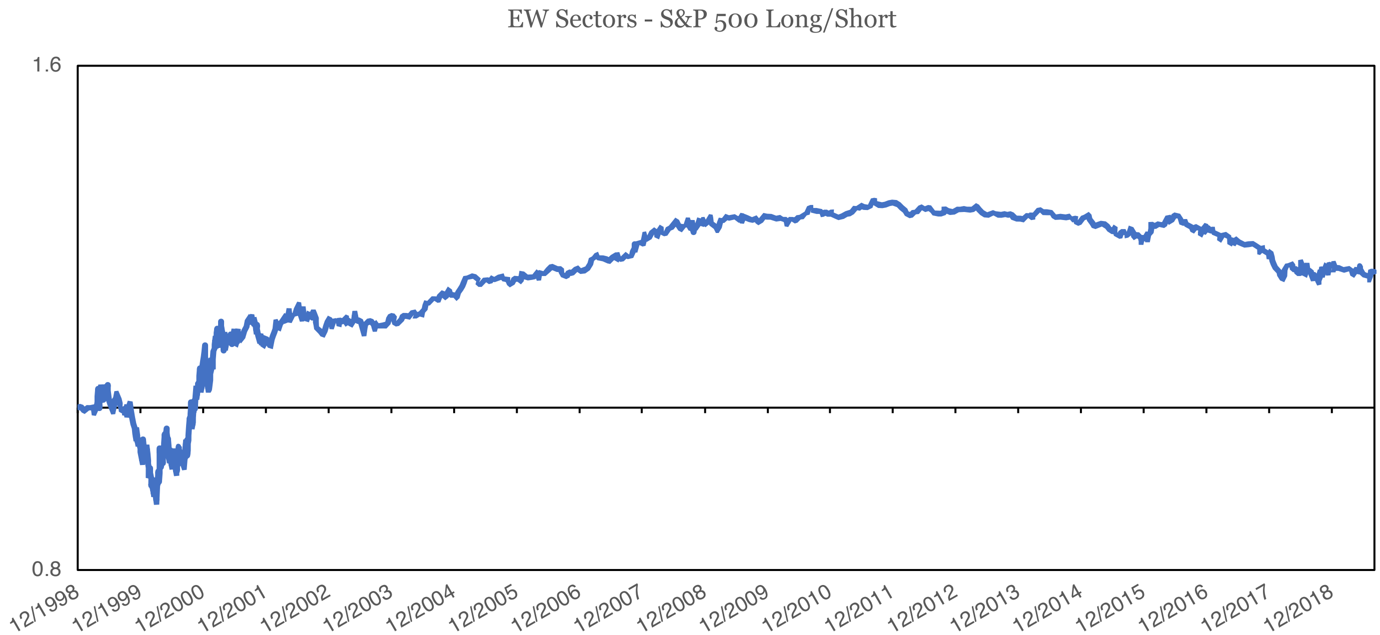
1. Start with the S&P 500 and tilt to equal-weight sector implementation;
2. Remove low-momentum sectors and re-allocate capital among the rest.

To more accurately demonstrate the impact of our momentum decisions, then, we really should be isolating the impact of each of these steps.

For example, plotting the weight differences over time, we can see that equal-weight sectors are typically overweight Utilities and Materials and underweight Technology. We would expect this to dramatically help in the dot-com fallout, but hurt us after 2008.



If we actually generate the performance of this long/short strategy, we can see the results almost perfectly line up with our expectations: it was a generally profitable trade in the new century through early 2012, at which point it began to reverse.



Source: CSI Data; S&P Dow Jones; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

If we plot the S&P 500, an equal-sector portfolio, and a top 4 momentum strategy, we can see that there is actually very little performance differential between the equal-sector approach and the momentum strategy. In fact, the momentum strategy does not appear to add much value at all other than active noise.

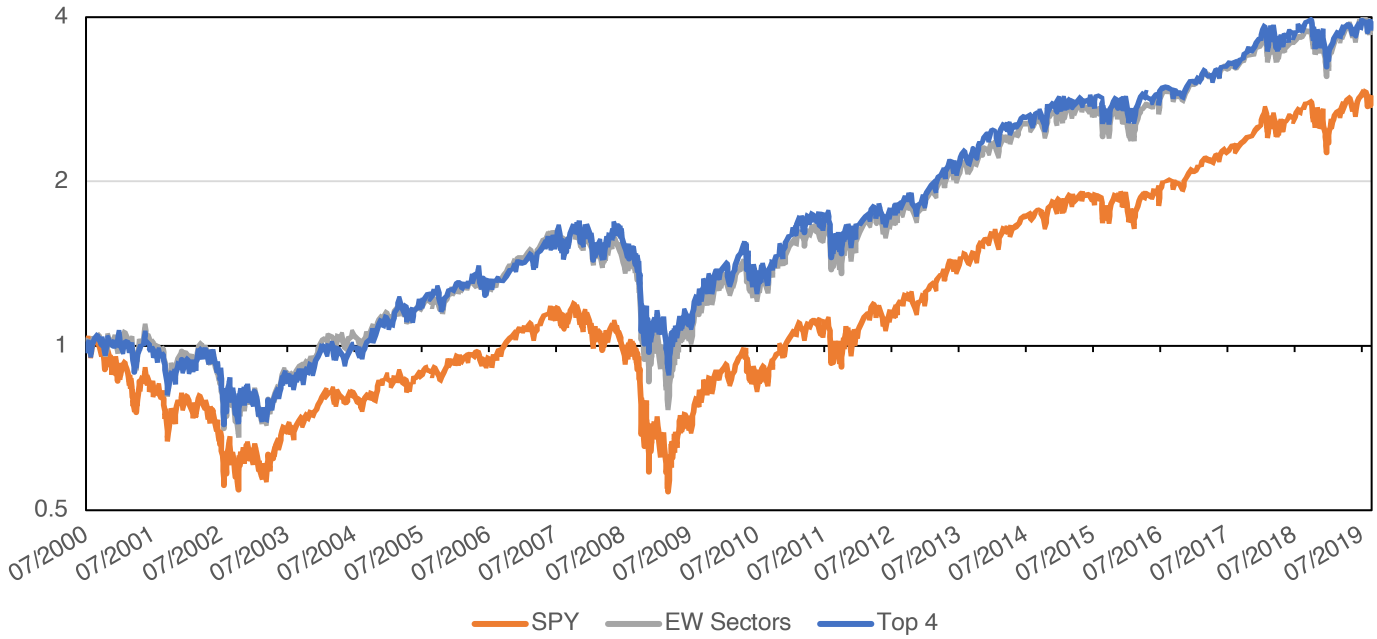
Which very much confirms the results we saw in our tracking-error-based solution: in the last century, the edge has not been in momentum.

Rather, it would appear that a top N momentum strategy was merely able to back its way into harvesting the return benefits of the equal-weight portfolio. The question we must now ask is whether that represents skill in the signal or merely luck.

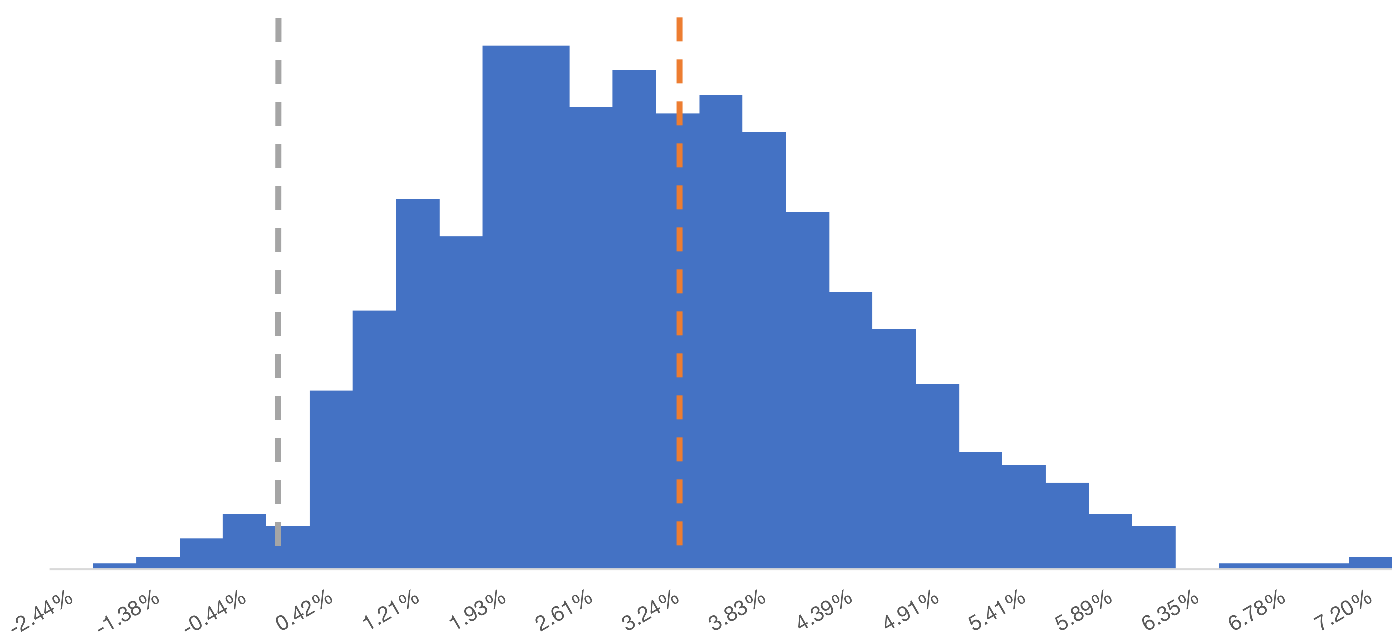
Given the abject failure of momentum to dislodge itself from the equal weight portfolio *after* 2012, our initial guess is "luck."

But to explore this idea further, we generate 1,000 randomly generated "top 4" strategies. At the end of each month, each strategy randomly selects four sectors and holds them in equal weight over the next month. Note that as the number of strategies increases, the average allocation across all strategies should approach an equally weighted sector portfolio.

We then plot the distribution of annualized excess returns (versus the S&P 500) of each randomly generated strategy from 7/2000 to 3/2012.



Source: CSI Data; S&P Dow Jones; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.



The grey bar highlights the zero excess return line and the orange bar highlights the annualized return of the top 4 momentum strategy.

Two things are immediately apparent:

- Simply throwing a dart to select sectors from 7/2000 to 3/2012 proved to be a very effective strategy. In fact, *97%* of the randomly generated strategies exhibited positive excess returns over this period.

- The ensemble top 4 momentum strategy finds itself almost exactly in the middle of the road of all the randomly generated portfolios. One positive interpretation is that the momentum strategy was likely not data mined, as it is not a significant positive outlier. A negative interpretation, however, is that momentum may not have provided much information to trade off of in the last twenty years.

It would appear that the real hero in our top N momentum strategy is not momentum, but rather the equal-weight sector tilt implied in the strategy.

Conclusion

As we have stressed in the past, we believe it is very important to evaluate the contribution of each step in a portfolio's construction. By treating each step as the overlay of a dollar-neutral long/short portfolio, we can not only isolate the allocation changes implied by that step, but also track the performance over time.

In the case of a top N momentum sector rotation strategy, we believe there are two key transformations: (1) the tilt from market-capitalization-weights to equal sector weights, and (2) the removal of low-ranking sectors.

With respect to performance over the last two decades, we find that the first step accounts for the vast majority of returns, while the second step has merely added noise.

To explore this further, we randomly generate 1000 different top 4 sector strategies and find that from the 2000-2012 period, 97% of the randomly generated strategies would have outperformed the S&P 500.

This fact has important implications for anyone evaluating a "top N"-style sector rotation track record from that era. The evidence suggests that it does not matter whether the process was based upon market cycle analysis, momentum, value, or dart-throwing monkeys: it would have been hard to under-perform.

While some might say that "top N" momentum strategies stopped working after 2012, we would suggest that they never appeared to work in this century at all.

BUILD YOUR OWN LONG/SHORT

September 10, 2019

SUMMARY

- We exploit the idea that long-only strategies are "long/short portfolios all the way down," we demonstrate how to isolate the active bets of portfolio managers.

- Using the example of a momentum / low-volatility barbell portfolio, we construct a simple long/short portfolio using ETFs and S&P 500 futures.

- Recognizing that not all investors will have access to S&P 500 futures, we argue that the capital efficiency provided by a fund like the PIMCO StocksPLUS Short Fund ("PSTIX") can potentially allow investors to achieve the same ends.

- By thinking about our aggregate exposure within a portfolio – and not line-item by line-item – we demonstrate how an investor might alter a standard 60/40 portfolio to introduce a 20% allocation to their own equity long/short portfolio

In this week's research note, we will explore a simple idea for how advisors and investors can "roll their own" long/short portfolio even if they do not have the ability to short individual securities.

The idea boils down to two simple concepts:

1. Active portfolios are long/short portfolios all the way down.
2. Capital efficiency.

The first concept can be quickly summarized as recognizing that a long-only, active portfolio can be thought of as the sum of two components: (1) a market-cap weighted index and (2) the over- and underweight decisions the active manager makes. The over- and underweight decisions form a dollar-neutral long/short portfolio and the notional size of the bets is equal to the active manager's active share. When we overlay the long/short portfolio on top of the market-cap weighted index, we are given the long-only portfolio.

We can, therefore, take this idea in reverse: shorting out the market-cap weighted index from a long-only active strategy leaves us with the long/short portfolio. This idea allows us to take any actively managed portfolio – whether it is a basket of securities we pick or an actively managed fund – and turn it into a long/short portfolio.

There are a number of ways in which this might be achieved. For example, an investor could short an S&P 500 ETF or could short futures contracts. Unfortunately, many advisors and individuals do not have this opportunity.

Another option is to hold a long position in an inverse ETF such as the ProShares Short S&P 500 ETF ("SH"). The downside here is that precious capital is now tied up in a position that effectively returns Treasury Bills minus the S&P 500. We could try to use a levered short ETF, such as the ProShares UltraShort Short S&P 500 ("SDS"). Only half the necessary capital would be tied up with this trade, but we'd be introducing the potential for significant tracking error due to compounding effects. And that's all to speak nothing of the cost.

For now, we are going to assume that an investor can simply short S&P 500 futures contracts and we'll return to how an advisor or investor might go about implementing this idea in a more constrained setting.

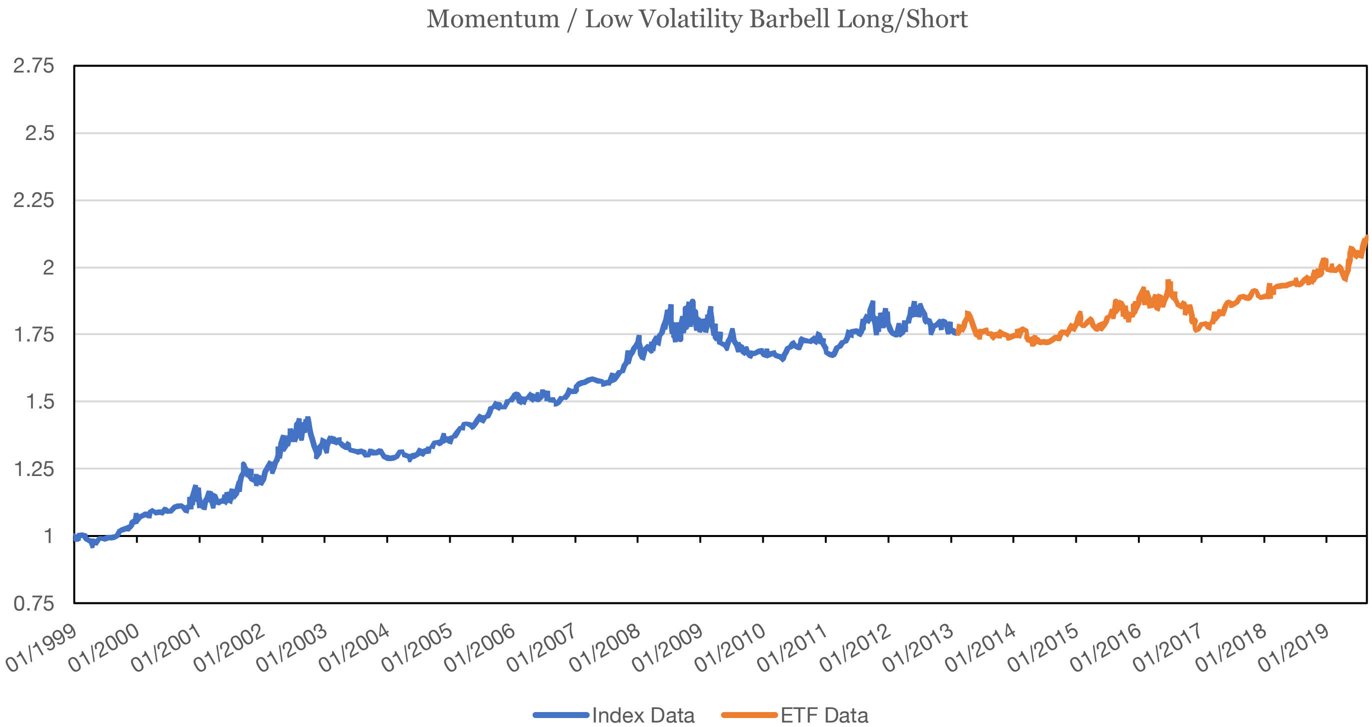
As an example of constructing a long/short portfolio, we will use the idea of a momentum / low-volatility barbell which has been proposed by Lawrence Hamtil (and who has written about the idea a number of times). We will use the iShares Edge MSCI USA Momentum Factor ETF ("MTUM") and the iShared Edge USA Min Vol Factor ETF ("USMV") for our long factor positions. Prior to ETF launch, we will use index returns.

We will construct our portfolio as a simple 50-50 barbell, holding 47.5% of the portfolio in each ETF and keeping 5% aside as collateral for a 95% short position in S&P 500 futures.

The results are rather strong, both prior to ETF launch and after. Over the full period, our long/short portfolio returns 3.6% annualized with a volatility of just 5.4% and a maximum drawdown of 11.5%.

Now we should acknowledge here that constructing our portfolio in such a manner means that the *notional* size of our long/short is constrained by the active share of the ETFs themselves. For example, MTUM has an active share of 68.1% and USMV has an active share of 57.2%, giving our barbell an active share of 62.6%. This implies that when we short out our exposure to S&P 500 futures, the long/short portfolio we are left with is really only using 62.6% of capital.

Though this speaks nothing to the volatility of those active bets, investors should be aware that a methodology like this, when used on a low active share strategy will leave very little residual exposure.

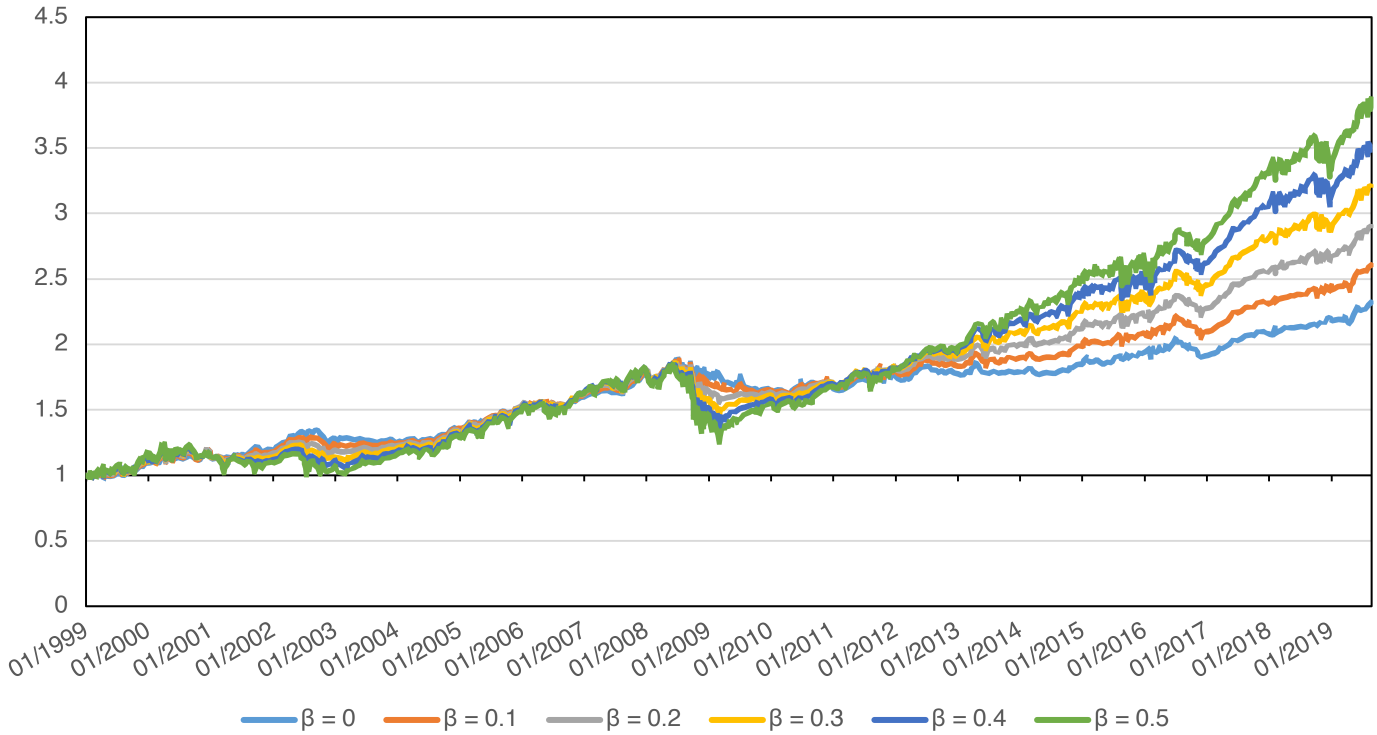


Source: CSI Data; MSCI; Stevens Futures; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

Unless they are explicitly identified as market neutral, most long/short portfolios retain some equity market beta to benefit from the positive expected long-term equity risk premium. Therefore, it may make more sense to actually construct this portfolio using a target beta approach.

For example, we might calculate the beta of our 50/50 momentum/low-volatility barbell portfolio to the S&P 500 and then increase or decrease our exposure to our short futures position accordingly. If our barbell has a beta of 1.0 and we only want a beta of 0.5, we could hold 95% of our portfolio in the barbell, 5% in cash, and short 45% S&P 500 futures exposure to hit our target beta.

Below we plot the equity curves for portfolios targeting a beta of 0 through 0.5. We estimate beta using the historical realized covariance matrix and use an ensemble approach (varying the lookback window of our covariance matrix calculation) to generate weights.



Source: CSI Data; MSCI; Stevens Futures; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

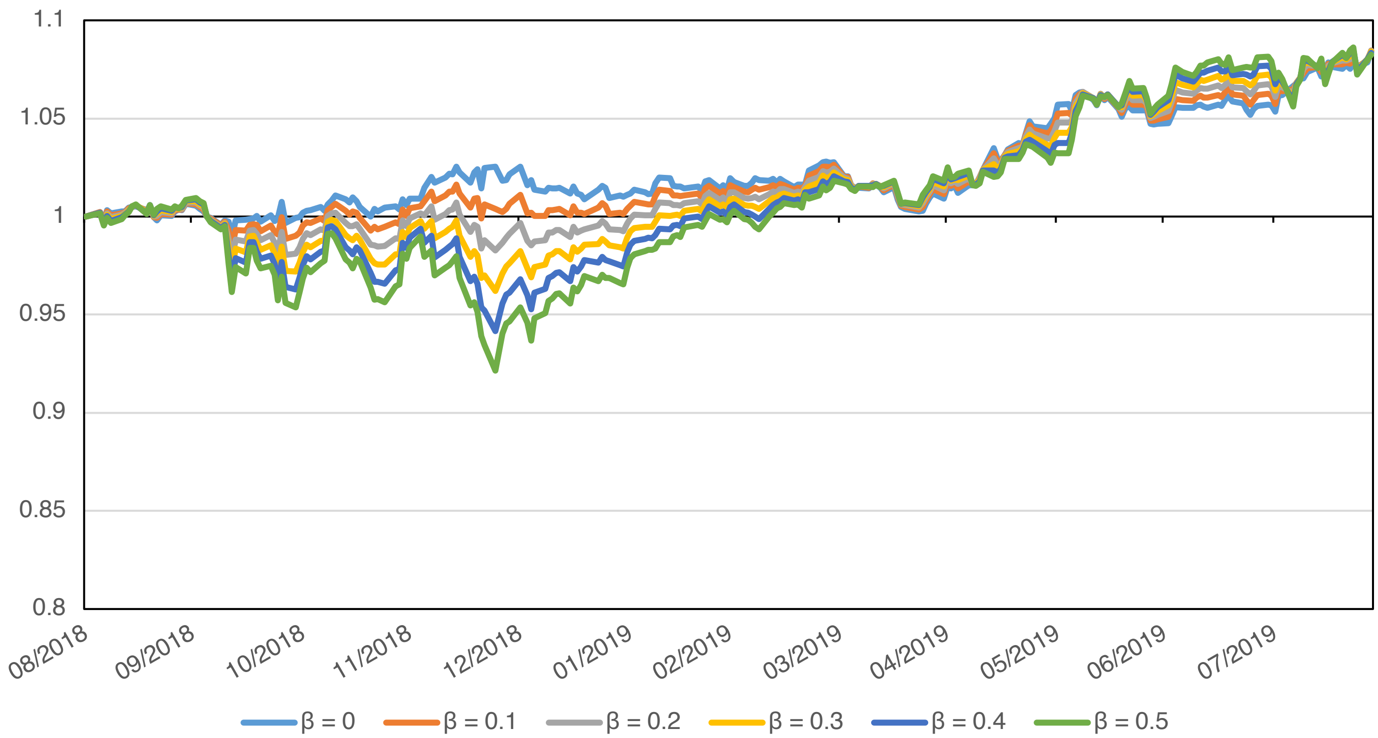
Unsurprisingly, given our knowledge of history, the highest beta long/short portfolio had the highest return over the last 20 years. Isolating different periods, however, provides a more nuanced perspective. For example, we can see that over the last 12 months the five variations all exhibited nearly identical total returns, while the portfolio that targeted zero beta was able to exhibit *positive* returns during the market's Q4 2018 drawdown.

Here we will pause to note a key difference in the first portfolio we constructed and the target beta portfolios.

In the former case, we used the basic math that a long-only active portfolio is equal to a market-capitalization-weighted portfolio plus a dollar-neutral long/short portfolio. By shorting a *dollar equivalent* amount of market beta, we could isolate the implied long/short.

The remaining long/short exposure may or may not have residual beta within in. For example, USMV has a beta of approximately 0.7, meaning that the long/short portfolio component must have a beta of -0.3. If we hold USMV and an equivalent dollar amount in short S&P 500 exposure, we would generally expect the residual exposure to be negatively correlated to the market (ignoring idiosyncratic returns for a moment).

If, however, we design our portfolio to explicitly be beta neutral, then we might not hold only $0.7 short in S&P 500 exposure for every $1 we hold in USMV. This results in a portfolio that may more explicitly reflect the idiosyncratic returns of the implied long/short active bets taken by USMV.



Source: CSI Data; MSCI; Stevens Futures; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

Unfortunately, up until this point, implementation of this idea requires shorting S&P 500 futures. This is where capital efficiency enters the equation.

The PIMCO StocksPLUS Short Fund ("PSTIX") provides access to an actively managed portfolio of fixed income securities and overlays that portfolio with a 100% S&P 500 futures short position. On its own, we would have to hold a near dollar-for-dollar amount in our long equity exposures as PSTIX to hedge out beta, making it inefficient.

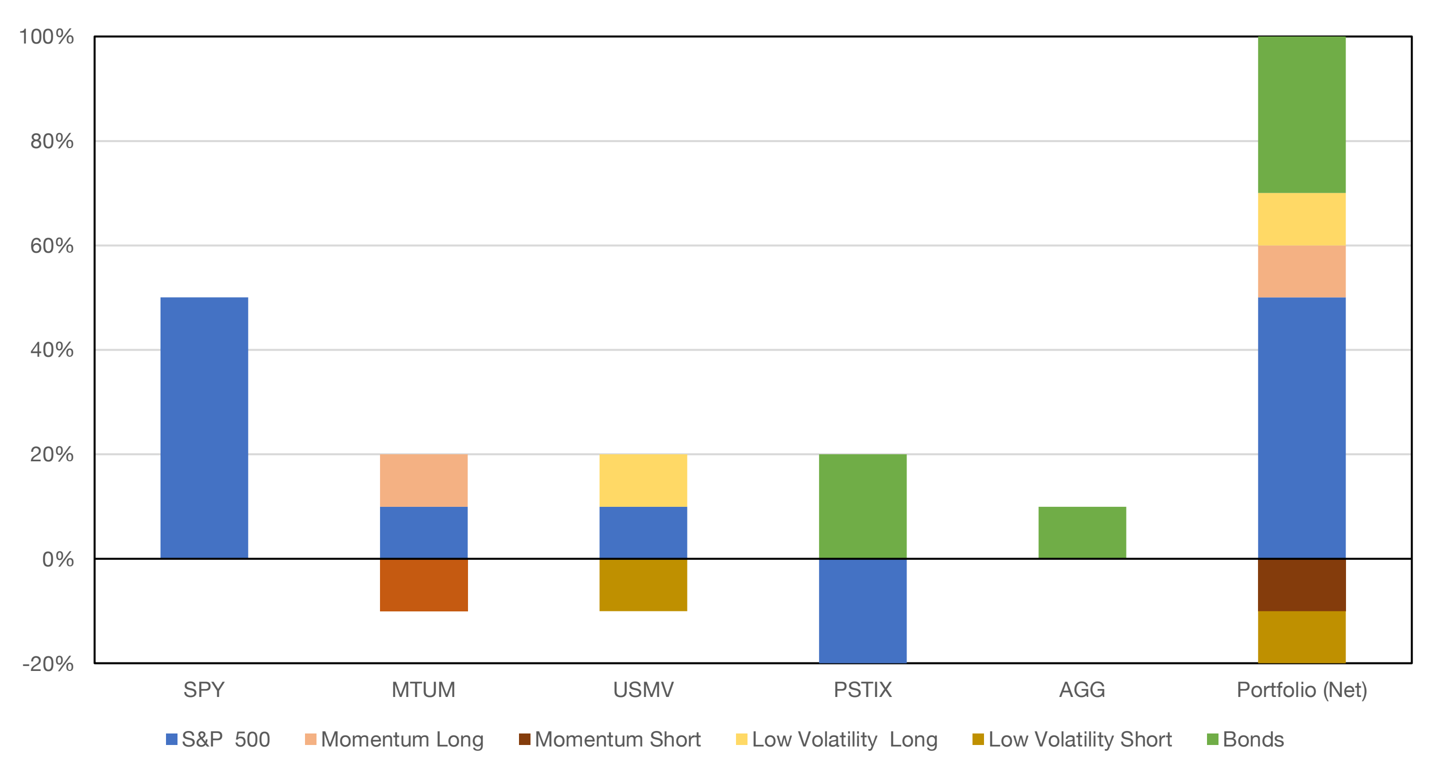
If, however, we think from a total portfolio balance sheet perspective, things become much more interesting.

Let's say, for example, that we currently hold a 60% S&P 500 / 40% Bloomberg Barclay's Aggregate US Bond portfolio and we would like to sell 10% of our equity exposure and 10% of our bond exposure to create a 20% allocation to a momentum/low-volatility barbell long/short portfolio.

To achieve this exposure, here is how we would construct our portfolio:

- Hold 50% of our portfolio in the S&P 500
- Hold 10% of our portfolio in MTUM
- Hold 10% of our portfolio in USMV
- Hold 20% of our portfolio in PSTIX
- Hold 10% of our portfolio in AGG

At first, this appears to be a massive reduction in bonds. However, when we consider that PSTIX provides us exposure to bonds and short S&P 500 exposure simultaneously, the aggregate picture makes more sense. Below we plot the exposure from each position as well as the net resulting exposure. We can see that we do indeed create a 50% S&P 500 / 30% Bond / 20% Equity Long/Short portfolio.



Conclusion

In this commentary, we introduce a simple idea for advisors and individuals to construct long/short strategies when they do not have the ability to explicitly short within their portfolios.

The idea is built from two foundational concepts: (1) active strategies can be thought of as a passive benchmark plus a dollar-neutral long/short strategy, and (2) exploiting capital efficiency. By combining long-only positions with short exposure to the S&P 500 (or other market index), we can isolate the implied active under- and overweight positions as a long/short portfolio.

To achieve our short exposure, we can take advantage of a capital-efficient portfolio such as the PIMCO StocksPLUS Short Fund ("PSTIX"), which provides simultaneous exposure to fixed income and a short S&P 500 position. By thinking holistically about our portfolio, we can replace existing fixed income with PSTIX and create an implied long/short strategy within our portfolio.

This approach is not without its trade-offs. In our example, creating a 20% long/short allocation results in PSTIX accounting for 2/3rds of the portfolio's fixed income exposure. This means we must not only be incredibly comfortable with the portfolio construction of PSTIX, but we must also be comfortable in the foregone opportunity cost to allocate to *other* fixed income managers.

This approach may also not be effective for investors who do not currently hold much fixed income. However, for conservative investors wanting to replace existing fixed income exposure with long/short equity, a combination of long-only exposure with PSTIX may be an effective way to take control of the long/short portfolio construction.

As of the date of this document, both Newfound Research and Corey Hoffstein hold positions MTUM and USMV, and Corey Hoffstein holds positions in PSTIX. Newfound Research and Corey Hoffstein do not take a position as to whether these securities should be recommended for any particular investor.

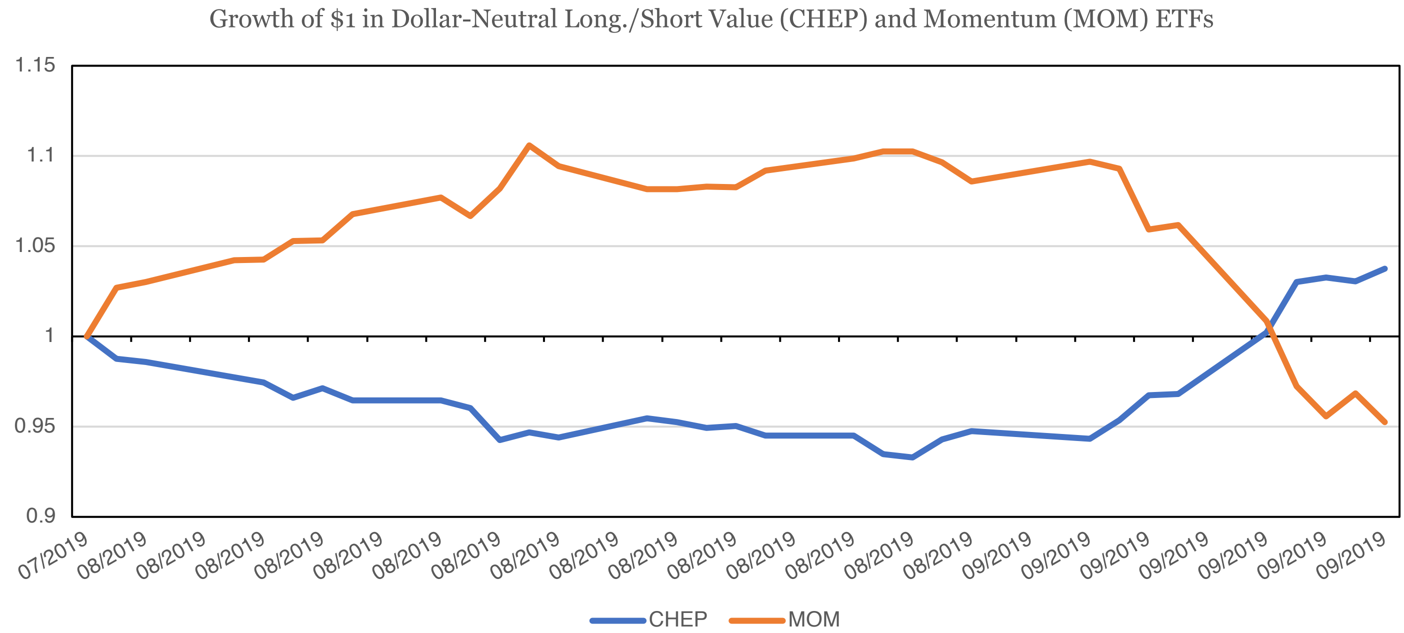
FACTORS AND THE GLIDE PATH

September 16, 2019

SUMMARY

- Value and momentum equities exhibited significant performance last week raising short-term questions about factor crowding and long-term questions about appropriate factor diversification.
- We explore the idea of appropriate factor diversification through the lens of a retiring investor, asking the question, "are all equity styles appropriate at all points in an investor's lifecycle?"
- Using a backwards induction method, we simulate portfolio decisions are derive optimal portfolios based upon an investor's age and net-worth relative to their desired spending level.
- We find that portfolios are split into four distinct zones – the zone of safety, the cone of balance, the triangle of growth, and the twilight zone – each representing a distinct asset allocation.

Calm headline returns last week belie a tumultuous undercurrent in factor equities. Specifically, value and momentum equities both performed a rapid turnaround relative to recent performance. When all was said and done, the rotation between the two factors resulted in a weekly performance spread of over 900 basis points.



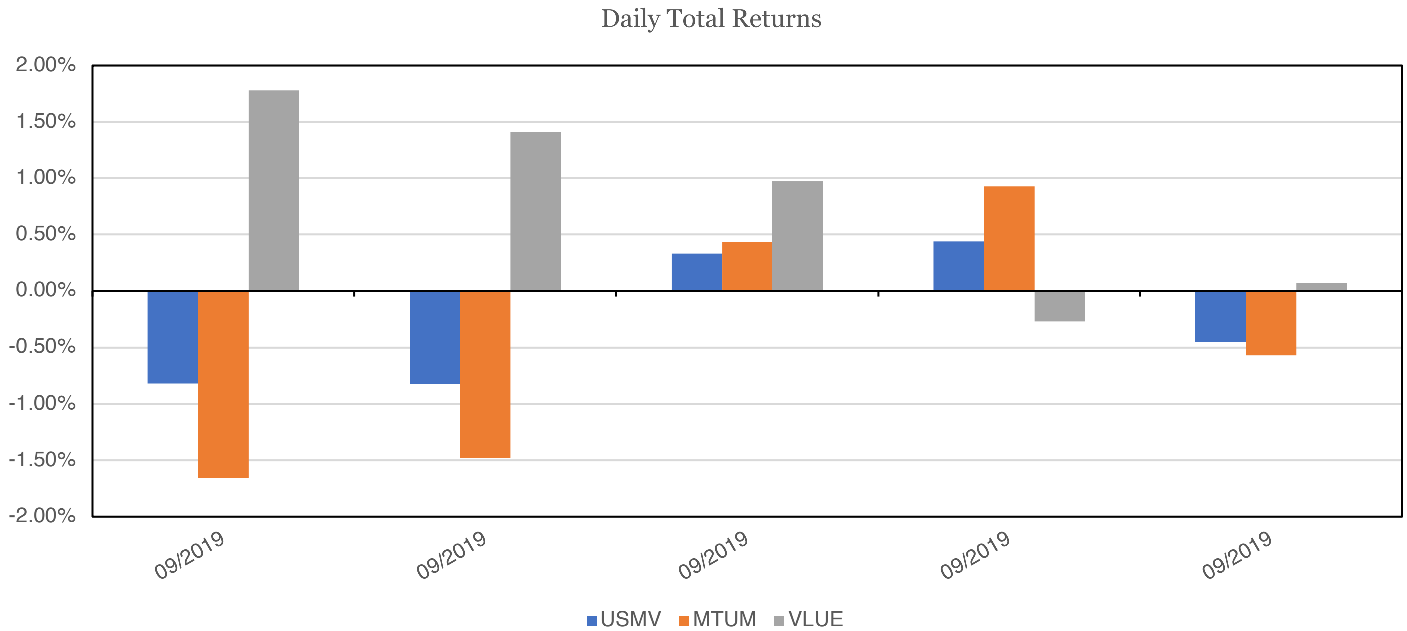
Source: CSI Data; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

This abrupt course correction raised eyebrows throughout the quant world. Some took to measuring the historical rarity of the event. Others opined about the impact of factor crowding and the potential ensuing risks to portfolio construction. Many reminisced about the 2007 quant quake and wondered, "who blew up this time?"

There were those who swore these tea leaves prophesized doom and gloom and those who crossed their fingers that value might finally be making a comeback, but the majority simply shrugged and said, "stay the course." Which, barring truly unusual circumstances, tends to be the prudent course of action.

It was, ironically, a particularly bad week for us to publish a research note that used a momentum / low-volatility portfolio barbell as a core example[49], as low-volatility equities and momentum moved in directional lockstep with one another all week.

[49] This unfortunate timing to be trumped only, perhaps, by the time we wrote a research note about XIV on February 5th, 2018: the very day it blew up.



Source: CSI Data; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

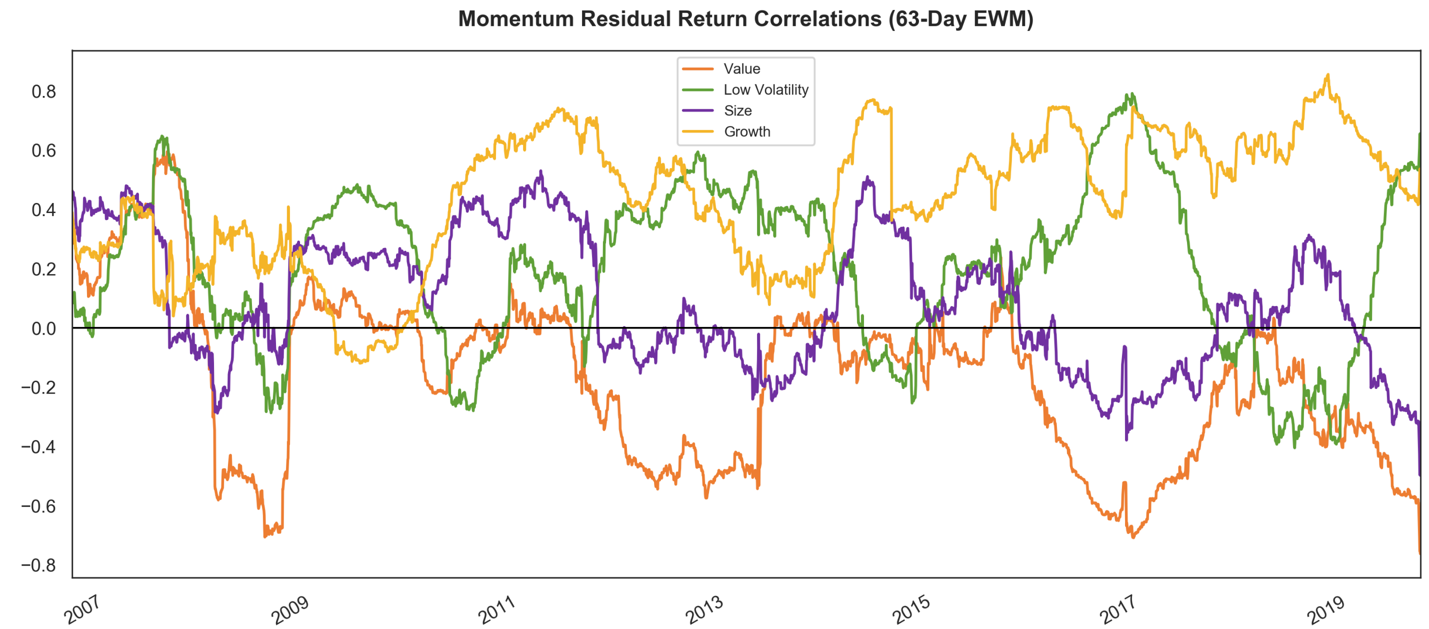
When quants talk about factor crowding, there are really two types of crowding that can occur:

- The first occurs when too many investors crowd into the same trade. The most common way to measure this type of crowding is by looking at valuation spreads between the two legs of the trade. For example, crowding into growth stocks may cause value spreads between growth and value equities to widen to a historically unusual level (e.g. the dot-com era). Similarly, crowding into value stocks would lead to an unusually compressed spread in relative valuation between value and growth stocks.

- The second is when multiple factors crowd into the same trade. This can occur when multiple factors – especially when implemented in an unconstrained fashion – identify the same securities as being attractive. For example, if low-volatility stocks are out-performing their peers, momentum portfolios may rotate into these winners. This can cause two sets of investors to rotate into the same securities, causing crowding.

The latter form of crowding can be measured through holdings-based analysis (e.g. determining positional overlaps) or statistical analysis (e.g. correlation-based).

Below we plot the rolling correlation between residual momentum style returns with other popular factors. We can see that while momentum has historically had a relatively high correlation with **–growth–**, its correlation with **–low volatility–** is significantly time-varying. Prior to this week, the correlation between momentum and low volatility had climbed from -0.3 to north of 0.5 in 2019.



Source: CSI Data; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

(Note that the above graph is available and updated daily on our Equity Style Dashboard.[50])

As proponents of style-based investing, we ask ourselves: what to do? Should we dynamically adjust our allocation to factors based upon crowding risk? If so, is the goal better performance (avoid highly crowded, and therefore lower return styles) or better risk management (maintain more consistent diversification across our style bets)?

In this pursuit, do we risk chasing signal only to realize noise?

These are non-trivial questions and their answers have important ramifications for portfolio design. While we believe these are questions worth exploring further, we want to start with perhaps one that is more trivial: are all styles even appropriate for all investors?

Factors and Financial Planning

This is not a new question for us. In June 2017 we wrote *Factors & Financial Planning*, employing capital market assumptions for stocks, bonds, and a variety of equity styles to construct different risk-based portfolios. We found that

[50] https://www.thinknewfound.com/style-dashboard

for growth-based investors, a barbell of value and momentum was preferred while for more conservative investors, a more diversified split of factors was prudent.

In our July 2018 note *The New Glide Path* we explored the use of trend equity strategies in designing a multi-asset glide path. Our process followed a backwards induction framework similar to Gordon Irlam's article *Portfolio Size Matters* (Journal of Portfolio Finance, Vol 13 Issue 2). The general process was:

1. Starting at age 100, assume a success rate of 100% for all wealth levels except for $0, which has a 0% success rate.

2. Move back in time T years and generate N real return simulations.

3. For each possible wealth level and each possible portfolio configuration of our assets, use the N simulations to generate N possible future wealth levels, subtracting the real annual spend level.

4. For a given simulation, use standard mortality tables to determine if the investor died during the year. If he did, set the success rate to 100% for that simulation. Otherwise, set the success rate to the success rate of the wealth bucket the simulation falls into at T+1.

5. For the given portfolio configuration, set the success rate as the average success rate across all simulations.

6. For the given wealth level, select the portfolio configuration that maximizes success rate. If multiple portfolios guarantee success, choose the most conservative portfolio. If multiple portfolios maximize the success rate, average them together.

7. Return to step 2.

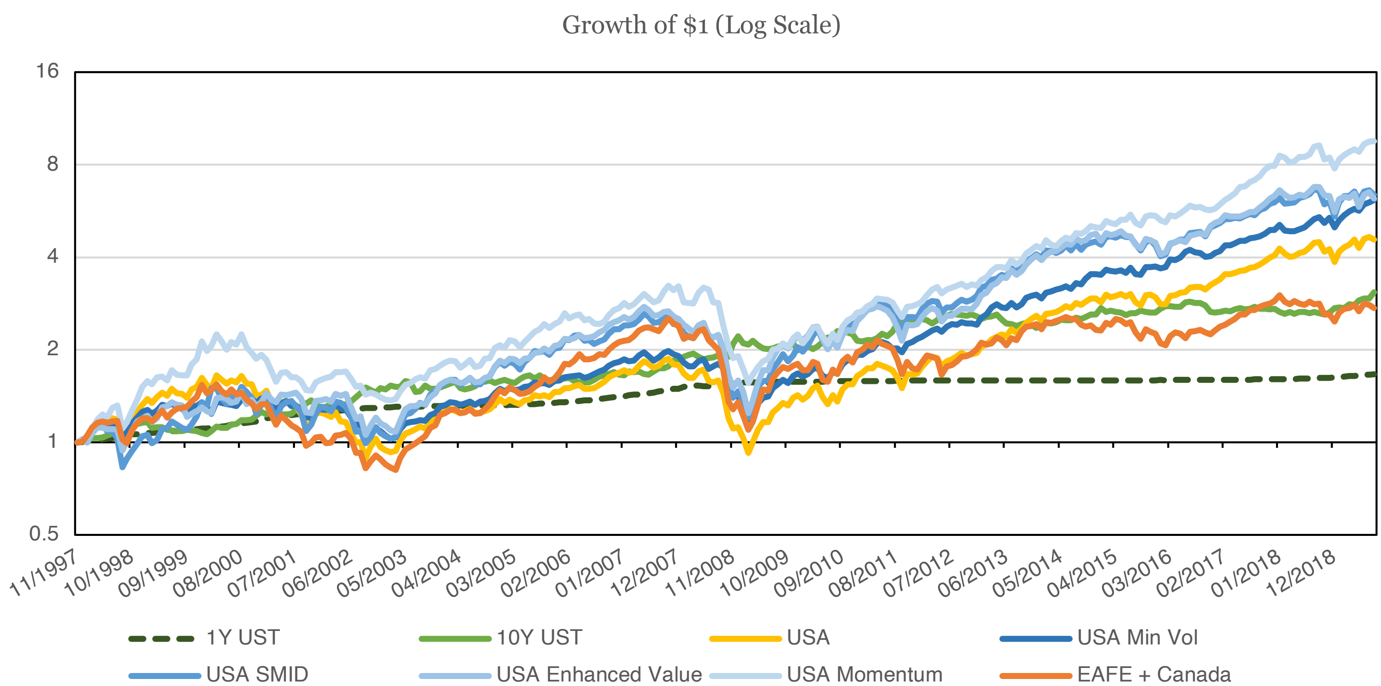
To quote *The New Glide Path*:

> *By working backwards, we can tackle what would be an otherwise computationally intractable problem. In effect, we are saying, "if we know the optimal decision at time T+1, we can use that knowledge to guide our decision at time T."*
>
> *This methodology also allows us to recognize that the relative wealth level to spending level is important. For example, having $2,000,000 at age 70 with a $40,000 real spending rate is very different than having $500,000, and we would expect that the optimal allocation would different.*
>
> *Consider the two extremes. The first extreme is we have an excess of wealth. In this case, since we are optimizing to maximize the probability of success, the result will be to take no risk and hold a significant amount of T-Bills. If, however, we had optimized to acknowledge a desire to bequeath wealth to the next generation, you would likely see the opposite extreme: with little risk of failure, you can load up on stocks and to try to maximize growth.*

The second extreme is having a significant dearth of wealth. In this case, we would expect to see the optimizer recommend a significant amount of stocks, since the safer assets will likely guarantee failure while the risky assets provide a lottery's chance of success.

In this note, we employ the process utilized in *The New Glide Path* but use an investment universe that includes a number of U.S. equity styles. Specifically, our investment universe will consist of:

- Equities: U.S. Total Market, EAFE + Canada.
- Bonds: 1-year US Treasuries, 10-year U.S. Treasuries.
- Long-Only Styles: Value, Size, Momentum, Low-Volatility.



Source: MSCI; Global Financial Data; Calculations by Newfound Research. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results.

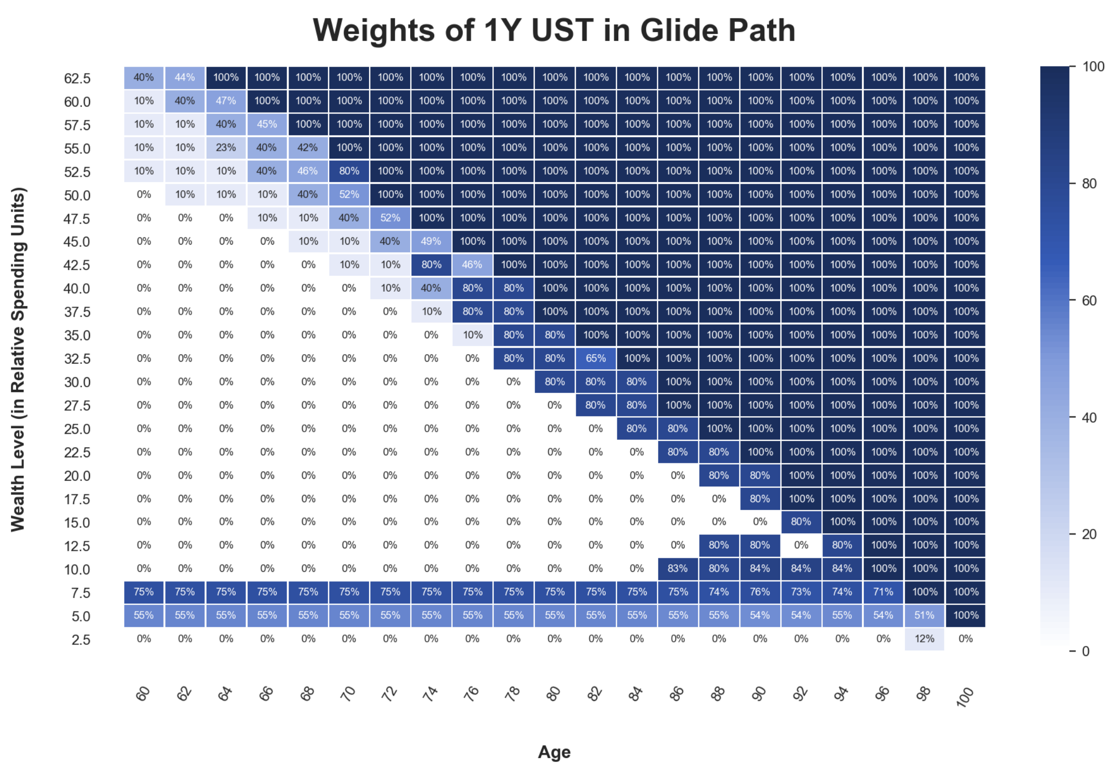
As a technical side note, we should note that generating all possible portfolio variations of these 8 assets would be computationally taxing, as would be an optimization-based approach. Therefore, we assume that each asset's allocation is one of the following values: 0%, 10%, 20%, 40%, 60%, 80%, or 100%. We also assume that portfolios all sum to 100%. While this leads to rather coarsely defined portfolios, the objective of our search is not precision, but rather directional guidance as to whether certain factors should be preferred at different points in an investor's lifecycle. These choices lead to 4,803 possible portfolio configurations.

Before we present our results, we should offer a few words of caution on reading too deeply into them.

- The joint dataset only goes back to 1997, severely biasing the test towards results realized over the last 20 years. Knowing that value has struggled, for example, we would expect value to be under-represented in our results. Further, being a low and stable inflation regime, we might see a larger representation of fixed income assets.

- A large proportion of the style data is backtested and therefore suffers all the usual risks of backtested results. Specifically, we risk that these results are overstated (and potentially in dramatic fashion), which may lead to an over-emphasis on certain factors.

With those risks in mind, results of our test are depicted graphically below. We will use the 1-year U.S. Treasury results as an example of how to read the graphs.

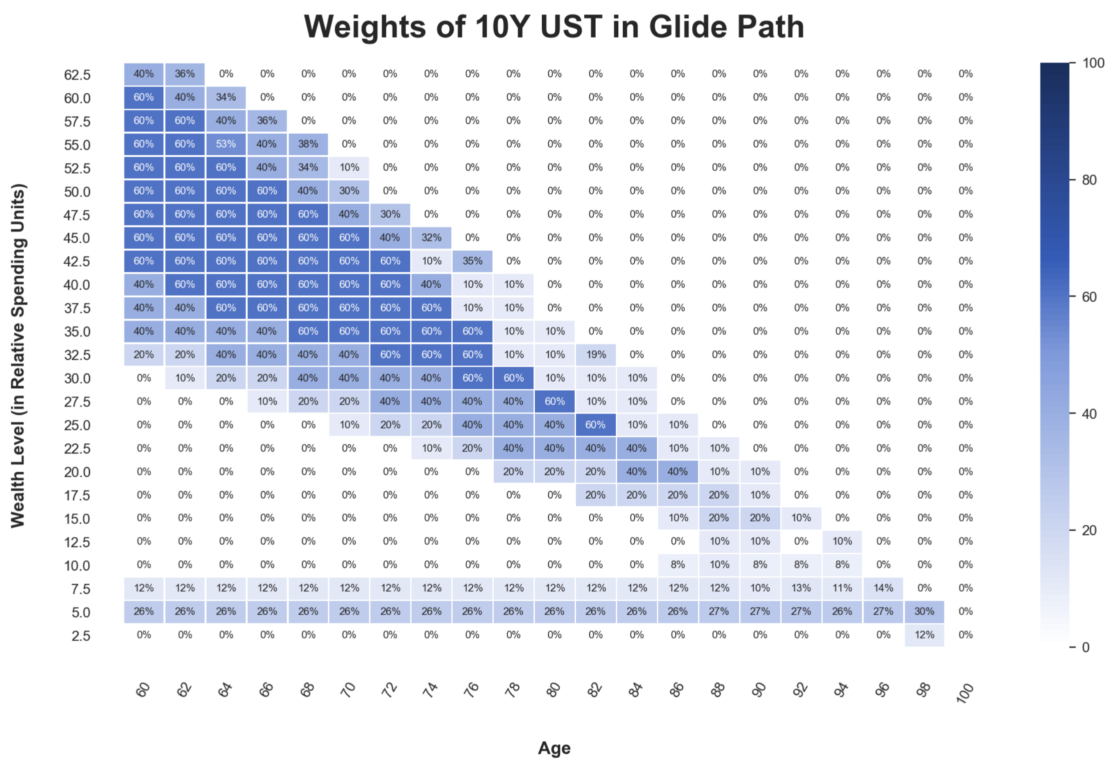


Each cell represents the optimal portfolio's allocation to 1-year U.S. Treasuries.

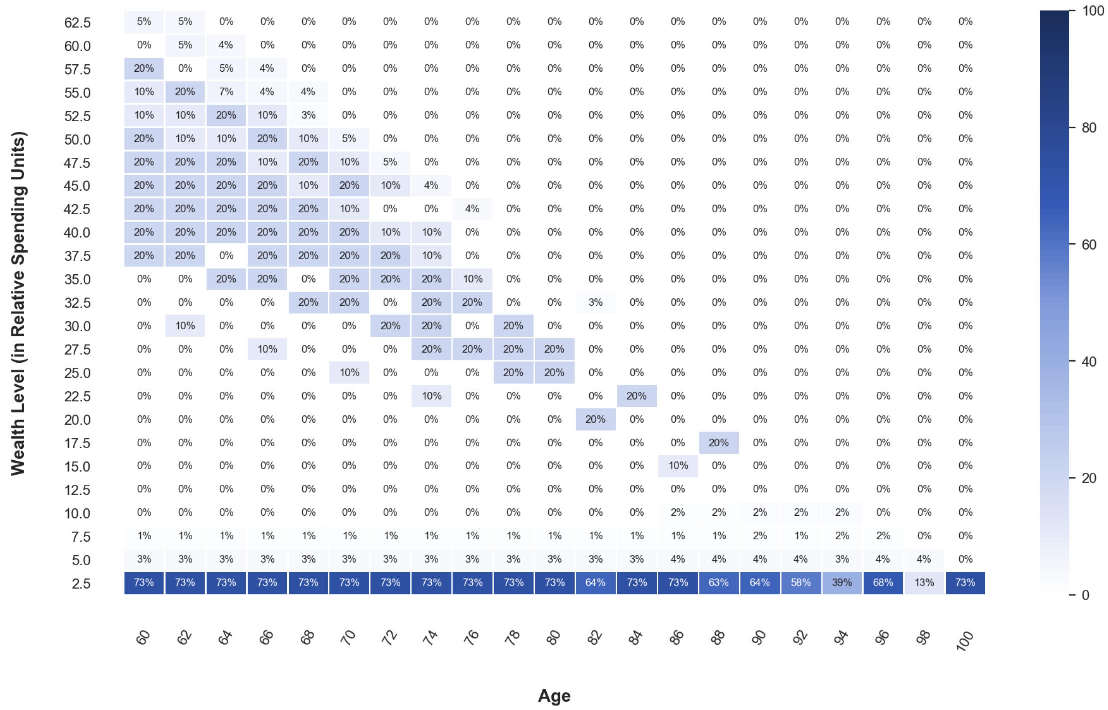
Along the bottom axis we have the investor's age.

The vertical axis represents "relative spending units": i.e. how much wealth an investor has relative to the amount they would like to spend each year in real terms. For example, if an investor has $1,000,000 and would like to spend $40,000, they have 25 relative spending units.

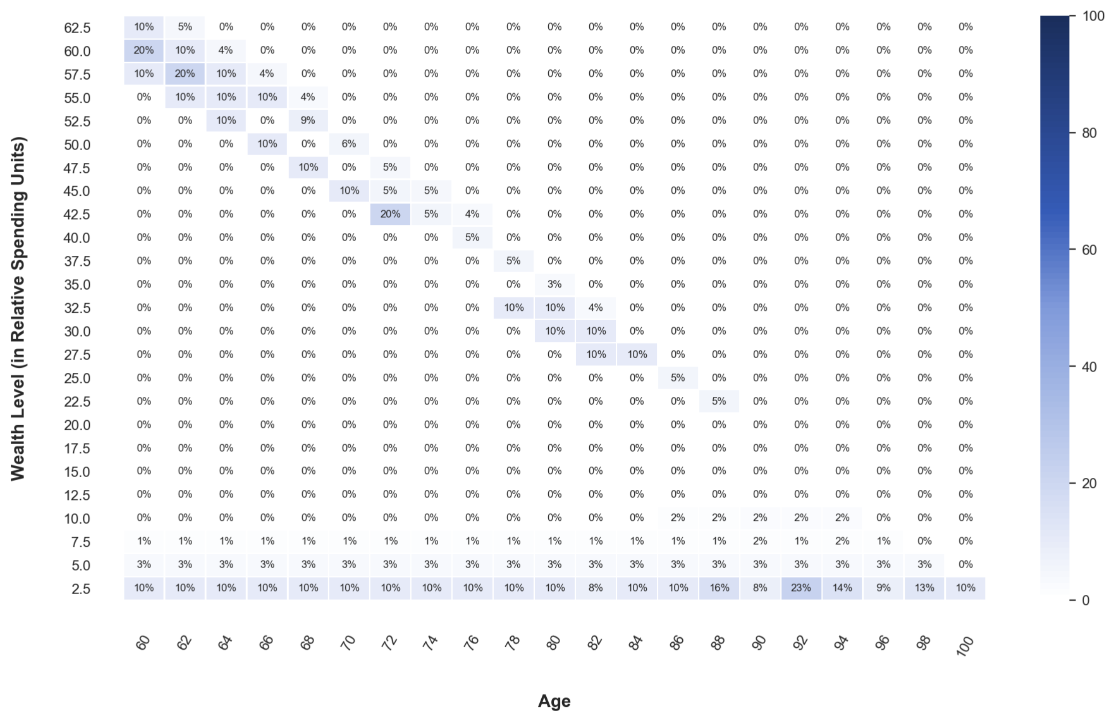
We can see that for high relative spending units, the investor can hold a significant proportion of their wealth in short-term U.S. Treasuries (bearing in mind our note above about understated inflation risk). This is because the investor is more likely to die before they out-spend their wealth. On the other hand, we can see no short-term U.S. Treasury allocation for lower relative spending units earlier in an investor's life, signaling the importance of growing assets in that time frame.

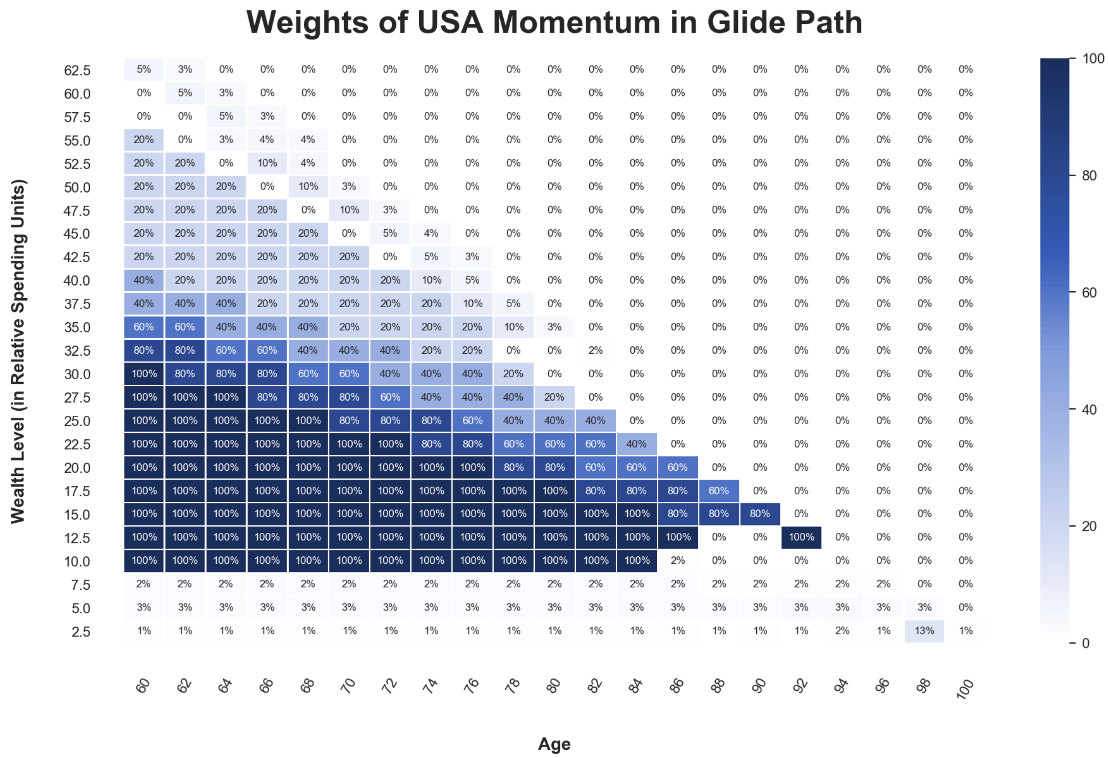


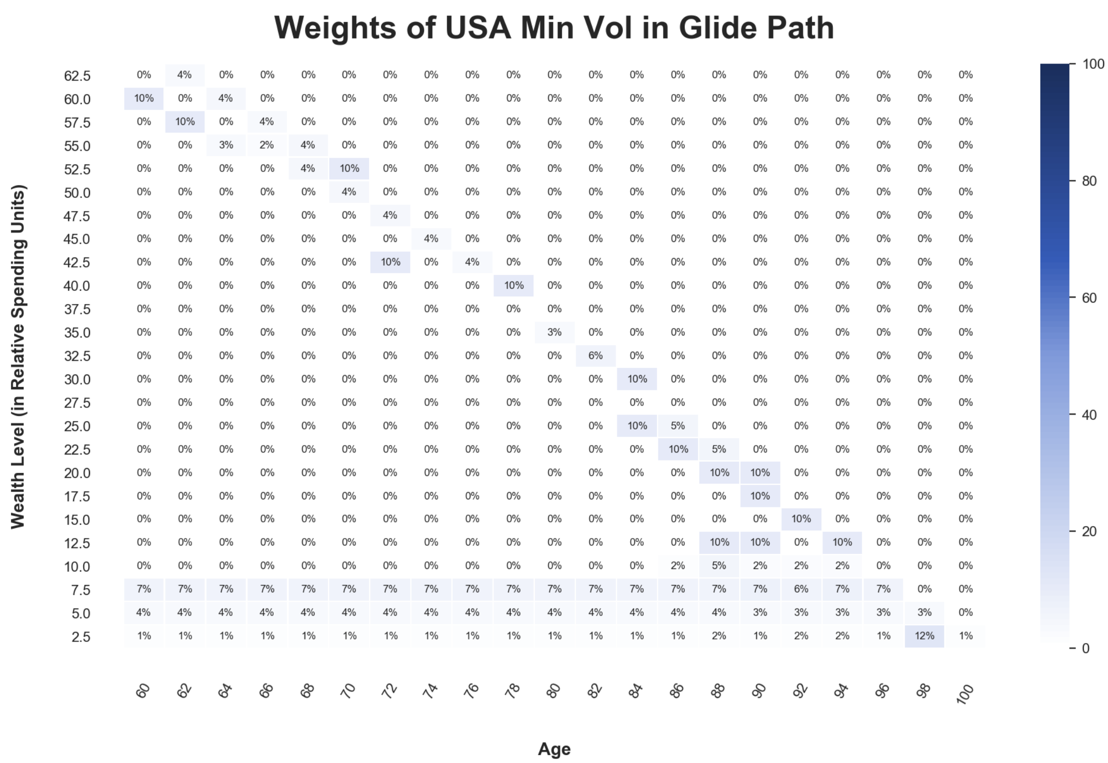
Weights of USA Enhanced Value in Glide Path



Weights of USA SMID in Glide Path




Weights of USA Momentum in Glide Path
Wealth Level (in Relative Spending Units)
Age
62.5 60.0 57.5 55.0 52.5 50.0 47.5 45.0 42.5 40.0 37.5 35.0 32.5 30.0 27.5 25.0 22.5 20.0 17.5 15.0 12.5 10.0 7.5 5.0 2.5
60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 100
100 80 60 40 20 0


Weights of USA Min Vol in Glide Path
Wealth Level (in Relative Spending Units)
Age
62.5 60.0 57.5 55.0 52.5 50.0 47.5 45.0 42.5 40.0 37.5 35.0 32.5 30.0 27.5 25.0 22.5 20.0 17.5 15.0 12.5 10.0 7.5 5.0 2.5
60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 100
100 80 60 40 20 0

Evaluating all the assets we can see:

- The upper right half of the graph largely represents the "zone of safety" where an investor has a sufficient amount of wealth relative to their spending plan that they can invest very conservatively and still succeed. Numerically, if relative spending units exceed -1.25 x Age + 142.5, the investor is generally very safe.

- Below the "zone of safety" is the "cone of balance." The cone is largest for the early retiree, likely representing the need to balance growth and capital preservation and declines in size as they age. Allocations to 10-year U.S. Treasuries are approximately 40-60% in this range. Equities within this range are largely a mixture between value and momentum.

- Below the "cone of balance" is the "triangle of growth." Early retirees with insufficient relative spending units need to focus on growing their capital (or decreasing their spending, thereby increasing their relative spending units) to avoid running out of money later.

- Finally, below the "triangle of growth" is the "twilight zone." This is a zone where investors need to both desperately growth their capital while simultaneously desperately avoiding losses. We can see that early in the zone, preservation is preferred. However, once deep in the zone – where an investor is guaranteed to withdraw the remainder of their capital – the strategy piles in on equities as a last ditch effort to grow.

- Most curiously, at least to us, is that minimum volatility securities fail to make a showing in this test. This may be due to several facts: (1) insufficient data; (2) poor test design; or (3) the specification of the minimum volatility portfolio we employed. Or, conversely, it may actually indicate that U.S. Treasury bonds were a better vehicle when total portfolio construction was considered.

Conclusion

In this research note, we ask the question, "should all equity styles be considered equally by investors in all stages of their lifecycle?" While we specifically focused on individual retirees, we believe this analysis extends naturally to pensions and other liability-driven institutions.

To explore this question, we designed a test that sought to maximize an individual investor's likelihood of achieving success in retirement, where success is (rather morbidly) defined as dying before running out of money.

The result of this test identified four critical zones:

- "Zone of Safety": Investors have sufficient capital to meet withdrawal needs and can invest as conservatively as they wish.

- "Cone of Balance": Investors likely have sufficient savings to retire but need to balance both long-term growth objectives with capital preservation. Here we found that 10-year U.S. Treasuries were balanced with value and momentum equities.

- "Triangle of Growth": Investors in this zone will outspend their net worth unless they grow their assets, so the portfolio tilts heavily towards momentum equities.

- "Twilight Zone": Investors in this zone have so few assets that they must protect them at all cost. However, if assets dwindle too far, the portfolio swings wildly back towards equities, as a bet on growth is the only opportunity for avoiding failure.

We should stress that the results of our study are specific to our study design and that alternative designs may lead to different conclusions. However, we believe that the general conclusions do line up well with our intuition and highlight that not all investment styles may necessarily be appropriate for investors at all points in their investment lifecycle.

TREND FOLLOWING ACTIVE RETURNS

September 23, 2019

SUMMARY

- Recent research suggests that equity factors exhibit positive autocorrelation, providing fertile ground for the application of trend-following strategies.
- In this research note, we ask whether the same techniques can be applied to the active returns of long-only style portfolios.
- We construct trend-following strategies on the active returns of popular MSCI style indices, including Value, Size, Momentum, Minimum Volatility, and Quality.
- A naïve, equal-weight portfolio of style trend-following strategies generates an information ratio of 0.57.
- The interpretation of this result is largely dependent upon an investor's pre-conceived views of style investing, as the diversified trend-following approach generally under-performs a naïve, equal-weight portfolio of factors except during periods of significant and prolonged factor dislocation.

There have been a number of papers published in the last several years suggesting that positive autocorrelation in factor returns may be exploitable through time-series momentum / trend following. For example,

- Ehsani and Linnainmaa (2017; revised 2019)[51] document that "most factors exhibit positive autocorrelation with the average factor earning a monthly return of 2 basis points following a year of losses but 52 basis points following a positive year."
- Renz (2018)[52] demonstrates that "risk premiums are significantly larger (lower) following recent uptrends (downtrends) in the underlying risk factor."
- Gupta and Kelly (2018; revised 2019)[53] find that, "in general, individual factors can be reliably timed based on their own recent performance."

[51] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3014521
[52] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3100165
[53] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3300728

- Babu, Levin, Ooi, Pedersen, and Stamelos (2019)[54] find “strong evidence of time-series momentum” across the 16 long/short equity factors they study.

While this research focuses mostly only long/short equity factors, it suggests that there may be opportunity for long-only style investors to improve their realized results as well. After all, long-only “smart beta” products can be thought of as simply a market-cap benchmark plus a dollar-neutral long/short portfolio of active bets.

Therefore, calculating the returns due to the active bets taken by the style is a rather trivial exercise: we can simply take the monthly returns of the long-only style index and subtract the returns of the long-only market-capitalization-weighted benchmark. The difference in returns will necessarily be due to the active bets.[55]

Below we plot the cumulative active returns for five popular equity styles: Value (MSCI USA Enhanced Value), Size (MSCI USA SMID), Momentum (MSCI USA Momentum), Minimum Volatility (MSCI USA Minimum Volatility), and Quality (MSCI USA Quality).

The active returns of these indices certainly rhyme with, but do not perfectly replicate, their corresponding long/short factor implementations. For example, while Momentum certainly exhibits strong, negative active returns from 6/2008 to 12/2009, the drawdown is nowhere near as severe as the “crash” that occurred in the pure long/short factor.

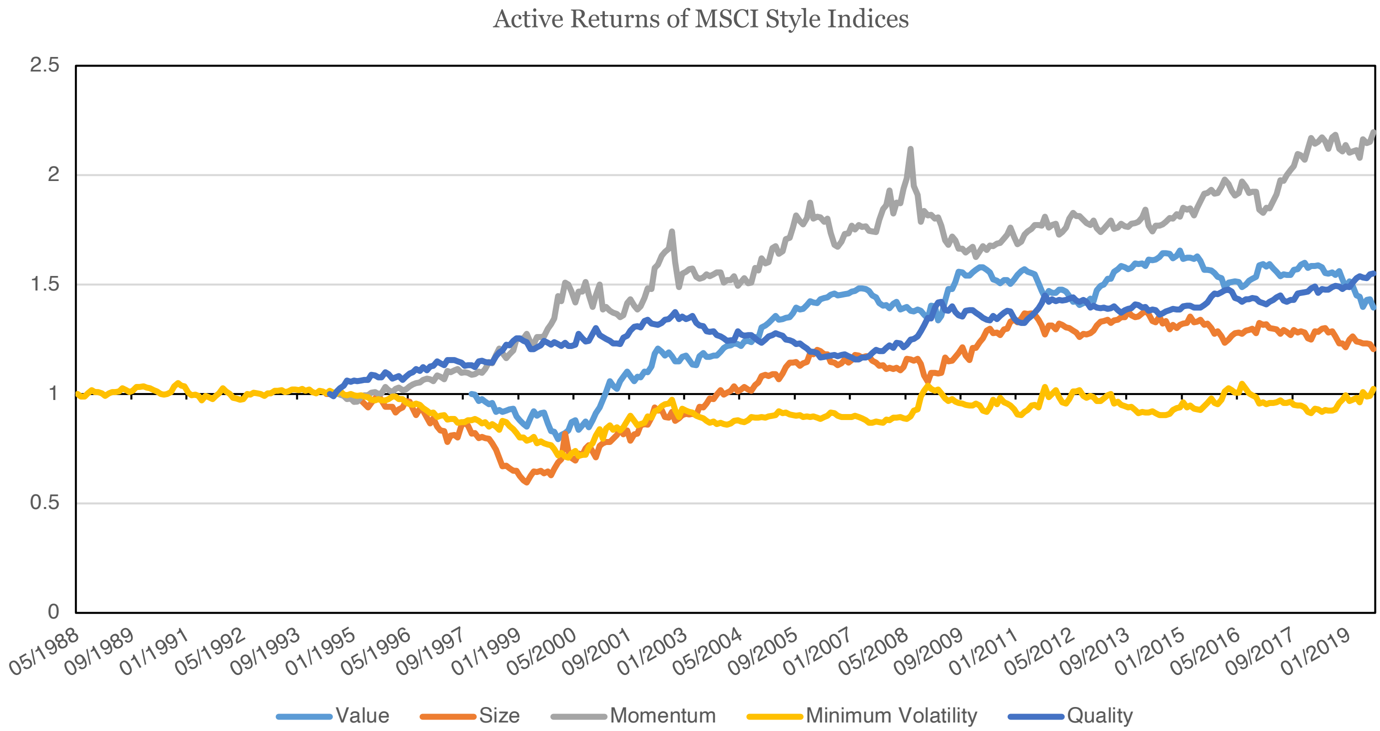
This is due to two facts:

1. The implied short side of the active bets is constrained by how far it can take certain holdings to zero. Therefore, long-only implementations tend to over-allocate towards top-quintile exposures rather than provide a balanced long/short allocation to top- and bottom-quintile exposures.

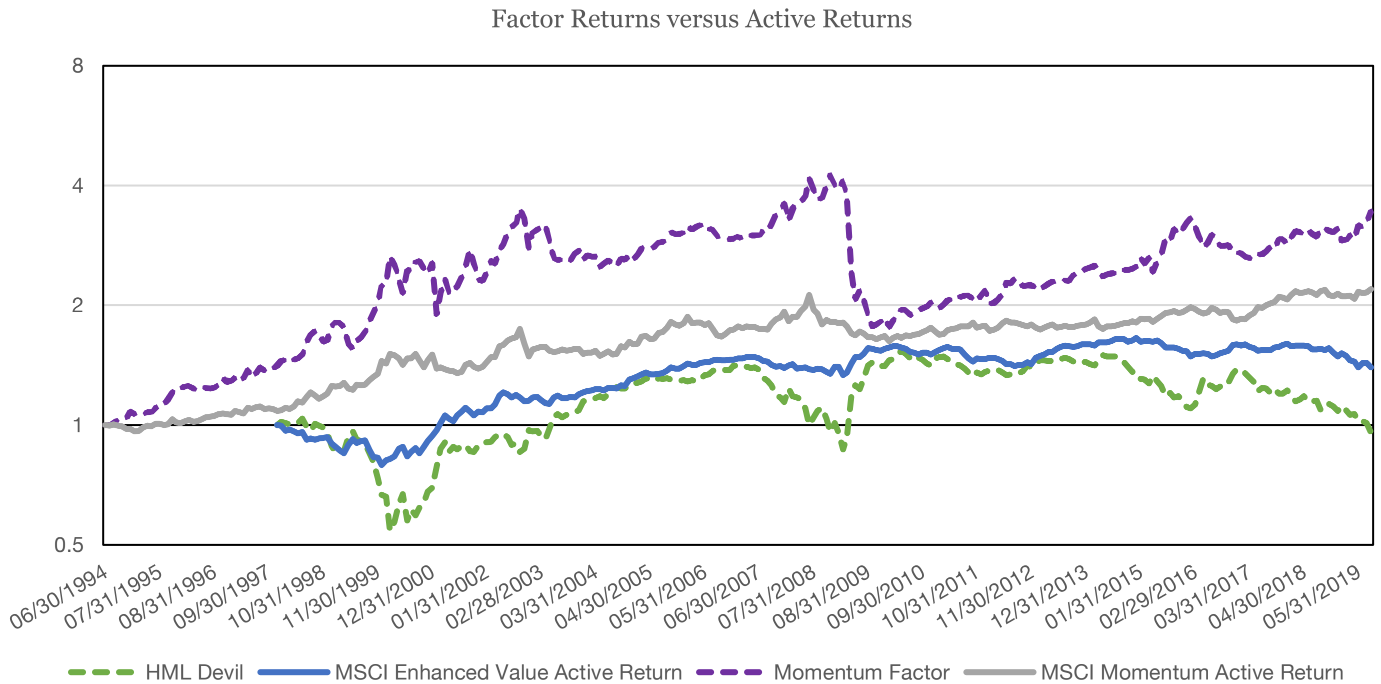
2. While the active bets form a long/short portfolio, the *notional* size of that portfolio is often substantially lower than the academic factor definitions (which, with the exception of betting-against-beta, more mostly assumed to have a notional exposure of 100% per leg). The active bets, on the other hand, have a notional size corresponding to the portfolio’s active share, which frequently hovers between 30-70% for most long-only style portfolios.

3. The implementation details of the long-only style portfolios and the long/short factor definitions may not perfectly match one another. As we have demonstrated a number of times in past research commentaries, these specification details can often swamp style returns in the short run, leading to meaningful cross-sectional dispersion in same-style performance.

[54] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3386035
[55] For indices, at least. For actively managed portfolios, implementation effects and costs will also come into play.



Source: MSCI. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.



Source: MSCI; AQR. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.

Nevertheless, "rhymes but does not replicate" may be sufficient for long-only investors to still benefit from trend-following techniques.

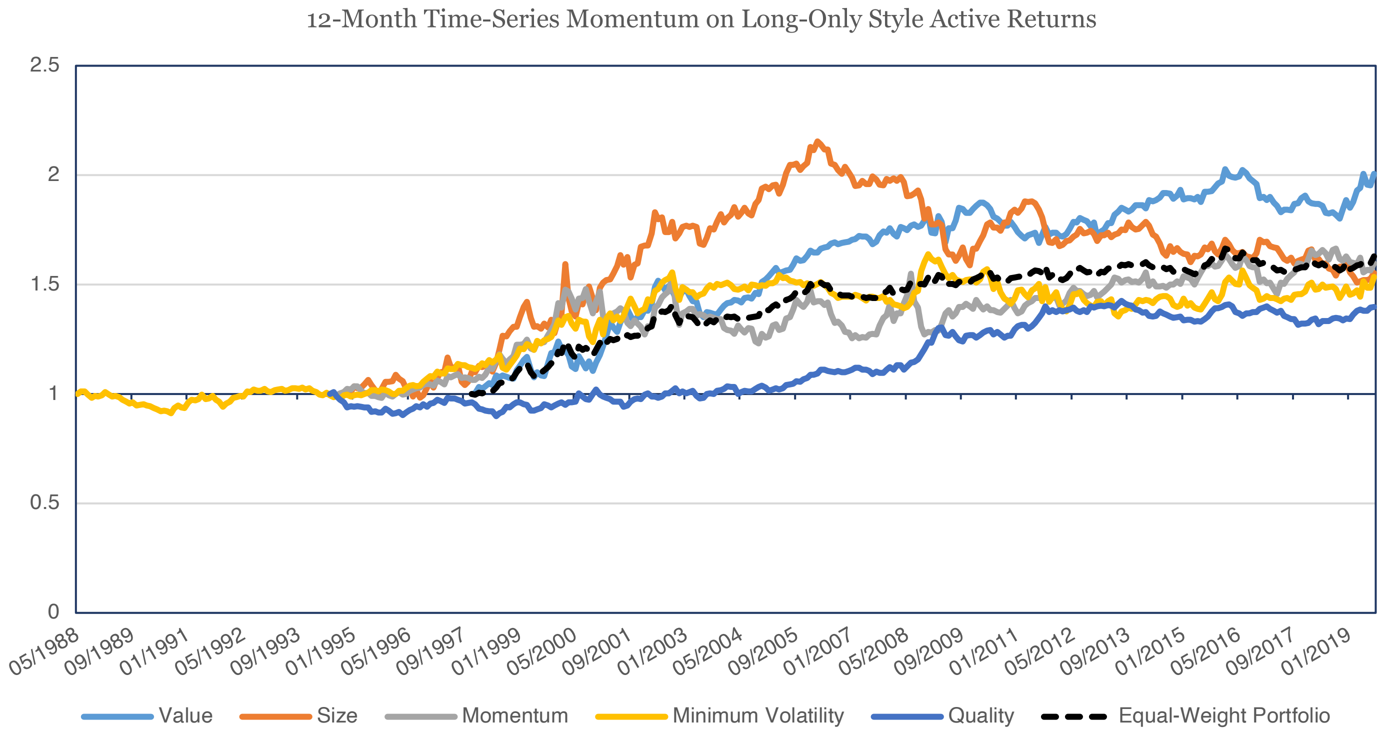
In our test, we will go long the style / short the benchmark (i.e. long active returns) when prior N-month returns are positive and short the style / long the benchmark (i.e. short active returns) when prior N-month returns are negative. Portfolios are formed monthly at the end of each month. Performance results are reported in the table below for 1, 3, 6, 9, and 12-month lookback periods.

		Annualized Return	Annualized Volatility	Information Ratio	Maximum Drawdown	Sample Size (Months)
	Value	1.7%	6.1%	0.28	-15.1%	261
	Size	-0.8%	8.2%	-0.10	-44.4%	303
1	*Momentum*	-0.2%	7.5%	-0.03	-21.3%	302
	Minimum Volatility	-0.1%	5.7%	-0.01	-25.0%	375
	Quality	1.3%	3.8%	0.35	-8.9%	302

3	*Value*	3.3%	6.0%	0.55	-15.5%	261
	Size	1.1%	8.2%	0.13	-34.5%	303
	Momentum	-0.8%	7.5%	-0.11	-38.0%	302
	Minimum Volatility	0.7%	5.7%	0.13	-19.4%	375
	Quality	0.9%	3.8%	0.24	-10.1%	302
6	*Value*	2.9%	6.0%	0.48	-21.0%	261
	Size	1.7%	8.2%	0.20	-20.8%	303
	Momentum	0.7%	7.5%	0.09	-28.8%	302
	Minimum Volatility	0.5%	5.7%	0.09	-27.8%	375
	Quality	0.6%	3.9%	0.16	-14.6%	302
9	*Value*	3.4%	6.0%	0.57	-14.8%	261
	Size	2.0%	8.2%	0.24	-27.1%	303
	Momentum	1.2%	7.5%	0.16	-23.4%	302
	Minimum Volatility	0.9%	5.7%	0.15	-20.8%	375
	Quality	0.3%	3.9%	0.07	-14.7%	302
12	*Value*	3.2%	6.0%	0.54	-11.2%	261
	Size	1.8%	8.2%	0.22	-29.9%	303
	Momentum	1.9%	7.5%	0.25	-20.0%	302
	Minimum Volatility	1.4%	5.7%	0.24	-17.3%	375
	Quality	1.3%	3.8%	0.34	-11.0%	302

Source: MSCI. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.

Below we plot the equity curves of the 12-month time-series momentum strategy. We also plot a portfolio that takes a naïve equal-weight position across all five trend-following strategies. The naïve blend has an annualized return of 2.3%, an annualized volatility of 4.0%, and an information ratio of 0.57.



Source: MSCI. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.

This analysis at least appears to provide a glimmer of hope for this idea. Of course, the analysis comes with several caveats:

1. We assume that investors can simultaneously generate signals and trade at month end, which may not be feasible for most.

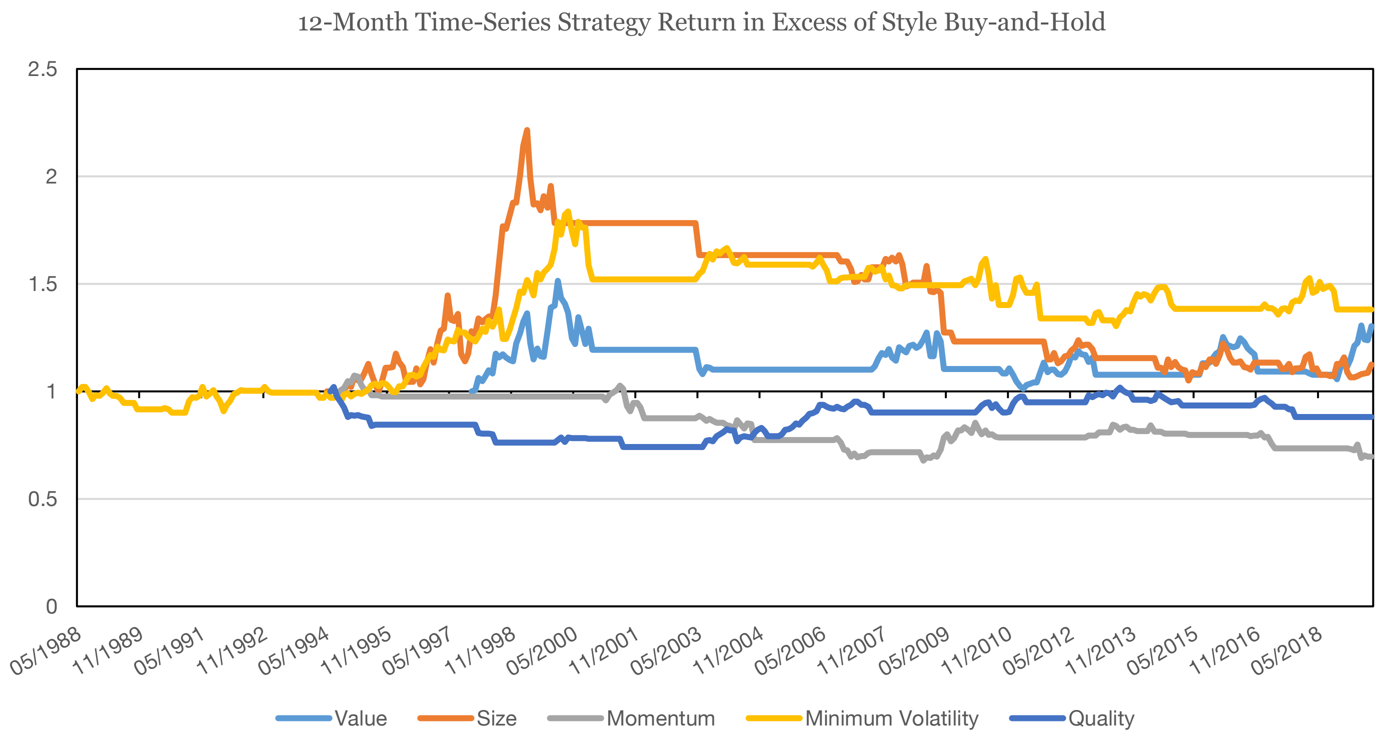
2. We are analyzing index data, which may be different than the realized results of index-tracking ETFs.

3. We do not factor in trading costs such as impact, slippage, or commissions.

It is also important to point out that the *per-style* results vary dramatically. For example, trend-following on the size style has been in a material drawdown since 2006. Therefore, attempting to apply time-series momentum onto of a single style to manage style risk may only invite further strategy risk; this approach may be best applied with an ensemble of factors (and, likely, trend signals).

What this commentary has conveniently ignored, however, is that the appropriate benchmark for this approach is not zero. Rather, a more appropriate benchmark would be the long-only active returns of the styles themselves, as our default starting point is simply holding the styles long-only.

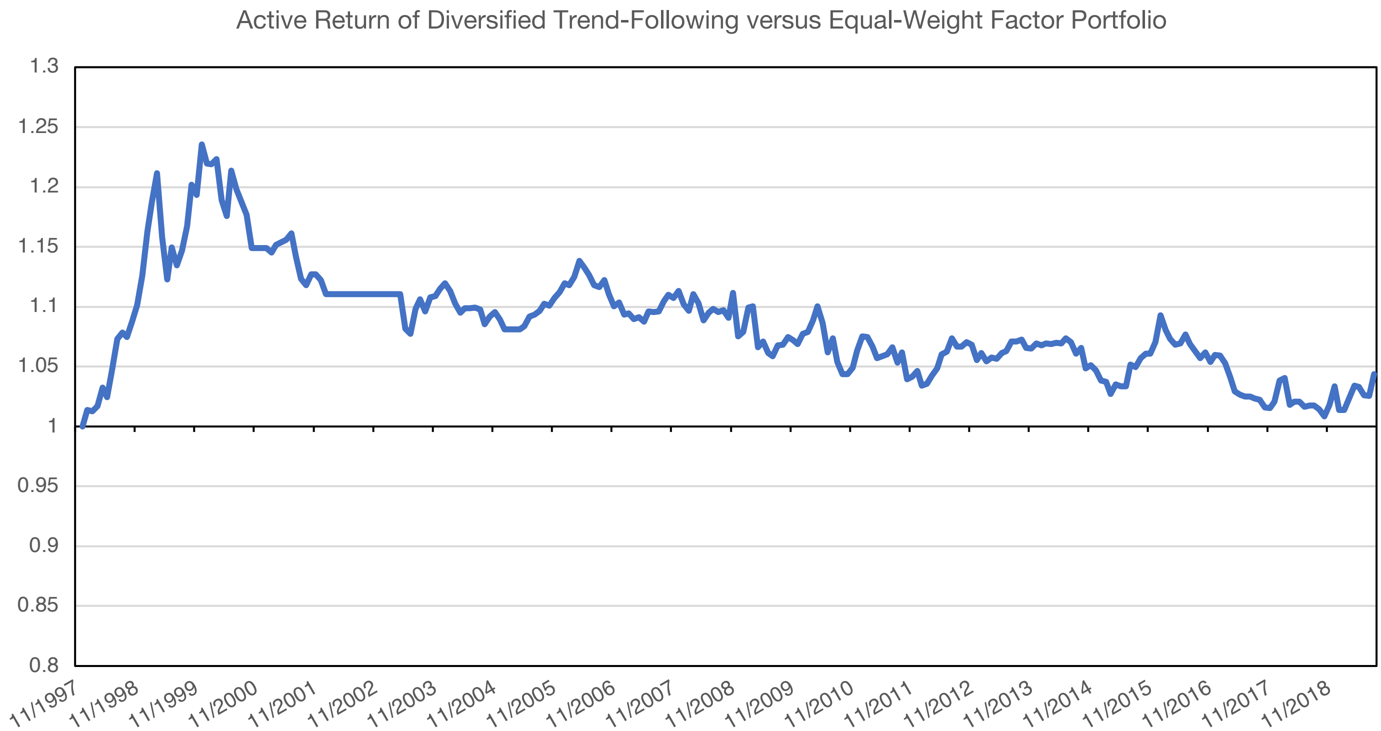
The results, when adjusted for our default of buy-and-hold, is much less convincing.



Source: MSCI. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.

What is clear is that the strategy can now only out-perform when the style is *under*-performing the benchmark. When the portfolio invests in the style, our relative return versus the style is flat.

When a diversified trend-following portfolio is compared against a diversified long-only factor portfolio, we see the general hallmarks of a trend-following approach: value-add during periods of sustained drawdowns with decay thereafter. Trend-following on styles, then, may be more appropriate as a hedge against prolonged style under-performance; but we should expect a cost to that hedge.



Source: MSCI. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.

For some styles, like Minimum Volatility, this appears to have helped relative performance drawdowns in periods like the dot-com bubble without too much subsequent give-up. Size, on the other hand, also benefited during the dot-com era, but subsequently suffered from significant trend-following whipsaw.

Conclusion

Recent research has suggested that equity style premia exhibit positive autocorrelation that can be exploited by trend followers. In this piece, we sought to explore whether this empirical evidence could be exploited by long-only investors by isolating the active returns of long-only style indices.

We found that a naïve 12-month time-series momentum strategy proved moderately effective at generating a timing strategy for switching between factor and benchmark exposure. Per-style results were fairly dramatic, and trend-following added substantial style risk of its own. However, diversification proved effective and an equal-weight portfolio of style trend-following strategies offered an information ratio of 0.57.

However, if we are already style proponents, a more relevant benchmark may be a long-only style portfolio. When our trend-following returns are taken in excess of this benchmark, results deflate dramatically, as the trend-following strategy can now only exploit periods when the style under-performs a market-capitalization-weighted index. Thus, for investors who already implement long-only styles in their portfolio, a trend-following overlay may serve to hedge periods of prolonged style drawdowns but will likely come with whipsaw cost which may drag down realized factor results.

MACRO AND MOMENTUM FACTOR ROTATION

September 30, 2019

SUMMARY

- While many investors have adopted a multi-factor approach to style investing, some have pushed these boundaries by advocating for an active, rotational approach to factor allocation.
- In a recent white paper, MSCI suggests several methods that might be conducive for performing style rotation, including macro-, momentum-, and value-based signals.
- In this commentary, we attempt to test the macro- and momentum-based approaches on (slightly) out-of-sample data.
- We find that both approaches have historically out-performed a naïve, equal-weight factor portfolio. However, the results for the macro-based approach are so good, they raise questions about hindsight bias. Momentum results, on the other hand, are far less compelling on U.S. equity factors than the World equity factors tested by MSCI.
- After appropriately discounting for fees, taxes, and other costs, as well as adequately discounting for testing biases, these methods may not offer much benefit over naïve, equal-weight approach.

While the empirical evidence suggests that factor investing has historically generated an excess (risk-adjusted) return premium over the long-run, short-term performance can be volatile. Because of this, an increasing number of investors are adopting a diversified approach to factor investing, holding multiple factors at once.

Although even simple allocation models – such as naïve equal-weight – have historically harvested these diversification benefits, some researchers believe that dynamic factor allocation can further enhance the returns.

One argument for taking a dynamic approach is that the performance variability is cyclical and linked to different stages of the economic cycle. The different underlying economic drivers lead to differentiated active returns and therefore lead to potential opportunities for cross-factor rotation.

For example, in a recent white paper, MSCI provides several dynamic models in what they call "Adaptive Multi-Factor Allocation."

In this commentary, we replicate two variations of the MSCI's Adaptive Multi-Factor models: the macro-cycle-based and momentum-based allocation methodologies. While the MSCI's methodology is tested on world equity data, we employ US equity data as a (slightly) out-of-sample test. In line with MSCI's research, six long-only US factors – Value, Size, Low Volatility, High Yield, Quality, and Momentum – were selected as our investible universe.

Macro Cycle-Based Allocation

Many research studies suggest that macro indicators have a strong explanatory power to systematic factor returns. One hypothesis is that the excess returns of factors could be compensation for bearing different forms of macroeconomic risk.

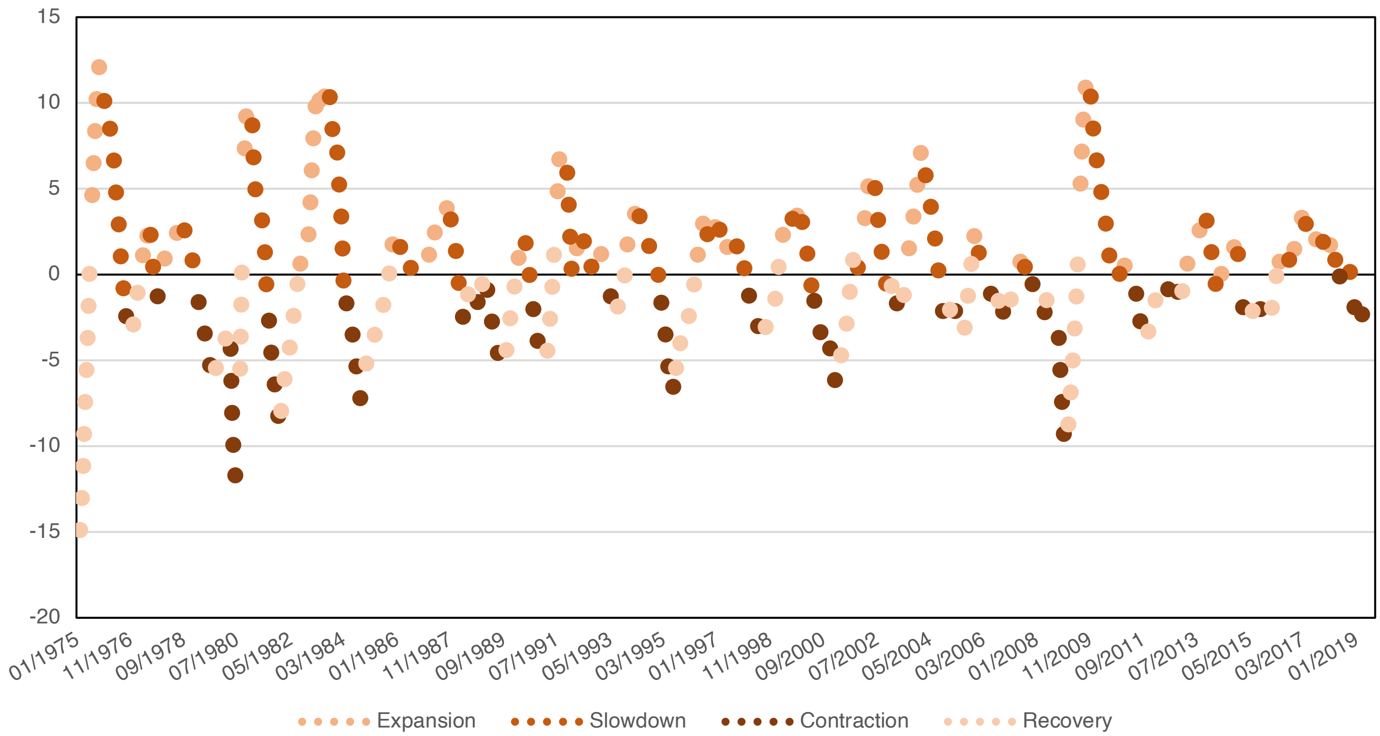
Historical performance of factor investing seems to prove this argument as the returns have been cyclical under different stages of the business cycle. For example, value and size factors tend to be most affected by negative economic growth (providing a risk-based argument for their long-term premium), while quality and low volatility are usually the most defensive factors due to their structurally lower equity betas. Therefore, it might make sense to invest in defensive factors during the periods of economic slowdown. On the contrary, cyclical factors could add more value during expansionary phases.

Following MSCI's methodology, economic cycles are classified into four primary states: Expansion, Slowdown, Contraction and Recovery. Each state is defined based upon the level and slope of a 3-month moving average ("MA") minus a 12-month MA. In this commentary, we will employ three macroeconomic indicators (which are the best 3 performing macro indicators in the MSCI's model):

- PMI (United States ISM Purchasing Managers Index)
- CFNAI (Chicago Fed National Activity Index)
- ADS (Aruoba-Diebold-Scotti Business Conditions Index)

If the 3-month MA is above the 12-month MA and the spread between the two is increasing, the economic state is labeled as an Expansion. If the spread is decreasing, however, the economic state is labeled as a Slowdown. On the other hand, if the 3-month MA is below the 12-month MA and the spread is declining, the state is Contraction. If the spread is negative but increasing, then the economy is in a Recovery

Exhibit 1: Economic Cycles of ISM PMI Index



Source: Quandl PMI Composite Index. Calculations by Newfound Research. Results are hypothetical and should not be used for investment purpose.

An equal-weight portfolio comprising of 3 of the 6 factors is constructed for each state. According to the MSCI paper, the 3 factors for each stage are predetermined based on their historical performance and the past studies on each factor. While we have described our skepticism of these choices in our previous commentary on Style Surfing the Business Cycle, we will assume that the intuition on these factor mixes is correct.

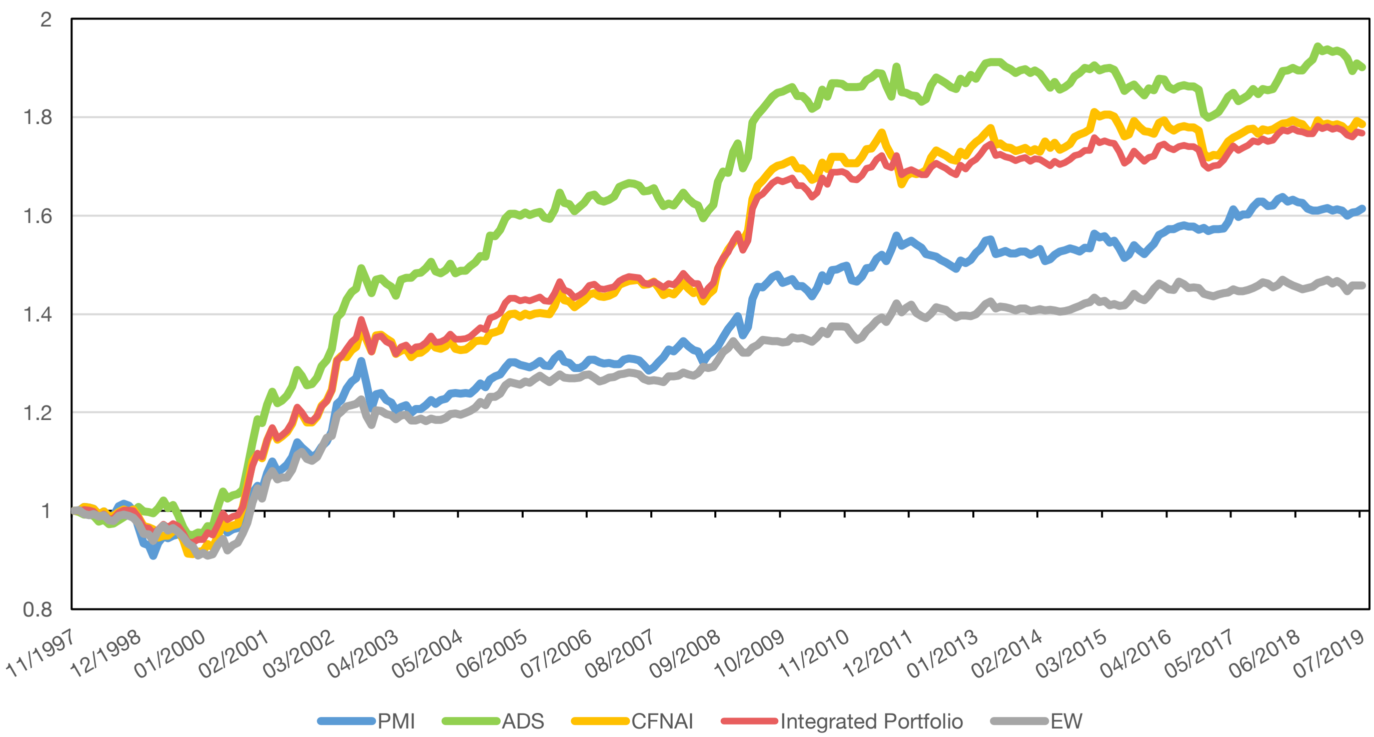
The given combinations are as follows:

Expansion:	Momentum, Size, Value
Slowdown:	Momentum, Quality, Low Volatility
Contraction:	Low Volatility, Quality, Value
Recovery:	Size, Value, High Yield

With the business cycle signals from the moving average cross-overs and the regime-based factor baskets, we can implement the dynamic factor strategies. We construct a portfolio for each indicator, rebalancing monthly based upon the identified economic regime. We also construct an blended portfolio by combining these three sub-portfolios together in hopes of benefiting from signal diversification.

Below we plot the relative returns for each portfolio against the MSCI USA Index and compare them to the equal-weighted portfolio across all 6 factors.

Exhibit 2: Relative Performance of Macro-Cycle Timing Portfolios vs. MSCI USA Index



Source: MSCI USA Standard Universe. Calculations by Newfound Research. Returns are hypothetical and are not intended to be interpreted as recommendation to any portfolio construction. Returns assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index. Sample Period is November 1997 - August 2019.

As we can observe from Exhibit 2, the blended portfolio (and each individual strategy) has meaningfully outperformed both a market-cap weighted benchmark as well as an equal-weight portfolio of the six underlying factors over the past 20+ years. It seems like the macro-cycle-based factor allocation provides a promising return.

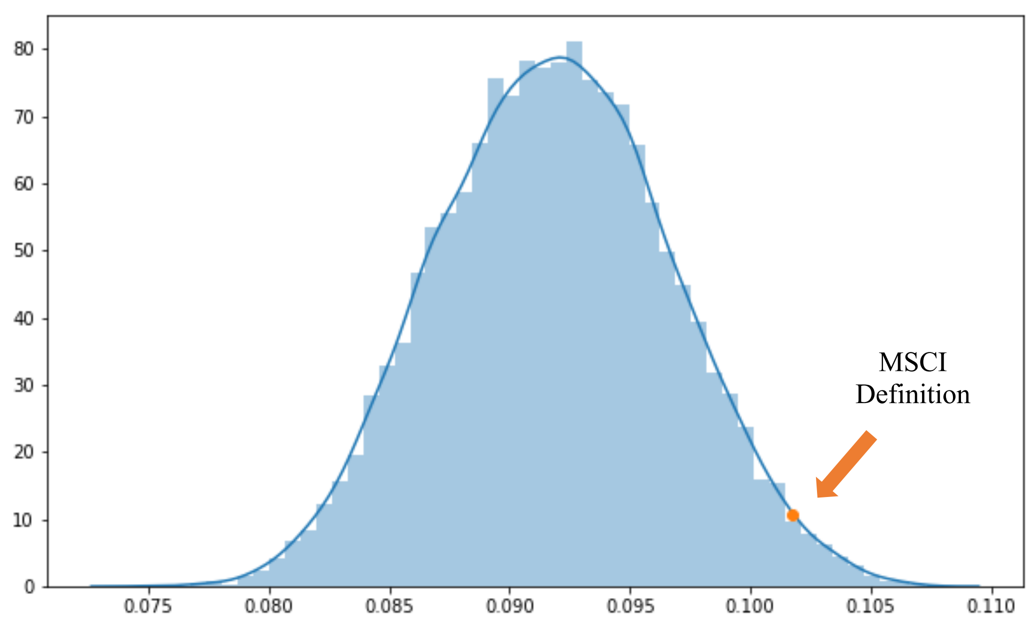
Why is that? One potential reason is that factor returns are linked to the economic cycle and that using monthly (or even daily) updated macro indicators can provide timely insights into the economic state. These higher-frequency signals allow us to potentially capture smaller business cycle fluctuations that are not announced by the NBER (National Bureau of Economic Research) or other institutions. The 3-month vs. 12-month moving average may also help filter undesired noise from the process.

Another potential reason is that the factors were well selected for each identified economic state. However, as we highlighted in our previous commentary, we know the factor allocations for each state were largely determined by the historical performances of each factor during that period of the business cycle.

This raises an important question: is the result a byproduct of data mining or the materialization of an unintentional hindsight bias?

To explore this question, we will perform a random sampling test. Specifically, we will look at the results of alternative portfolio choices we could have made. With four economic states and 3 (out of 6) factors selected for each, there are 160,000 possible economic state / factor portfolio configurations. Below we plot the annualized return distribution of these different configurations and highlight where the MSCI selection falls:

Exhibit 3: Distribution of Random Sampling Test



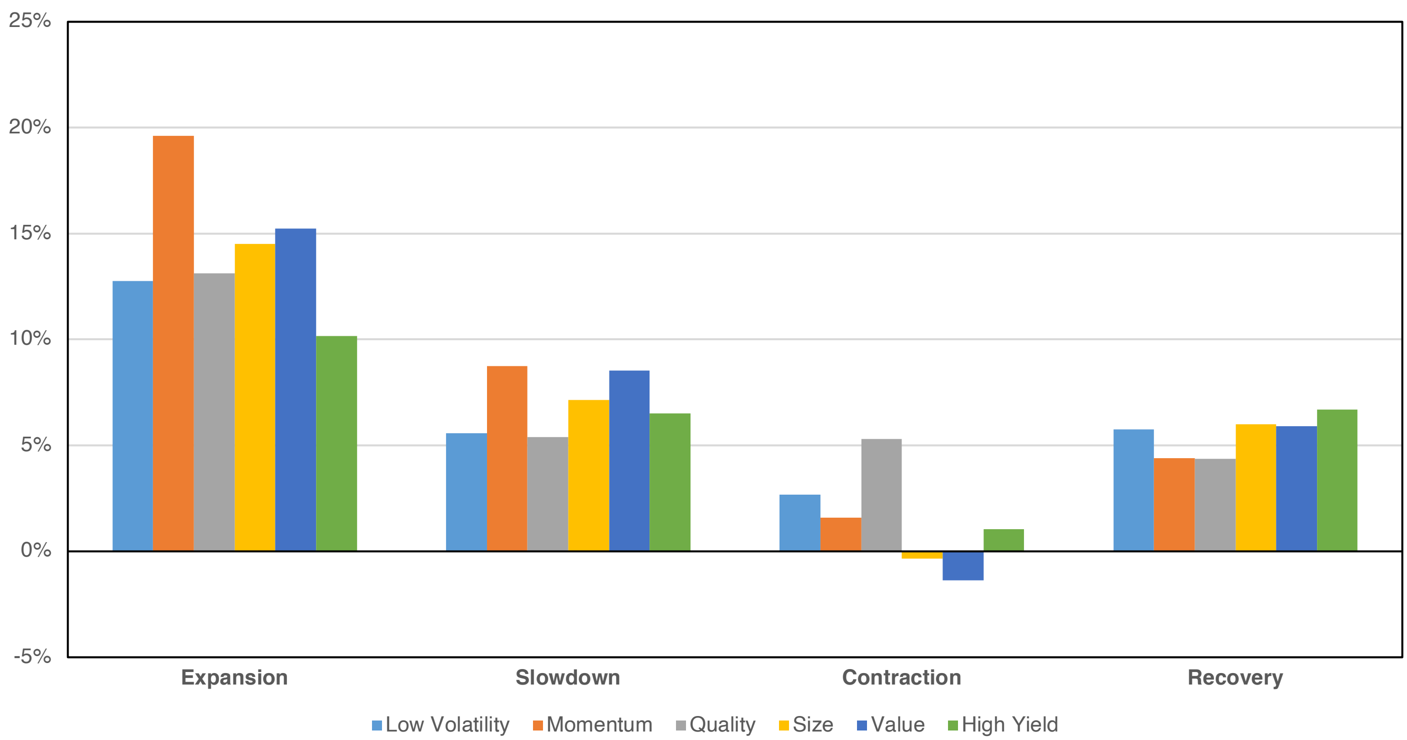
Source: MSCI USA Standard Universe. Calculations by Newfound Research. Returns assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.

Among all the possible combinations, the portfolio defined by MSCI's factor choices lies at the 98th percentile on the annualized return distribution curve. While we would certainly want the choice to perform better than a random selection, such strong performance might suggest the choice was impacted by the benefit of hindsight.

However, it is interesting to note that in the MSCI's definition, value is held in three states of the business cycle (contraction, recovery, and expansion) while the value factor in MSCI's construction, either intuitionally or historically, may not actually be the most appropriate factor for all three periods. For example, during periods of expansion, some argue that market tends to favor companies with high growth potential instead of firms with low intrinsic value.

Below we plot the annualized returns for reach factor during each macro-economic state (as defined by the CFNAI indicator).

Exhibit 4: Annualized Returns for Each Factor During Each Macro State

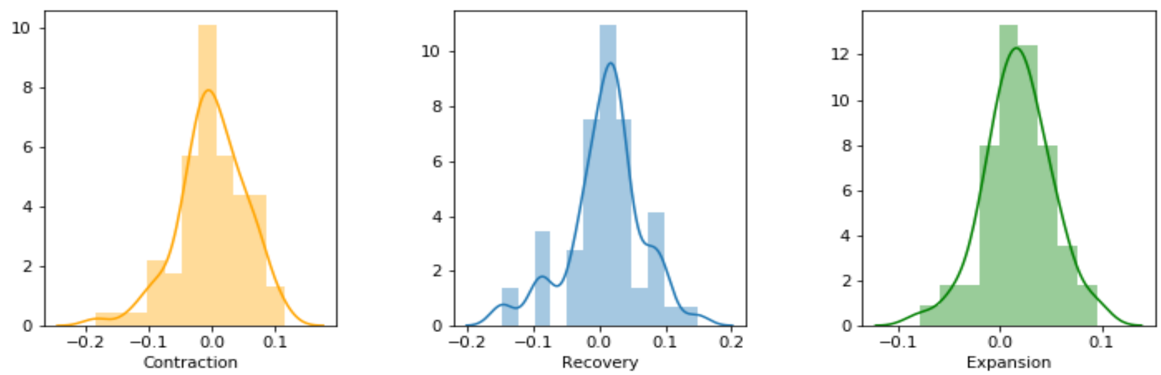


Source: MSCI USA Standard Universe. Calculations by Newfound Research. Returns assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.

Of the 12 possible factor choices to match up with the fully data-mined allocations, MSCI aligns with 9. Still, we cannot assert that the MSCI's predefined rotation rule is a byproduct of pure data mining. It could be a mix of data mining and prevalent beliefs (e.g. the defensive nature of value prior to the 2008 crisis). We should also remember that we are testing U.S. equity factors while the original MSCI research was performed on world equity data, which might lead to subtly different factor choices.

We should also be careful to consider how market unpredictability might negatively skew the returns. As there is no set reason for how or why a financial crisis might unfold, the reliability of using predetermined definitions based solely upon past history may be questionable for future performance.

Exhibit 5: Monthly Return Distribution of MSCI US Value Factor in Contraction, Recovery, and Expansion Phases



Source: MSCI USA Standard Universe. Calculations by Newfound Research. Returns assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index. Sample Period is November 1997 - August 2019.

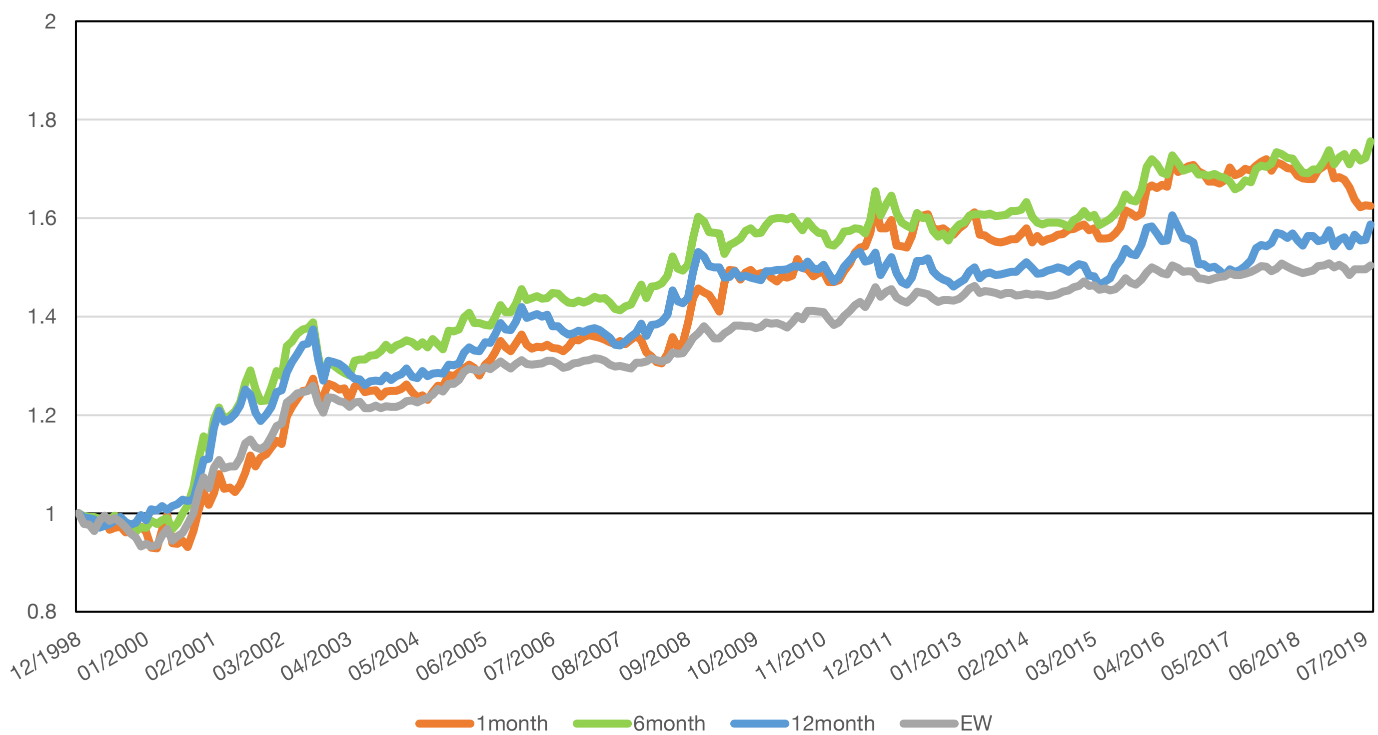
Momentum-Based Allocation

MSCI also tests a dynamic factor construction based upon momentum signals. This approach is also not without academic basis. For example, Research Affiliates performed a study on the momentum effect amongst 51 factors and found that factors exhibit stronger momentum than both individual stocks and industries. They found that momentum is a prevailing property of almost all factors.

To test the viability of momentum-based allocation, we follow MSCI's methodology and rank the factors based upon their prior returns, rebalancing monthly and holding the top 3 ranked factors. The ranking is calculated based upon the last 1-month, 6-month, and 12-month total returns for each factor.

Below we plot the relative performance for each formation period versus a benchmark index. We also plot the relative performance of a naïve, equal-weight factor portfolio. Exhibit 6 plots the approach applied to U.S. equity factors while Exhibit 7 attempts to replicate MSCI's original results with World equity factors.

Exhibit 6: Relative Performance of Momentum-Based Multi-Factor Portfolios vs MSCI USA Index



Source: MSCI USA Standard Universe. Calculations by Newfound Research. Returns are hypothetical and are not intended to be interpreted as recommendation to any portfolio construction. Returns assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index. Sample Period is December 1998 - August 2019.

Exhibit 7: Relative Performance of Momentum-Based Multi-Factor Portfolios vs MSCI World Index

Source: MSCI World Standard Universe. Calculations by Newfound Research. Returns are hypothetical and are not intended to be interpreted as recommendation to any portfolio construction. Returns assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index. Sample Period is December 1998 - August 2019.

In line with the MSCI's results, all three momentum-based indicators generate excess returns over the benchmark as well as the equal-weighted portfolio. The best performing indicator is the 6-month variation. However, our out-of-sample test using US equity factors failed to generate similar returns compared to the World equity factors. It is probably because that there is a stronger factor momentum on the global level.

Comparing the Two Methodologies

Now that we have introduced our test results for macro-based and momentum-base dynamic factor allocation, we want to compare their performance of . Summary performance information is reported below:

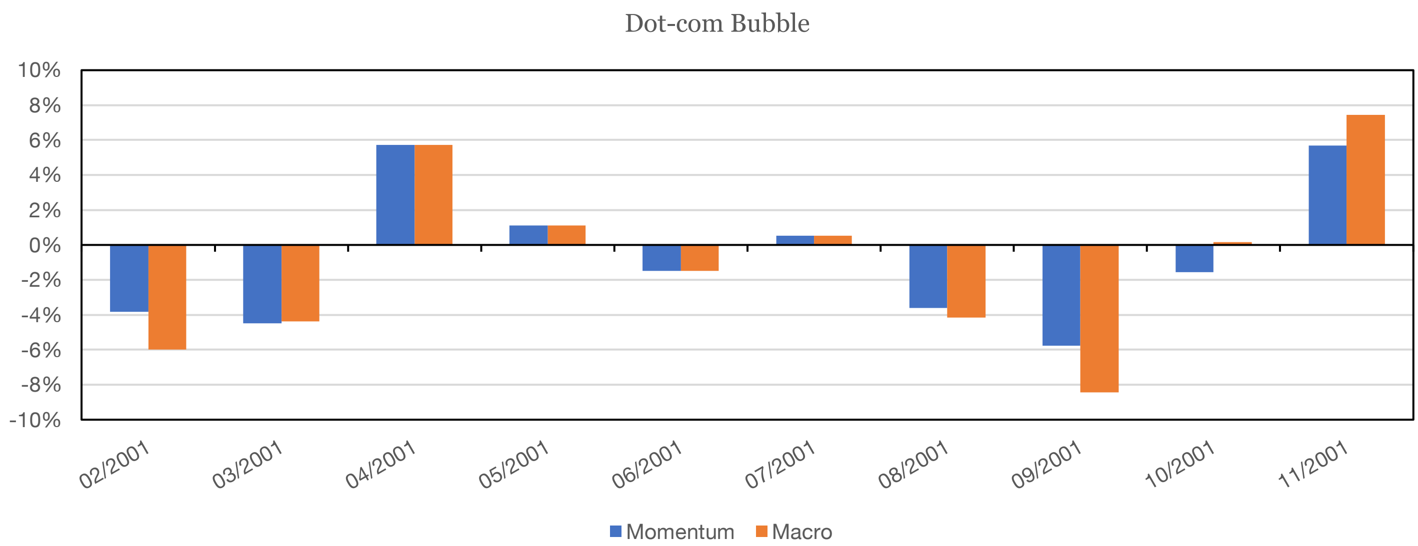
	MSCI US Index	Macro-Based (Blended)	Momentum-Based (6-Month)
Annualized Return	7.3%	10.2%	9.1%
Annualized Volatility	15.0%	14.1%	13.2%
Sharpe Ratio	0.55	0.76	0.75
Rebalance Freq.		8.7 / year	5.2 / year

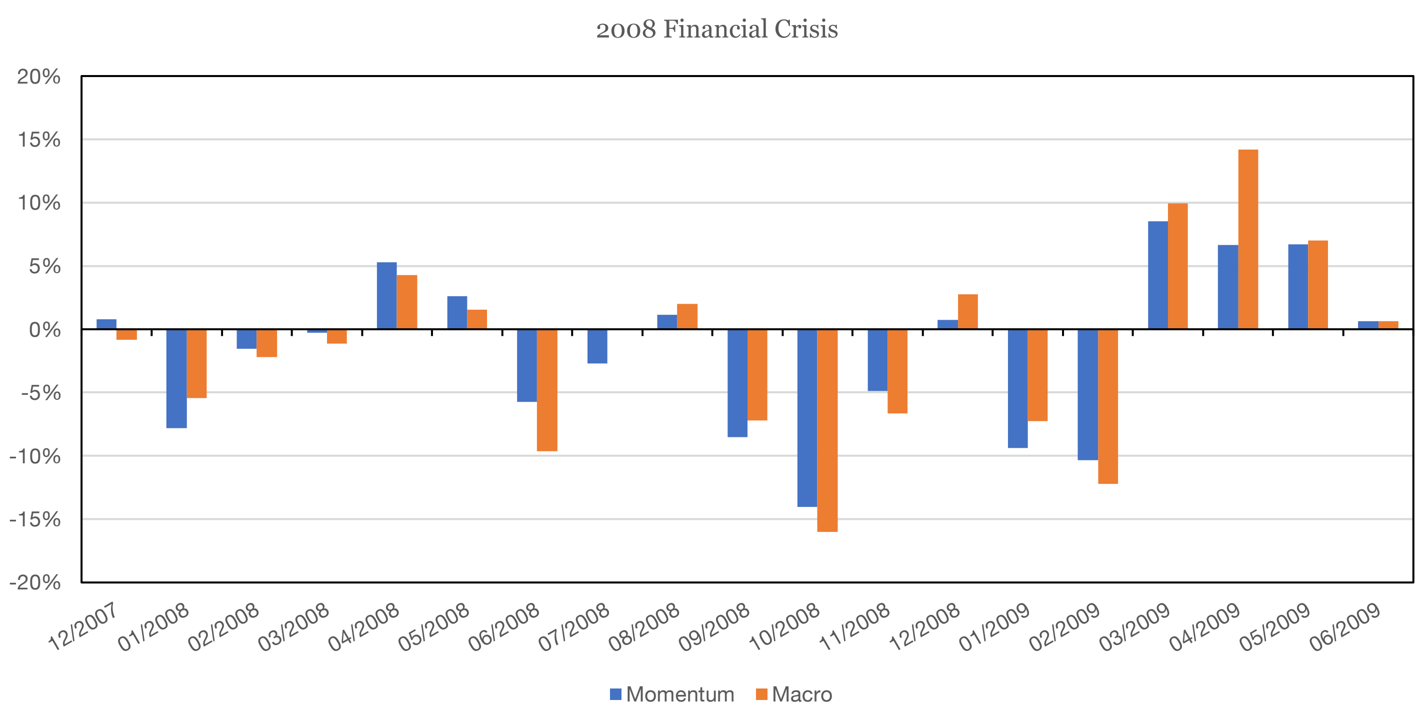
Source: MSCI USA Standard Universe. Calculations by Newfound Research. Returns assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.

It is worth noting that momentum-based timing rotation provides a more stable annualized return with lower volatility while also maintaining a similar Sharpe ratio compared to the macro-based allocation. The trading frequency is also lower on annual basis.

Since these are long-only strategies, a key risk is underperforming during large equity market drawdowns, adding insult to injury. To see these effects, we can perform a scenario test for these two allocation methodologies during the dot-com bubble and the 2008 Financial Crisis. Defined periods are based on NBER's listed history.

Exhibit 8: Monthly Returns of Macro-Based Allocation vs. Momentum-Based Allocation Under Two Recessions





Source: MSCI USA Standard Universe. Calculations by Newfound Research. Returns assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.

In general, momentum-based allocation provides better combinations of factors during recessions but tends to react slower when market starts to exit and enter into recovery. This is likely due to the inherent lag from the lookback periods that the momentum strategy has to undertake before generating signals.

Conclusion

Rotating among factors can be very tempting, and in this commentary, we examined two potential ways to implement this strategy: macro cycle-based and momentum-based.

Regardless of which strategy we select, it is important to remember that risk is not destroyed but rather shifted into different forms.

With the momentum-based allocation, the embedded assumption is that the near future will look like the recent past.

On the other hand, dynamic allocation based upon macro-economic regimes requires us to estimate both the current regime we are in as well as which factors will do well during that regime. In sacrificing more dynamic combinations, the fixed structure of the regime allocations may help reduce the impact of short-term noise that might lead to whipsaw trades in a momentum-based approach.

One positive about the momentum-based methodology is that the factor selection is inherently dynamic whereas the macro cycle-based method prespecified regime-dependent factor baskets. We could expect future returns to remain consistent with the historical hypothetical performance, but this is an assumption that may be informed by hindsight-based "intuition". The trade-off with using momentum to be dynamic is that the lag of the signals may fail to capitalize on potential opportunities during the transitions between business cycle states. This was the case in recovery state during the last two recessions.

We should also keep in mind that there are only 6 factors in our investible universe. What would the returns look like if we add more factors to our universe? What if we use different constructions of the same factors? Will the momentum-based timing rotation still outperform the benchmark? This is an open question for future research.

Both macro-based and momentum-based dynamic factor allocation proved successful in our (slightly) out-of-sample test. However, we should stress that all tests were performed gross of any fees and costs, which can have a substantial impact upon results (especially for high turnover strategies). Furthermore, the success of the macro-based test was highly dependent upon the factors selected for each macro regime, and there is a risk those factors were determined with hindsight bias.

Nevertheless, we believe this evidence suggests that further research is warranted, perhaps incorporating a blend of the approaches as well as other specifications to provide further signal diversification.

MACRO TIMING WITH TREND FOLLOWING

October 7, 2019

SUMMARY

- While it may be tempting to time allocations to active strategies, it is generally best to hold them as long-term allocations.

- Despite this, some research has shown that there may be certain economic environments where trend following equity strategies are better suited.

- In this commentary, we replicate this data and find that a broad filter of recessionary periods does indeed show this for certain trend equity strategies but not for the style of trend equity in general.

- However, further decomposing the business cycle into contractions, recoveries, expansions, and slowdowns using leading economic indicators such as PMI and unemployment does show some promising relationships between the forecasted stage of the business cycle and trend following's performance relative to buy-and-hold equities.

- Even if this data is not used to time trend equity strategies, it can be beneficial to investors for setting expectations and providing insight into performance differences.

Systematic active investing strategies are a way to achieve alternative return profiles that are not necessarily present when pursuing standard asset allocation and may therefore play an important role in developing well-diversified portfolios.

But these strategies are best viewed as allocations rather than trades.[56] This is a topic we've written about a number of times with respect to factor investing over the past several years, citing the importance of weathering short-term pain for long-term gains. For active strategies to outperform, some underperformance is necessary. Or, as we like to say, "no pain, no premium."

That being said, being tactical in our allocations to active strategies *may* have some value in certain cases. In one sense, we can view the multi-layered active decisions simply as another active strategy, distinct from the initial one.

An interesting post on *Philosophical Economics* looked at using a variety of recession indicators (unemployment, earnings growth, industrial production, etc.) as ways to systematically invest in either U.S. equities or a trend following strategy on U.S. equities. If the economic indicator was in a favorable trend, the strategy was 100% invested in equities. If the economic indicator was in an unfavorable trend, the strategy was invested in a trend following strategy applied to equities, holding cash when the market was in a downtrend.

[56] https://blog.thinknewfound.com/2016/02/active-strategies-allocation-not-trade/

The reasoning behind this strategy is intuitively appealing. Even if a recession indicator flags a likely recession, the market may still have room to run before turning south and warranting capital protection. On the other hand, when the recession indicator was favorable, purely investing in equities avoids some of the whipsaw costs that are inherent in trend following strategies.

In this commentary, we will first look at the general style of trend equity in the context of recessionary and non-recessionary periods and then get a bit more granular to see when trend following has worked historically through the economic cycle of Expansion, Slowdown, Contraction, and Recovery.

Replicating the Data

To get our bearings, we will first attempt to replicate some of the data from the *Philosophical Economics* post using only the classifications of “recession” and “not-recession”.

Keeping in line with the Philosophical Economics method, we will use whether the economic metric is above or below its 12-month moving average as the recession signal for the next month. We will use market data from the Kenneth French Data Library for the total U.S. stock market returns and the risk-free rate as the cash rate in the equity trend following model.

The following table shows the results of the trend following timing models using the United States ISM Purchasing Managers Index (PMI) and the Unemployment Rate as indicators.

	U.S. Equities	**12mo MA Trend Equity**	**12m MA Trend Timing Model (PMI)**	**12mo MA Trend Timing Model (Unemployment)**
Annualized Return	11.3%	11.1%	11.3%	12.2%
Annualized Volatility	14.7%	11.2%	11.9%	12.4%
Maximum Drawdown	50.8%	24.4%	32.7%	30.0%
Sharpe Ratio	0.49	0.62	0.61	0.66

Source: Quandl and U.S. Bureau of Labor Statistics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index. Data is from Jan 1948 – Sep 2019.

With the trend timing model, we see an improvement in the absolute returns compared to the trend equity strategy alone. However, this comes at the expense of increasing the volatility and maximum drawdown.

In the case of unemployment, which was the strongest indicator that Philosophical Economics found, there is an improvement in risk-adjusted returns in the timing model.

Still, while there is a benefit, it may not be robust.

If we remove the dependence of the trend following model on a single metric or lookback parameter, the benefit of the macro-timing decreases. Specifically, if we replace our simple 12-month moving average trend equity rule with the ensemble approach utilized in the Newfound Trend Equity Index, we see very different results. This may indicate that one specific *variant* of trend following did well in this overall model, but the *style* of trend following might not lend itself well to this application.

	U.S. Equities	Newfound Trend Equity Index	Trend Equity Index Blend (PMI)	Trend Equity Index Blend (Unemployment)
Annualized Return	11.3%	10.7%	10.9%	10.9%
Annualized Volatility	14.7%	11.1%	11.8%	13.5%
Maximum Drawdown	50.8%	25.8%	36.1%	36.0%
Sharpe Ratio	0.49	0.59	0.58	0.50

Source: Quandl and U.S. Bureau of Labor Statistics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index. Data is from Jan 1948 – Sep 2019.

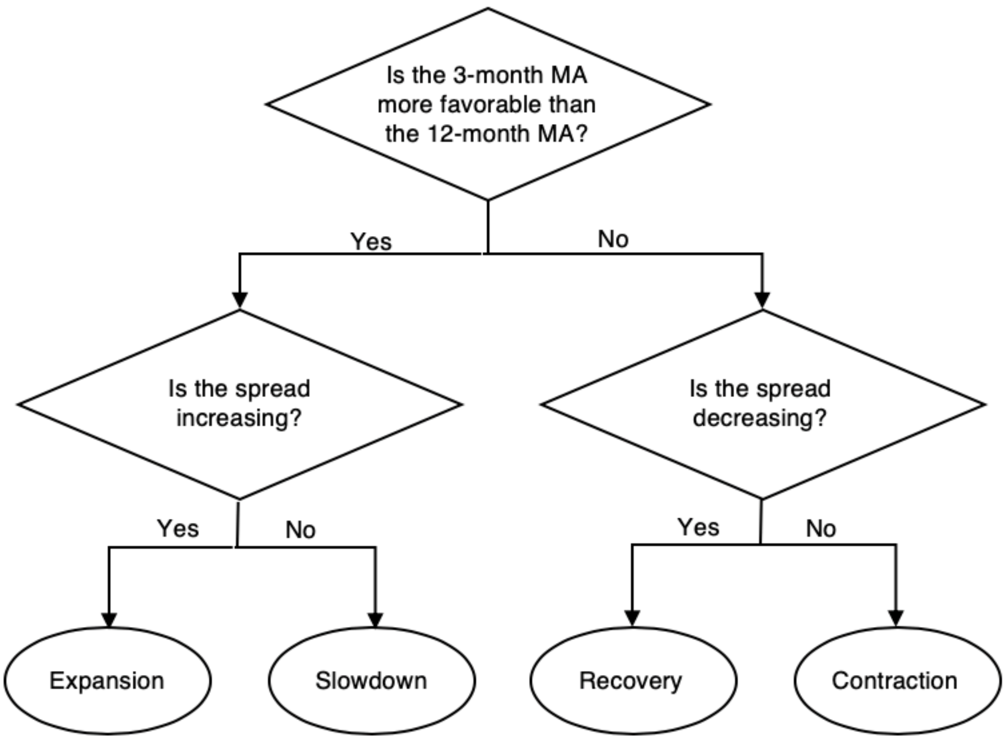
A more robust trend following model may already provide more upside capture during non-recessionary periods but at the expense of more downside capture during recessions. However, we cannot confidently assert that the lower level of down-capture in the single specification of the trend model is not partially due to luck.

If we desire to more thoroughly evaluate the style of trend following, we must get more granular with the economic cycles.

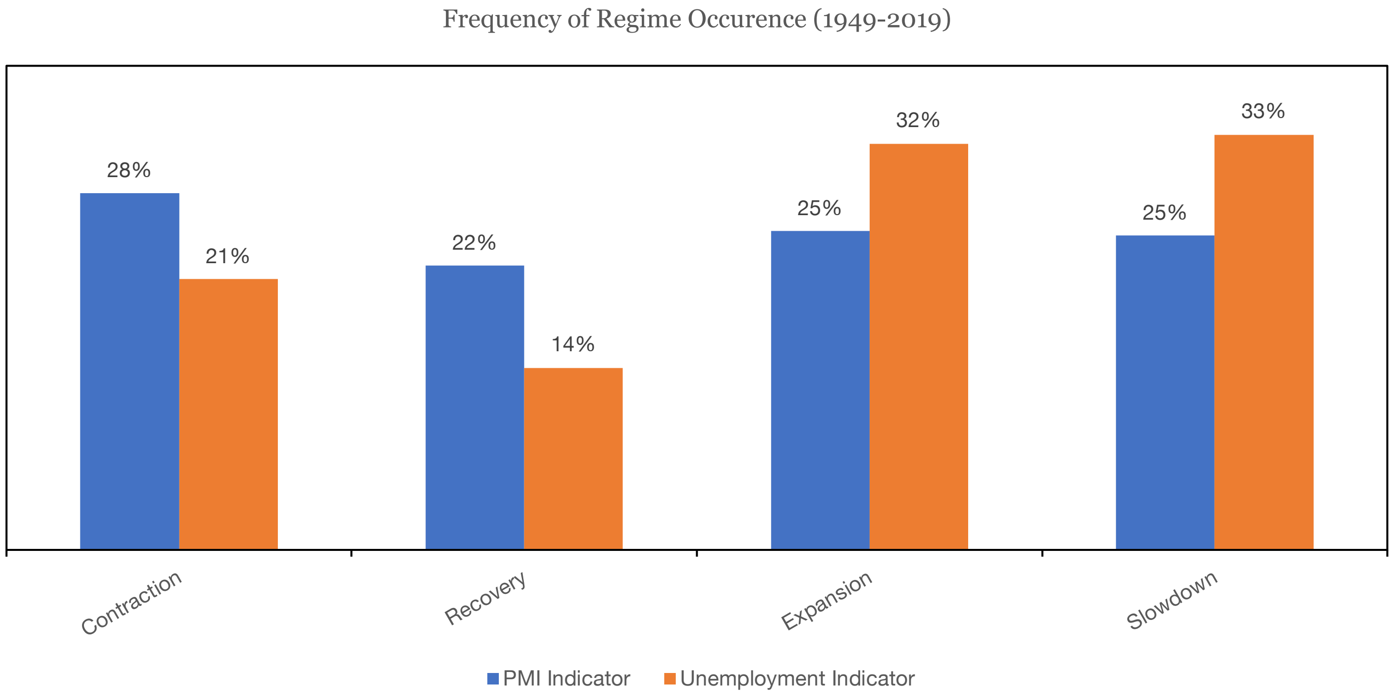
Breaking Down the Economic Cycle

Moving beyond the simple classification of “recession” and “not-recession”, we can follow MSCI’s methodology, which we used here previously, to classify the economic cycle into four primary states: Expansion, Slowdown, Contraction and Recovery.

We will focus on the 3-month moving average (“MA”) minus the 12-month MA for each indicator we examine according to the decision tree below. In the tree, we use the terms better or worse since lower unemployment rate and higher PMI values signal a stronger economy.

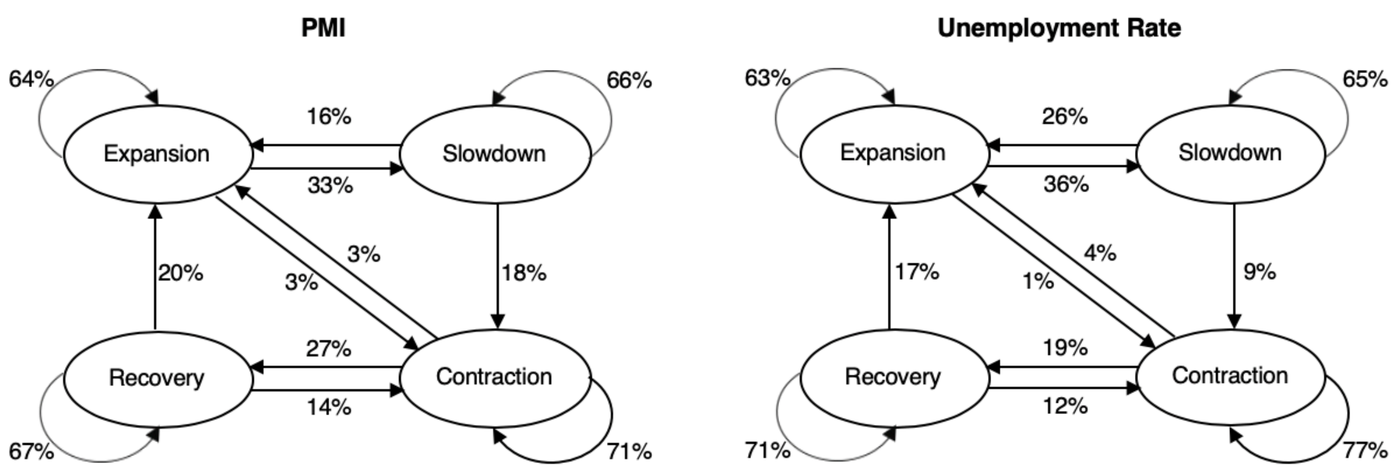


There is a decent amount of difference in the classifications using these two indicators, with the unemployment indicator signaling more frequent expansions and slowdowns. This should be taken as evidence that economic regimes are difficult to predict.



Source: Quandl and U.S. Bureau of Labor Statistics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index. Data is from Jan 1948 – Sep 2019.

Once each indicator is in each state the transition probabilities are relatively close.



Source: Quandl and U.S. Bureau of Labor Statistics. Calculations by Newfound Research. Results are hypothetical. Past performance is not an indicator of future results.

This agrees with intuition when we consider the cyclical nature of these economic metrics. While not a perfect mathematical relationship, these states generally unfold sequentially without jumps from contractions to expansions or vice versa.

Trend Following in the Economic Cycle

Applying the four-part classification to the economic cycle shows where trend equity outperformed.

	PMI Indicator		Unemployment Indicator	
	U.S. Equities	Trend Equity	U.S. Equities	Trend Equity
Contraction	7.6%	10.3%	1.0%	7.3%
Recovery	12.2%	9.3%	15.4%	15.0%
Expansion	14.3%	14.4%	13.9%	11.3%
Slowdown	7.2%	5.4%	10.5%	8.0%

Source: Quandl and U.S. Bureau of Labor Statistics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index. Data is from Jan 1948 – Sep 2019.

During contraction phases, regardless of indicators, trend equity outperformed buy-and-hold.

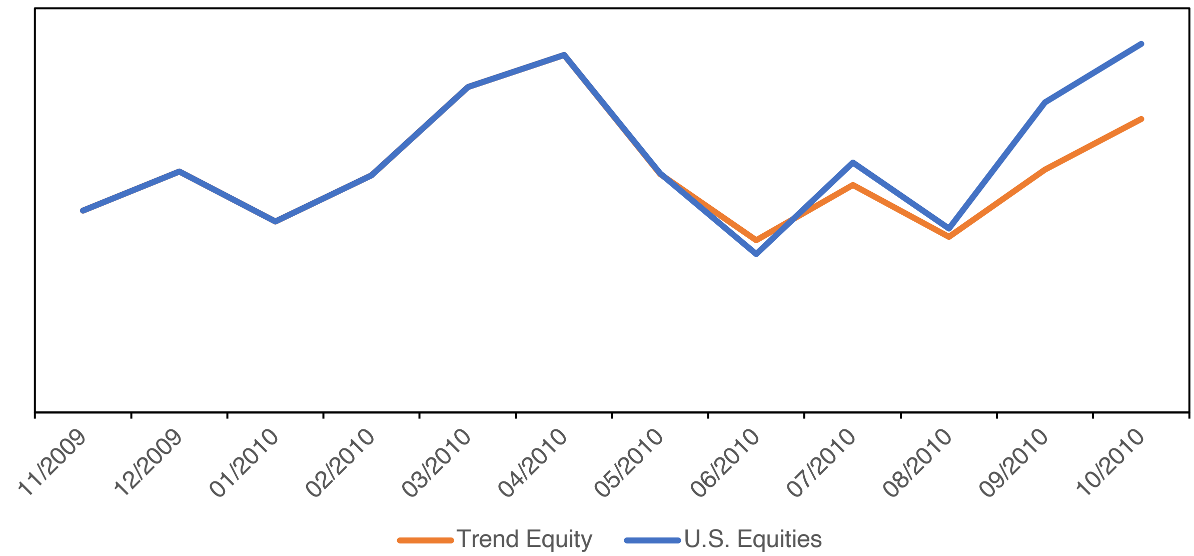
For the PMI indicator, trend equity was able to keep up during expansions, but this was not the case with the unemployment indicator. The reverse of this was true for recoveries: trend following was close to keeping up in the periods denoted by the unemployment indicator but not by the PMI indicator.

For both indicators, trend following underperformed during slowdowns.

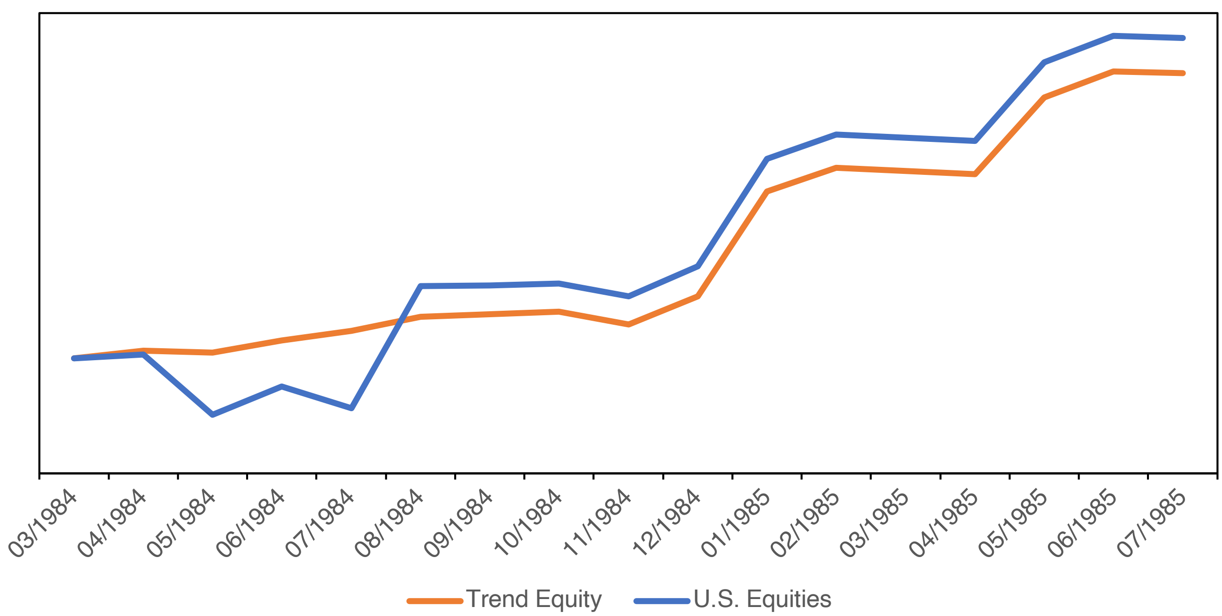
This may seem contradictory at first, but these may be periods of more whipsaw as markets try to forecast future states. And since slowdowns typically occur after expansions and before contractions (at least in the idealized model), we may have to bear more of this whipsaw risk for the strategy to be adaptable enough to add value during the contraction.

The following two charts show the longest historical slowdowns for each indicator: the PMI indicator was for 11 months in late 2009 through much of 2010 and the unemployment rate indicator was for 16 months in 1984-85.

Longest Slowdown Based on PMI (Nov 2009 - Nov 2010)



Longest Slowdown Based on Unemployment (Mar 1984 - Aug 1985)



Source: Quandl and U.S. Bureau of Labor Statistics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index.

In the first slowdown period, the trend equity strategy rode in tandem with equities as they continued to climb and then de-risked when equities declined. Equities quickly rebounded leaving the trend equity strategy underexposed to the rally.

In the second slowdown period, the trend equity strategy was heavily defensive going into the slowdown. This protected capital initially but then caused the strategy to lag once the market began to increase steadily.

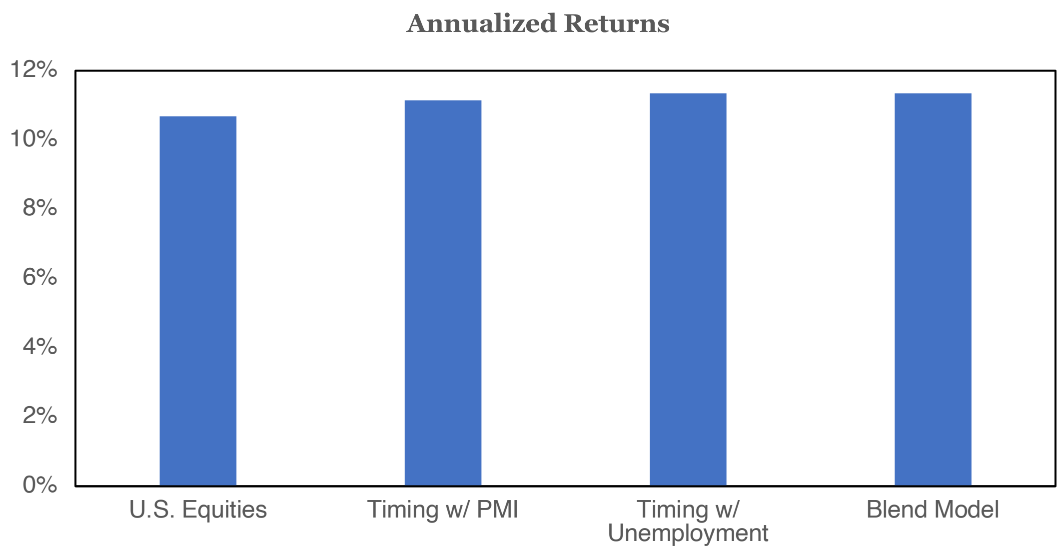
The first period illustrates a time when the trend equity strategy was ready to adapt to changing market conditions and was unfortunately whipsawed. The second period illustrates a time when the trend equity strategy was already adapted to a supposedly oncoming contraction that did not materialize.

Using these historical patterns of performance, we can now explore how a strategy that systematically allocates to trend equity strategies might be constructed.

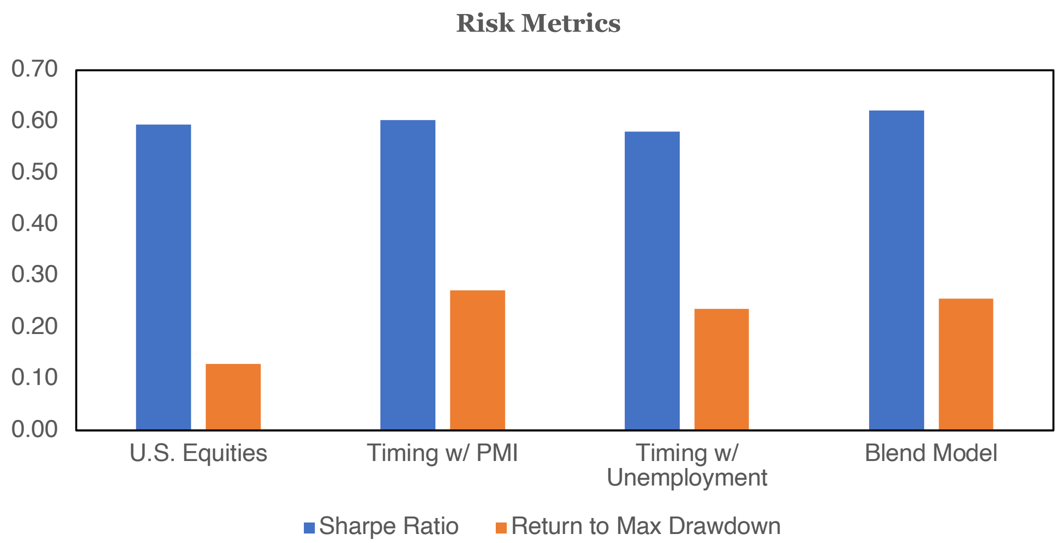
Timing Trend Following with the Economic Cycle

One simple way to apply a systematic timing strategy for shifting between equities and trend following is to only invest in equities when a slowdown is signaled.

The charts below show the returns and risk metrics for models using the PMI and unemployment rate individually and a model that blends the two allocations.



Source: Quandl and U.S. Bureau of Labor Statistics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index. Data is from Jan 1948 – Sep 2019.



Source: Quandl and U.S. Bureau of Labor Statistics. Calculations by Newfound Research. Results are hypothetical. Results assume the reinvestment of all distributions. Results are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Past performance is not an indicator of future results. You cannot invest in an index. Data is from Jan 1948 – Sep 2019.

The returns increased slightly in every model relative to buy-and-hold, and the blended model performed consistently high across all metrics.

Blending multiple models generally produces benefits like these shown here, and in an actual implementation, utilizing additional economic indicators may make the strategy even more robust. There may be other ways to boost performance across the economic cycle, and we will explore these ideas in future research.

Conclusion

Should investors rotate in and out of active strategies?

Not in most cases, since the typical drivers are short-term underperformance that is a necessary component of active strategies.

However, there may be opportunities to make allocation tweaks based on the economic cycle.

The historical data suggests that a specification-neutral trend-equity strategy has outperformed buy-and-hold equities during economic contractions for both economic indicators. The performance during recoveries and expansions was mixed across indicators. It kept up with the buy-and-hold strategy during expansions denoted by PMI but not unemployment. This relationship was reversed for recoveries denoted by unemployment. In both models, trend equity has also lagged during economic slowdowns as whipsaw becomes more prevalent.

Based on the most recent PMI data, the current cycle is a contraction, indicating a favorable environment for trend equity under both cycle indicators. However, we should note that December 2018 through March 2019 was also labeled as a contraction according to PMI. Not all models are perfect.

Nevertheless, there may be some evidence that trend following can provide differentiated benefits based on the prevailing economic environment.

While an investor may not use this knowledge to shift around allocations to active trend following strategies, it can still provide insight into performance difference relative to buy-and-hold and set expectations going forward.

YIELD CURVE TRADES WITH TREND AND MOMENTUM

October 14, 2019

SUMMARY

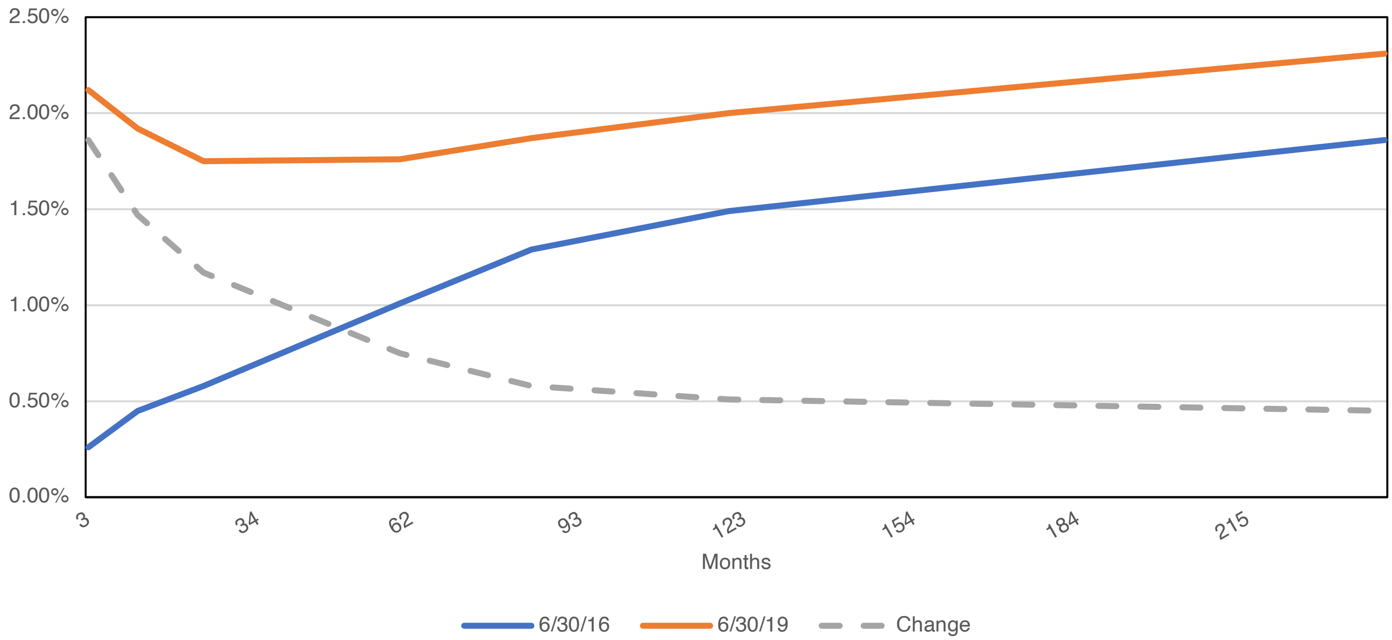
- Yield curve changes over time can be decomposed into Level, Slope, and Curvature changes, and these changes can be used to construct portfolios.
- Market shocks, monetary policy, and preferences of different segments of investors (e,g. pensions) may create trends within these portfolios that can be exploited with absolute and relative momentum.
- In this commentary, we investigate these two factors in long/short and long/flat implementations and find evidence of success with some structural caveats.
- Despite this, we believe the results have potential applications as either a portable beta overlay or for investors who are simply trying to figure out how to position their duration exposure.
- Translating these quantitative signals into a forecast about yield-curve behavior may allow investors to better position their fixed income portfolios.

It has been well established in fixed income literature that changes to the U.S. Treasury yield curve can be broken down into three primary components: a level shift, a slope change, and a curvature twist.

A level change occurs when rates increase or decrease across the entire curve at once. A slope change occurs when short-term rates decrease (increase) while long-term rates increase (decrease). Curvature defines convexity and concavity changes to the yield curve, capturing the bowing that occurs towards the belly of the curve.

Obviously these three components do not capture 100% of changes in the yield curve, but they do capture a significant portion of them. From 1962-2019 they explain 99.5% of the variance in daily yield curve changes.

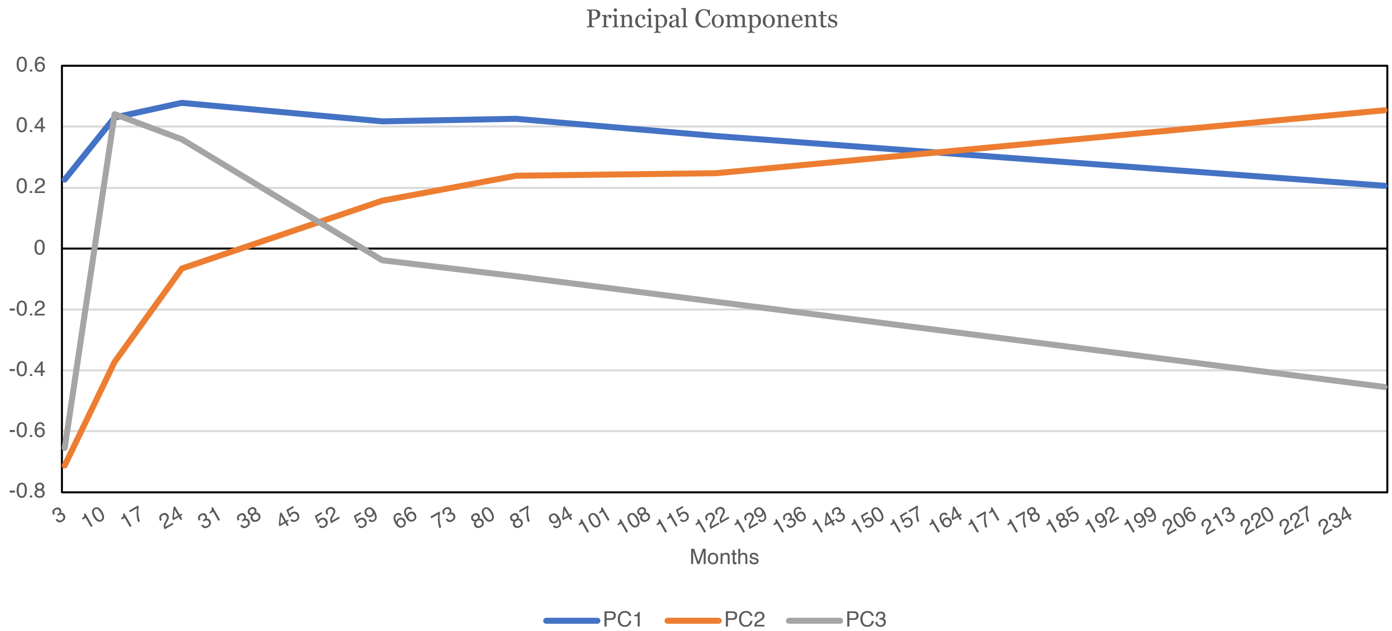
We can even decompose longer-term changes in the yield curve into these three components. For example, consider how the yield curve has changed in the three years from 6/30/2016 to 6/30/2019.



Source: Federal Reserve of St. Louis.

We can see that there was generally a positive increase across the entire curve (i.e. a positive level shift), the front end of the curve increased more rapidly (i.e. a flattening slope change) and the curve flipped from concave to convex (i.e. an inverted bowing of the curve).

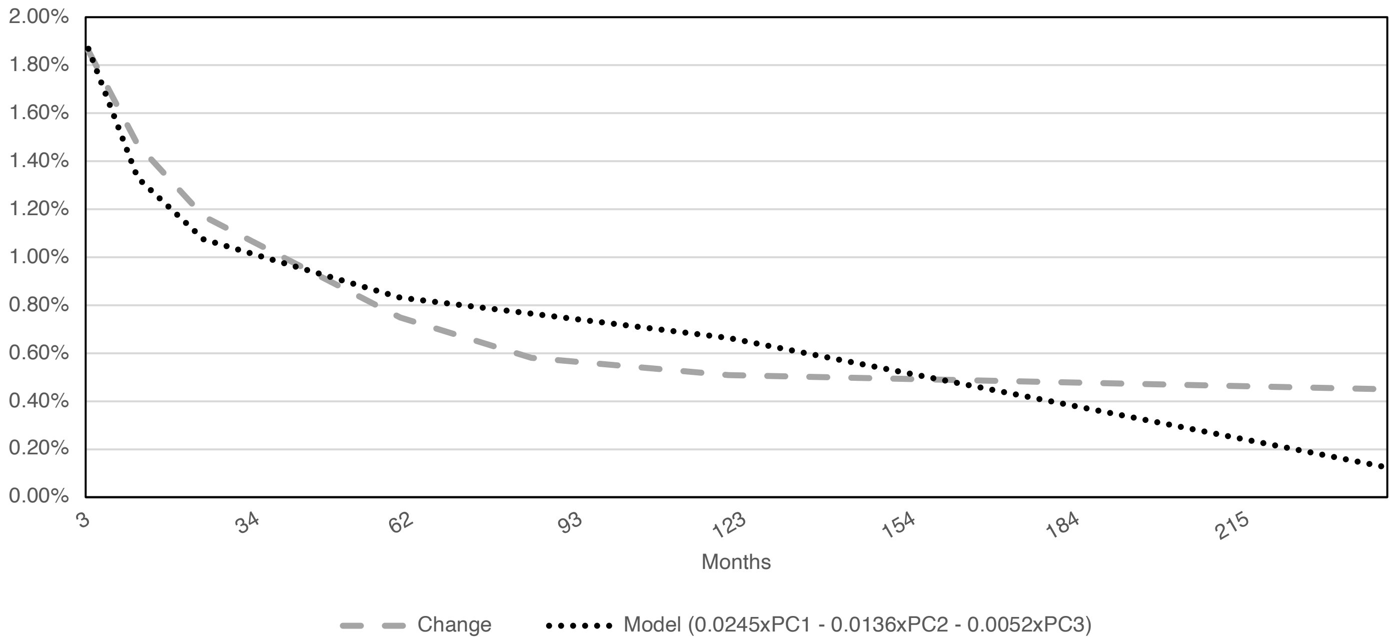
Using the historical yield curve changes, we can mathematically estimate these stylized changes using principal component analysis. We plot the loadings of the first three components below for this three-year change.



Source: Federal Reserve of St. Louis. Calculations by Newfound Research.

We can see that –PC1– has generally positive loadings across the entire curve, and therefore captures our level shift component. –PC2– exhibits negative loadings on the front end of the curve and positive loadings on the back, capturing our slope change. Finally, –PC3– has positive loadings from the 1-to-5-year part of the curve, capturing the curvature change of the yield curve itself.

Using a quick bit of linear algebra, we can find the combination of these three factors that closely matches the change in the curve from 6/30/2016 to 6/30/2019. Comparing our model versus the actual change, we see a reasonably strong fit.



Source: Federal Reserve of St. Louis. Calculations by Newfound Research.

So why might this be useful information?

First of all, we can interpret our principal components as if they are portfolios. For example, our first principal component is saying, "buy a portfolio that is long interest rates across the entire curve." The second component, on the other hand, is better expressed as, "go short rates on the front end of the curve and go long rates on the back end."

Therefore, insofar as we believe changes to the yield curve may exhibit absolute or relative momentum, we may be able to exploit this momentum by constructing a portfolio that profits from it.

As a more concrete example, if we believe that the yield curve will generally steepen over the next several years, we might short 2-year U.S. Treasury futures and buy 10-year U.S. Treasury futures. The biggest wrinkle we need to deal with is the fact that 2-year U.S. Treasury futures will exhibit very different sensitivity to rate changes than 10-year U.S. Treasury futures, and therefore we must take care to duration-adjust our positions.

Why might such changes exhibit trends or relative momentum?

- During periods where arbitrage capital is low, trends may emerge. We might expect this during periods of extreme market shock (e.g. recessions) where we might also see the simultaneous influence of monetary policy.

- Effects from monetary policy may exhibit autocorrelation. If investors exhibit any anchoring to prior beliefs, they might discount future policy changes.

- Segmented market theory suggests that different investors tend to access different parts of the curve (e.g. pensions may prefer the far end of the curve for liability hedging purposes). Information flow may therefore be segmented, or even impacted by structural buyers/sellers, creating autocorrelation in curve dynamics.

In related literature, Fan et al (2019) find that net hedging or speculative positions has strong cross-sectional explanatory power for agricultural and currency futures returns, but *not* in fixed income markets. To quote,

> *"In sharp contrast, we find no evidence of a significant speculative pressure premium in the interest rate and fixed income futures markets. Thus, albeit from the lens of different research questions, our paper reaffirms Bessembinder (1992) and Moskowitz et al. (2012) in establishing that fixed income futures markets behave differently from other futures markets as regards the information content of the net positions of hedgers or speculators. A hedgers-to-speculators risk transfer in fixed income futures markets would be obscured if agents choose to hedge their interest rate risk with other strategies (i.e. immunization, temporary change in modified duration)."*

Interestingly, Markowitz et al. (2012) suggest that speculators may be profiting from time-series momentum at the expense of hedgers, suggesting that they earn a premium for providing liquidity. Such does not appear to be the case for fixed income futures, however.

As far as we are aware, it has not yet been tested in the literature whether the net speculator versus hedger position has been tested for yield curve trades, and it may be possible that a risk transfer does not exist at the individual maturity basis, but rather exists for speculators willing to bear level, slope, or curvature risk.

Stylized Component Trades

While we know the exact loadings of our principal components (i.e. which maturities make up the principal portfolios), to avoid the risk of overfitting our study we will capture level, slope, and curvature changes with three different stylized portfolios.

To implement our portfolios, we will buy a basket of 2-, 5-, and 10-year U.S. Treasury futures contracts ("UST futures"). We will assume that the 5-year contract has 2.5x the duration of the 2-year contract and the 10-year contract has 5x the duration of the 2-year contract.

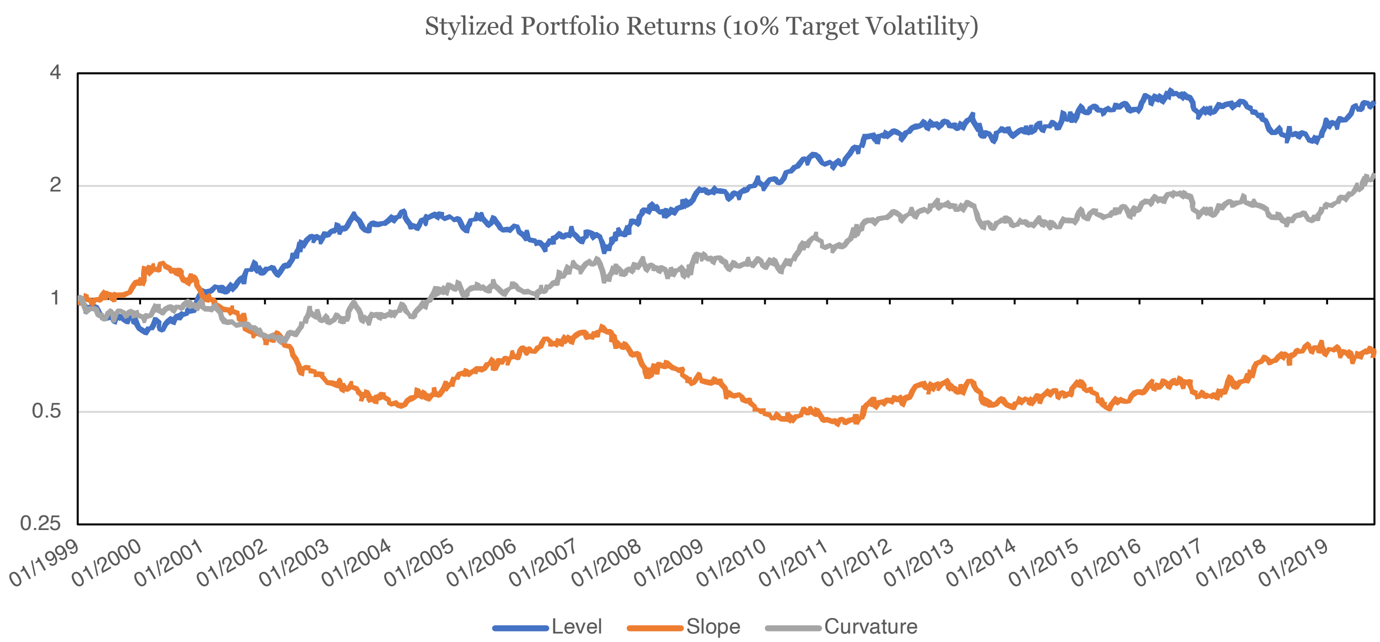
To capture a level shift in the curve, we will go long across all the contracts. Specifically, for every dollar of 2-year UST futures exposure we purchase, we will buy $0.4 of 5-year UST futures and $0.20 of 10-year UST futures. This creates equal duration exposure across the entire curve.

To capture slope change, we will go short 2-year UST futures and long the 10-year UST futures, holding zero position in the 5-year UST futures. As before, we will duration-adjust our positions such that for each $1 short of the 2-year UST futures position, we are $0.20 long the 10-year UST futures.

Finally, to capture curvature change we will construct a butterfly trade where we short the 2- and 10-year UST futures and go long the 5-year UST futures. For each $1 long in the 5-year UST futures, we will short $1.25 of 2-year UST futures and $0.25 of 10-year UST futures.

Note that the slope and curvature portfolios are implemented such that they are duration neutral (based upon our duration assumptions) so a level shift in the curve will generate no profit or loss.

An immediate problem with our approach arises when we actually construct these portfolios. Unless adjusted, the volatility exhibited across these trades will be meaningfully different. Therefore, we target a constant 10% volatility for all three portfolios by adjusting the notional exposure of each portfolio based upon an exponentially-weighted estimate of prior 3-month realized volatility.



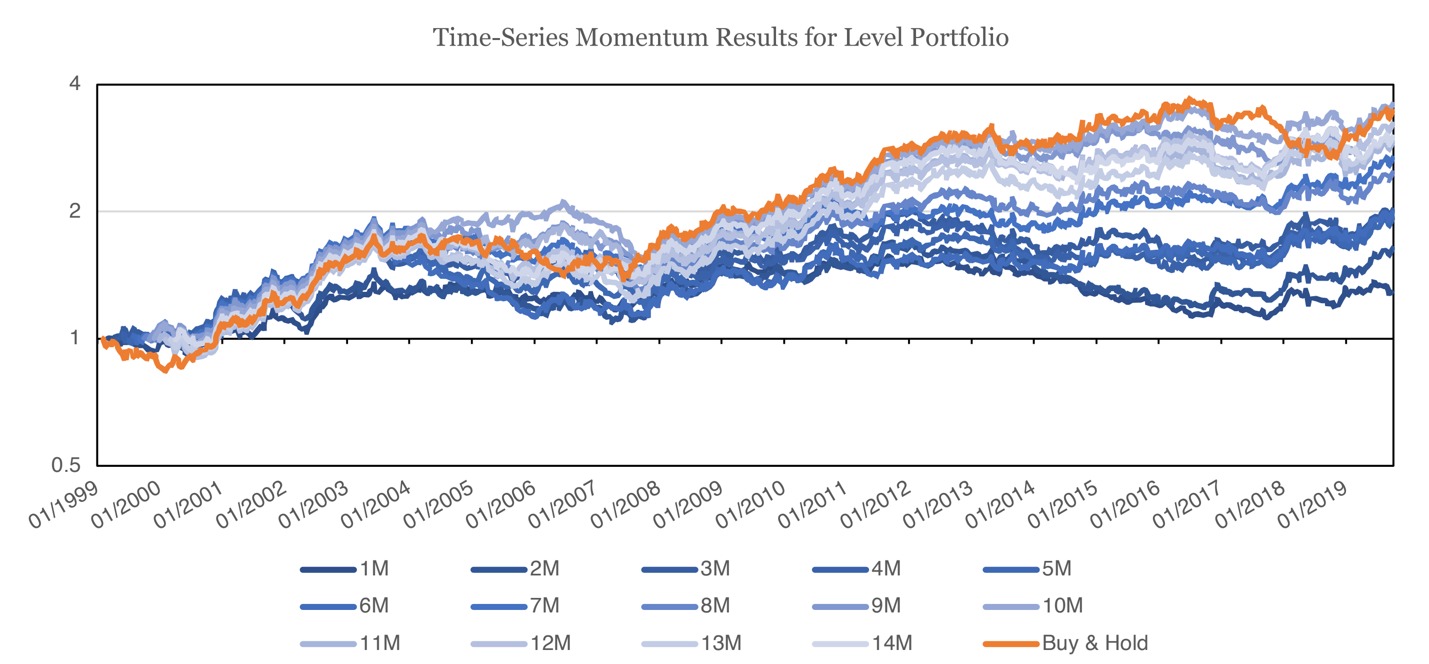
Source: Stevens Futures. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

It appears, at least to the naked eye, that changes in the yield curve – and therefore the returns of these portfolios – may indeed exhibit positive autocorrelation. For example, –Slope– appears to exhibit significant trends from 2000-2004, 2004-to 2007, and 2007-2012.

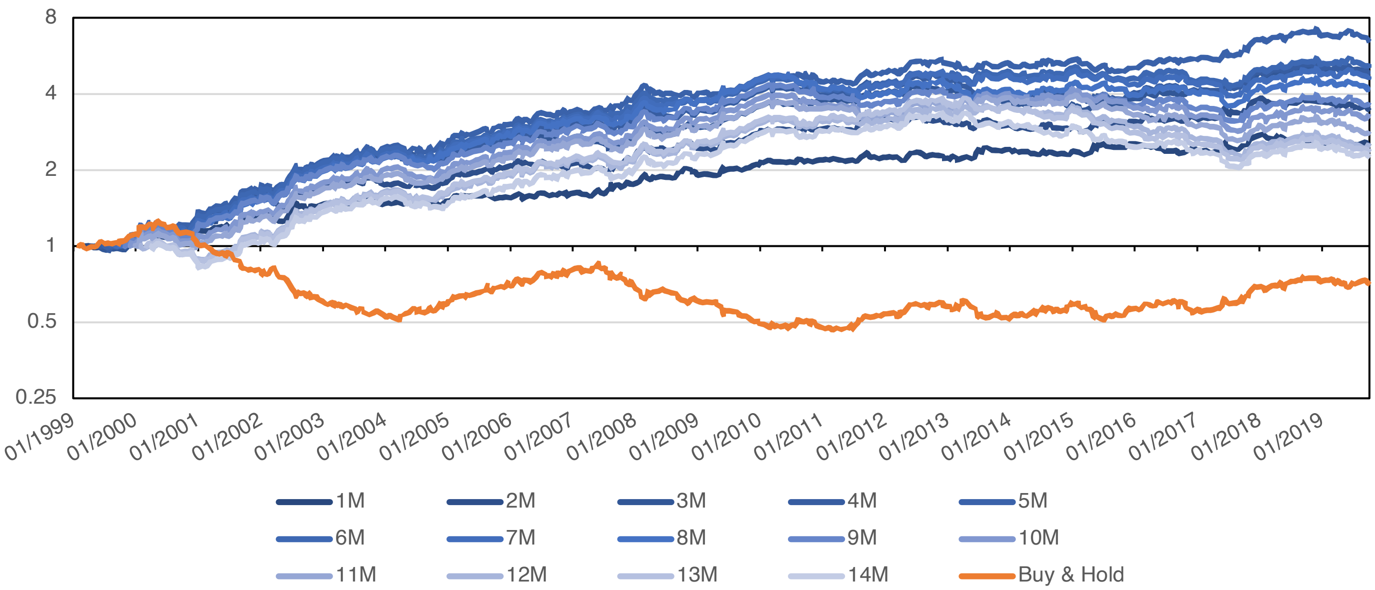
Whether those trends can be identified and exploited is another matter entirely. Thus, with our stylized portfolios in hand, we can begin testing.

Trend Signals

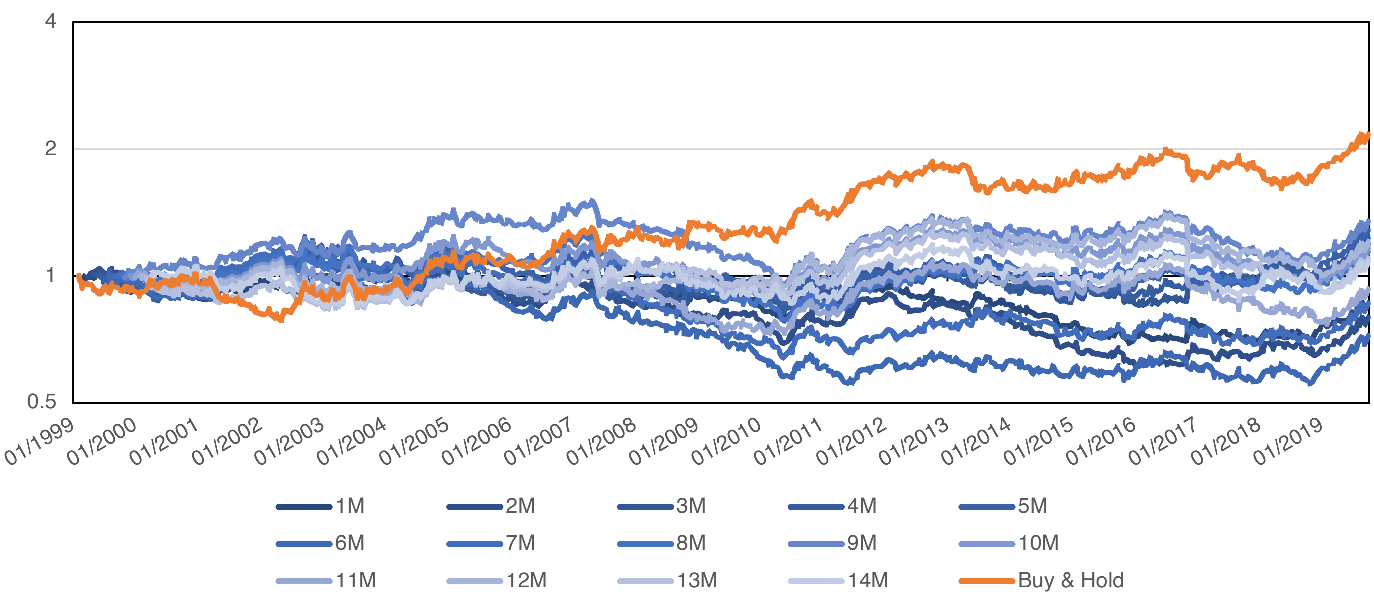
We begin our analysis by exploring the application of time-series momentum signals across all three of the portfolios. We evaluate lookback horizons ranging from 21-to-294 trading days (or, approximately 1-to-14 months). Portfolios assume a a 21-trading-day holding period and are implemented using 21 overlapping portfolios to control for timing luck.



Time-Series Momentum Results for Slope Portfolio



Time-Series Momentum Results for Curve Portfolio



Source: Stevens Futures. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

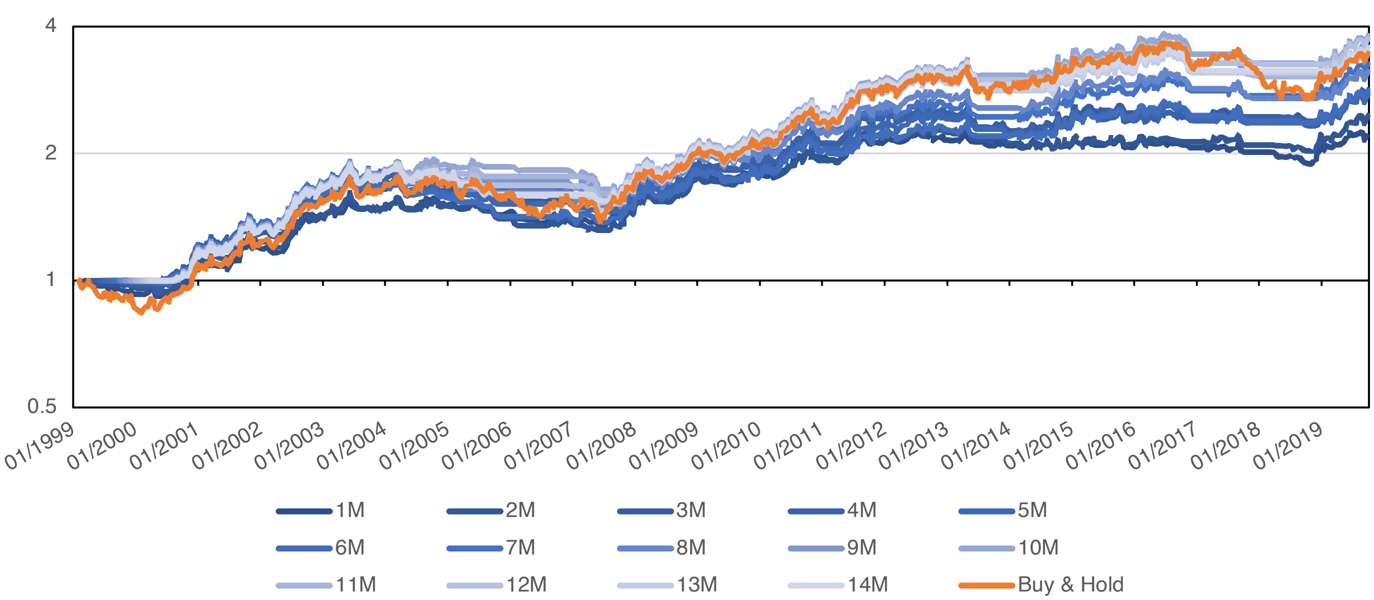
Some observations:

- Time-series momentum appears to generate positive returns for the Level portfolio. Over the period tested, longer-term measures (e.g. 8-to-14-month horizons) offer more favorable results.
- Time-series momentum on the Level portfolio does, however, underperform naïve buy-and-hold. The returns of the strategy also do not offer a materially improved Sharpe ratio or drawdown profile.
- Time-series momentum also appears to capture trends in the Slope portfolio. Interestingly, both short- and long-term lookbacks are less favorable over the testing period than intermediate-term (e.g. 4-to-8 month) ones.
- Finally, time-series momentum appeared to offer no edge in timing curvature trades.

Here we should pause to acknowledge that we are blindly throwing strategies at data without much forethought. If we consider, however, that we might reasonably expect duration to be a positively compensated risk premium, as well as the fact that we would expect the futures to capture a generally positive roll premium (due to a generally upward sloping yield curve), then explicitly shorting duration risk may not be a keen idea.

In other words, it may make more sense to implement our level trade as a long/flat rather than a long/short. When implemented in this fashion, we see that the annualized return versus buy-and-hold is much more closely maintained while volatility and maximum drawdown are significantly reduced.

Long/Flat Time-Series Momentum Results for Level Portfolio



Source: Stevens Futures. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

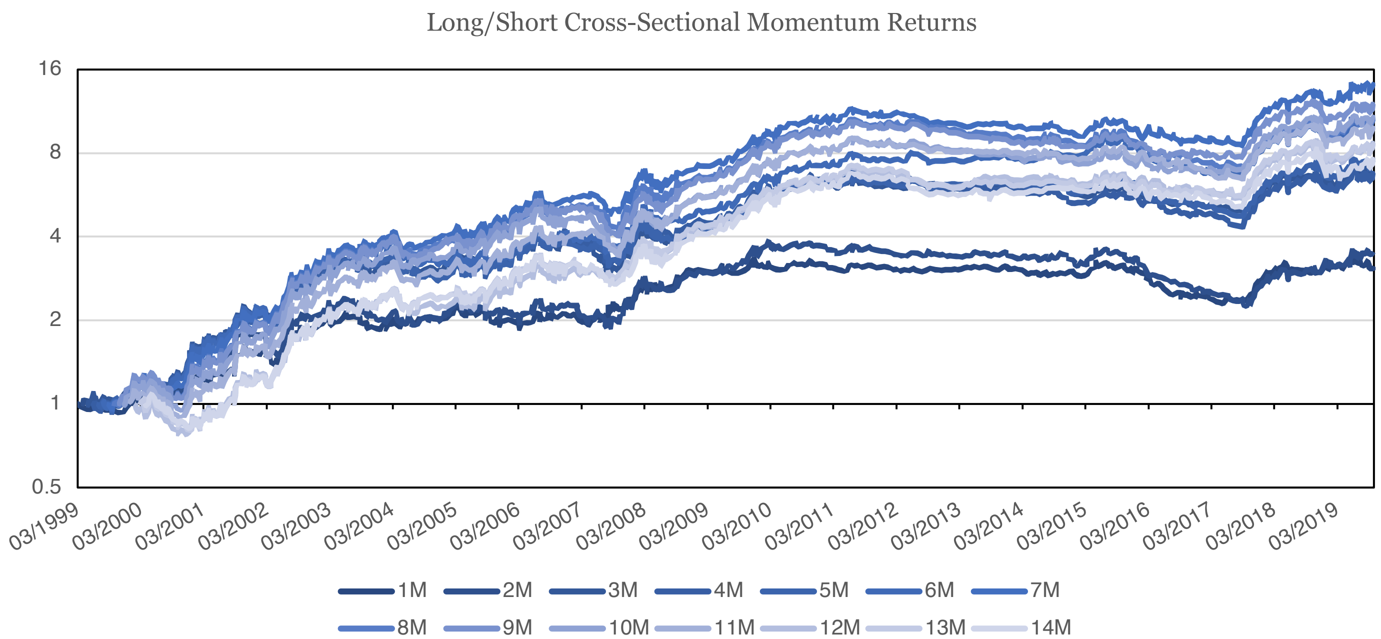
Taken together, it would appear that time-series momentum may be effective for trading the persistence in Level and Slope changes, though not in Curvature.

Momentum Signals

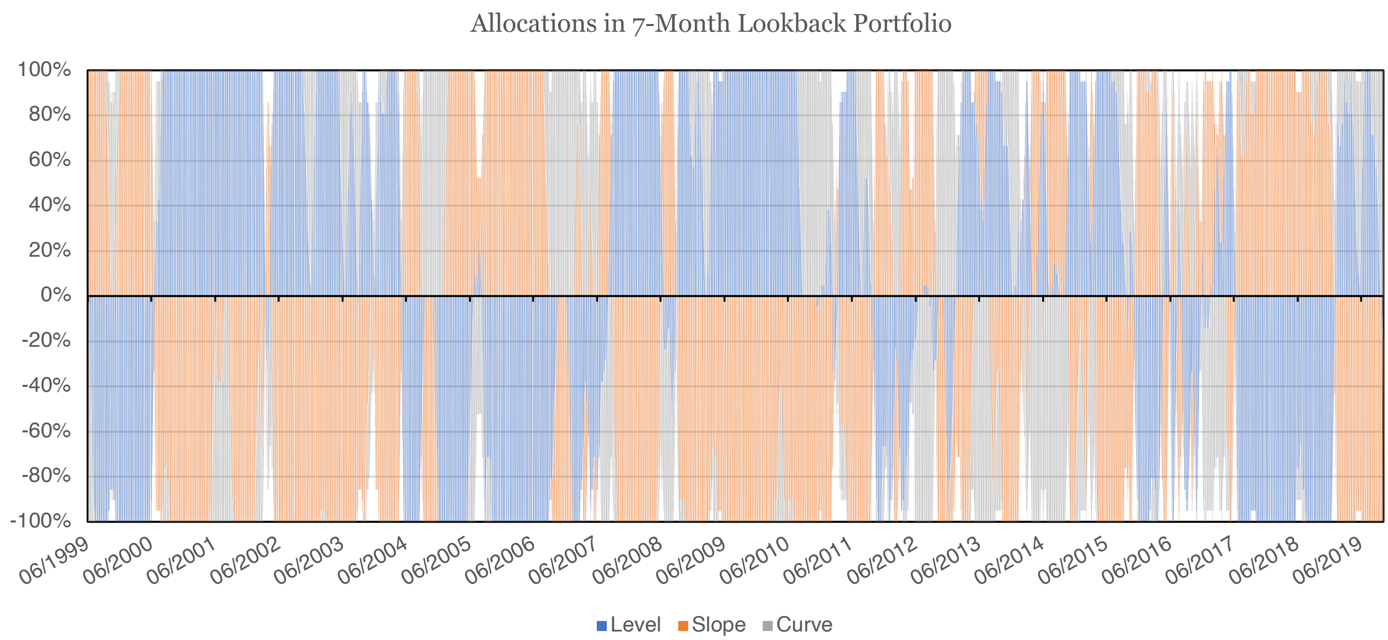
If we treat each stylized portfolio as a separate asset, we can also consider the returns of a cross-sectional momentum portfolio. For example, each month we can rank the portfolios based upon their prior returns. The top-ranking portfolio is held long; the 2nd ranked portfolio is held flat; and the 3rd ranked portfolio is held short.

As before, we will evaluate lookback horizons ranging from 21-to-294 trading days (approximately 1-to-14 months) and assuming a 21-trading-day holding period, implemented with 21 overlapping portfolios.

Results – as well as example allocations from the 7-month lookback portfolio – are plotted below.



Source: Stevens Futures. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.



Here we see very strong performance results except in the 1- and 2-month lookback periods. The allocation graph appears to suggest that results are not merely the byproduct of consistently being long or short a particular portfolio and the total return level appears to suggest that the portfolio is able to simultaneously profit from both legs.

If we return back to the graph of the stylized portfolios, we can see a significant negative correlation between the Level and Slope portfolios from 1999 to 2011. The negative correlation appears to disappear after this point, almost precisely coinciding with a 6+ year drawdown in the cross-sectional momentum strategy.

This is due to a mixture of construction and the economic environment.

From a construction perspective, consider that the Level portfolio is long the 2-, the 5-, and the 10-year UST futures while the Slope portfolio is short 2-year and long the 10-year UST futures. Since the positions are held in a manner that targets equivalent duration exposure, when the 2-year rate moves more than the 10-year rate, we end up in a scenario where the two trades have negative correlation, since one strategy is short and the other is long the 2-year position. Conversely, if the 10-year rate moves more than the 2-year rate, we end up in a scenario of positive correlation, since both strategies are long the 10-year.

Now consider the 1999-2011 environment. We had an easing cycle during the dot-com bust, a tightening cycle during the subsequent economic expansion, and another easing cycle during the 2008 crisis. This caused significantly more directional movement in the 2-year rate than the 10-year rate. Hence, negative correlation.

After 2008, however, the front end of the curve became pinned to zero. This meant that there was significantly more movement in the 10-year than the 2-year, leading to positive correlation in the two strategies. With positive correlation there is less differentiation among the two strategies and so we see a considerable increase in strategy turnover – and effectiveness – as momentum signals become less differentiated.

With that in mind, had we designed our Slope portfolio to be long 2-year UST futures and short 10-year UST futures (i.e. simply inverted the sign of our allocations), we would have seen *positive* correlation between Level and Slope from 1999 to 2011, resulting in a very different set of allocations and returns. In actually testing this step, we find that the 1999-2011 period is no longer dominated by Level versus Slope trades, but rather Slope versus Curvature. Performance of the strategy is still largely positive, but the spread among specifications widens dramatically.

Taken all together, it is difficult to conclude that the success of this strategy was not, in essence, driven almost entirely by autocorrelation in easing and tightening cycles with a relatively stable back end of the curve.[57] Given that there have only been a handful of full rate cycles in the last 20 years, we'd be reluctant to rely too heavily on the equity curve of this strategy as evidence of a robust strategy.

Conclusion

In this research note, we explored the idea of generating stylized portfolios designed to isolate and profit from changes to the form of the yield curve. Specifically, using 2-, 5-, and 10-year UST futures we design portfolios that aim to profit from level, slope, and curvature changes to the US Treasury yield curve.

With these portfolios in hand, we test whether we can time exposure to these changes using time-series momentum.

We find that while time-series momentum generates positive performance for the Level portfolio, it fails to keep up with buy & hold. Acknowledging that level exposure may offer a positive long-term risk premium, we adjust the strategy from long/short to long/flat and are able to generate a substantially improved risk-adjusted return profile.

Time-series momentum also appears effective for the Slope portfolio, generating meaningful excess returns above the buy-and-hold portfolio.

Applying time-series momentum to the Curvature portfolio does not appear to offer any value.

We also tested whether the portfolios can be traded employing cross-sectional momentum. We find significant success in the approach but believe that the results are an artifact of (1) the construction of the portfolios and (2) a market regime heavily influenced by monetary policy. Without further testing, it is difficult to determine if this approach has merit.

Finally, even though our study focused on portfolios constructed using U.S. Treasury futures, we believe the results have potential application for investors who are simply trying to figure out how to position their duration exposure. For example, a signal to be short (or flat) the Level portfolio and long the Slope portfolio may imply a view of rising rates with a flattening

[57] From 1999 to 2008, the 2-year rate moved from 7% down to 1%, up to 5% and back down to 1% while the 10-year rate was largely rangebound in the 4-5% area (with the exception of early 2001 and late 2008).

curve. Translating these quantitative signals into a forecast about yield-curve behavior may allow investors to better position their fixed income portfolios.

Since this study utilized U.S. Treasury futures, these results translate well to implementing a portable beta strategy. For example, if you were an investor with a desired risk profile on par with 100% equities, you could add bond exposure *on top* of the higher risk portfolio. This would add a (generally) diversifying return source with only a minor cash drag to the extent that margin requirements dictate.

Risk-Adjusted Momentum: A Momentum and Low-Volatility Barbell?

October 21, 2019

Summary

- After the Great Financial Crisis, the Momentum factor has exhibited positive returns, but those returns have been largely driven by the short side of the portfolio.

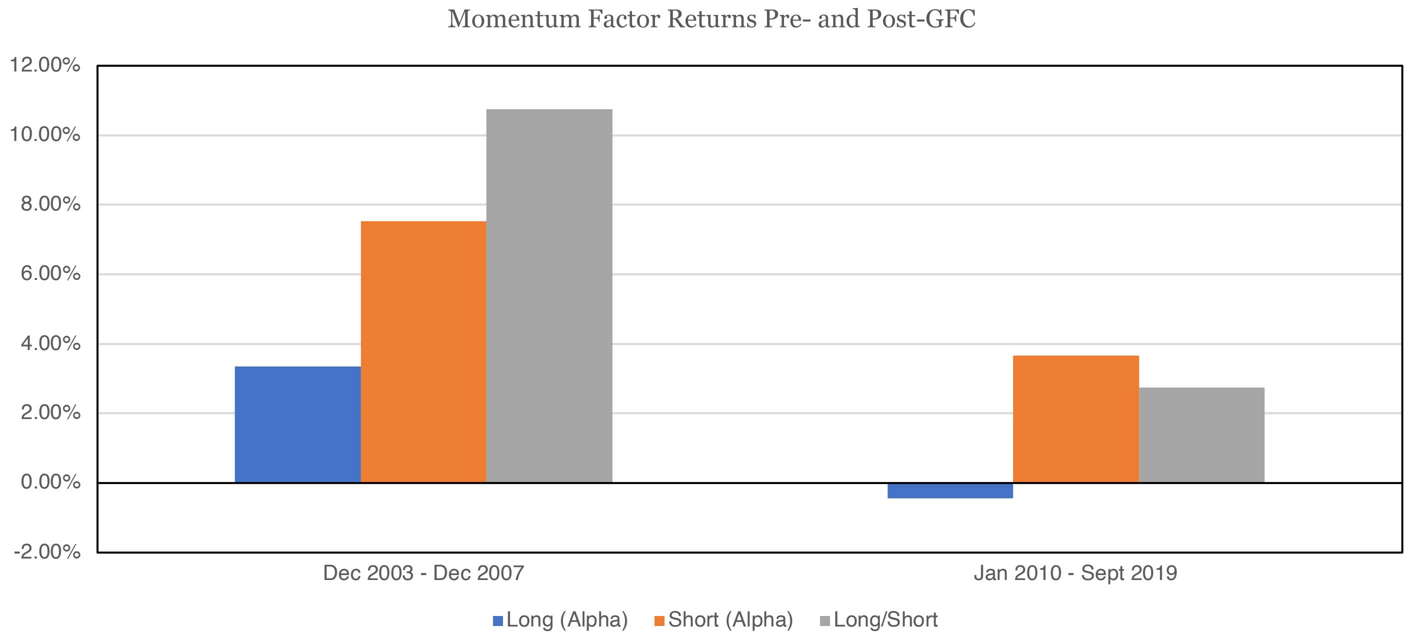
- One research note suggests that this is driven by increased risk aversion among investors, using the correlation of high volatility and low momentum baskets as evidence.

- In contradiction to this point, the iShares Momentum ETF (MTUM) has generated positive excess annualized returns against its benchmark since inception. The same note suggests that this is due to the use of risk-adjusted momentum measures.

- We explore whether risk-adjusting momentum scores introduces a meaningful and structural tilt towards low-volatility equities.

- For the examples tested, we find that it does not, and risk-adjusted momentum portfolios behave very similarly to momentum portfolios.

A research note recently crossed my desk that aimed to undress the post-Global Financial Crisis (GFC) performance of the momentum factor in U.S. equities. Not only have we witnessed a significant reduction in the factor's return, but the majority of the return has been generated by the short side of the strategy, which can be more difficult for long-only investors to access.

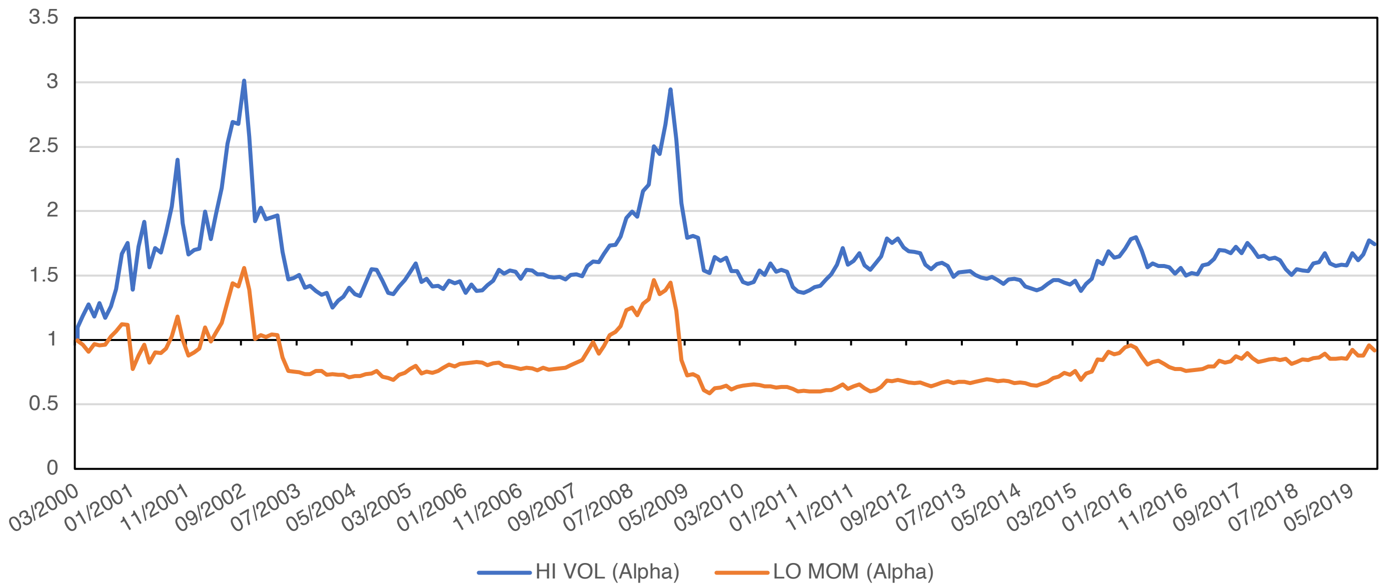


Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. The Long (Alpha) strategy is a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on 12-1 month momentum and shorts an equal-weight S&P 500 portfolio. The Short (Alpha) strategy is a monthly rebalanced portfolio that goes long an equal-weight S&P 500 portfolio and shorts, in equal weight, the bottom 50 securities in the S&P 500 ranked on 12-1 month momentum.

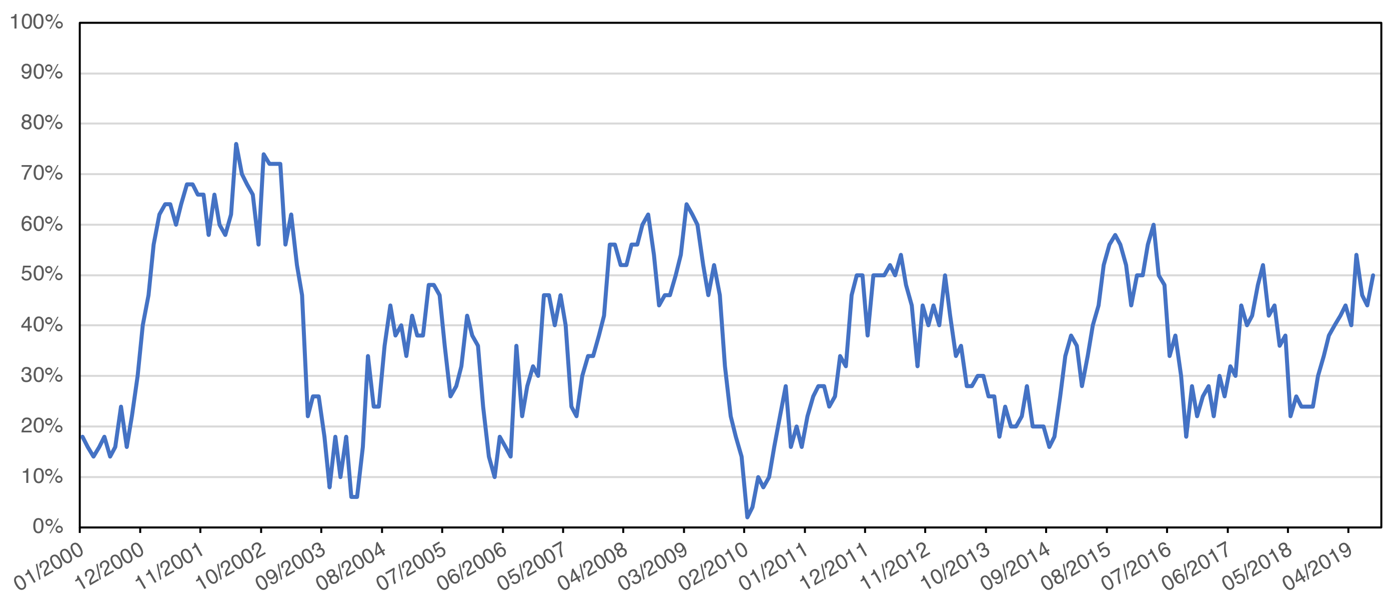
The note makes the narratively-appealing argument that the back-to-back recessions of the dot-com bubble and the Great Financial Crisis amplified investor risk aversion to downside losses. The proposed evidence of this fact is the correlation of the cumulative alpha generated from shorting low momentum stocks and the cumulative alpha generated from shorting high volatility stocks.

While correlation does not imply causation, one argument might be that in a heightened period of risk aversion, investors may consistently punish higher risk stocks, causing them to become persistent losers. Or, conversely, losers may be rapidly sold, creating both persistence and high levels of volatility. We can arguably see this in the convergence of holdings in low momentum and high volatility stocks during “risk off” regimes.

Cumulative Alpha



Percent Holdings Overlap in Low Momentum and High Volatility Decile Portfolios



Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. The HI VOL (Alpha) strategy is a monthly rebalanced portfolio that goes long an equal-weight S&P 500 portfolio and shorts, in equal weight, the bottom 50 securities in the S&P 500 ranked on trailing 252-day realized volatility. The LO MOM (Alpha) strategy is a monthly *rebalanced portfolio that goes long an equal-weight S&P 500 portfolio and shorts, in equal weight, the bottom 50 securities in the S&P 500 ranked on 12-1 month momentum.*

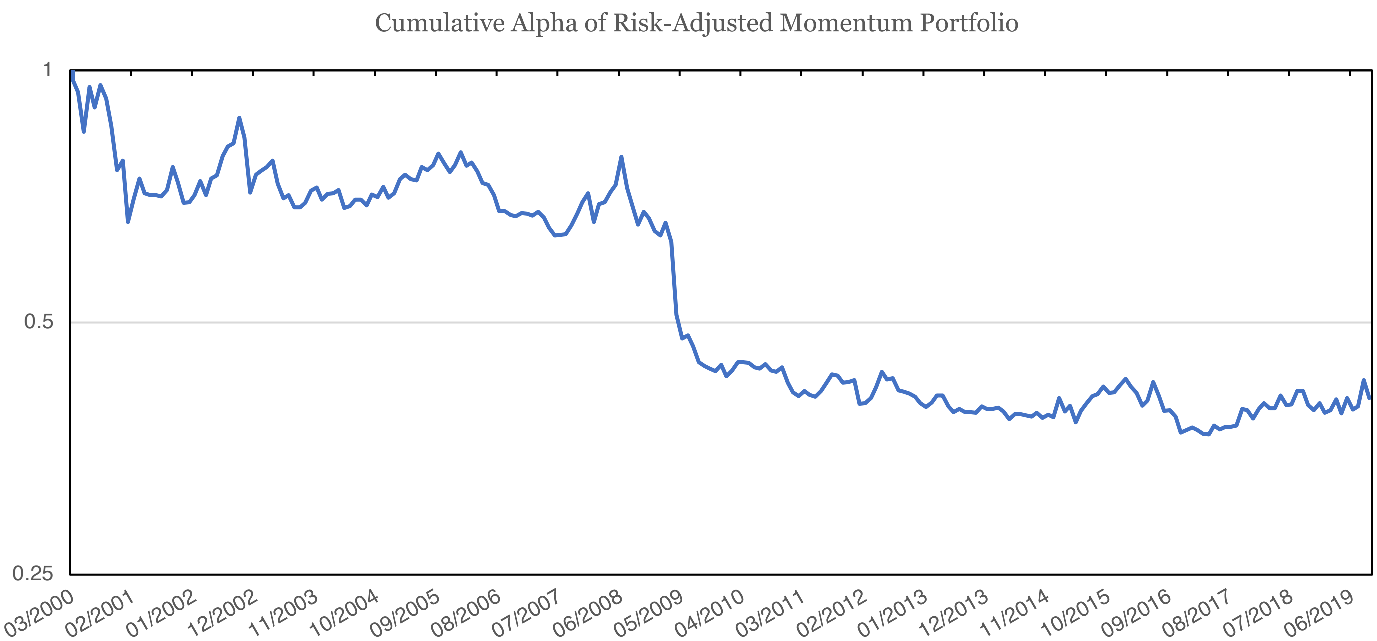
Given these facts, we would expect long-only momentum investors to have harvested little out-performance in recent years. Yet we find that the popular iShares Momentum ETF (MTUM) has out-performed the S&P 500 by 290 basis points per year since its inception in 2013.

The answer to this conundrum, as proposed by the research note, is that MTUM's use of *risk-adjusted* momentum is the key.

If we think of risk-adjusted momentum as simply momentum dividend by volatility (which is how MTUM defines it), we might interpret it as an integrated signal of both the momentum and low-volatility factors. Therefore, risk-adjusting creates a multi-factor portfolio that tilts away from high volatility stocks.

And hence the out-performance.

Except if we actually create a risk-adjusted momentum portfolio, that does not appear to really be the case at all.

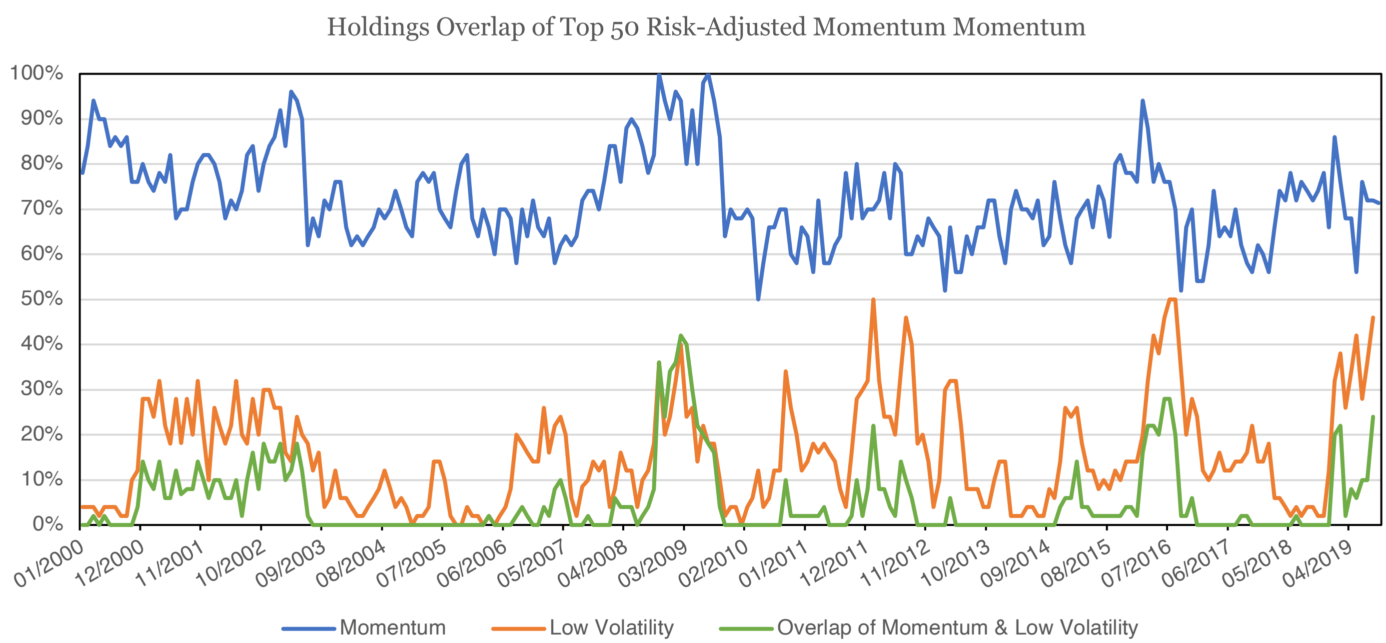


Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. The alpha of the risk-adjusted momentum strategy is defined as the return of a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on risk-adjusted momentum (12-1 month momentum divided by 252-day realized volatility) and shorts an equal-weight S&P 500 portfolio.

To be fair, MTUM's construction methodology differs quite a bit from that employed herein. We are simply equally-weighting the top 50 stocks in the S&P 500 when ranked by risk-adjusted momentum, whereas MTUM uses a blend of 6- and 12-month risk-adjusted momentum scores and then tilts market-capitalization weights based upon those scores.

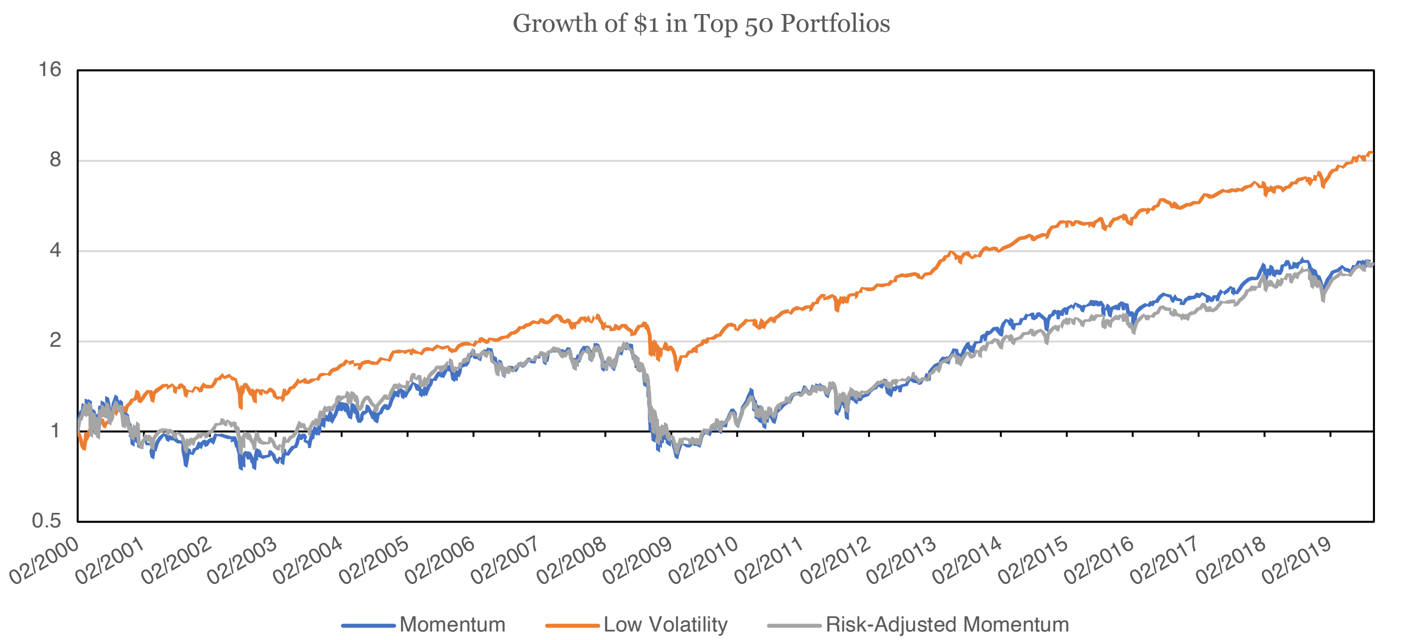
Nevertheless, if we look at actual holdings overlap over time of our Risk-Adjusted Momentum portfolio versus Momentum and Low Volatility portfolios, not only do we see persistently higher overlap with the Momentum portfolio, but we see fairly low average overlap with the Low Volatility portfolio.

For the latter point, it is worth first anchoring ourselves to the standard overlap between Momentum and Low Volatility (green line below). While we can see that the Risk-Adjusted Momentum portfolio does indeed have a higher average overlap with Low Volatility than does the Momentum portfolio, the *excess* tilt to Low Volatility due to the use of risk-adjusted momentum (i.e. the orange line minus the green line) appears rather small. In fact, on average, it is just 10%.



Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. The risk-adjusted momentum strategy is a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on risk-adjusted momentum (12-1 month momentum divided by 252-day realized volatility). The momentum strategy is a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on 12-1 month momentum. The low volatility strategy is a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on trailing 252-day realized volatility.

This is further evident by looking at the actual returns of the strategies themselves:



Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. The risk-adjusted momentum strategy is a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on risk-adjusted momentum (12-1 month momentum divided by 252-day realized volatility). The momentum strategy is a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on 12-1 month momentum. The low volatility strategy is a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on trailing 252-day realized volatility.

The Risk-Adjusted Momentum portfolio performance tracks that of the Momentum portfolio very closely.

As it turns out, the step of adjusting for risk creates far less of a low volatility factor tilt in our top-decile portfolio than one might initially suspect. (Or, at least, I'll speak for myself: it created far less of a tilt than *I* expected.)

To understand this point, we will first re-write our risk-adjusted momentum signal as:

$$MOM_{RA} = \frac{MOM}{VOL} = MOM * \left(\frac{1}{VOL}\right) = MOM * INVVOL$$

While trivial algebra, re-writing risk-adjusted momentum as the product of momentum and inverse volatility is informative to understanding why risk-adjusted momentum appears to load much more heavily on momentum than low volatility.

At a given point in time, it would appear as if Momentum and Low Volatility should have an equal influence on the rank of a given security. However, we need to dig a level deeper and consider how changes in these variables impact change in risk-adjusted momentum.

Fortunately, the product makes this a trivial exercise: holding INVVOL constant, changes in MOM are scaled by INVVOL and vice versa. This scaling effect can cause large changes in risk-adjusted momentum – and therefore ordinal ranking – particularly as MOM crosses the zero level.

Consider a trivial example where INVVOL is a very large number (e.g. 20) due to a security having a very low volatility profile (e.g. 5%). This would appear, at first glance, to give a security a structural advantage and hence create a low volatility tilt in the portfolio. However, a move from positive prior returns to negative prior returns would shift the security from ranking among the best to ranking among the worst in risk-adjusted momentum.[58]

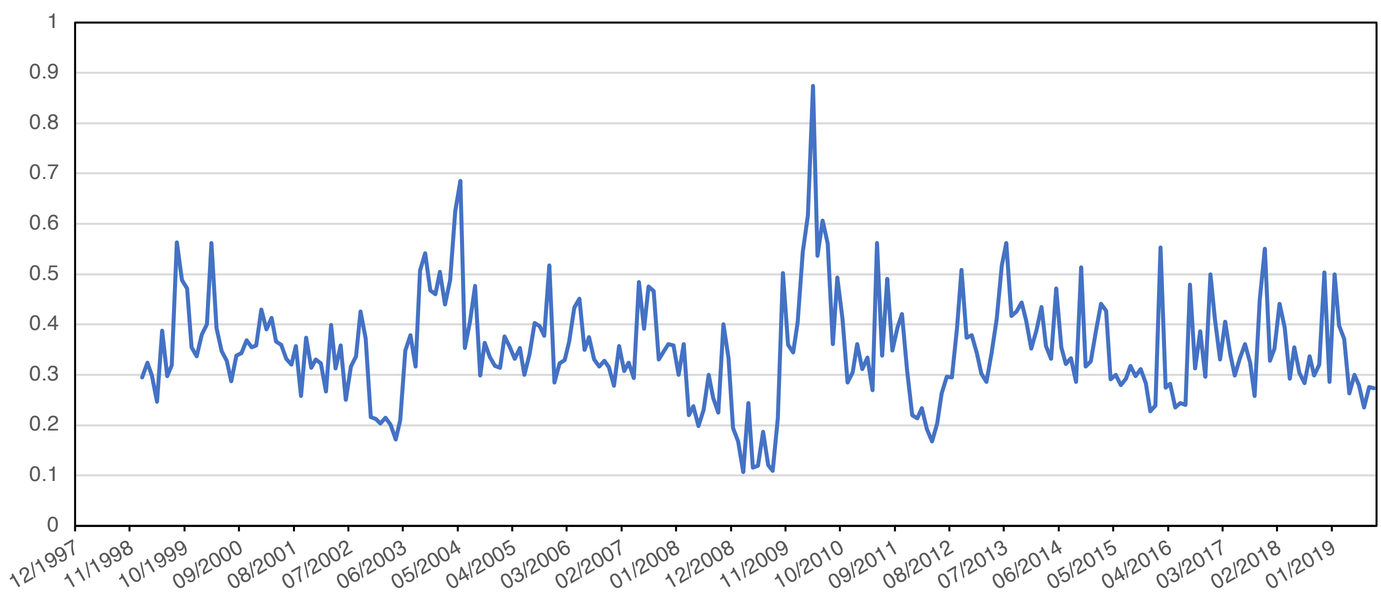
A first order estimate of change in risk-adjusted momentum is:

$$\Delta(MOM * INVVOL) \approx \Delta MOM * INVVOL + MOM * \Delta INVVOL$$

So which term ultimately has more influence on the change in scores over time?

To get a sense of relative scale, we plot the cross-sectional mean absolute difference between the two terms over time. This should, at least partially, capture interaction effects between the two terms.

Mean Absolute Difference of $\Delta MOM * INVVOL$ and $MOM * \Delta INVVOL$



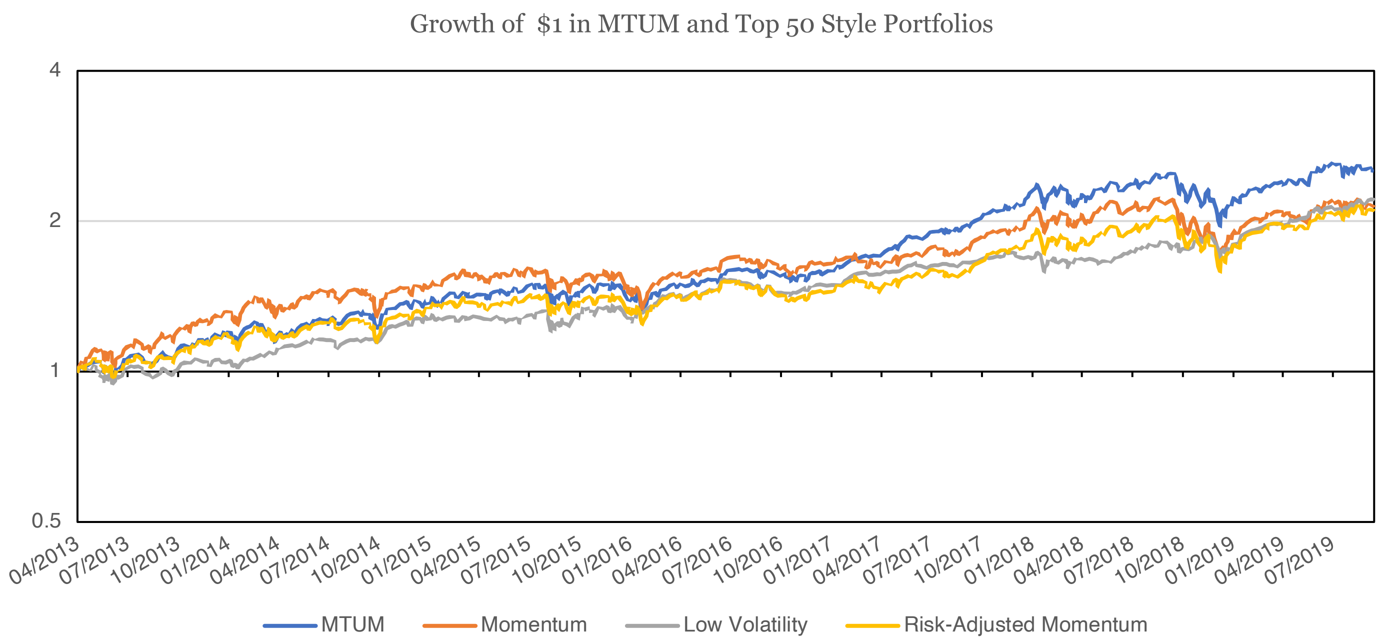
Source: Sharadar. Calculations by Newfound Research.

---

58 While it does not make much of a difference for our "top 50" portfolios discussed herein, an appropriate fix for this issue is to divide by volatility when momentum is positive but multiply by volatility when momentum is negative. This fix keeps negative risk-adjusted returns orderable. MSCI does not appear to take this step in their index methodology.

We can see that the term including the change in MOM has a much more significant influence on changes in risk-adjusted momentum than changes in INVVOL do. Thus, we might expect a portfolio driven entirely by changes in momentum to share more in common with our risk-adjusted momentum portfolio than one driven entirely by changes in volatility.

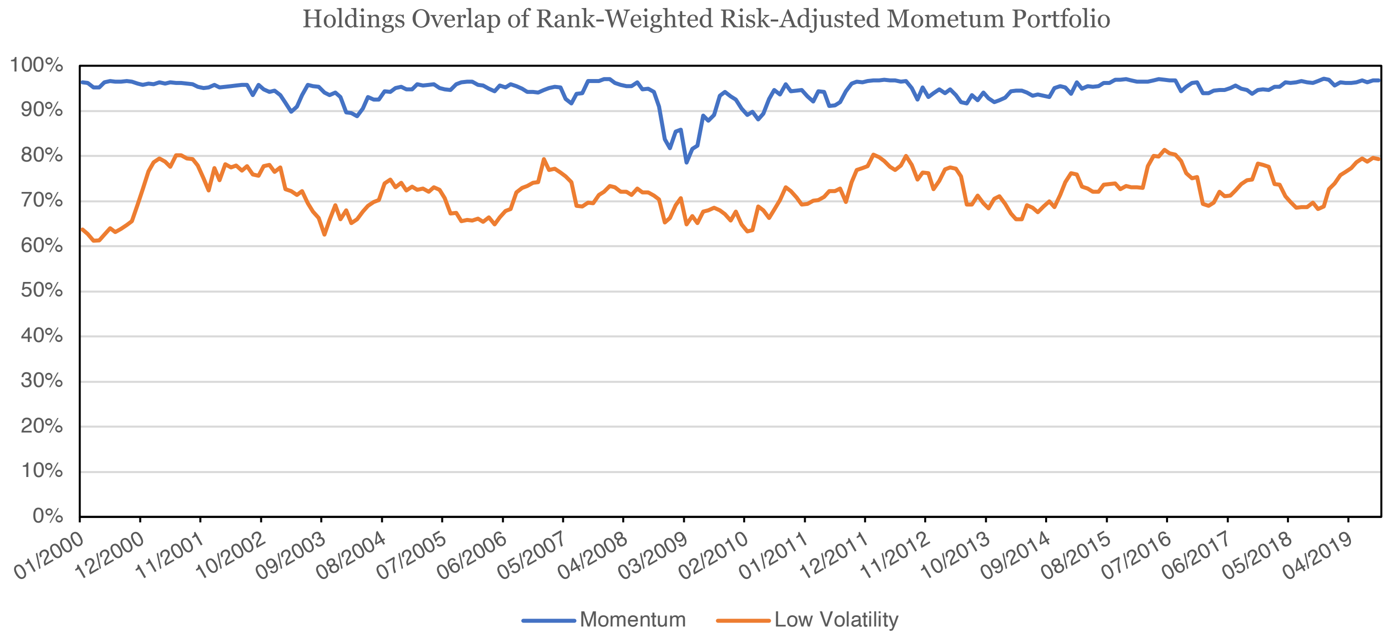
This is somewhat evident when we plot the return of MTUM versus our top 50 style portfolios. The correlation of daily returns between MTUM and our Momentum, Low Volatility, and Risk-Adjusted Momentum portfolios is 0.93, 0.72, and 0.93 respectively, further suggesting that MTUM is driven more by momentum than volatility.



Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. The risk-adjusted momentum strategy is a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on risk-adjusted momentum (12-1 month momentum divided by 252-day realized volatility). The momentum strategy is a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on 12-1 month momentum. The low volatility strategy is a monthly rebalanced portfolio that goes long, in equal weight, the top 50 securities in the S&P 500 ranked on trailing 252-day realized volatility.

This is only one part of the equation, however, as it is possible that changes to the risk-adjusted momentum score are so small – despite being largely driven by momentum – that relative rankings never actually change. Or, because we have constructed our portfolios by choosing only the top 50 ranked securities, that momentum does drive the majority of change across the entire universe, but the top 50 are always structurally advantaged by the non-linear scaling of low volatility.

To create a more accurate picture, we can rank-weight the entire S&P 500 and evaluate the holdings overlap over time.



Source: Sharadar. Calculations by Newfound Research.

Note that by now including all securities, and not just selecting the top 50, the overlap with both the Momentum and Low Volatility portfolios naturally appears higher on average. Nonetheless, we can see that the overlap with the Momentum portfolio is consistently higher than that of the Low Volatility portfolio, again suggesting that momentum has a larger influence on the overall portfolio composition than volatility does.

Conclusion

Without much deep thought, it would be easy to assume that a risk-adjusted momentum measure – i.e. prior returns divided by realized volatility – would tilt a portfolio towards both prior winners and low-volatility securities, resulting in a momentum / low-volatility barbell.

Upon deeper consideration, however, the picture complicates quickly. For example, momentum can be both positive and negative; dividing by volatility creates a non-linear impact; and momentum tends to change more rapidly than volatility.

We do not attempt to derive a precise, analytical equation that determines which of the two variables ultimately drives portfolio composition, but we do construct long-only example portfolios for empirical study. We find that a high-concentration risk-adjusted momentum portfolio has significantly more overlap in holdings with a traditional momentum portfolio than a low-volatility portfolio, resulting in a more highly correlated return stream.

The most important takeaway from this note is that intuition can be deceiving: it is important to empirically test our assumptions to ensure we truly understand the impact of our strategy construction choices.

Factor Orphans

October 28, 2019

Summary

- To generate returns which are different than the market, we must adopt a positioning that is different than the market.
- With the increasing adoption of systematic factor portfolios, we explore whether an anti-factor stance can generate contrarian-based profits.
- Specifically, we explore the idea of factor orphans: stocks that are not included in any factor portfolio at a given time.
- To identify these stocks, we replicate four popular factor indices: the S&P 500 Enhanced Value index, the S&P 500 Momentum index, the S&P 500 Low Volatility index, and the S&P 500 Quality index.
- On average, there are over 200 stocks in the S&P 500 that are orphaned at any given time.
- Generating an equal-weight portfolio of these stocks does not exhibit meaningfully different performance than a naïve equal-weight S&P 500 portfolio.

Contrarian investing is nothing new. Holding a variant perception to the market is often cited as a critical component to generating differentiated performance. The question in the details is, however, "contrarian to what?"

In the last decade, we've witnessed a dramatic rise in the popularity of systematically managed active strategies. These so-called "smart beta" portfolios seek to harvest documented risk premia and market anomalies and implement them with ruthless discipline.

But when massively adopted, do these strategies become the commonly-held view and therefore more efficiently priced into the market? Would this mean that the variant perception would actually be buying those securities totally ignored by these strategies?

This is by no means a new idea. Morningstar has long maintained its Unloved strategy that purchases the three equity categories that have witnessed the largest outflows at the end of the year. A few years ago, Vincent Deluard constructed a "DUMB" beta portfolio that included all the stocks shunned by popular factor ETFs. In the short out-of-sample period the performance of the strategy was tested, it largely kept pace with an equal-factor portfolio. More recently, a Bank of America research note claimed that a basket of most-hated securities – as defined by companies neglected by mutual funds and shorted by hedge funds hedge funds – had tripled the S&P 500's return over the past year.

The approach certainly has an appealing narrative: as the crowd zigs to adopt smart beta, we zag. But has it worked?

To test this concept, we wanted to identify what we call “factor orphans”: those securities not held by any factor portfolio. Once identified, we can build a portfolio holding these stocks and track its performance over time.

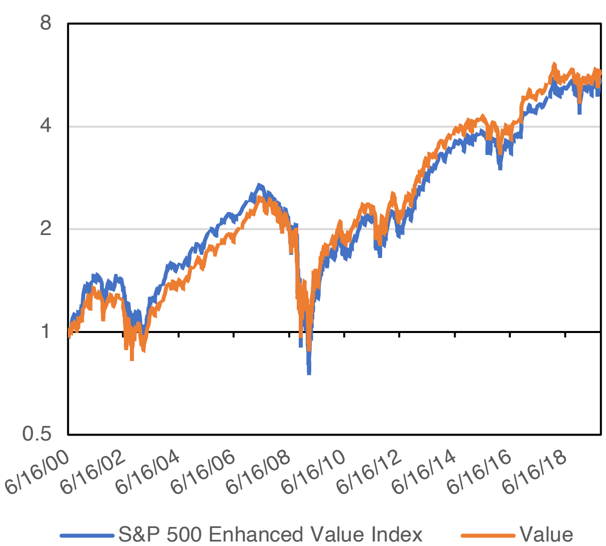
As a quant, this idea strikes us as a little crazy. A stock *not* held in a value, momentum, low volatility, or quality index is likely one that is expensive, highly volatile, with poor fundamentals and declining performance. Precisely the type of stock factor investing would tell us not to own.

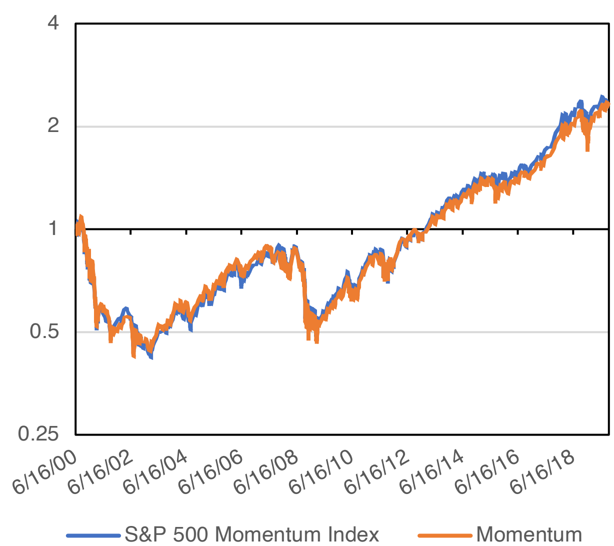
But perhaps the fact that these securities are orphaned means that there are no more sellers: the major cross-section of market strategies have already abandoned the stock. Thus, stepping in to buy them may allow us to offload them later when they are picked back up by these systematic approaches.

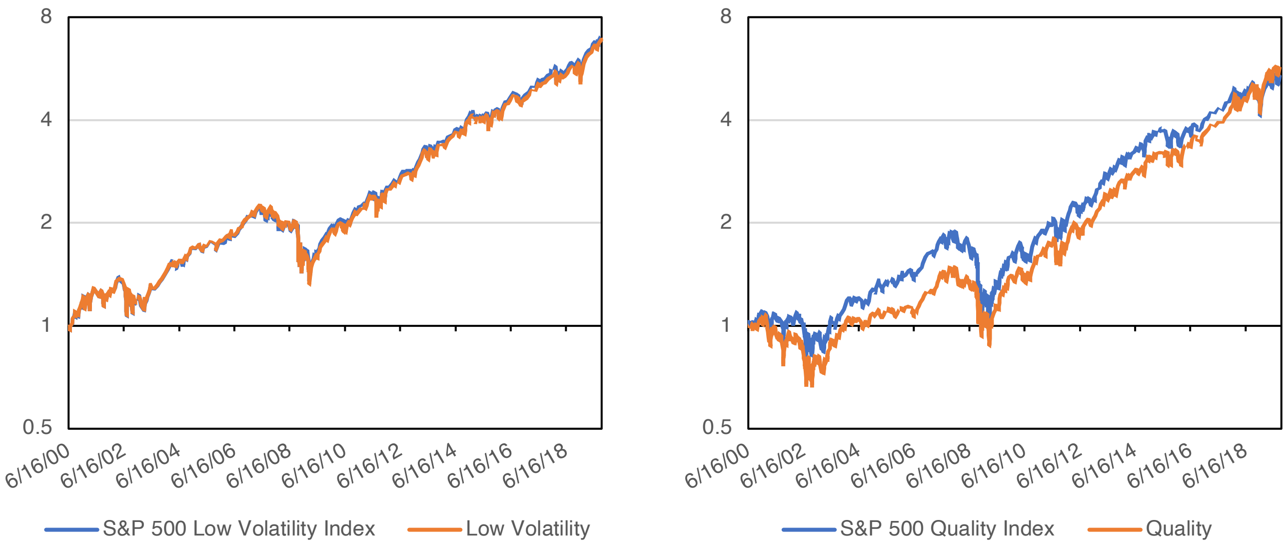
Perhaps this idea is crazy enough it just might work…

To test this idea, we first sought to replicate four common factor benchmarks: the S&P 500 Enhanced Value index, the S&P 500 Momentum index, the S&P 500 Low Volatility index and the S&P 500 Quality index. Once replicated, we can use the underlying baskets as being representative of the holdings for factor portfolios is general.

Results of our replication efforts are plotted below. We can see that our models fit the shape of most of the indices closely, with very close fits for the Momentum and Low Volatility portfolios.





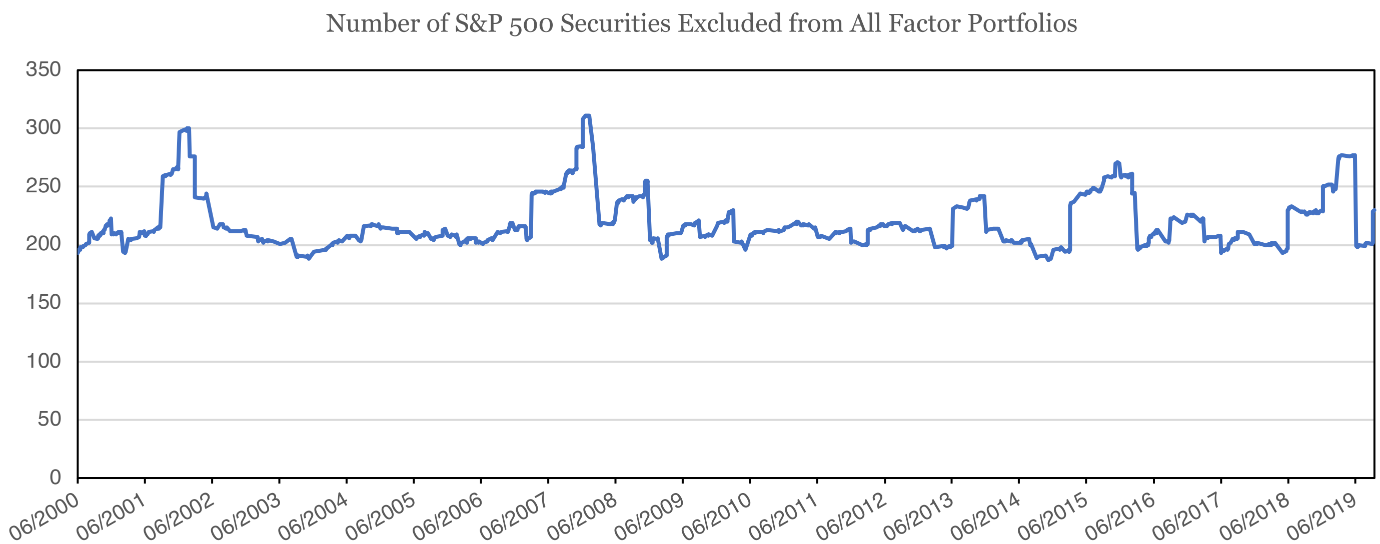


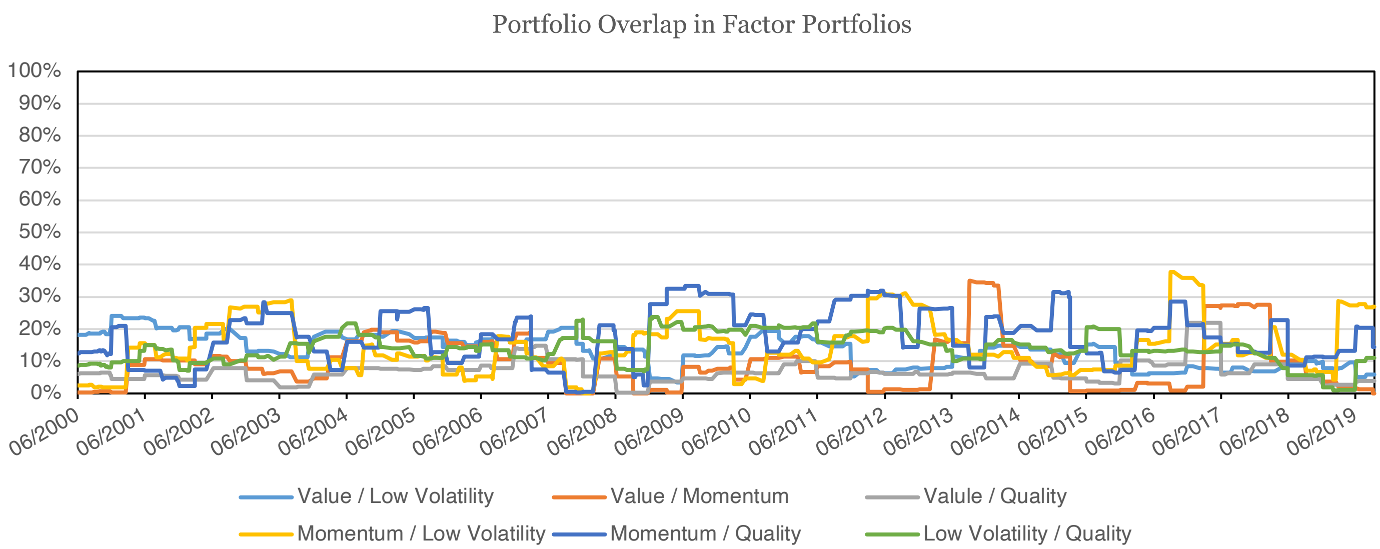
Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

The Quality replication represents the largest deviation from the underlying index, but still approximates the shape of the total return profile rather closely. This gives us confidence that the portfolio we constructed is a quality portfolio (which should come as no surprise, as securities were selected based upon common quality metrics), but the failure to more closely replicate this index may represent a thorn in our ability to identify truly orphaned stocks.

At the end of each month, we identify the set of all securities held by any of the four portfolios. The securities in the S&P 500 (at that point in time) but not in the factor basket are the orphaned stocks. Somewhat surprisingly, we find that approximately 200 names are orphaned at any given time, with the number reaching as high as 300 during periods when underlying factors converge.

Also interesting is that the actual overlap in holdings in the factor portfolios is quite low, rarely exceeding 30%. This is likely due to the rather concentrated nature of the indices selected, which hold only 100 stocks at a given time.

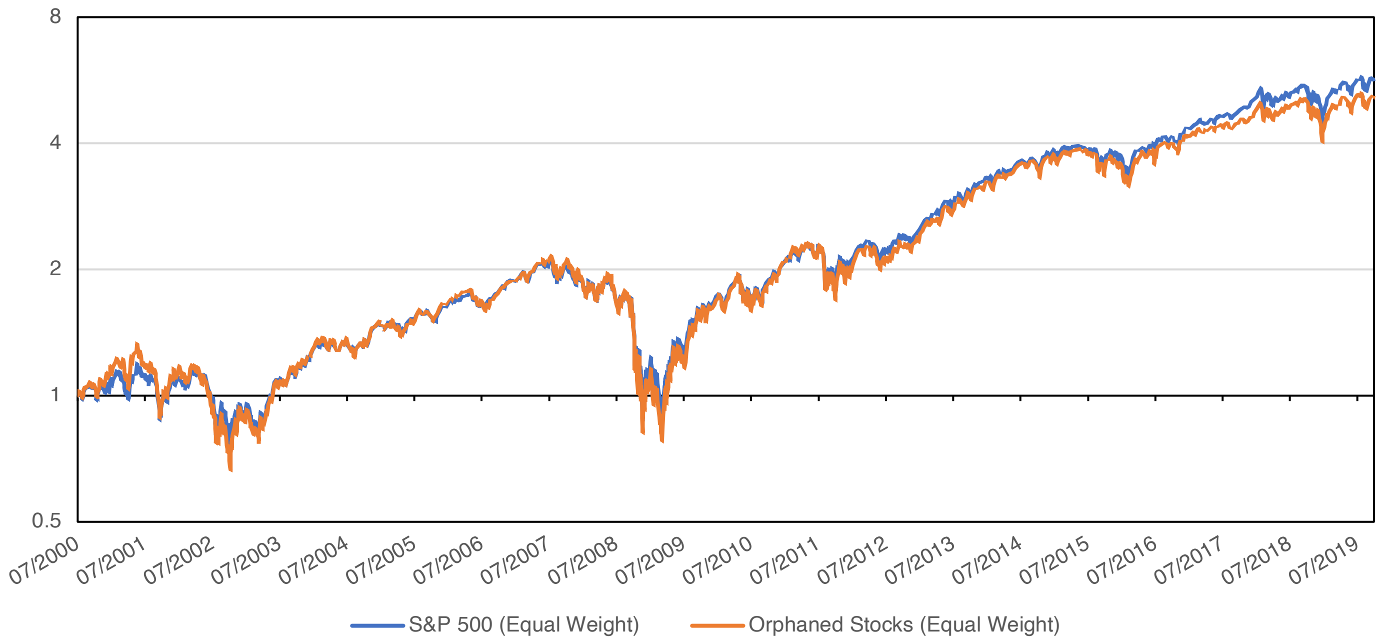




Source: Sharadar. Calculations by Newfound Research.

Once our orphaned stocks are identified, we construct a portfolio that holds them in equal weight. We rebalance our portfolio monthly to sell those stocks that have been acquired by a factor portfolio and roll into those securities that have been abandoned.

We plot the results of our exercise below as well as an equally weighted S&P 500 benchmark.

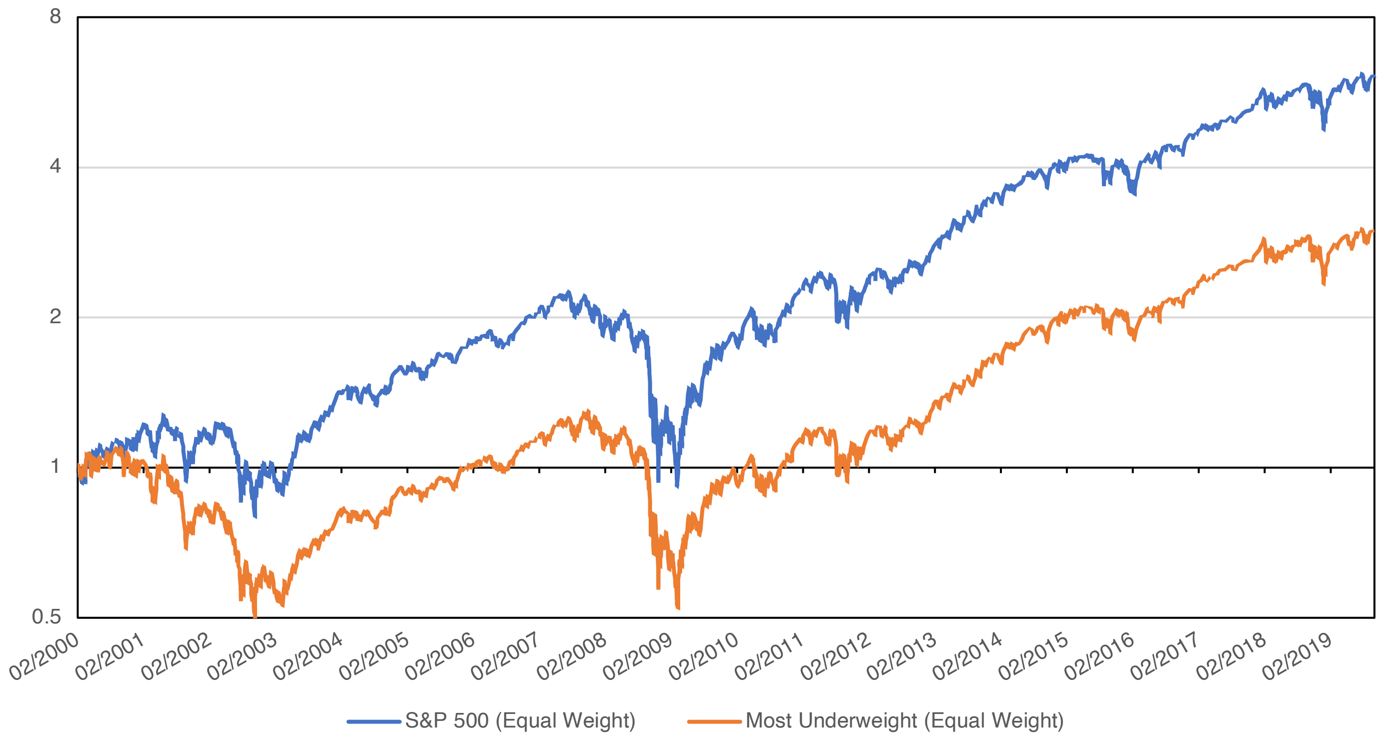


Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

While the total return is modestly less (but certainly not statistically significantly so), what is most striking is how little deviation there is in the orphaned stock portfolio versus the equal-weight benchmark.

However, as we have demonstrated in the past, the construction choices in a portfolio can have a significant impact upon the realized results. As we look at the factor portfolios themselves, we must acknowledge that they represent *relative* tilts to the benchmark, and that the absence of one security might actually represent a significantly smaller relative underweight to the benchmark than the absence of another. Or the absence of one security may actually represent a smaller relative underweight than another that is actually included.

Therefore, as an alternative test we construct an equal-weight factor portfolio and subtract the S&P 500 market-capitalization weights. The result is the implied over- and under-weights of the combined factor portfolios. We then rank securities to select the 100 most under-weight securities each month and hold them in equal weight.



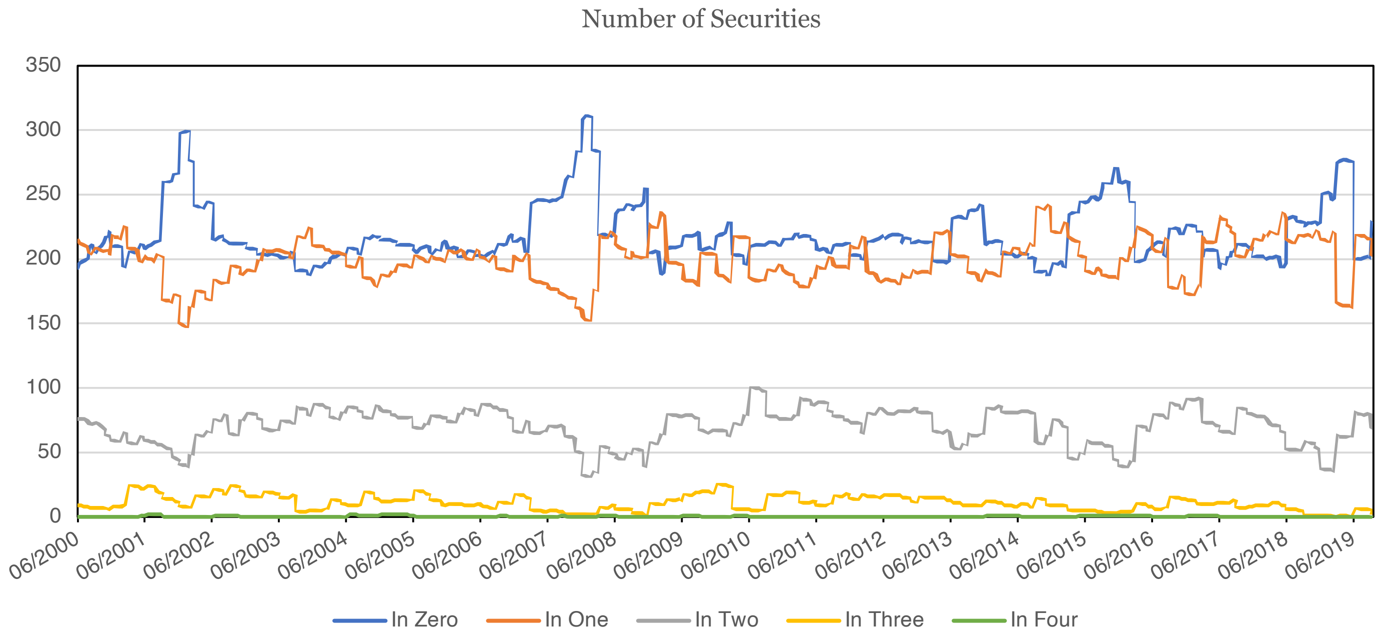
Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

Of course, we didn't actually have to perform this exercise had we stepped back to think for a moment. We generally know that these (backtested) factors have out-performed the benchmark. Therefore, selecting stocks that they are underweight means we're taking the opposite side of the factor trade, which we know has not worked.

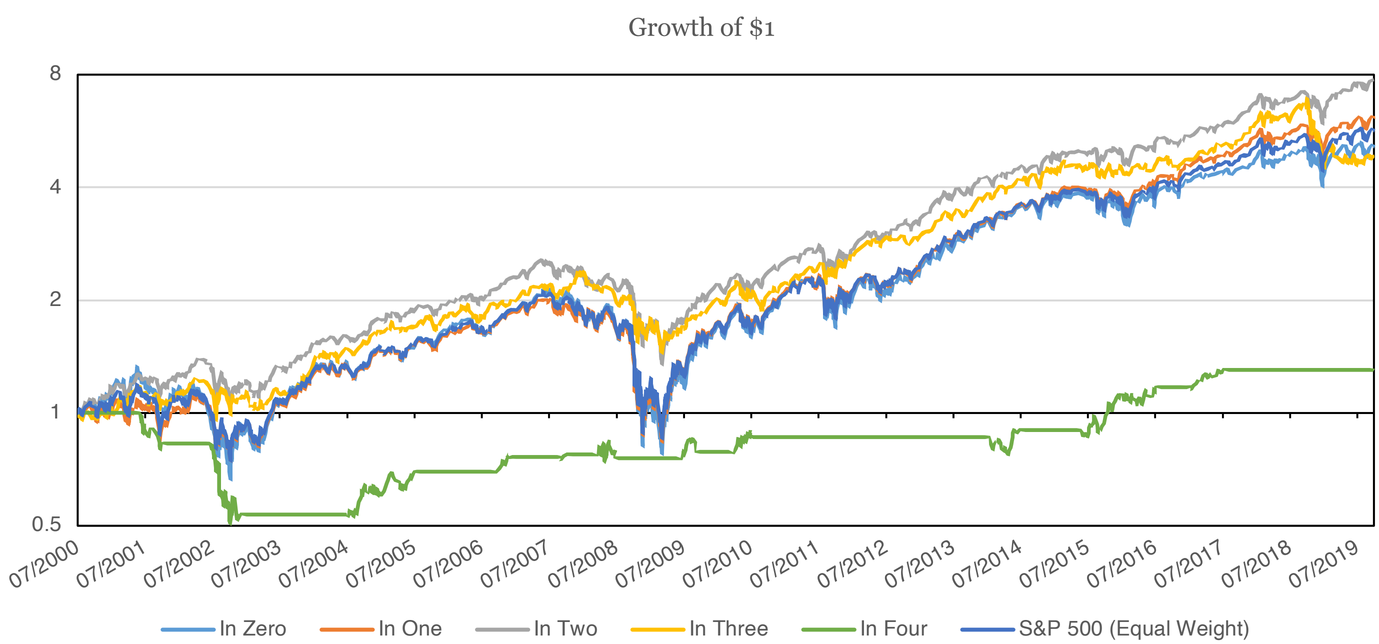
Which does draw an important distinction between *most underweight* and *orphaned.* It would appear that factor orphans do not necessarily create the strong anti-factor tilt the way that the most underweight portfolio does.

For the sake of completion, we can also evaluate the portfolios containing securities held in just one of the factor portfolios, two of the factor portfolios, three of the factor portfolios, or all of the factor portfolios at a given time.

Below we plot the count of securities in such portfolios over time. We can see that it is very uncommon to identify securities that are simultaneously held by all the factors, or even three of the factors, at once.



Source: Sharadar. Calculations by Newfound Research.



Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

We can see that the portfolio built from stocks held in just one factor ("In One") closely mimics the portfolio built from stocks held in no factor ("In Zero"), which in turn mimics the S&P 500 Equal Weight portfolio. This is likely because the portfolios include so many securities that they effectively bring you back to the index.

On the other end of the spectrum, we see the considerable risks of concentration manifest in the portfolios built from stocks held in three or four of the factors. The portfolio comprised of stocks held in all four factors simultaneously ("In Four") not only goes long stretches of holding nothing at all, but is also subject to large bouts of volatility due to the extreme concentration.

We also see this for the portfolio that holds stocks held by three of the factors simultaneously ("In Three"). While this portfolio has modestly more diversification – and even appears to out-perform the equal-weight benchmark – the concentration risk finally materializes in 2018-2019, causing a dramatic drawdown.

The portfolio holding stocks held in just two of the factors ("In Two"), though, appears to offer some out-performance opportunity. Perhaps by forcing just two factors to agree, we strike a balance between confirmation among signals and portfolio diversification.

Unfortunately, our enthusiasm quickly wanes when we realize that this portfolio closely matches the results achieved just by naively equally-weighting exposure among the four factor portfolios themselves, which is far more easily implemented.

Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

Conclusion

To achieve differentiated results, we must take a differentiated stance from the market. As systematic factor portfolios are more broadly adopted, we should consider asking ourselves if taking an anti-factor stance might lead to contrarian-based profits.

In this study, we explore the idea of factor orphans: stocks not held by any factor portfolio at a given time. Our hypothesis is that these orphaned securities may be systematically over-sold, leading to an opportunity for future out-performance if they are re-acquired by the factor portfolios at a later date.

We begin by replicating four factor indices: the S&P 500 Enhanced Value index, the S&P 500 Momentum index, the S&P 500 Low Volatility index, and the S&P 500 Quality index. Replicating these processes allows us to identify historical portfolio holdings, which in turn allows us to identify stocks *not* held by the factors.

We are able to closely replicate the S&P 500 Momentum and Low Volatility portfolios, create meaningful overlap with the S&P 500 Enhanced Value method, and generally capture the S&P 500 Quality index. The failure to more closely replicate the S&P 500 Quality index may have a meaningful impact on the results herein, though we believe our methodology still captures the generic return of a quality strategy.

We find that, on average, there are over 200 factor orphans at a given time. Constructing an equal-weight portfolio of these orphans, however, only seems to lead us back to an S&P 500 Equal Weight benchmark. While there does not appear to be an edge in this strategy, it is interesting that there does not appear to be a *negative* edge either.

Recognizing that long-only factor portfolios represent active bets expressed as over- and underweights relative to the S&P 500, we also construct a portfolio of the most underweight stocks. Not surprisingly, as this portfolio actively captures a negative factor tilt, the strategy meaningfully underperforms the S&P 500 Equal Weight benchmark. Though the relative underperformance meaningfully dissipates in recent years.

Finally, we develop portfolios to capture stocks held in just one, two, three, or all four of the factors simultaneously. We find the portfolios comprised stocks held in either three or four of the factors at once exhibit significant concentration risk. As with the orphan portfolio, the portfolio of stocks held by just one of the factors closely tracks the S&P 500 Equal Weight benchmark, suggesting that it might be *over*-diversified.

The portfolio holding stocks held by just two factors at a time appears to be the Goldilocks portfolio, with enough concentration to be differentiated from the benchmark but not so much as to create significant concentration risk.

Unfortunately, this portfolio also almost perfectly replicates a naïve equal-weight portfolio among the four factors, suggesting that the approach is likely a wasted effort.

In conclusion, we find no evidence that factor orphans have historically offered a meaningful excess return opportunity. Nor, however, do they appear to have been a drag on portfolio returns either. We should acknowledge, however, that the adoption of factor portfolios accelerated rapidly after the Great Financial Crisis, and that backtests may not capture current market dynamics. More recent event studies of orphaned stocks being added to factor portfolios may provide more insight into the current environment.

GLOBAL GROWTH TREND TIMING

November 4, 2019

SUMMARY

- While trend following may help investors avoid prolonged drawdowns, it is susceptible to whipsaw where false signals cause investors to either buy high and sell low (realizing losses) or sell low and buy high (a missed opportunity).
- Empirical evidence suggests that using economic data in the United States as a filter of when to employ trend-following – a "growth-trend timing" model – has historically been fruitful.
- When evaluated in other countries, growth-trend timing has been historically successful in mitigating whipsaw losses without sacrificing the ability to avoid large drawdowns. However, we see mixed results on whether this actually improves upon naïve trend-following.
- We find that countries that can be influenced by factors originating outside of their borders might not benefit from an introspective economic signal.

We apologize in advance, as this commentary will be fairly graph- and table-heavy.

We have written fairly extensively on the topic of factor-timing in the past, and much of the success has been proven to be both hard to implement and recreate out of sample.

One of the inherent pains of trend following is the existence of whipsaws, or more precisely, the misidentification of perceived market trends, which turn out to be more noise than signal. An article from Philosophical Economics proposed using several economic indicators to tune down the noise that might affect price-driven signals such as trend following. Generally, this strategy imposed an overlay that turned trend following "on" when the change in the economic indicators were negative year-over-year signaling a higher likelihood of recession, and conversely, adopted a buy-and-hold stance when the economic indicators were not flashing warning lights.

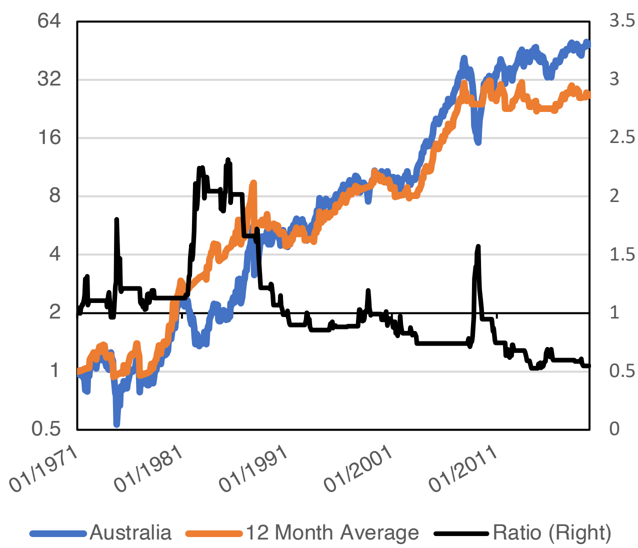
This strategy presents a certain appeal as leading economic indicators may, as their name implies, lead the market for some time until capital preservation is warranted. Switching to a trend-following approach may allow a strategy to continue to participate in market appreciation while it lasts. On the other hand, using economic confirmation as a filter may help a strategy avoid the whipsaw costs generated from noisy market dips while positive economic conditions persist.

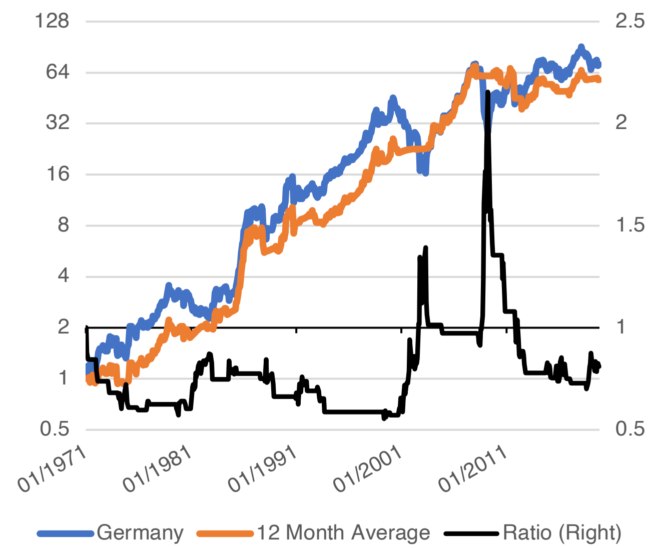
In an effort to test such a strategy out-of-sample, we took the approach global, hoping to capture a broader cross-section of economic and market environments.

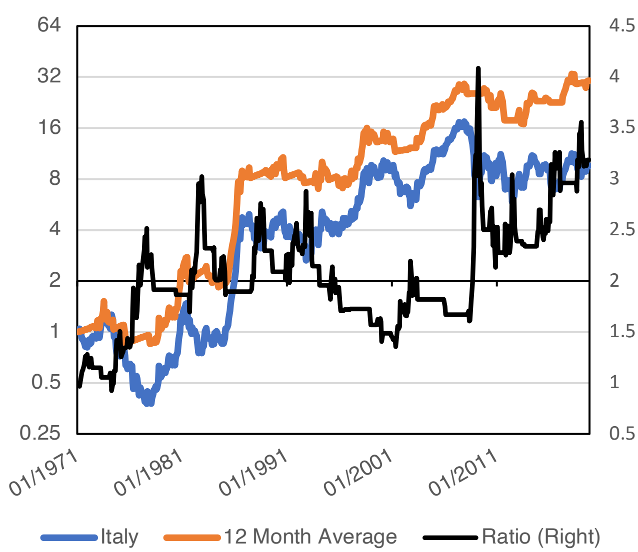
First, we will consider trend following with no timing using the economic indicators. [59]

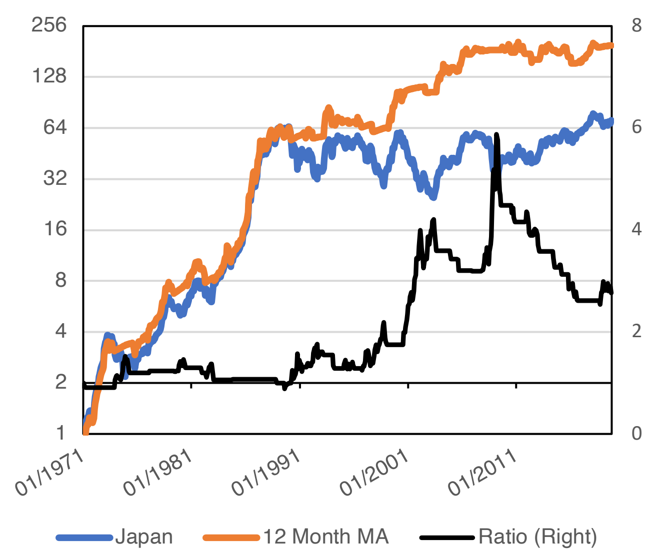
[59] Alpha Architect has an excellent post further exploring global trend-following and whipsaw costs.

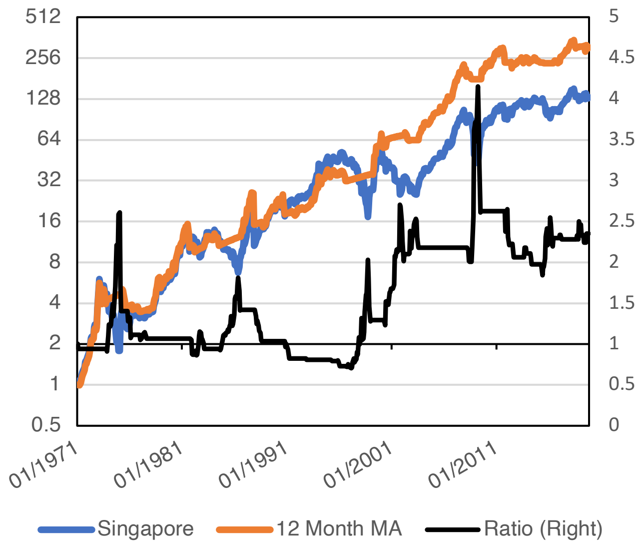
Below we plot the equity curves for Australia, Germany, Italy, Japan, Singapore, the United Kingdom, and the United States, alongside a strategy that is long the market when the market is above the trailing twelve-month average (“12 Month average”) and steps to cash when the price is below it. The ratio between the two is also included to show the relative cumulative performance between the trend strategy and the respective market. An increasing ratio means that the trend following strategy is adding value over buy-and-hold.

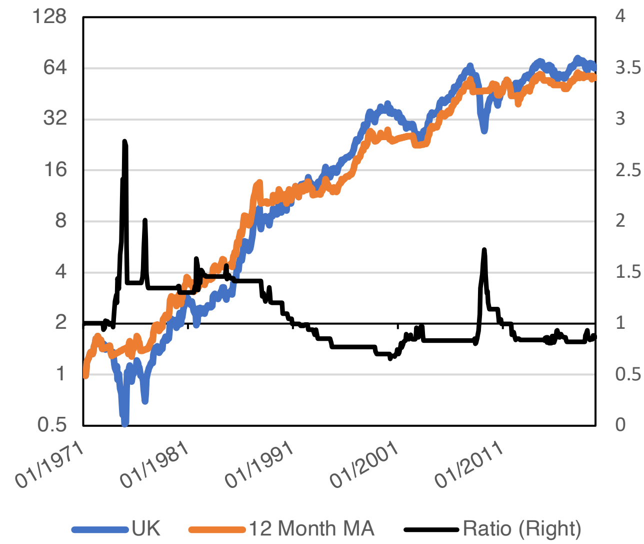


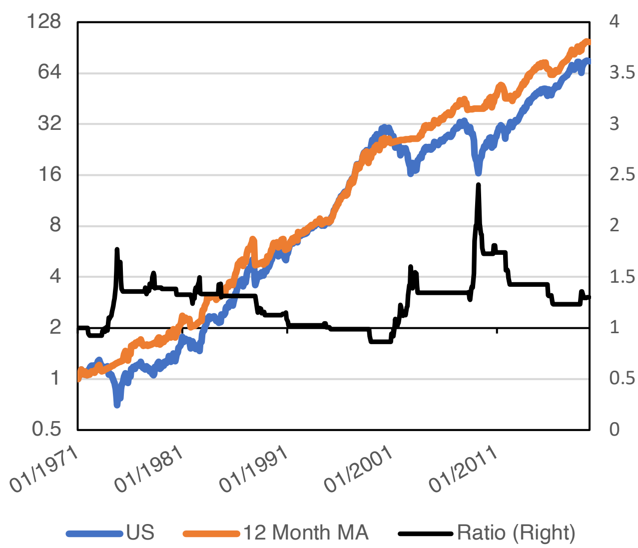












Source: MSCI, Global Financial Data. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

Through the graphs above, it becomes clear that much of the trend premium is realized by avoiding the large, prolonged bear markets that tend to occur during economic distress. In between these periods, however, the trend strategy lags the market. It makes sense, then, that a potential improvement to this strategy would be to implement an augmentation that could better distinguish between real price break-outs and those that lead to a whipsaw in the portfolio.

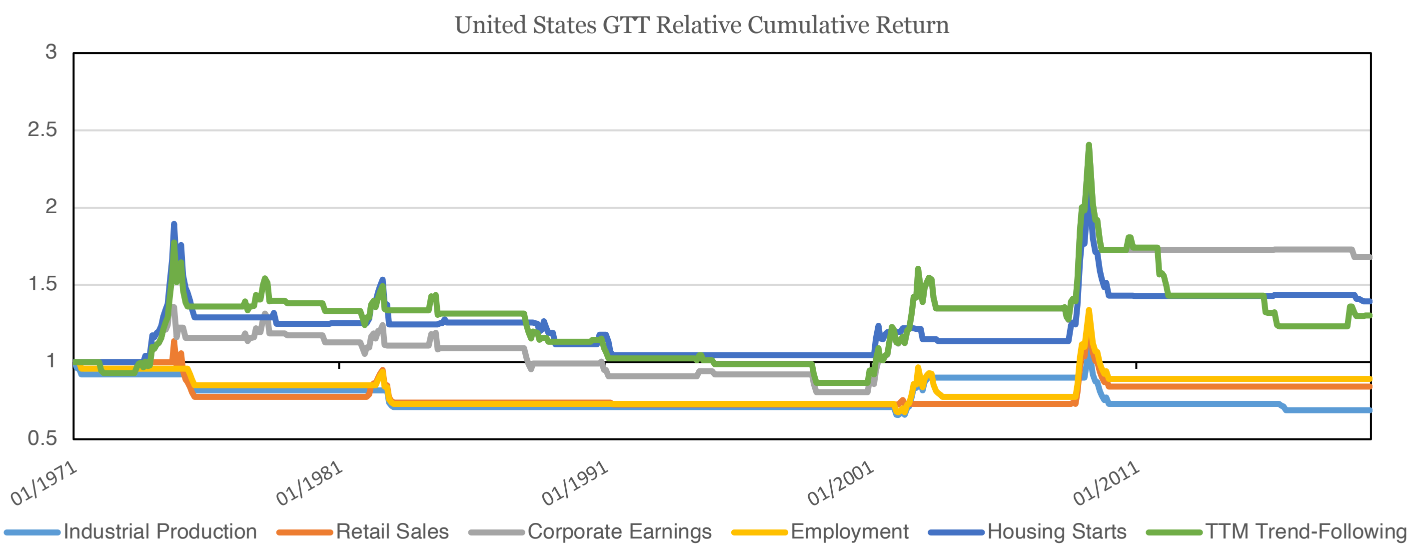
Growth-Trend Timing

For each country, we look at a number of economic indicators, including: corporate earnings growth, employment, housing starts, industrial production, and retail sales growth.[60] The strategy then followed the same rules as described above: if the economic indicator in question displays a negative percentage change over the previous twelve-month period, a position is taken in a trend following strategy utilizing a twelve-month moving average signal. Otherwise, a buy-and-hold position is established.

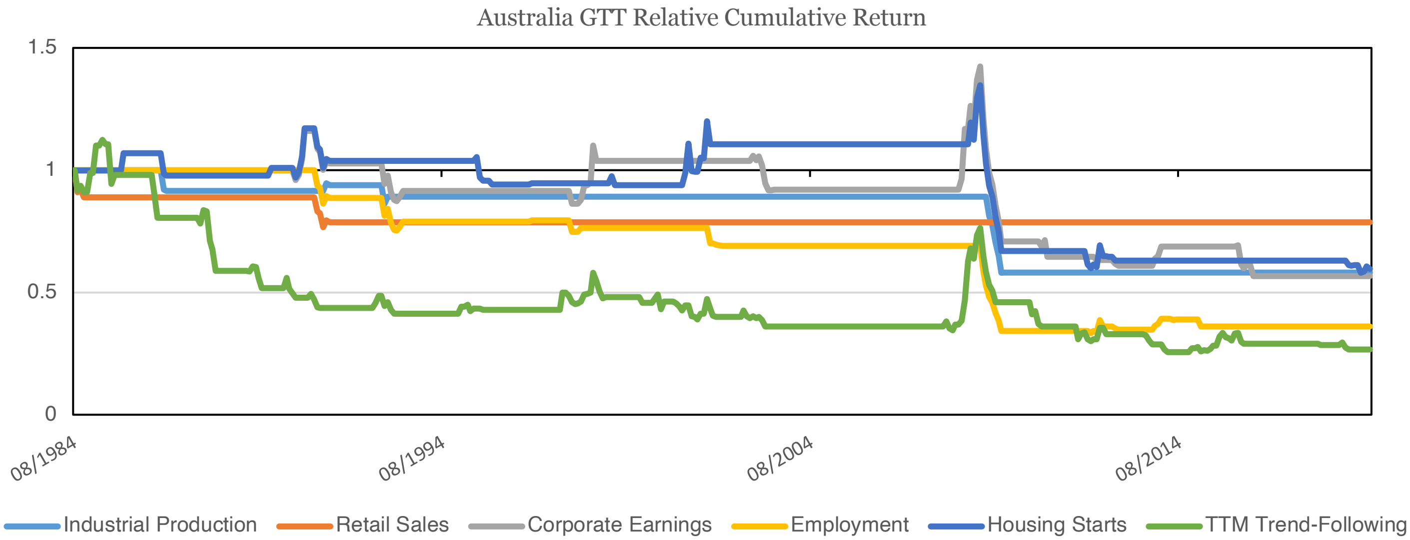
To ensure that we are not benefitting from look-ahead bias, a lag of three months was imposed on each of the economic indicators, as it would be unrealistic to assume that the economic levels would be known at the end of each month.

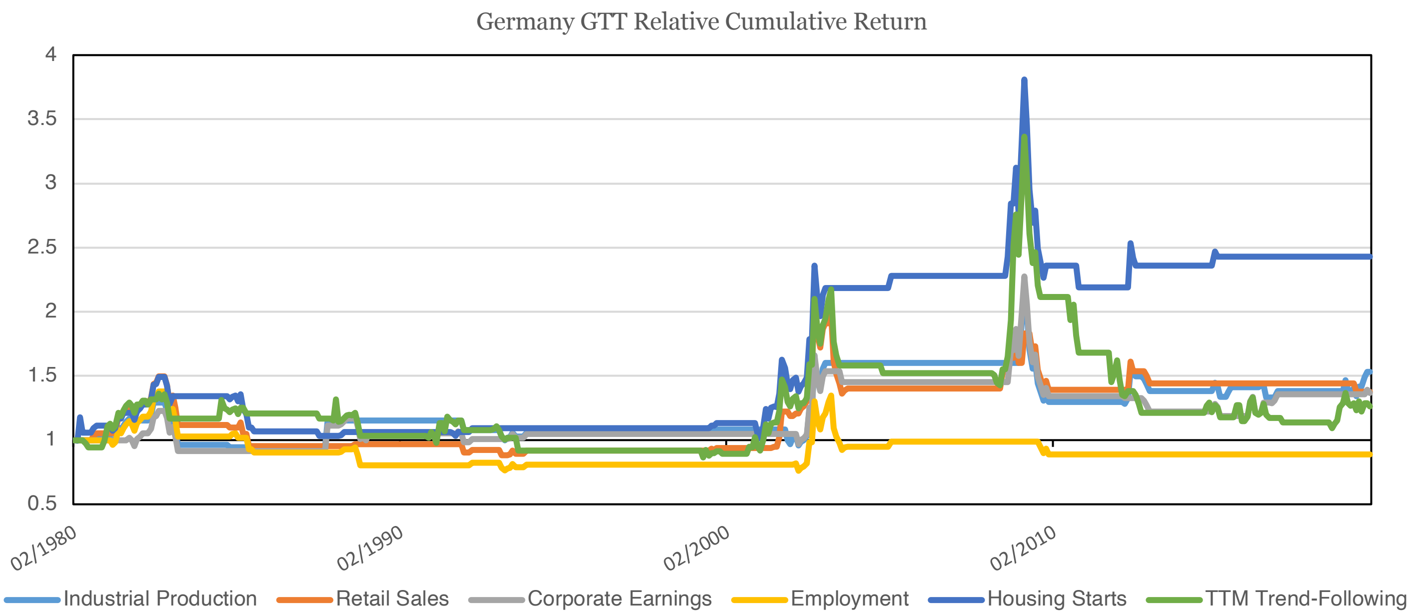
Unfortunately, some of the economic data points could not be found for the entire period in which prices are available, though the analysis can still prove beneficial by indicating what economic regimes trend following is benefitted by growth-trend timing, or the potential identification where one indicator may work when another does not.[61]

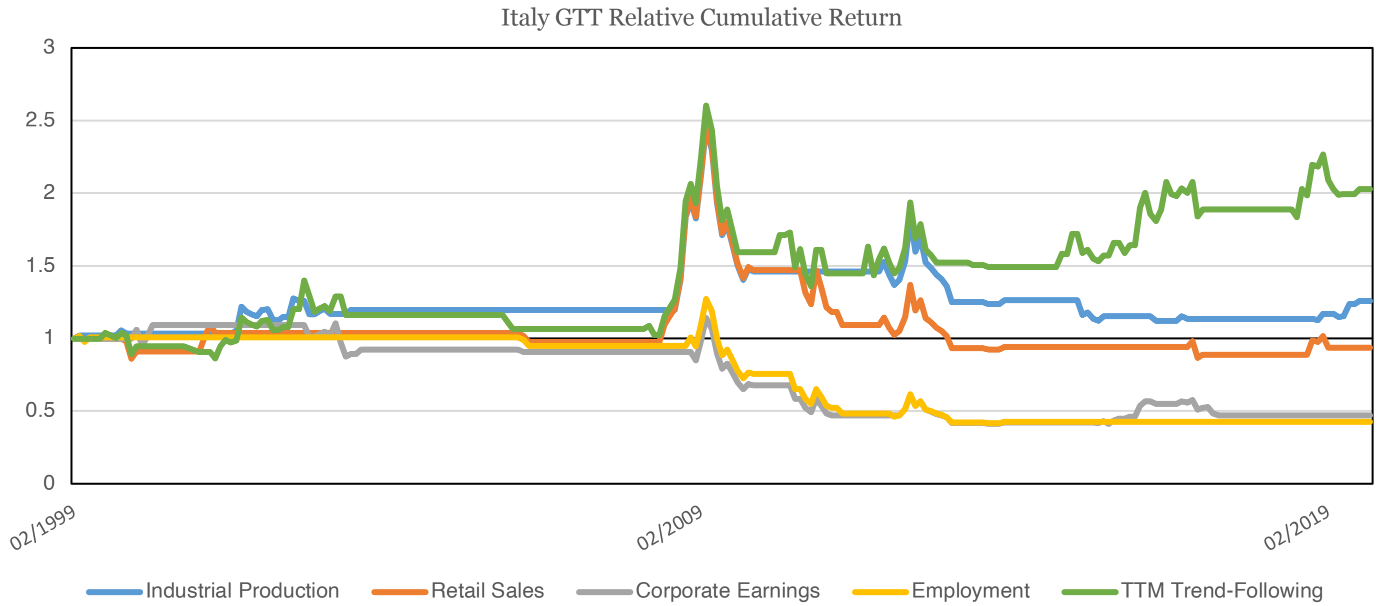
In the charts below, we plot the growth-trend timing (referred to as GTT for the remainder of this commentary) for each country utilizing the available signals. The charts represent the relative cumulative performance over the respective country's market return. For example, when the lines remain flat, the GTT approach has adopted buy-and-hold exposure and therefore matches the respective market's returns. Any changes in the ratios are due to the GTT strategy investing in the trend following strategy.

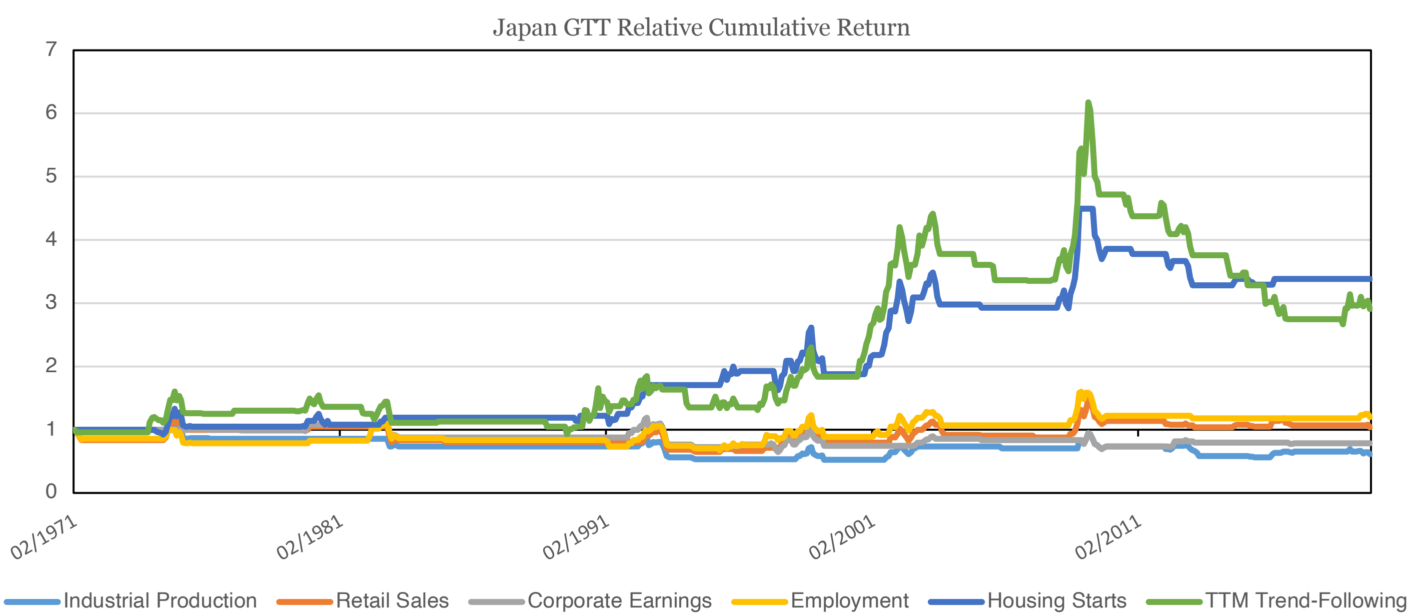


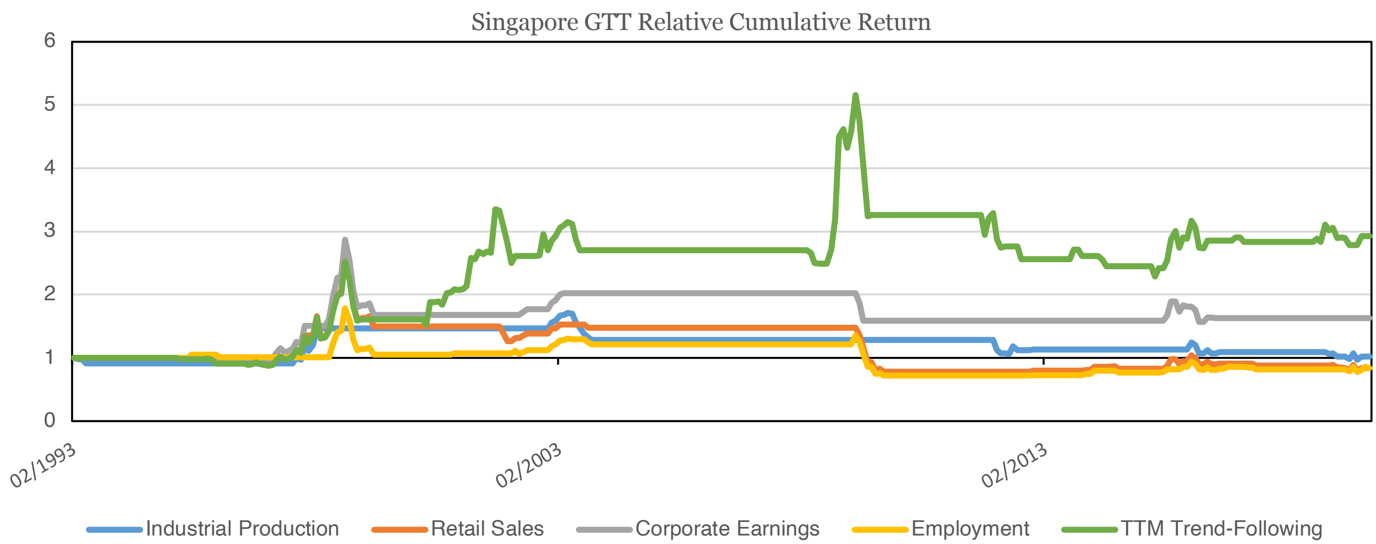
[60] Data was sourced from the St. Louis Fed, Global Financial Data, and Bloomberg.
[61] Data could not be reliably found for housing starts in Singapore. Reliable housing start data for Italy began in 2008, so it was removed for the majority of the analysis.

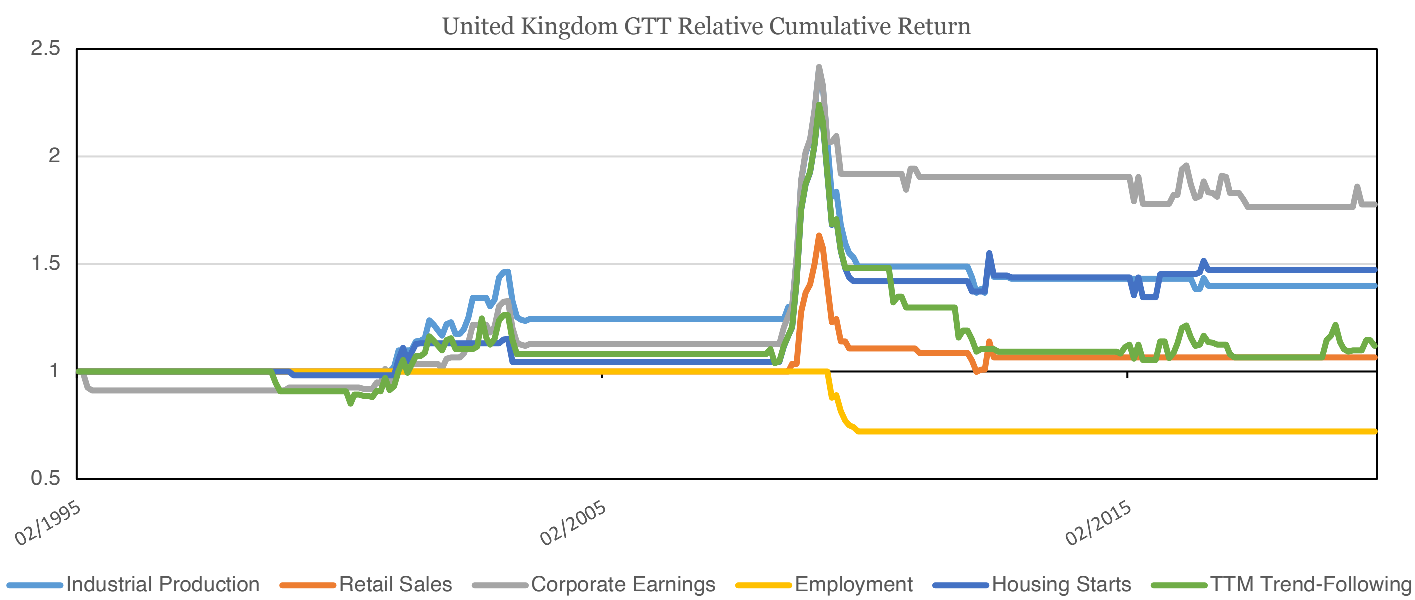

Australia GTT Relative Cumulative Return
1.5
1
0.5
0
08/1984
08/1994
08/2004
08/2014
Industrial Production
Retail Sales
Corporate Earnings
Employment
Housing Starts
TTM Trend-Following


Germany GTT Relative Cumulative Return
4
3.5
3
2.5
2
1.5
1
0.5
02/1980
02/1990
02/2000
02/2010
Industrial Production
Retail Sales
Corporate Earnings
Employment
Housing Starts
TTM Trend-Following


Italy GTT Relative Cumulative Return
3
2.5
2
1.5
1
0.5
0
02/1999
02/2009
02/2019
Industrial Production
Retail Sales
Corporate Earnings
Employment
TTM Trend-Following


Japan GTT Relative Cumulative Return
7
6
5
4
3
2
1
0
02/1971
02/1981
02/1991
02/2001
02/2011
Industrial Production
Retail Sales
Corporate Earnings
Employment
Housing Starts
TTM Trend-Following





Source: MSCI, Global Financial Data, St. Louis Fed, Bloomberg. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

What we see from the above figures is a mixed bag of results.

The overlay of economic indicators was by far successful in the mitigation of whipsaw losses, as each country reaped the benefits of being primarily long the market during bull markets. As the 12-month moving average strategy tended to slowly give up a portion of the gains realized from severe market environments, the majority of the GTT strategies remained relatively stagnant until the next major correction.

There are some instances, however, where the indicator was late to the economic party. It is worth remembering that the market is, in theory, a forward-looking measure, and therefore sudden economic shocks may not be captured in economic data as quickly as it is in market returns. This created cases where the strategy either missed the chance to be out of the market during a correction or was sitting on the sidelines during the subsequent recoveries. Notably, the employment signal in Australia, Italy, Singapore, and the United Kingdom tended to be a poor leading indicator as the strategy tended to be invested longer in the bear markets than the trend strategy.

A Candidate for Ensembling

The implicit assumption in the analysis above is that the included indicators behave in similar ways. For example, by using a twelve-month lookback period for the indicators, we are assuming that each indicator will begin to trend in roughly the same way.

That may not be a particularly fair assumption. Whereas housing starts and retail sales are generally considered leading indicators, employment (unemployment) rates are normally categorized as lagging indicators. For this reason, it may be more beneficial to use a shorter lookback period so as to pick up on potential problems in the economy as they begin to present themselves. Further, some signals tend to be more erratic than others, suggesting that a meaningful lookback period for one indicator may not be meaningful for another. With no perfect reason to prefer one lookback over another, we might consider different lookback periods so as to diversify any specification risk that may exist within the strategy.

With the benefit of hindsight, we know that not all recessions occur for the same reasons, so being reliant on one signal that has worked in the past may not be as beneficial in the future. With this in mind, we should consider that all indicators hold some information as to the state of the economy since one indicator may be signaling the all-clear while another may be flashing warning lights.

For the same reason medical professionals take multiple readings to gain insight into the state of the body, we should also consider any available signals to ascertain the health of the economy.

To ensemble this strategy, we will vary the lookbacks from six to eighteen months, while holding the lag at three months, as well as combine the available economic signals for each country. For the sake of brevity, we will hold the trend-following strategy the same with a twelve-month moving average.

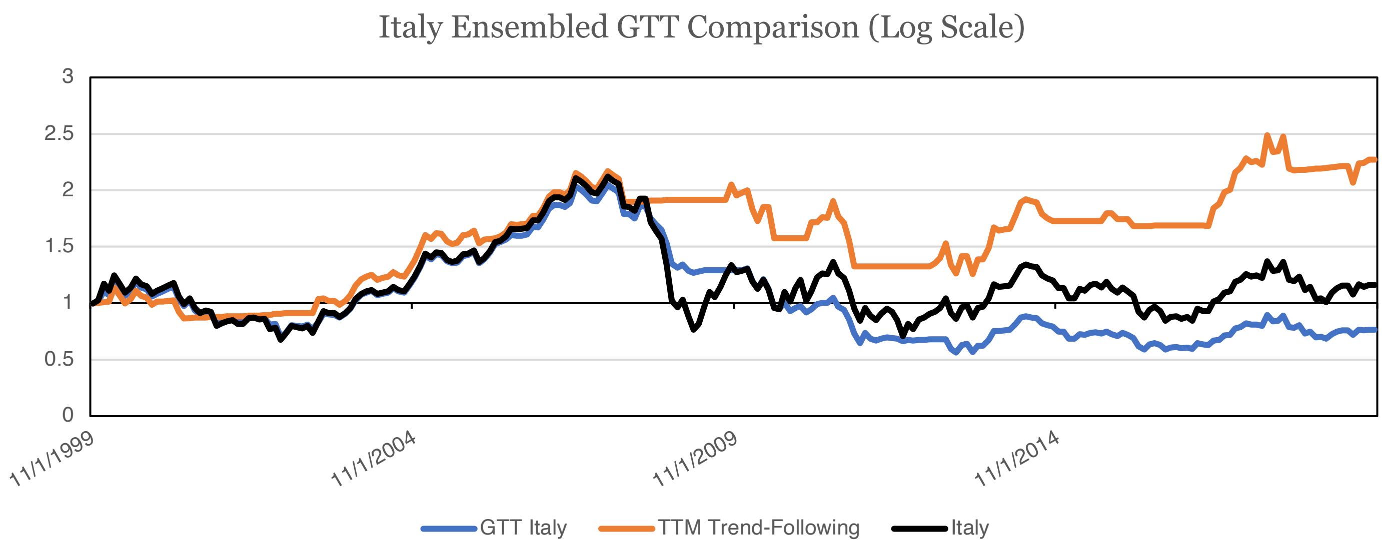
Remember, if the economic signal is negative, it does not mean that we are immediately out of the market: a negative economic signal simply moves the strategy into a trend-following approach. With 5 economic indicators and 13 lookback periods, we have 65 possible strategies for each country. As an example, if 40 of these 65 models were positive and 25 were negative, we would hold 62% in the market and 38% in the trend following strategy.

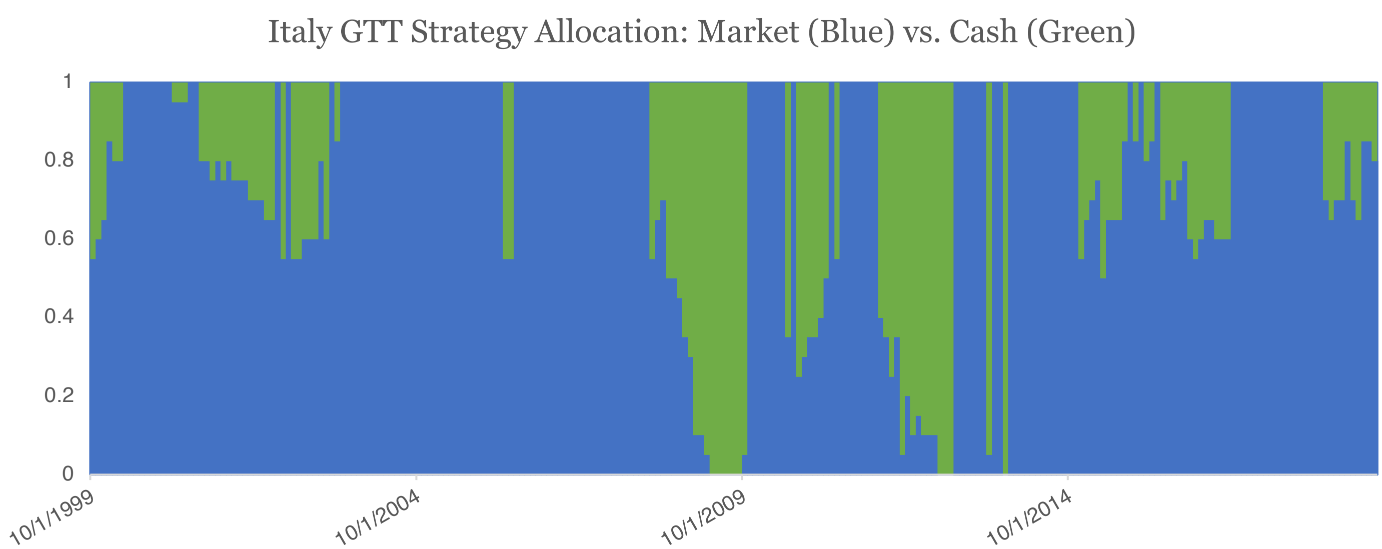
The resulting performance statistics can be seen in the table below.

		United-States	Germany	Japan	UK	Australia	Italy	Singapore
Market	*Annualized Return*	9.29%	8.59%	8.76%	5.41%	9.97%	0.76%	5.20%
	Volatility	14.99%	22.40%	20.46%	15.61%	23.22%	23.68%	24.54%
	Sharpe Ratio (Rf = 0)	0.62	0.38	0.43	0.35	0.43	0.03	0.21
	Max. Drawdown	51.12%	64.12%	62.01%	59.05%	63.26%	66.51%	66.43%
Twelve-Month Moving Average	*Annualized Return*	9.89%	9.41%	11.31%	5.91%	5.46%	4.21%	9.65%
	Volatility	11.55%	16.33%	14.76%	10.67%	18.14%	14.38%	14.82%
	Sharpe Ratio (Rf = 0)	0.86	0.58	0.77	0.55	0.3	0.29	0.65
	Max. Drawdown	29.61%	44.67%	28.35%	28.65%	52.85%	42.02%	29.25%
Ensembled 6-18 Month GTT	*Annualized Return*	9.30%	9.19%	9.56%	6.06%	8.28%	-1.33%	4.84%
	Volatility	13.01%	18.76%	17.23%	12.51%	21.69%	17.57%	19.80%
	Sharpe Ratio (Rf = 0)	0.71	0.49	0.55	0.48	0.38	-0.08	0.24
	Max. Drawdown	36.62%	48.37%	41.46%	33.26%	56.91%	72.50%	57.69%
	Beginning Date	Jan-1971	Nov-1980	Nov-1971	Nov-1995	May-1985	Nov-1999	Nov-1993

Source: MSCI, Global Financial Data, St. Louis Fed, Bloomberg. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

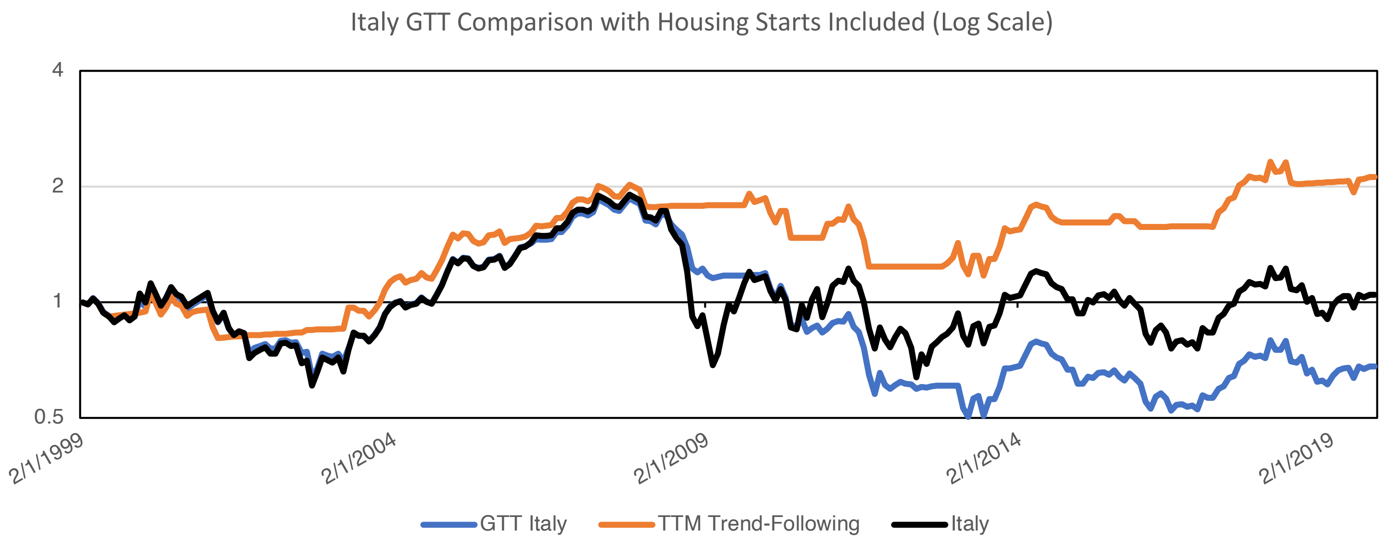
From the table above, we see that there are, again, mixed results. One country that particularly stands out is Italy in that the sign on its return flipped to negative and the drawdown was actually deeper with GTT than with a simple buy-and-hold strategy.





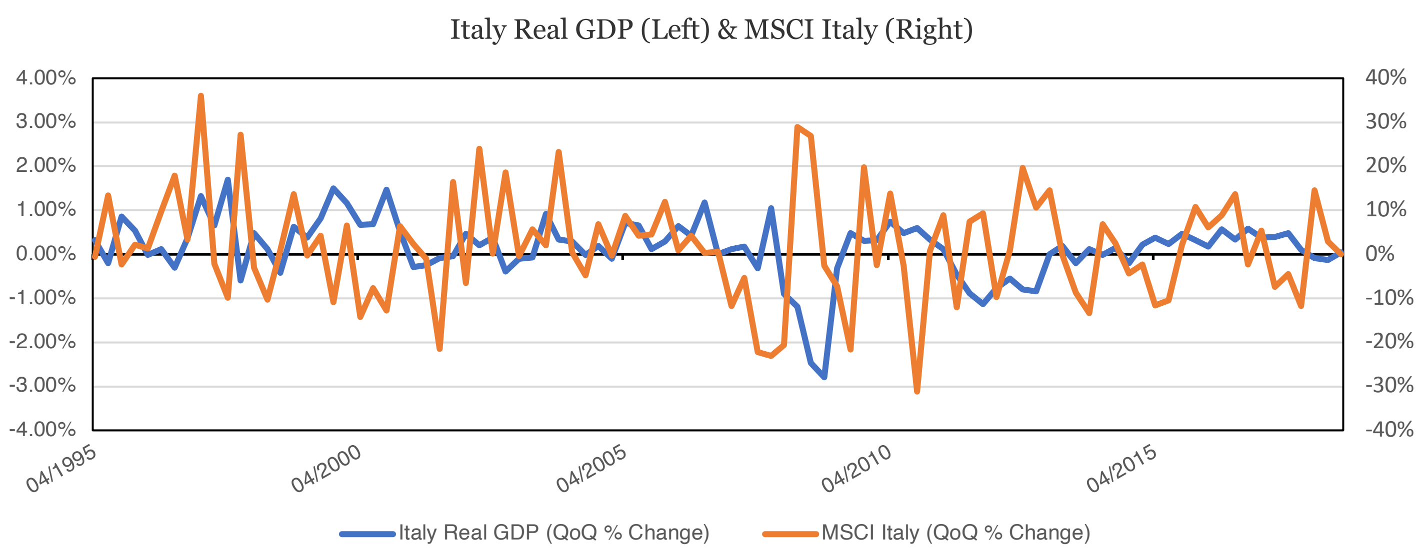
Source: MSCI, Global Financial Data, St. Louis Fed, Bloomberg. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

Digging deeper, it appears that the GTT strategy for Italy was actually whipsawed by more than just trend-following. Housing start data for Italy was not readily available until December 2008, so Italy may have been at a relative disadvantage when compared against the other countries. Since the reliable data we could find begins at the end of 2008 and the majority of the whipsaw losses occur post-Great Financial Crisis, we can run the analysis again, but with housing start data being added in upon its availability.



Source: MSCI, Global Financial Data, St. Louis Fed, Bloomberg. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

Adding housing starts in as an indicator did not meaningfully alter the results over the period. One hypothesis is that the indicators included could not fully encapsulate the complex state of Italy's economy over the period. Italy has weathered three technical recessions over the past decade, so this could be a regime where the market is looking to sources outside the country for indications of distress or where the economic indicator is not reflective of the pressures driving the market.



Source: MSCI, St. Louis Fed. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

Above, we can see several divergences between the market movement and changes in real GDP. Specifically, in the past decade, we see that the market reacted to information that didn't materialize in the country's real GDP. More likely, the market was reacting to regional financial distress driven by debt concerns.

The MSCI Italy index is currently composed of 24 constituents with multinational business operations. Additionally, the index maintains large concentrations in financials, utilities, and energy: 33%, 25%, and 14%, respectively.[62] Because of this sector concentration, utilizing the economic indicators may overly focus on the economic health of Italy while ignoring external factors such as energy prices or broader financial distress that could be swaying the market needle.

A parallel explanation could be that the Eurozone is entangled enough that signals could be interfering with each other between countries. Further research could seek to disaggregate signals between the Eurozone and the member-countries, attempting to differentiate between zone, regional, and country signals to ascertain further meaning.

Additionally, economic indicators are influenced by both the private and public sector so this could represent a disconnect between public company health and private company health.

Conclusion

In this commentary, we sought to answer the question, "can we improve trend-following by drawing information from a country's economy". It intuitively makes sense that an investor would generally opt for remaining in the market unless there are systemic issues that may lead to market distress. A strategy that successfully differentiates between market choppiness and periods of potential recession would drastically mitigate any losses incurred from whipsaw, thereby capturing a majority of the equity premium as well as the trend premium.

We find that growth-trend timing has been relatively successful in countries such as the United States, Germany, and Japan. However, the country that is being analyzed should be considered in light of their specific circumstances.

Peeking under the hood of Italy, it becomes clear that market movements may be influenced by more than a country's implicit economic health. In such a case, we should pause and ask ourselves whether a macroeconomic indicator is truly reflective of that country's economy or if there are other market forces pulling the strings.

[62] Source: MSCI.

THE LIMIT OF FACTOR TIMING

November 11, 2019

SUMMARY

- We have shown previously that it is possible to time factors using value and momentum but that the benefit is not large.

- By constructing a simple model for factor timing, we examine what accuracy would be required to do better than a momentum-based timing strategy.

- While the accuracy required is not high, finding the system that achieves that accuracy may be difficult.

- For investors focused on managing the risks of underperformance – both in magnitude and frequency – a diversified factor portfolio may be the best choice.

- Investors seeking outperformance will have to bear more concentration risk and may be open to more model risk as they forego the diversification among factors.

A few years ago, we began researching factor timing – moving among value, momentum, low volatility, quality, size etc. – with the hope of earning returns in excess not only of the equity market, but also of buy-and-hold factor strategies.

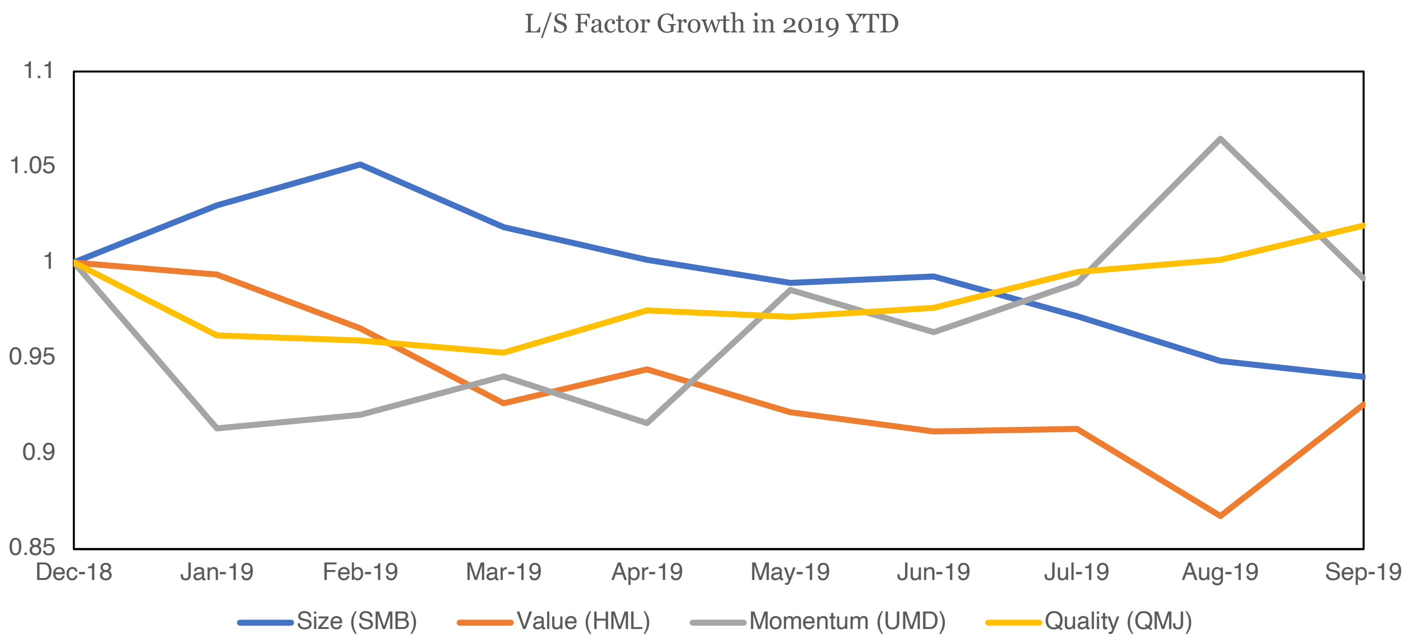
To time the factors, our natural first course of action was to exploit the behavioral biases that may create the factors themselves. We examined value and momentum across the factors and used these metrics to allocate to factors that we expected to outperform in the future.

The results were positive. However, taking into account transaction costs led to the conclusion that investors were likely better off simply holding a diversified factor portfolio.

We then looked at ways to time the factors using the business cycle.

The results in this case were even less convincing and were a bit too similar to a data-mined optimal solution to instill much faith going forward.

But this evidence does not necessarily remove the temptation to take a stab at timing the factors, especially since explicit transactions costs have been slashed for many investors accessing long-only factors through ETFs.



Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

After all, there is a lot to gain by choosing the right factors. For example, in the first 9 months of 2019, the spread between the best (Quality) and worst (Value) performing factors was nearly 1,000 basis points ("bps"). One month prior, that spread had been *double*!

In this research note, we will move away from devising a systematic approach to timing the factors (as AQR asserts, this is deceptively difficult) and instead focus on what a given method would have to overcome to achieve consistent outperformance.

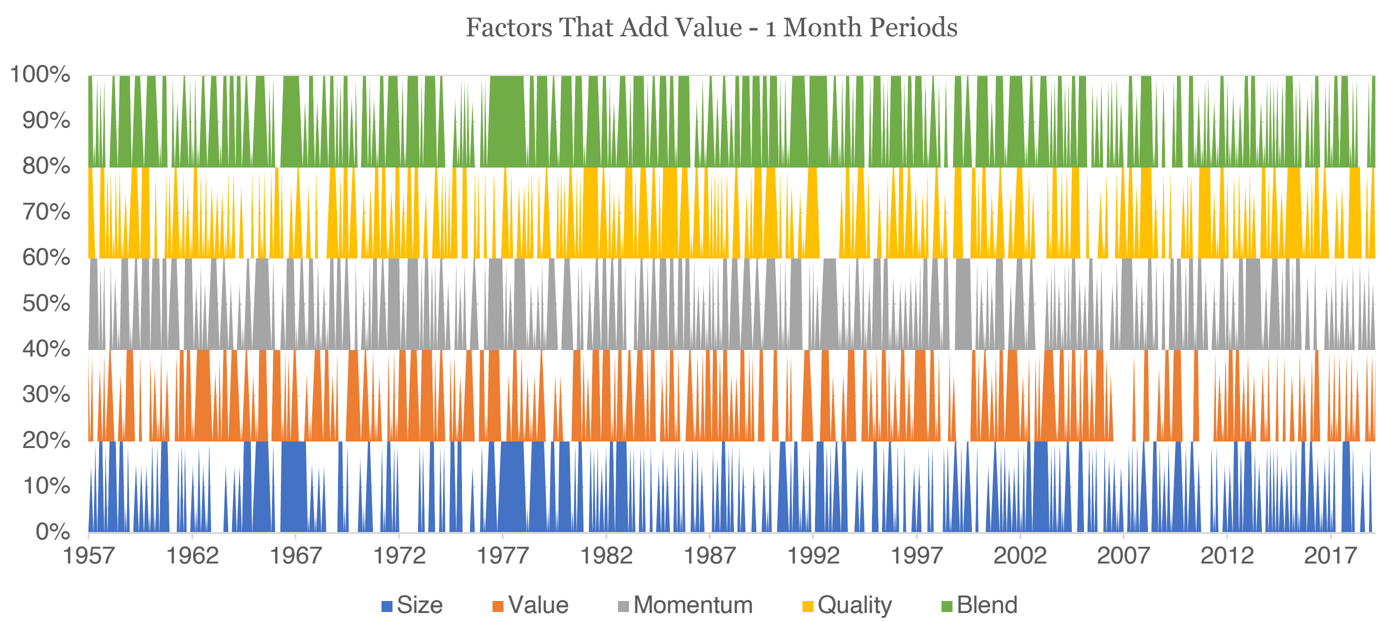
Benchmarking Factor Timing

With all equity factor strategies, the goal is usually to outperform the market-cap weighted equity benchmark.

Since all factor portfolios can be thought of as a market cap weighted benchmark plus a long/short component that captures the isolated factor performance, we can focus our study solely on the long/short portfolio.

Using the common definitions of the factors (from Kenneth French and AQR), we can look at periods over which these self-financing factor portfolios generate positive returns to see if overlaying them on a market-cap benchmark would have added value over different lengths of time.[63]

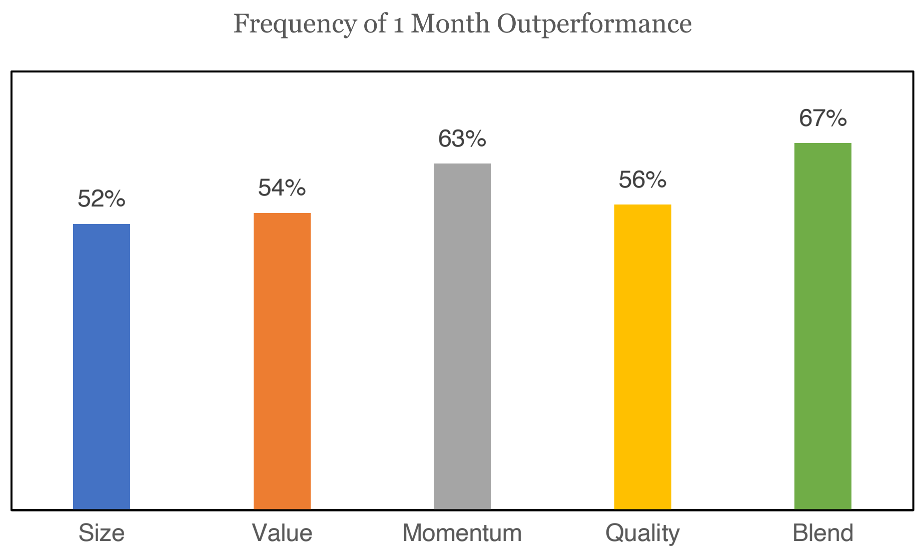
We will also include the performance of an equally weighted basket of the four factors ("Blend").

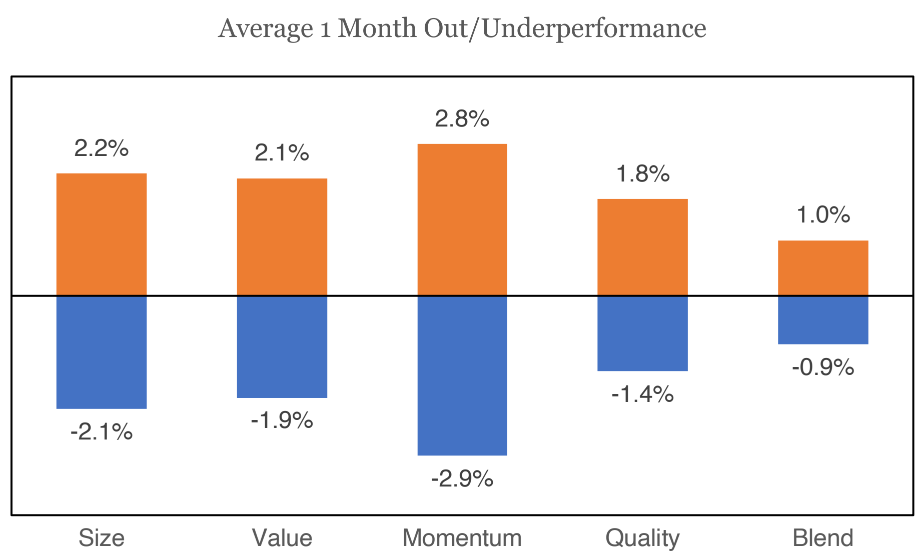


Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. Data from July 1957 – September 2019.

The persistence of factor outperformance over one-month periods is transient. If the goal is to outperform the most often, then the blended portfolio satisfies this requirement, and any timing strategy would have to be accurate enough to overcome this already existing spread.

[63] We should note here that the long/short portfolios implicitly derived from long-only factor portfolios are *not* the same as the academic long/short portfolios we are using here. In fact, they can be substantially different at times. Nevertheless, we do not believe this fact meaningfully alters the results of our study.





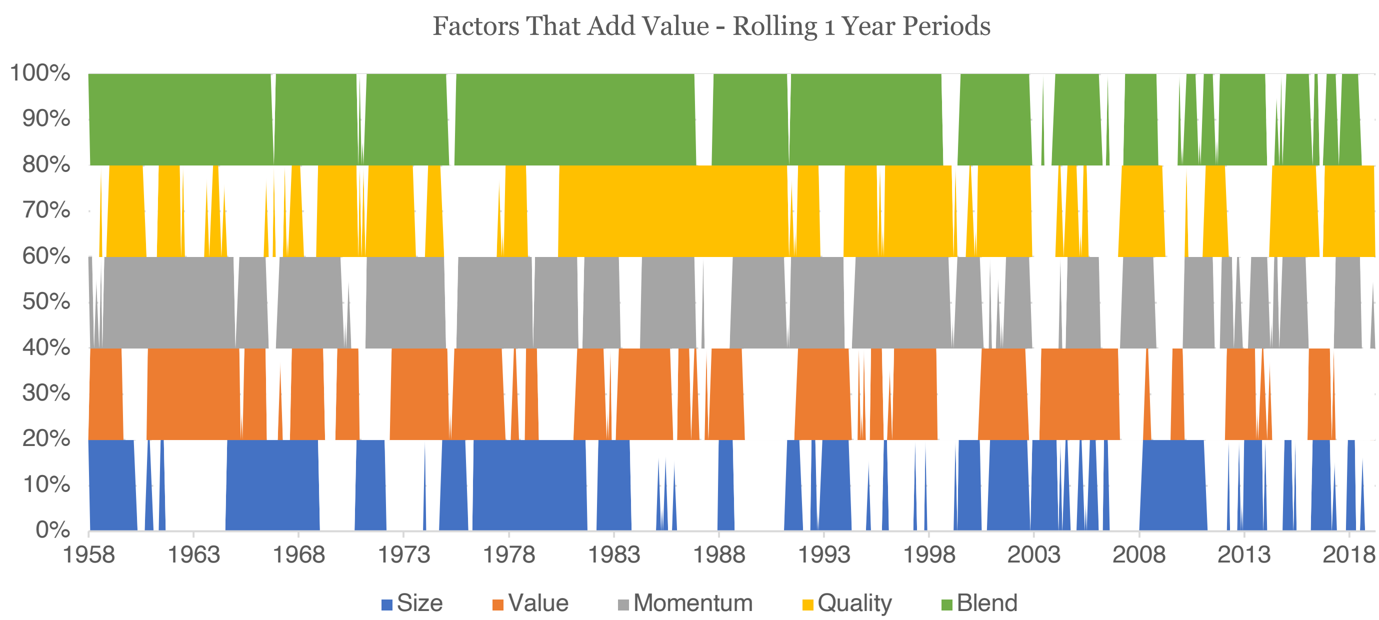
Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. Data from July 1957 – September 2019.

The results for the blended portfolio are so much better than the stand-alone factors because the factors have correlations much lower than many other asset classes, allowing even naïve diversification to add tremendous value.

The blended portfolio also cuts downside risk in terms of returns. If the timing strategy is wrong, and chooses, for example, momentum in an underperforming month, then it could take longer for the strategy to climb back to even. But investors are used to short periods of underperformance and often (we hope) realize that some short-term pain is necessary for long-term gains.

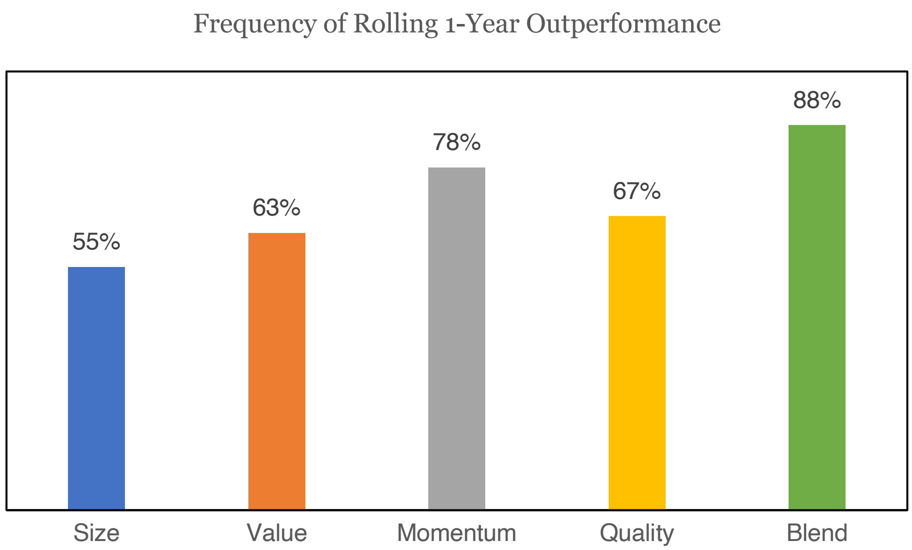
Looking at the same analysis over rolling 1-year periods, we do see some longer periods of factor outperformance. Some examples are quality in the 1980s, value in the mid-2000s, momentum in the 1960s and 1990s, and size in the late-1970s.

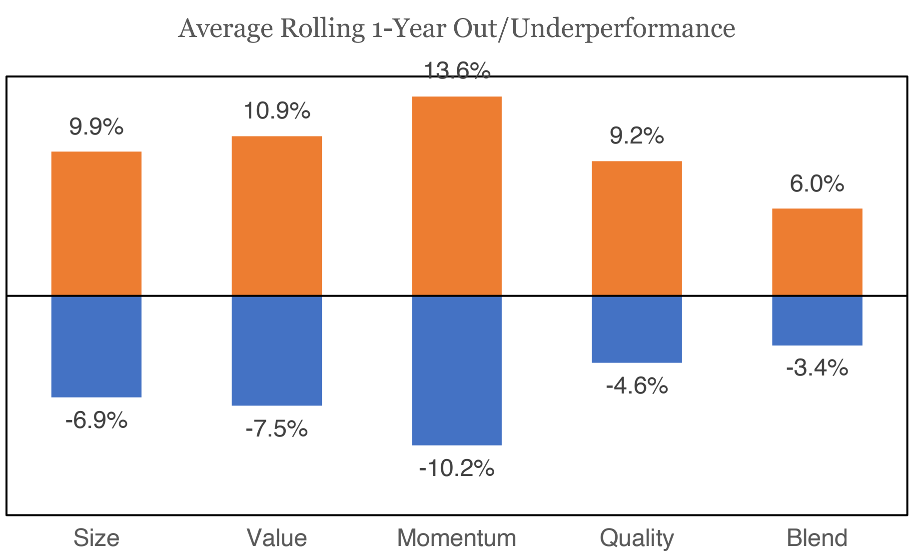
However, there are also decent stretches where the factors underperform. For example, the recent decade for value, quality in the early 2010s, momentum sporadically in the 2000s, and size in the 1980s and 1990s. If the timing strategy gets stuck in these periods, then there can be a risk of abandoning it.



Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. Data from July 1957 – September 2019.

Again, a blended portfolio would have addressed many of these underperforming periods, giving up some of the upside with the benefit of reducing the risk of choosing the wrong factor in periods of underperformance.

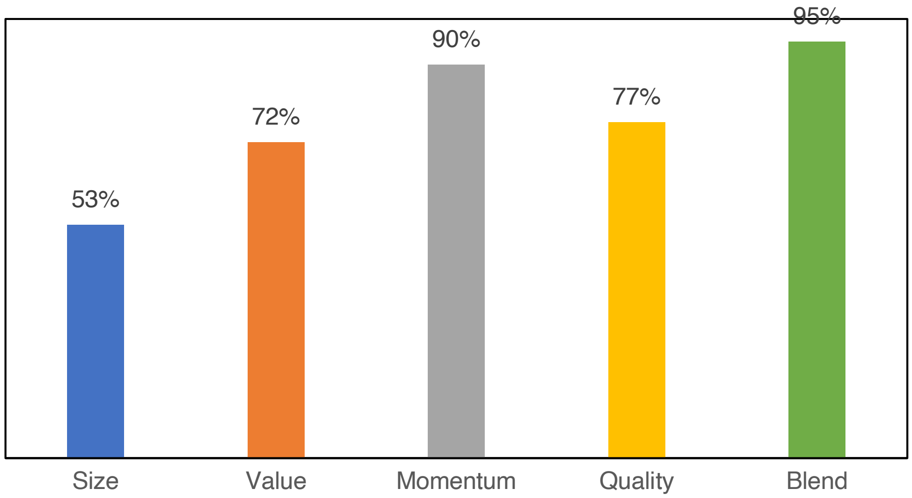




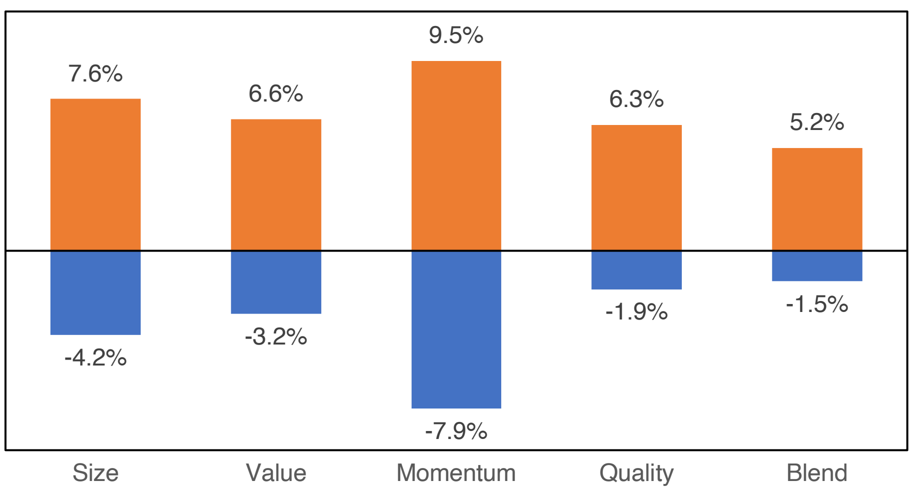
Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. Data from July 1957 – September 2019.

And finally, if we extend our holding period to three years, which may be used for a slower moving signal based on either value or the business cycle, we see that the diversified portfolio still exhibits outperformance over the most rolling periods and has a strong ratio of upside to downside.

Frequency of Rolling 3-Year Outperformance



Average Rolling 3-Year Out/Underperformance



Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. Data from July 1957 – September 2019.

The diversified portfolio stands up to scrutiny against the individual factors but could a generalized model that can time the factors with a certain degree of accuracy lead to better outcomes?

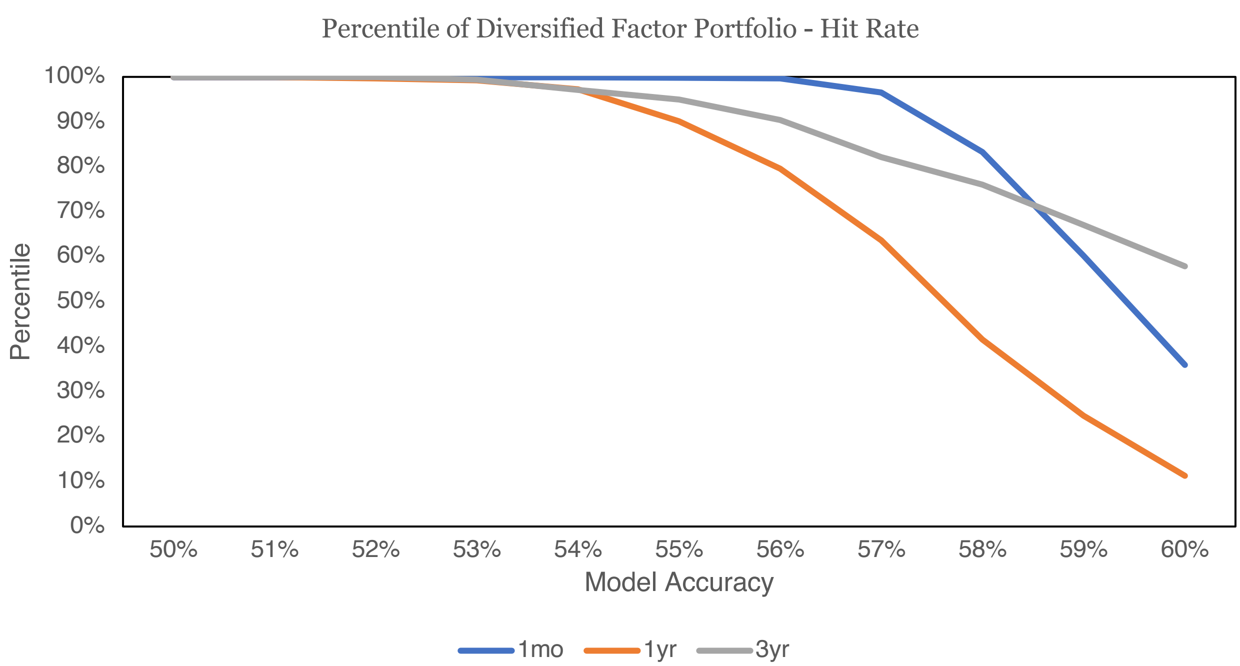
Generic Factor Timing

To construct a generic factor timing model, we will consider a strategy that decides to hold each factor or not with a certain degree of accuracy.

For example, if the accuracy is 50%, then the strategy would essentially flip a coin for each factor. Heads and that factor is included in the portfolio; tails and it is left out. If the accuracy is 55%, then the strategy will hold the factor with a 55% probability when the factor return is positive and not hold the factor with the same probability when the factor return is negative. Just to be clear, this strategy is constructed with look-ahead bias as a tool for evaluation.

All factors included in the portfolio are equally weighted, and if no factors are included, then the returns is zero for that period.

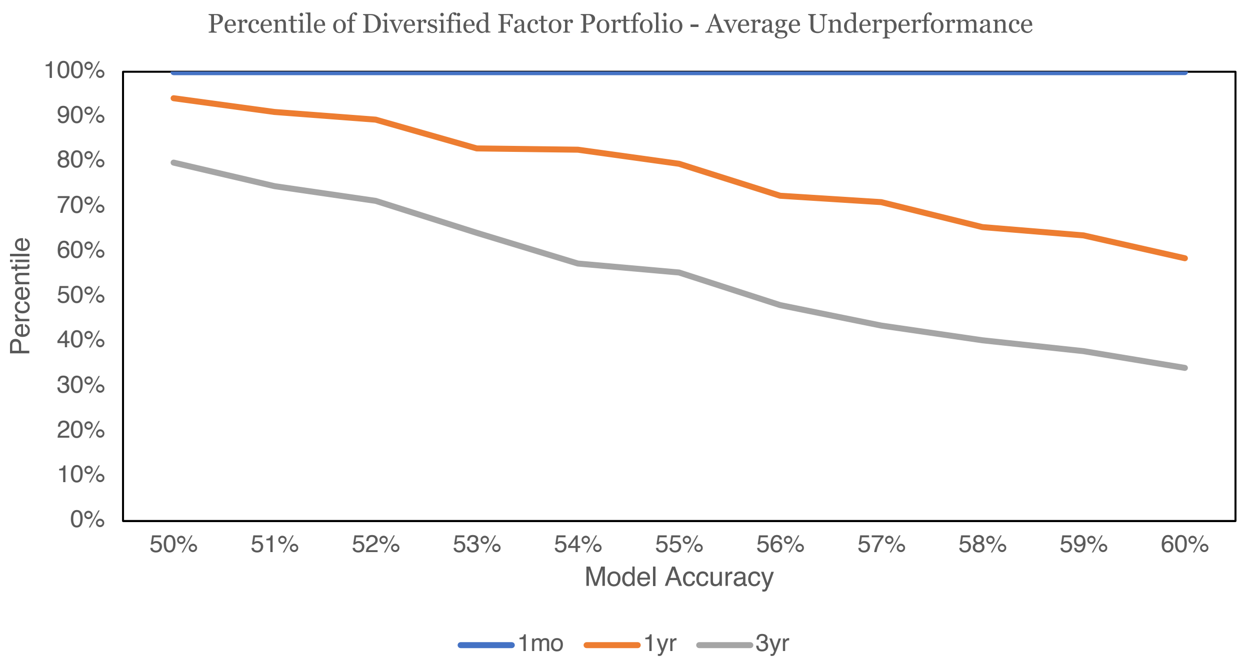
This toy model will allow us to construct distributions to see where the blended portfolio of all the factors falls in terms of frequency of outperformance (hit rate), average outperformance, and average underperformance. The following charts show the percentiles of the diversified portfolio for the different metrics and model accuracies using 1,000 simulations.[64]



[64] We typically see accuracies in the range of 50%-60% for many different strategies that take frequent bets with the hope of realizing a premium over a longer time horizon. When there can be large positive tail events, the accuracy would not need to be as high.

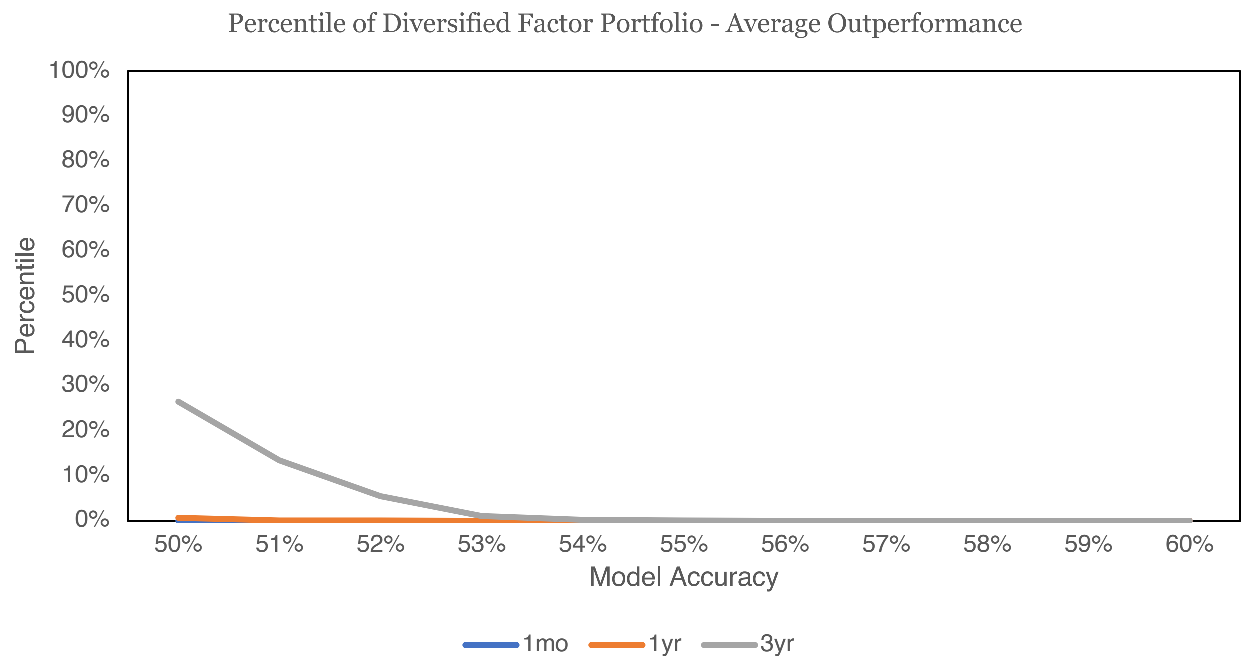
Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. Data from July 1957 – September 2019.

In terms of hit rate, the diversified portfolio behaves in the top tier of the models over all time periods for accuracies up to about 57%. Even with a model that is 60% accurate, the diversified portfolio was still above the median.



Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. Data from July 1957 – September 2019.

For average underperformance, the diversified portfolio also did very well in the context of these factor timing models. The low correlation between the factors leads to opportunities for the blended portfolio to limit the downside of individual factors.



Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. Data from July 1957 – September 2019.

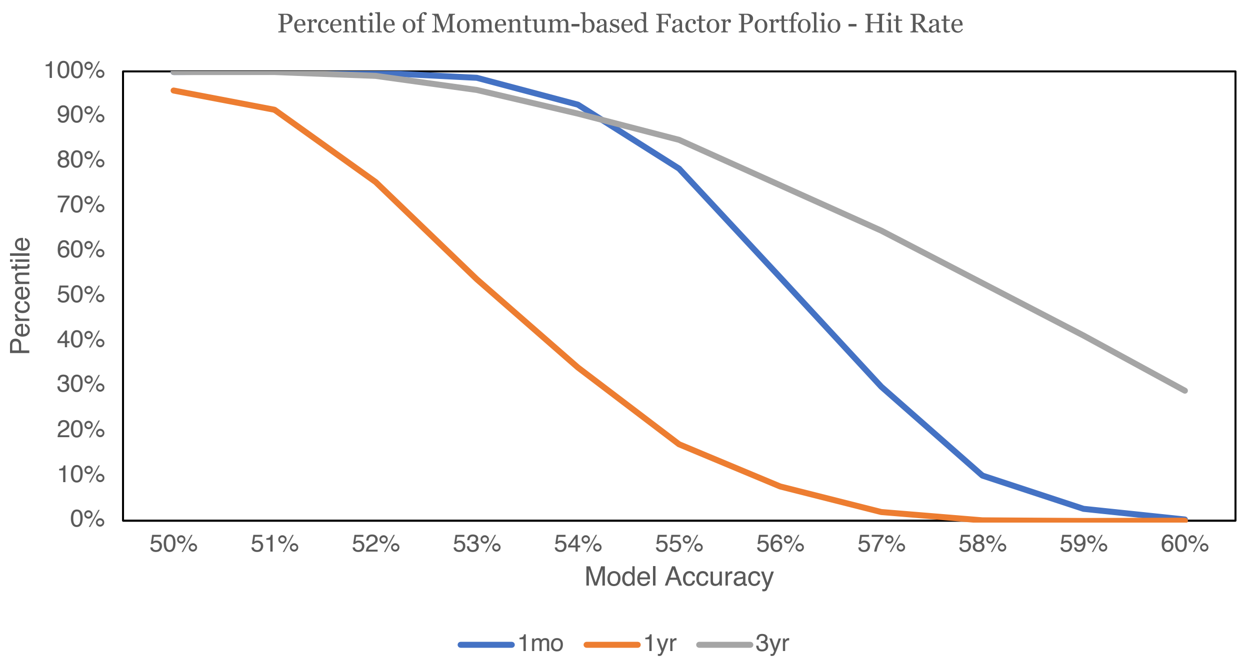
For average outperformance, the diversified portfolio did much worse than the timing model over all time horizons. We can attribute this also to the low correlation between the factors, as choosing only a subset of factors and equally weighting them often leads to more extreme returns.

Overall, the diversified portfolio manages the risks of underperformance, both in magnitude and in frequency, at the expense of sacrificing outperformance potential. We saw this in the first section when we compared the diversified portfolio to the individual factors.

But if we want to have increased return potential, we will have to introduce some model risk to time the factors.

Checking in on Momentum

Momentum is one model-based way to time the factors. Under our definition of accuracy in the toy model, a 12-1 momentum strategy on the factors has an accuracy of about 56%. While the diversified portfolio exhibited some metrics in line with strategies that were even more accurate than this, it never bore concentration risk: it always held all four factors.

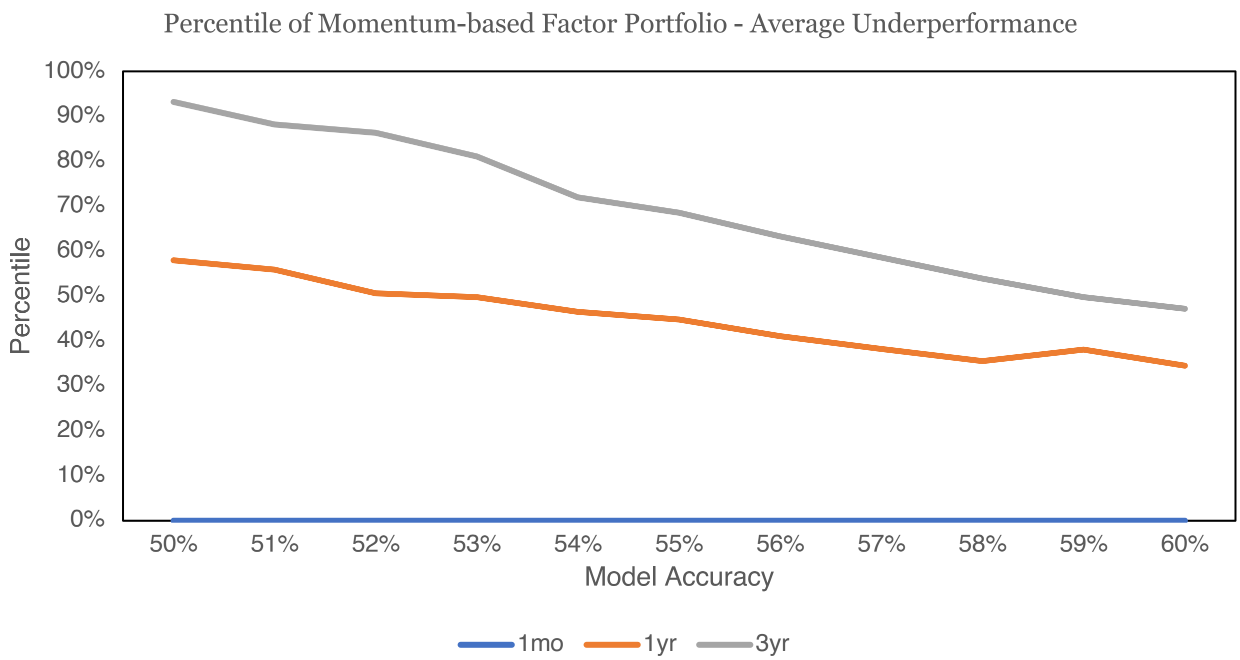


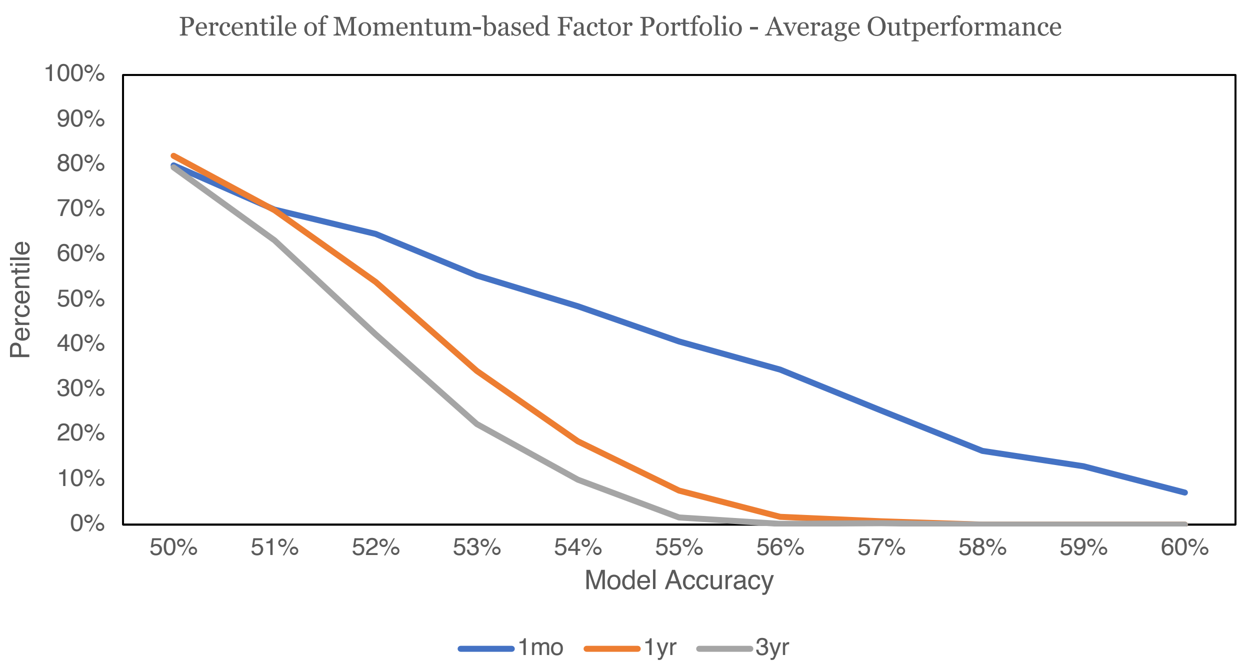
Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. Data from July 1957 – September 2019.

For the hit rate percentiles of the momentum strategy, we see a more subdued response. Momentum does not win as much as the diversified portfolio over the different time periods.

But not winning as much can be fine if you win bigger when you do win.

The charts below show that momentum does indeed have a higher outperformance percentile but with a worse underperformance percentile, especially for 1-month periods, likely due to mean reversionary whipsaw.





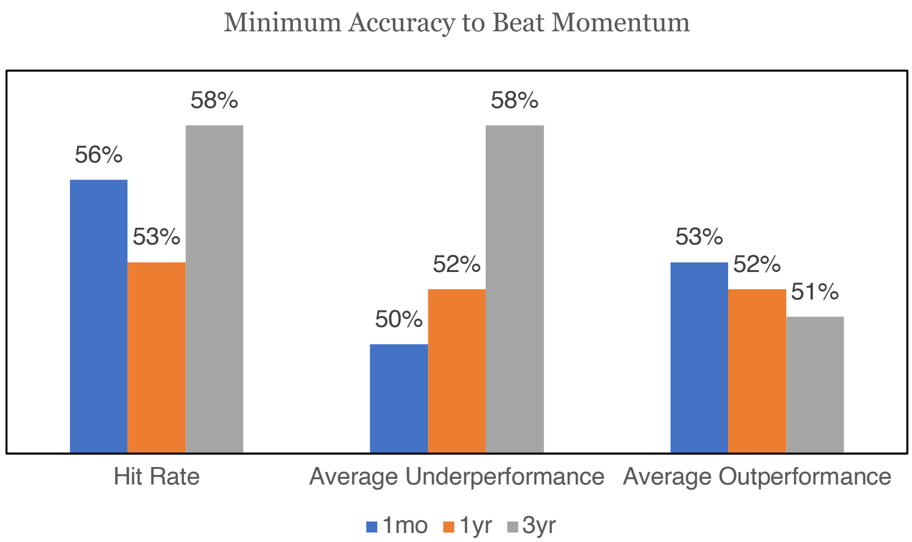
Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions. Data from July 1957 – September 2019.

While momentum is definitely not the only way to time the factors, it is a good baseline to see what is required for higher average outperformance.

Now, turning back to our generic factor timing model, what accuracy would you need to beat momentum?

Sharpening our Signal

The answer is: not a whole lot. Most of the time, we only need to be about 53% accurate to beat the momentum-based factor timing.



Source: Kenneth French Data Library, AQR. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

The caveat is that this is the median performance of the simulations. The accuracy figure climbs closer to 60% if we use the 25th percentile as our target.

While these may not seem like extremely high requirements for running a successful factor timing strategy, it is important to observe that not many investors are doing this. True accuracy may be hard to discover, and sticking with the system may be even harder when the true accuracy can never be known.

Conclusion

If you made it this far looking for some rosy news on factor timing or the Holy Grail of how to do it skillfully, you may be disappointed.

However, for most investors looking to generate some modest benefits relative to market-cap equity, there is good news. Any signal for timing factors does not have to be highly accurate to perform well, and in the absence of a signal for timing, a diversified portfolio of the factors can lead to successful results by the metrics of average underperformance and frequency of underperformance.

For those investors looking for higher outperformance, concentration risk will be necessary.

Any timing strategy on low correlation investments will generally forego significant diversification in the pursuit of higher returns.

While this may be the goal when constructing the strategy, we should always pause and determine whether the potential benefits outweigh the costs. Transaction costs may be lower now. However, there are still operational burdens and the potential stress caused by underperformance when a system is not automated or when results are tracked too frequently.

Factor timing may be possible, but timing and tactical rotation may be better suited to scenarios where some of the model risk can be mitigated.

The Dumb (Timing) Luck of Smart Beta

November 18, 2019

Summary

- In past research notes we have explored the impact of rebalance timing luck on strategic and tactical portfolios, even using our own Systematic Value methodology as a case study.
- In this note, we generate empirical timing luck estimates for a variety of specifications for simplified value, momentum, low volatility, and quality style portfolios.
- Relative results align nicely with intuition: higher concentration and less frequent rebalancing leads to increasing levels of realized timing luck.
- For more reasonable specifications – e.g. 100 stock portfolios rebalanced semi-annually – timing luck ranges between 100 and 400 basis points depending upon the style under investigation, suggesting a significant risk of performance dispersion due only to when a portfolio is rebalanced and nothing else.
- The large magnitude of timing luck suggests that any conclusions drawn from performance comparisons between smart beta ETFs or against a standard style index may be spurious.

We've written about the concept of rebalance timing luck *a lot*. It's a cowbell we've been beating for over half a decade, with our first article going back to August 7th, 2013.

As a reminder, rebalance timing luck is the performance dispersion that arises from the choice of a particular rebalance date (e.g. semi-annual rebalances that occur in June and December versus March and September).

We've empirically explored the impact of rebalance timing luck as it relates to strategic asset allocation, tactical asset allocation, and even used our own Systematic Value strategy as a case study for smart beta. All of our results suggest that it has a highly non-trivial impact upon performance.

This summer we published a paper in the Journal of Index Investing that proposed a simple solution to the timing luck problem: diversification. If, for example, we believe that our momentum portfolio should be rebalanced every quarter – perhaps as an optimal balance of cost and signal freshness – then we proposed splitting our capital across the three portfolios that spanned different three-month rebalance periods (e.g. JAN-APR-JUL-OCT, FEB-MAY-AUG-NOV, MAR-JUN-SEP-DEC). This solution is referred to either as "tranching" or "overlapping portfolios."

The paper also derived a formula for estimating timing luck ex-ante, with a simplified representation of:

$$L = \left(\frac{T}{2F}\right)S$$

Where L is the timing luck measure, T is turnover rate of the strategy, F is how many times per year the strategy rebalances, and S is the volatility of a long/short portfolio that captures the difference of what a strategy is currently invested in versus what it *could* be invested in if the portfolio was reconstructed at that point in time.

Without numbers, this equation still informs some general conclusions:

- Higher turnover strategies have higher timing luck.
- Strategies that rebalance more frequently have lower timing luck.
- Strategies with a less constrained universe will have higher timing luck.

Bullet points 1 and 3 may seem similar but capture subtly different effects. This is likely best illustrated with two examples on different extremes. First consider a very high turnover strategy that trades within a universe of highly correlated securities. Now consider a very low turnover strategy that is either 100% long or 100% short U.S. equities. In the first case, the highly correlated nature of the universe means that differences in specific holdings may not matter as much, whereas in the second case the perfect inverse correlation means that small portfolio differences lead to meaningfully different performance.

L, in and of itself, is a bit tricky to interpret, but effectively attempts to capture the potential dispersion in performance between a particular rebalance implementation choice (e.g. JAN-APR-JUL-OCT) versus a timing-luck-neutral benchmark.

After half a decade, you'd would think we've spilled enough ink on this subject.

But given that just about every single major index still does not address this issue, and since our passion for the subject clearly verges on fever pitch, here comes some more cowbell.

Equity Style Portfolio Definitions

In this note, we will explore timing luck as it applies to four simplified smart beta portfolios based upon holdings of the S&P 500 from 2000-2019:

- Value: Sort on earnings yield.
- Momentum: Sort on prior 12-1 month returns.
- Low Volatility: Sort on realized 12-month volatility.
- Quality: Sort on average rank-score of ROE, accruals ratio, and leverage ratio.

Quality is a bit more complicated only because the quality factor has far less consistency in accepted definition. Therefore, we adopted the signals utilized by the S&P 500 Quality Index.

For each of these equity styles, we construct portfolios that vary across two dimensions:

- Number of Holdings: 50, 100, 150, 200, 250, 300, 350, and 400.
- Frequency of Rebalance: Quarterly, Semi-Annually, and Annually.

For the different rebalance frequencies, we also generate portfolios that represent each possible rebalance variation of that mix. For example, Momentum portfolios with 50 stocks that rebalance annually have 12 possible variations: a January rebalance, February rebalance, et cetera. Similarly, there are 12 possible variations of Momentum portfolios with 100 stocks that rebalance annually.

By explicitly calculating the rebalance date variations of each Style x Holding x Frequency combination, we can construct an overlapping portfolios solution. To estimate empirical annualized timing luck, we calculate the standard deviation of monthly return dispersion between the different rebalance date variations of the overlapping portfolio solution and annualize the result.

Empirical Timing Luck Results

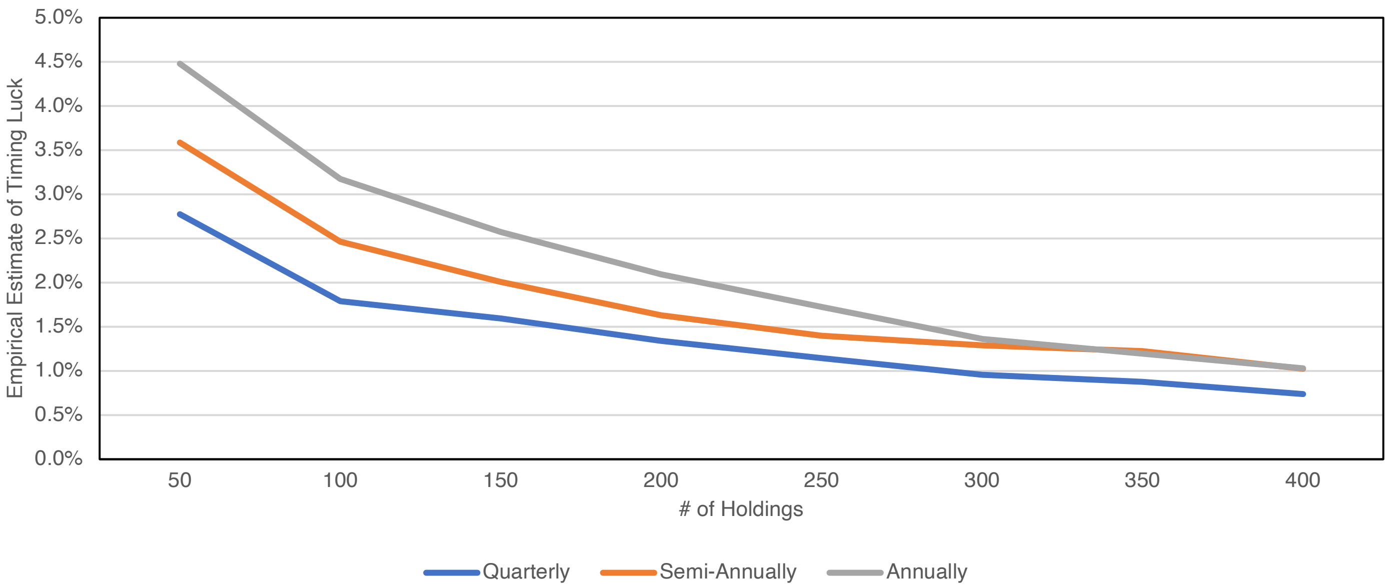
Before looking at the results plotted below, we would encourage readers to hypothesize as to what they expect to see. Perhaps not in absolute magnitude, but at least in relative magnitude.

For example, based upon our understanding of the variables affecting timing luck, would we expect an annually rebalanced portfolio to have more or less timing luck than a quarterly rebalanced one?

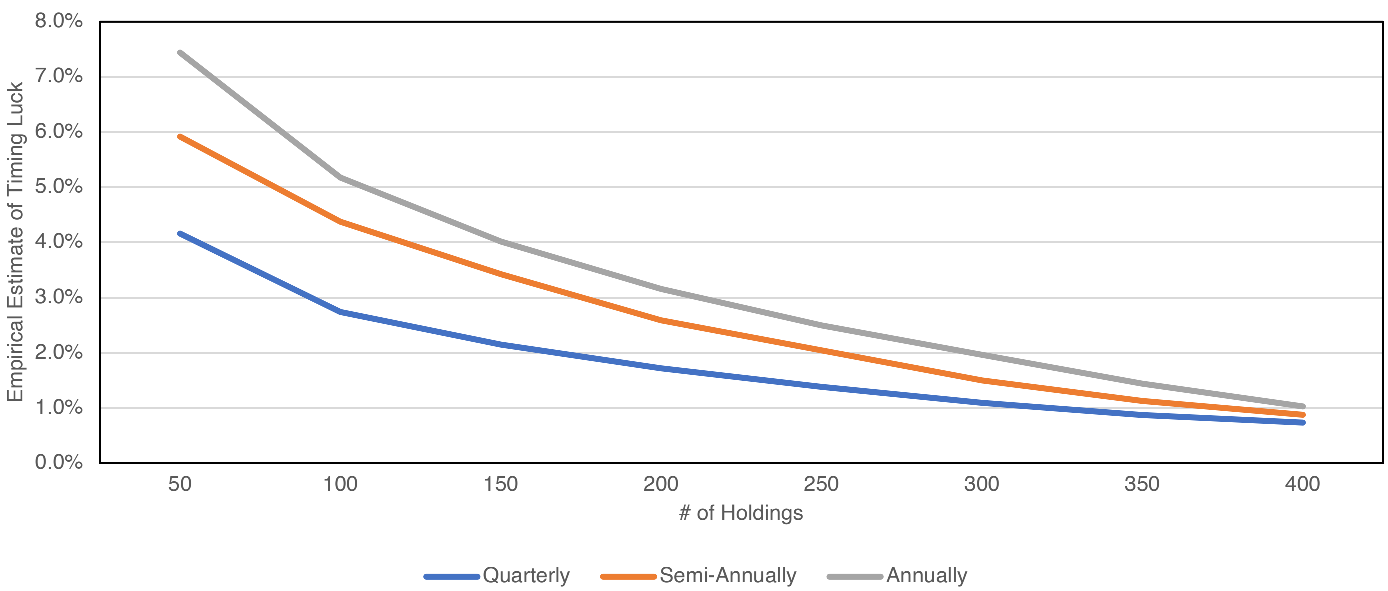
Should a more concentrated portfolio have more or less timing luck than a less concentrated variation?

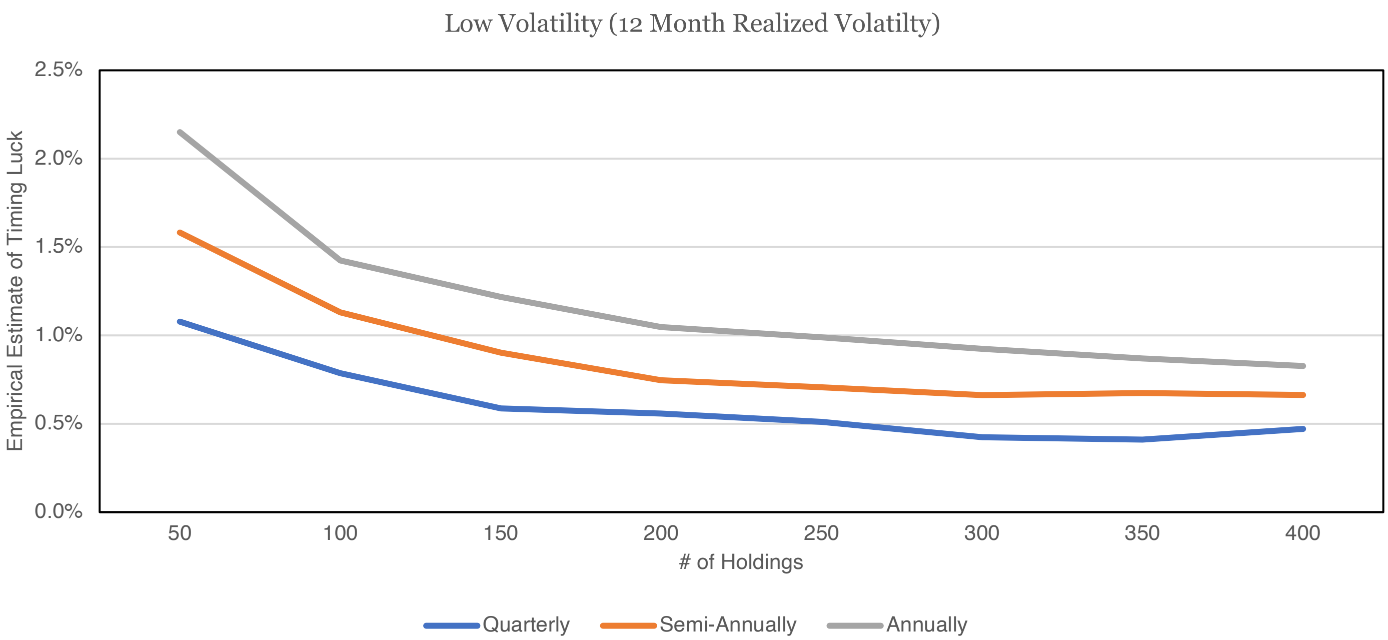
Which factor has the greatest risk of exhibiting timing luck?

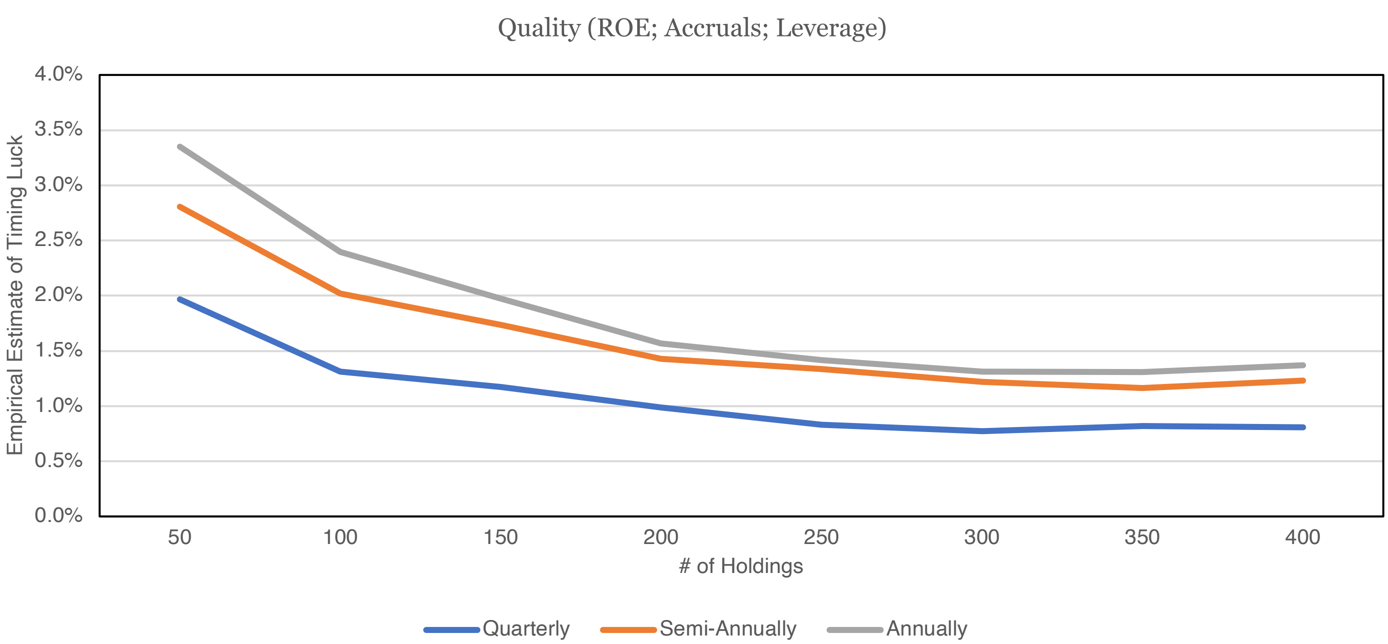
Value (Earnings Yield)



Momentum (12-1 Month Total Return)

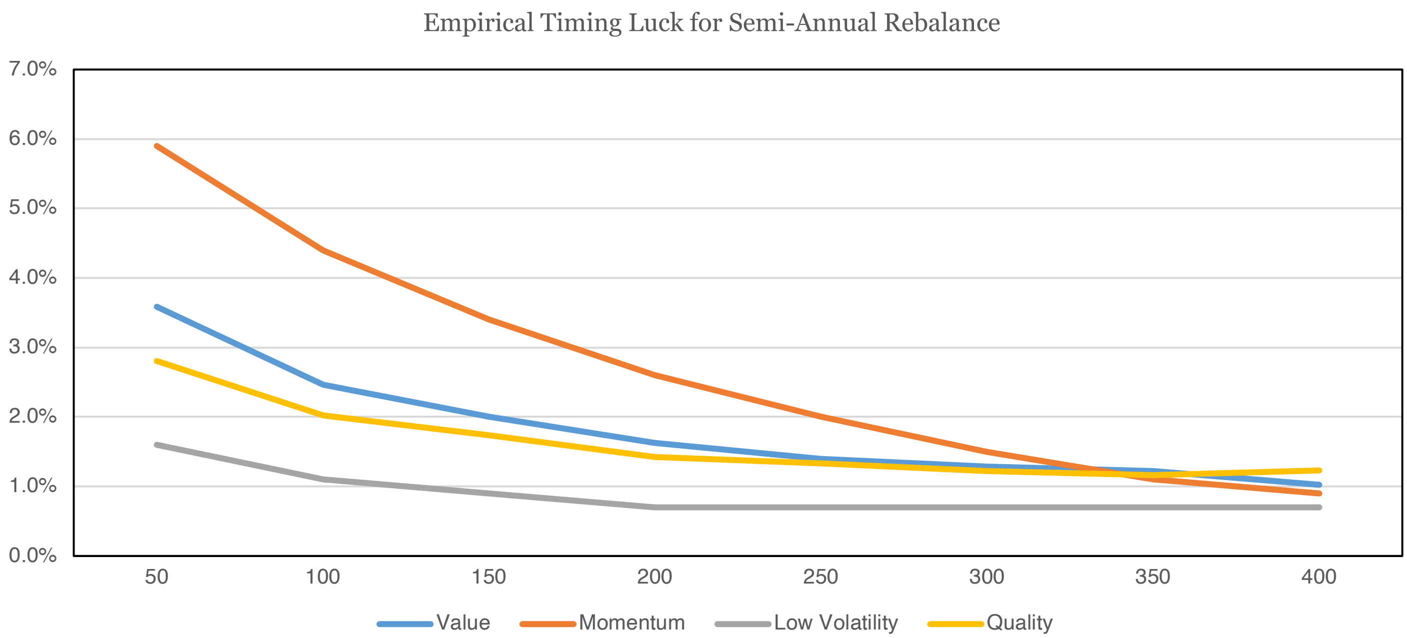






Source: Sharadar. Calculations by Newfound Research.

To create a sense of scale across the styles, below we isolate the results for semi-annual rebalancing for each style and plot it.



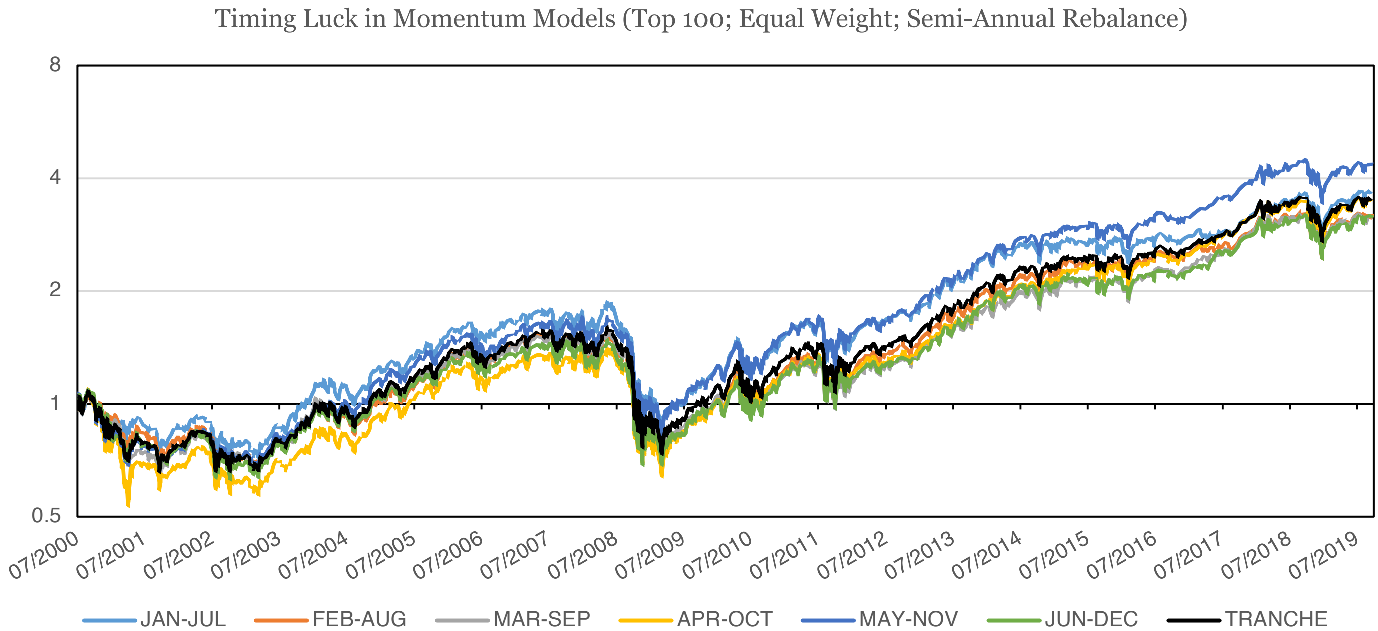
Source: Sharadar. Calculations by Newfound Research.

In relative terms, there is no great surprise in these results:

- More frequent rebalancing limits the risk of portfolios changing significantly between rebalance dates, thereby decreasing the impact of timing luck.
- More concentrated portfolios exhibit larger timing luck.
- Faster-moving signals (e.g. momentum) tend to exhibit more timing luck than more stable, slower-moving signals (e.g. low volatility).

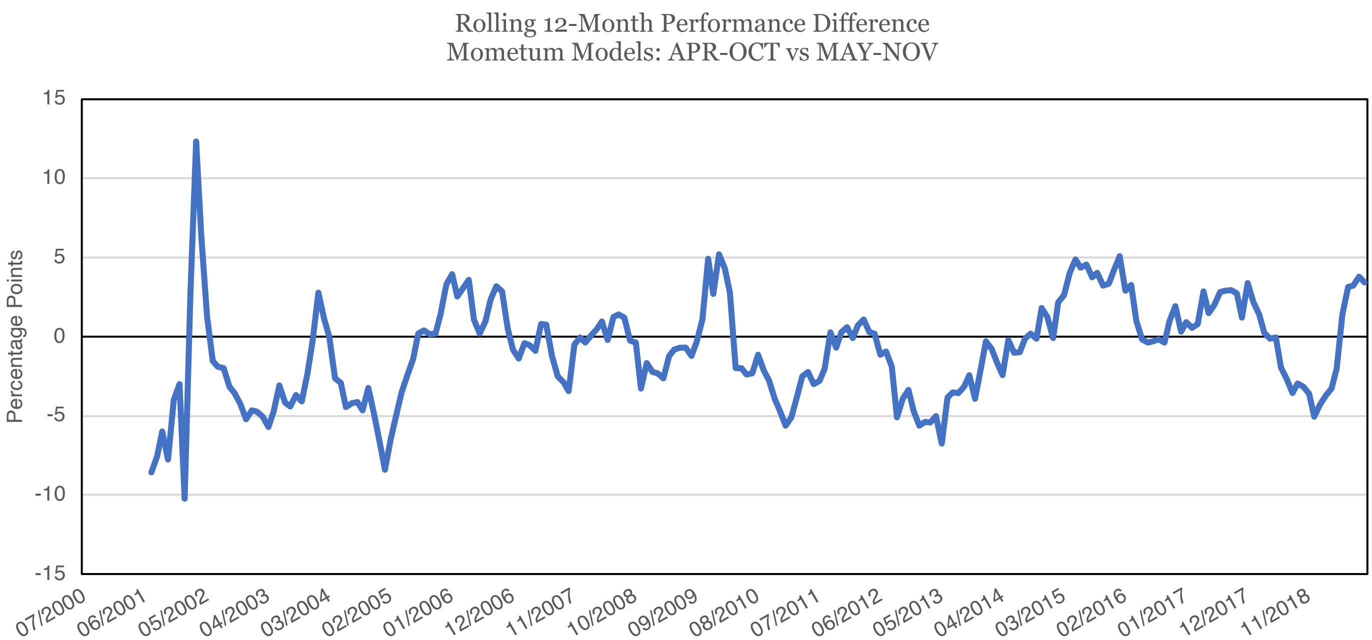
What is perhaps the most surprising is the sheer magnitude of timing luck. Consider that the S&P 500 Enhanced Value, Momentum, Low Volatility, and Quality portfolios all hold 100 securities and are rebalanced semi-annually. Our study suggests that timing luck for such approaches may be as large as 2.5%, 4.4%, 1.1%, and 2.0% respectively.

But what does that really mean? Consider the realized performance dispersion of different rebalance date variations of a Momentum portfolio that holds the top 100 securities in equal weight and is rebalanced on a semi-annual basis.



Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

The 4.4% estimate of annualized timing luck is a measure of dispersion between each underlying variation and the overlapping portfolio solution. If we isolate two sub-portfolios and calculate rolling 12-month performance dispersion, we can see that the difference can be far larger, as one might exhibit positive timing luck while the other exhibits negative timing luck. Below we do precisely this for the APR-OCT and MAY-NOV rebalance variations.



Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

In fact, since these variations are identical in every which way *except* for the date on which they rebalance, a portfolio that is long the APR-OCT variation and short the MAY-NOV variation would explicitly capture the effects of rebalance timing luck. If we assume the rebalance timing luck realized by these two portfolios is independent (which our research suggests it is), then the volatility of this long/short is approximately the rebalance timing luck estimated above scaled by the square-root of two.[65]

Thus, if we are comparing two identically-managed 100-stock momentum portfolios that rebalance semi-annually, our 95% confidence interval for performance dispersion due to timing luck is +/- 12.4% (2 x SQRT(2) x 4.4%).

Even for more diversified, lower turnover portfolios, this remains an issue. Consider a 400-stock low-volatility portfolio that is rebalanced quarterly. Empirical timing luck is still 0.5%, suggesting a 95% confidence interval of 1.4%.

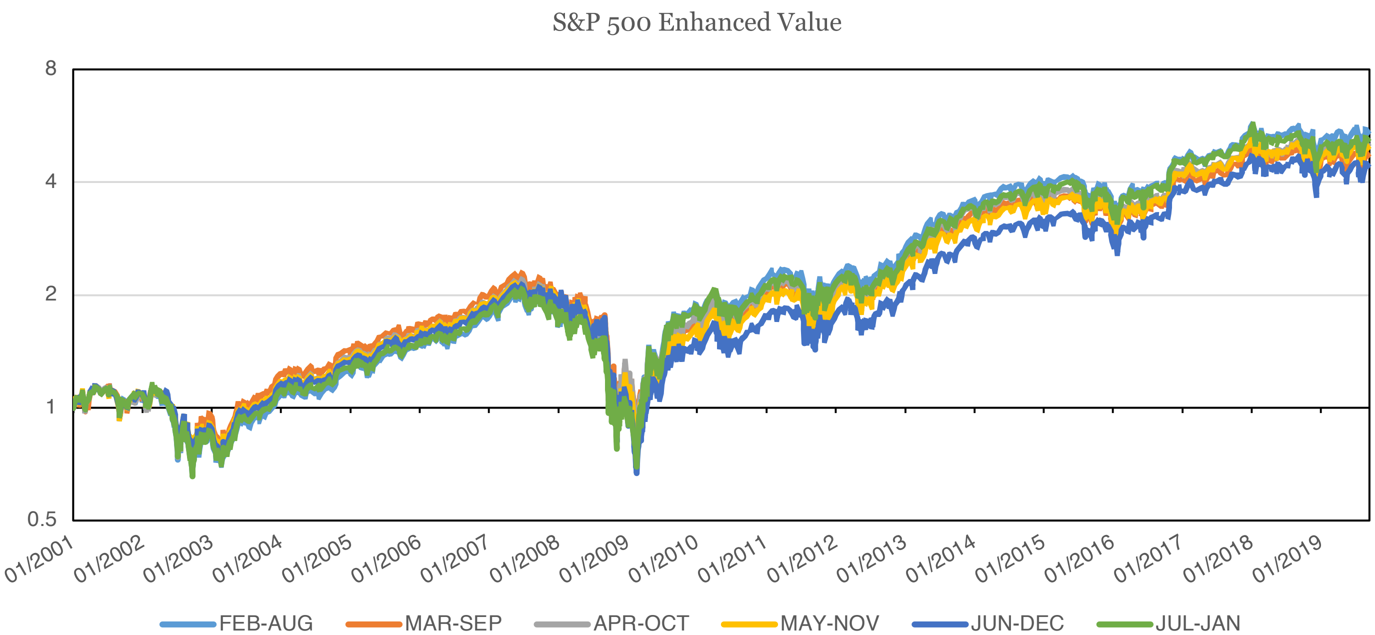
[65] For variations v_i and v_j and overlapping-portfolio solution V, then:

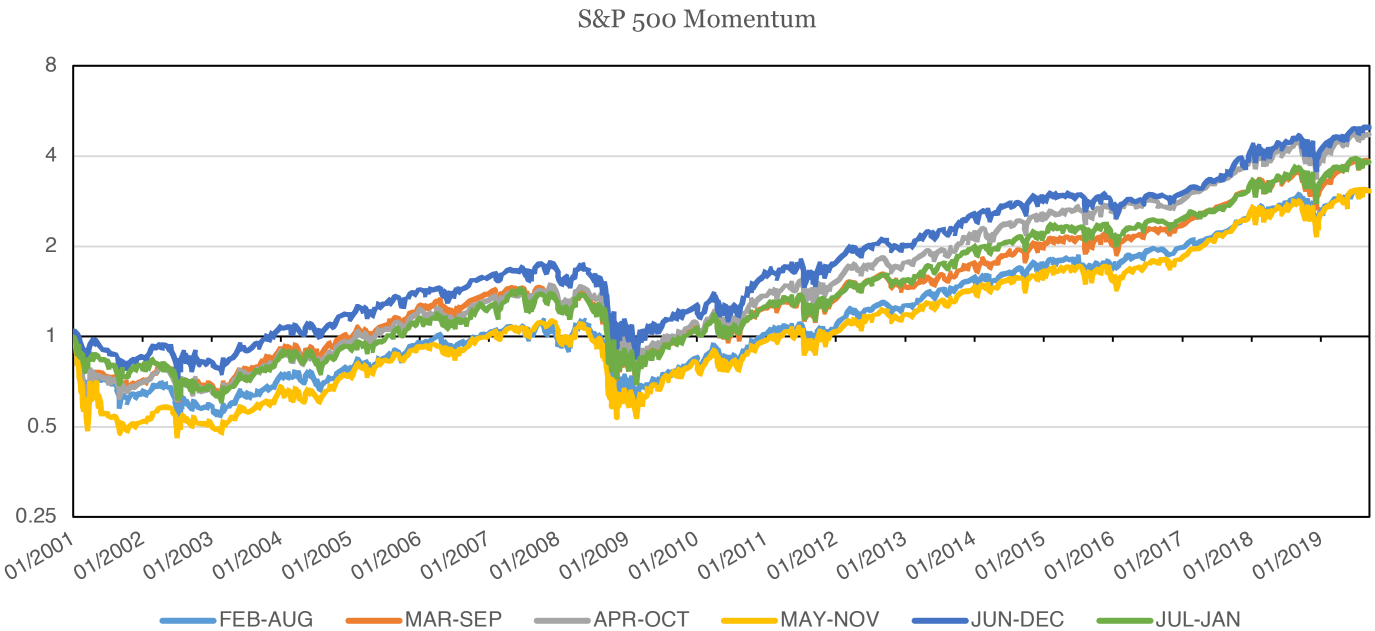
$$\sigma(v_i - v_j) = \sigma\left((v_i - V) - (v_j - V)\right) = \sqrt{\sigma^2(v_i - V) + \sigma^2(v_j - V)} = \sqrt{2}L$$

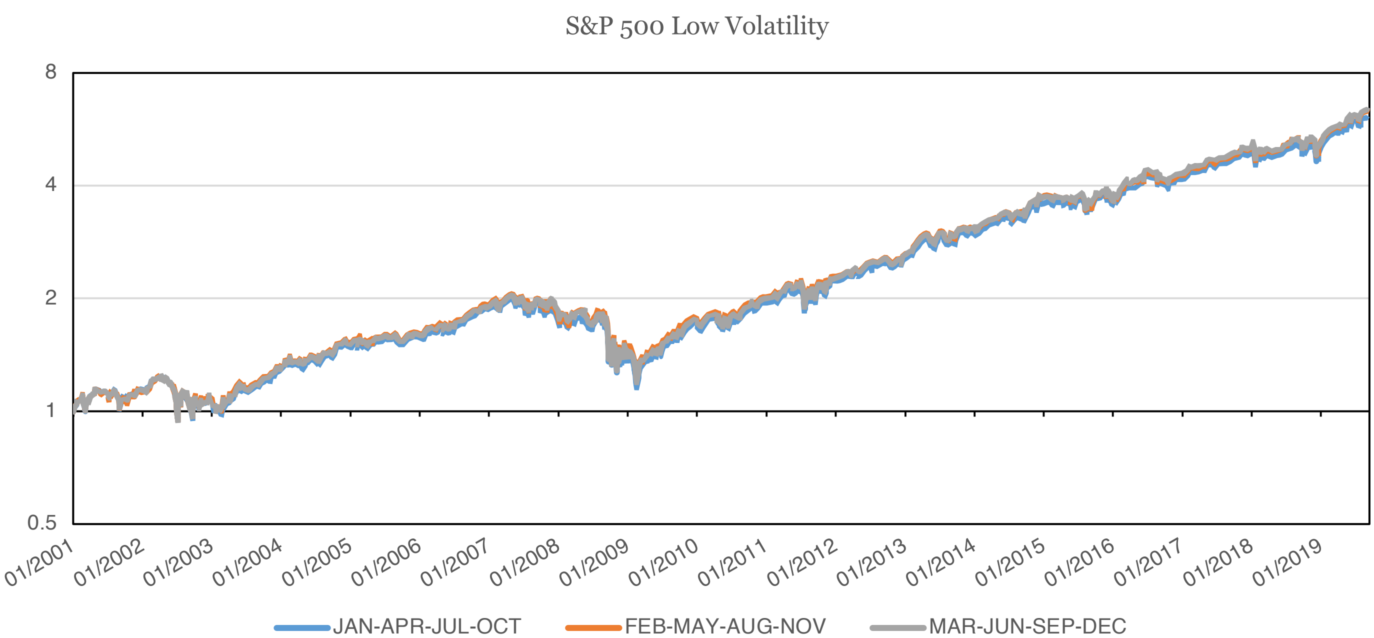
S&P 500 Style Index Examples

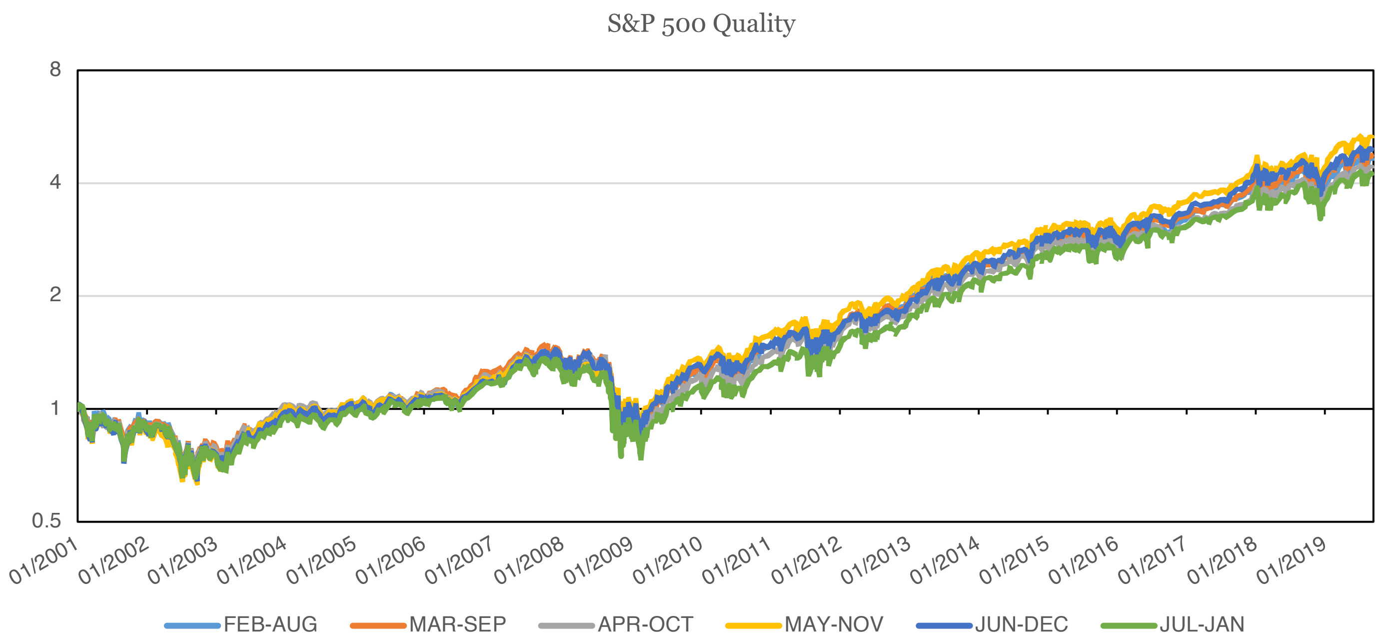
One critique of the above analysis is that it is purely hypothetical: the portfolios studied above aren't really those offered in the market today.

We will take our analysis one step further and replicate (to the best of our ability) the S&P 500 Enhanced Value, Momentum, Low Volatility, and Quality indices. We then created different rebalance schedule variations. Note that the S&P 500 Low Volatility index rebalances quarterly, so there are only three possible rebalance variations to compute.




S&P 500 Momentum
8
4
2
1
0.5
0.25
01/2001
01/2002
01/2003
01/2004
01/2005
01/2006
01/2007
01/2008
01/2009
01/2010
01/2011
01/2012
01/2013
01/2014
01/2015
01/2016
01/2017
01/2018
01/2019
FEB-AUG
MAR-SEP
APR-OCT
MAY-NOV
JUN-DEC
JUL-JAN


S&P 500 Low Volatility
8
4
2
1
0.5
01/2001
01/2002
01/2003
01/2004
01/2005
01/2006
01/2007
01/2008
01/2009
01/2010
01/2011
01/2012
01/2013
01/2014
01/2015
01/2016
01/2017
01/2018
01/2019
JAN-APR-JUL-OCT
FEB-MAY-AUG-NOV
MAR-JUN-SEP-DEC



Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

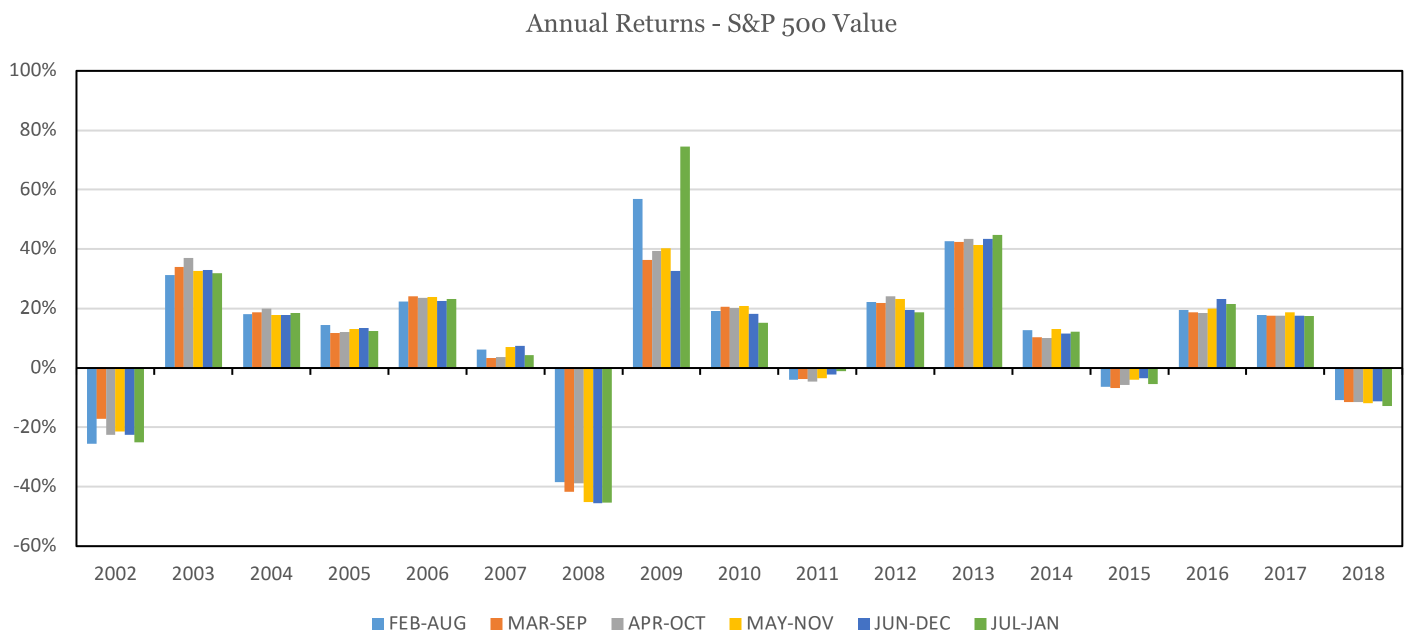
We see a meaningful dispersion in terminal wealth levels, even for the S&P 500 Low Volatility index, which appears at first glance in the graph to have little impact from timing luck.

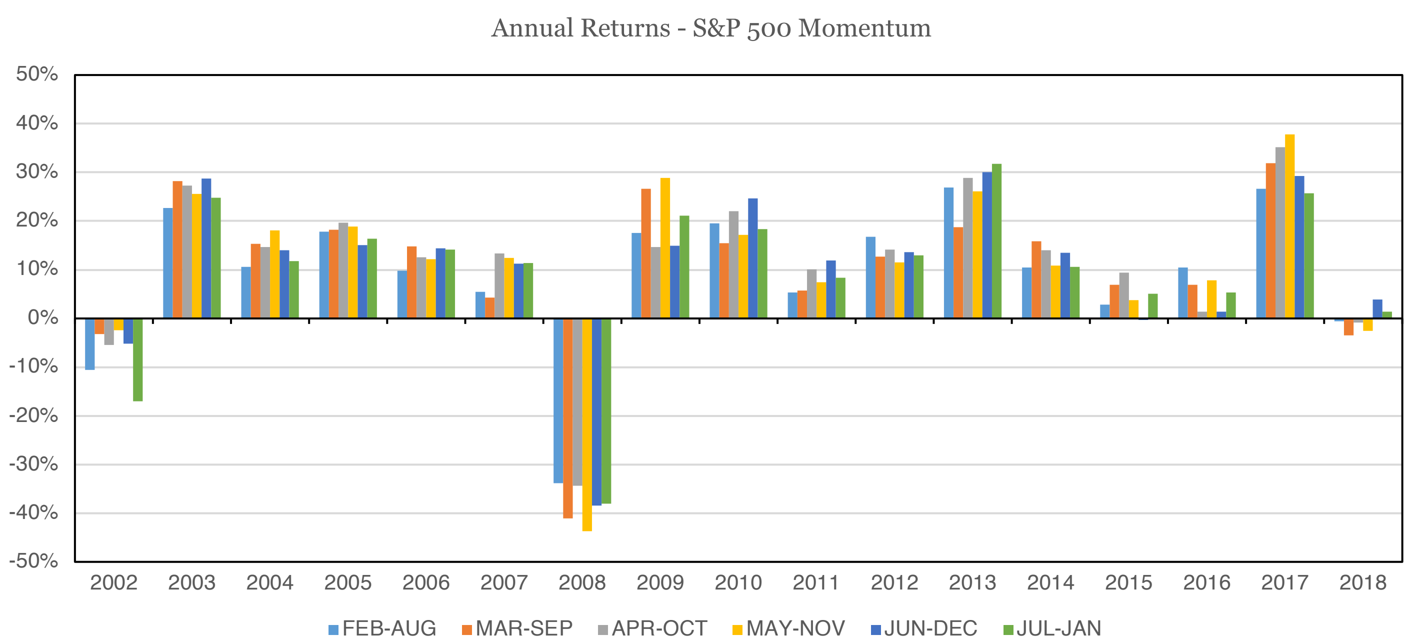
	Minimum Terminal Wealth	Maximum Terminal Wealth
Enhanced Value	$4.45	$5.45
Momentum	$3.07	$4.99
Low Volatility	$6.16	$6.41
Quality	$4.19	$5.25

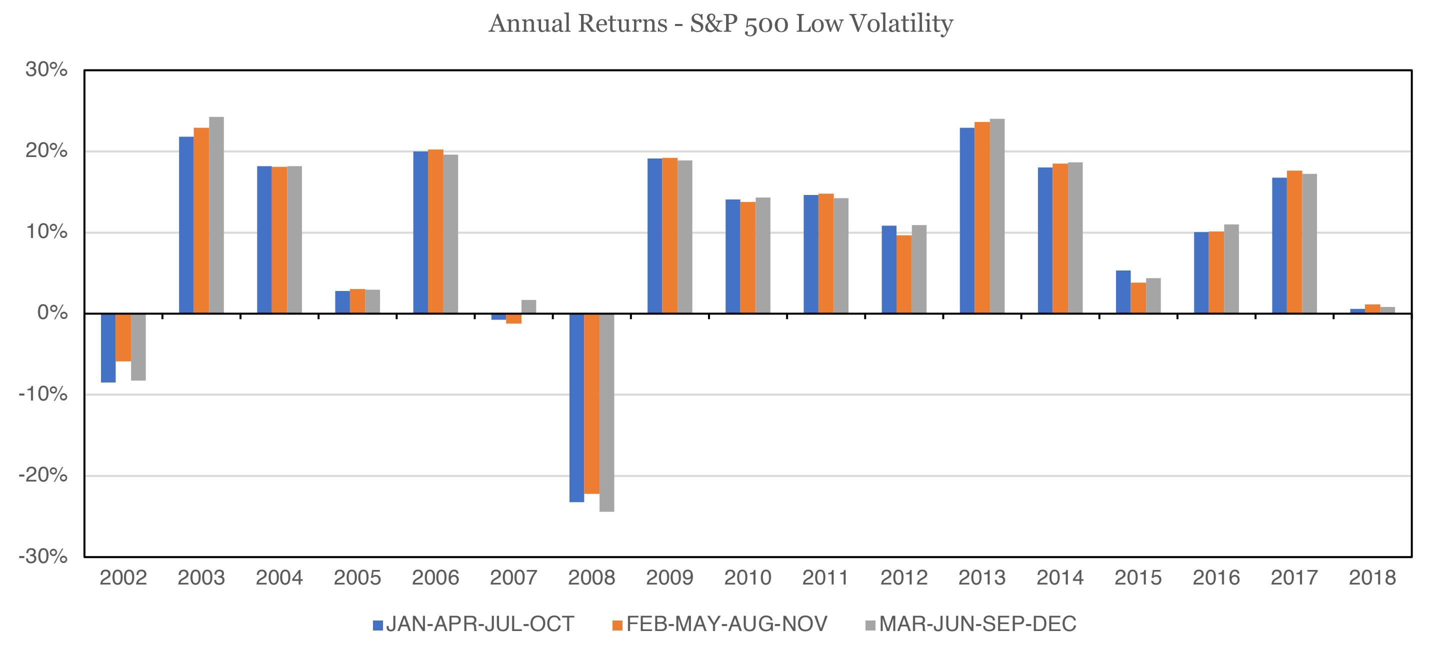
We should further note that there does not appear to be one set of rebalance dates that does significantly better than the others. For Value, FEB-AUG looks best while JUN-DEC looks the worst; for Momentum it's almost precisely the opposite.

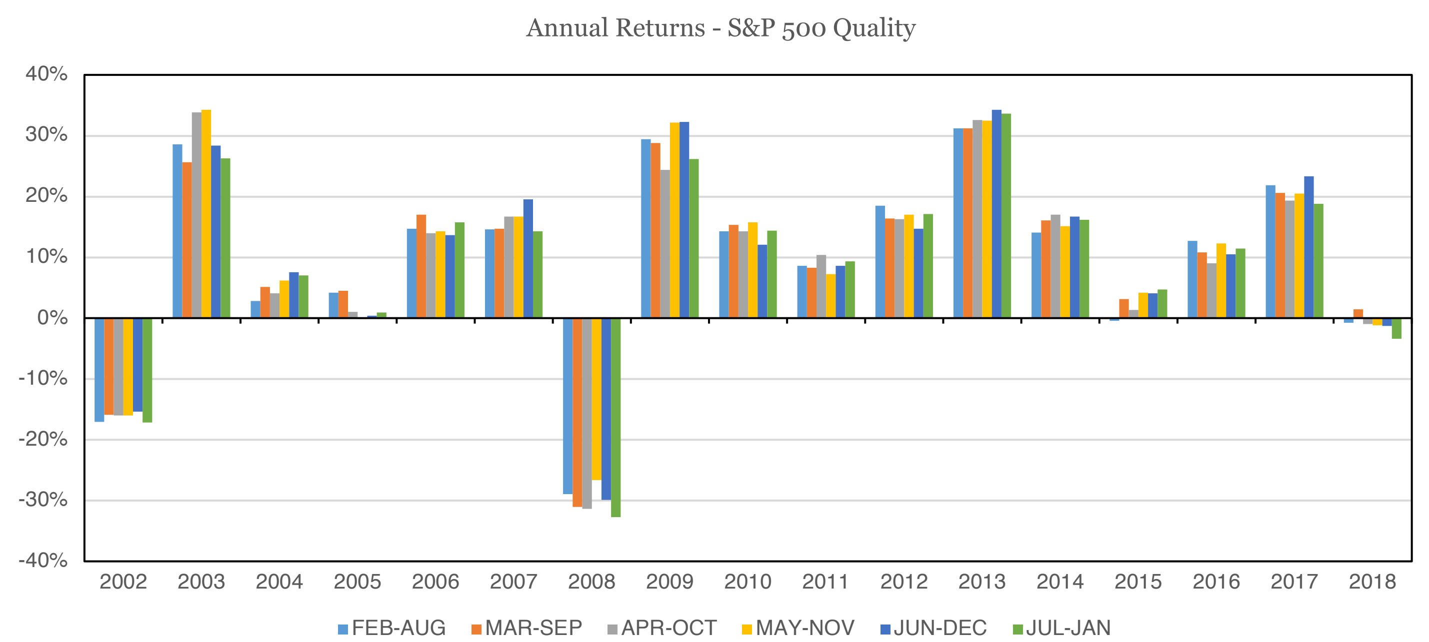
Furthermore, we can see that even seemingly closely related rebalances can have significant dispersion: consider MAY-NOV and JUN-DEC for Momentum. Here is a real doozy of a statistic: at one point, the MAY-NOV implementation for Momentum is down -50.3% while the JUN-DEC variation is down just -13.8%.

These differences are even more evident if we plot the annual returns for each strategy's rebalance variations. Note, in particular, the extreme differences in Value in 2009, Momentum in 2017, and Quality in 2003.









Source: Sharadar. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

Conclusion

In this study, we have explored the impact of rebalance timing luck on the results of smart beta / equity style portfolios.

We empirically tested this impact by designing a variety of portfolio specifications for four different equity styles (Value, Momentum, Low Volatility, and Quality). The specifications varied by concentration as well as rebalance frequency. We then constructed all possible rebalance variations of each specification to calculate the realized impact of rebalance timing luck over the test period (2000-2019).

In line with our mathematical model, we generally find that those strategies with higher turnover have higher timing luck and those that rebalance more frequently have less timing luck.

The sheer magnitude of timing luck, however, may come as a surprise to many. For reasonably concentrated portfolios (100 stocks) with semi-annual rebalance frequencies (common in many index definitions), annual timing luck ranged from 1-to-4%, which translated to a 95% confidence interval in annual performance dispersion of about +/-1.5% to +/-12.5%.

The sheer magnitude of timing luck calls into question our ability to draw meaningful relative performance conclusions between two strategies.

We then explored more concrete examples, replicating the S&P 500 Enhanced Value, Momentum, Low Volatility, and Quality indices. In line with expectations, we find that Momentum (a high turnover strategy) exhibits significantly higher realized timing luck than a lower turnover strategy rebalanced more frequently (i.e. Low Volatility).

For these four indices, the amount of rebalance timing luck leads to a staggering level of dispersion in realized terminal wealth.

“But Corey,” you say, “this only has to do with systematic factor managers, right?”

Consider that most of the major equity style benchmarks are managed with annual or semi-annual rebalance schedules. Good luck to anyone trying to identify manager skill when your benchmark might be realizing hundreds of basis points of positive or negative performance luck a year.

Re-specifying the Fama French 3-Factor Model

December 16, 2019

Summary

- The Fama French three-factor model provides a powerful tool for assessing exposures to equity risk premia in investment strategies.
- In this note, we explore alternative specifications of the value (HML) and size (SMB) factors using price-to-earnings, price-to-cash flow, and dividend yield.
- Running factor regressions using these alternate specifications on a suite of value ETFs and Newfound's Systematic Value strategy, lead to a wide array of results, both numerically and directionally.
- While many investors consider the uncertainty of the parameter estimates from the regression using the three-factor model, most do not consider the uncertainty that comes from the assumption of how you construct the equity factors in the first place.
- Understanding the additional uncertainty is crucial for manager and investors who must consider what risks they are trying to measure and control by using tools like factor regression and make sure their assumptions align with their goals.

In their 1992 paper, *The Cross-Section of Expected Stock Returns*, Eugene Fama and Kenneth French outlined their three-factor model to explain stock returns.

While the Capital Asset Pricing Model (CAPM) only describes asset returns in relation to their exposure to the market's excess return through the stock's beta and identifies any return beyond that as alpha, Fama and French's three-factor model reattributed some of that supposed alpha to exposures to a value factor (High-minus-low or HML) based on returns stratified by price-to-book ratios and a size factor (small-minus-big or SMB) based on returns stratified by market capitalization.

This gave investors a tool to judge investment strategies based on the loadings to these risk factors. A manager with a seemingly high alpha may have simply been investing in value and small-cap stocks historically.

The notion of compensated risk premia has also opened the floodgate of many additional factors from other researchers (such as momentum, quality, low beta, etc.) and even two more factors from Fama and French (investment and profitability).

A richer factor universe opens up a wide realm of possibilities for analysis and attribution. However, setting further developments aside and going back to the original three-factor model, we would be remiss if we didn't dive a bit further into its specification.

At the highest level, we agree with treating "value" and "size" as risk factors, but there is more than one way to skin a factor.

What is "value"?

Fama and French define it using the price-to-book ratio of a stock. This seems legitimate for a broad swath of stocks, especially those that are very capital intensive – such as energy, manufacturing, and financial firms – but what about industries that have structurally lower book values and may have other potential price drivers? For example, a technology company might have significant intangible intellectual property and some utility companies might employ leverage, which decreases their book value substantially.

To determine value in these sectors, we might utilize ratios that account for sales, dividends, or earnings. But then if we analyzed these strategies using the Fama French three-factor model as it is specified, we might misjudge the loading on the value factor.

"Size" seems more straightforward. Companies with low market capitalizations are small. However, when we consider how the size factor is defined based on the value factor, there might even be some differences in SMB using different value metrics.

$$SMB = 1/3\ (Small\ Value + Small\ Neutral + Small\ Growth) \\ -\ 1/3\ (Big\ Value + Big\ Neutral + Big\ Growth).$$

In this commentary, we will explore what happens when we alter the definition of value for the value factor (and hence the size factor) and see how this affects factor regressions of a sample of value ETFs along with our Systematic Value strategy.

HML Factor Definitions

In the standard version of the Fama French 3-fatcor model, HML is constructed as a self-financing long/short portfolio using a 2x3 sort on size and value. The investment universe is split in half based on market capitalization and in three parts (30%/40%/30%) based on valuation, in this base case, price-to-book ratio.

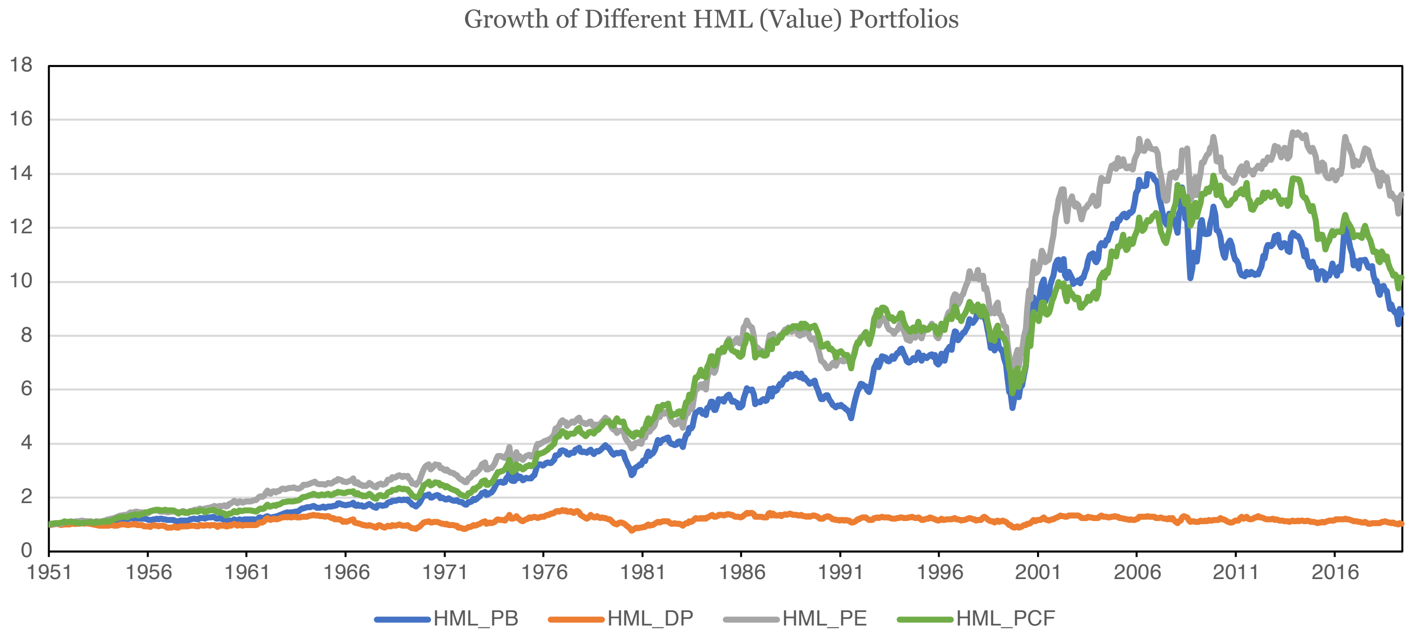
$$HML = 1/2\ (Small\ Value + Big\ Value) \\ -\ 1/2\ (Small\ Growth + Big\ Growth).$$

Using additional data from the Kenneth French Data Library and the same methodology, we will construct HML factors using sorts based on size and:

- Price-to-earnings ratios
- Price-to-cash flow ratios
- Dividend yields

The common inception date for all the factors is June 1951.

The chart below shows the growth of each of the four value factor portfolios.



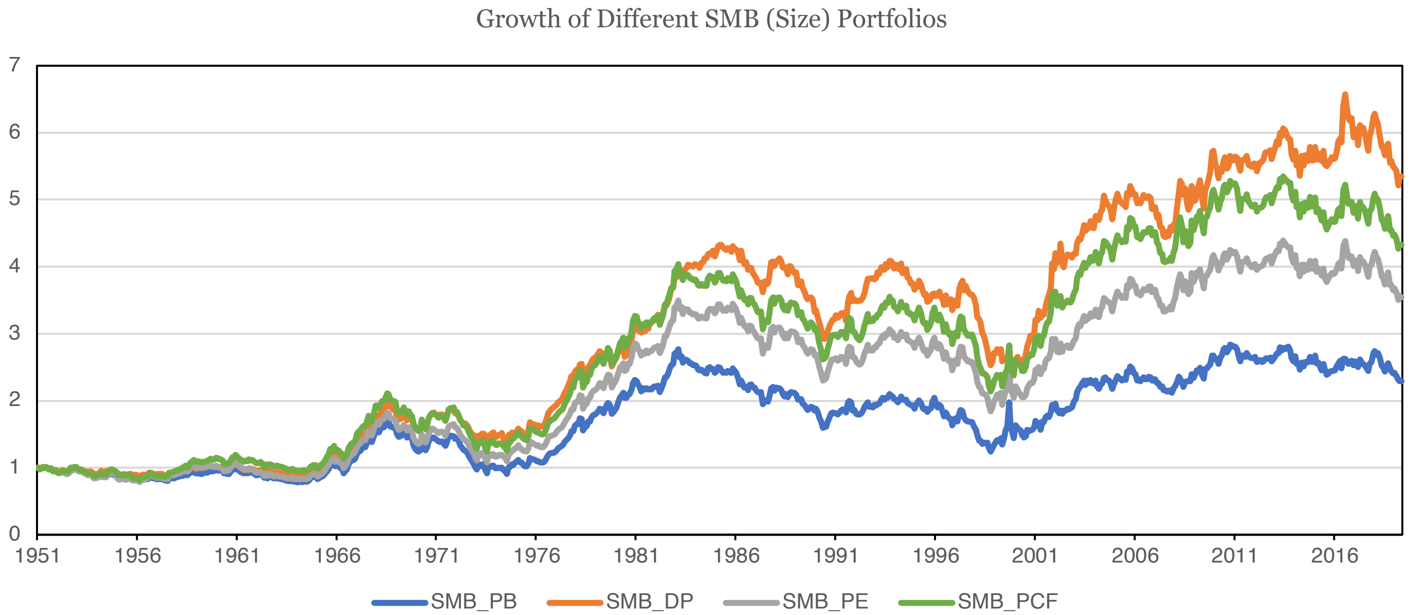
Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

Over the entire time period – and for many shorter time horizons – the standard HML factor using price-to-book does not even have the most attractive returns. Price-to-earnings and price-to-cash flow often beat it out.

On the other hand, the HML factor formed using dividend yields doesn't look so hot.

One of the reasons behind this is that the small, low dividend yield companies performed much better than the small companies that were ranked poorly by the other value factors. We can see this effect borne out in the SMB chart for each factor, as the SMB factor for dividend yield performed the best.

(Recall that we mentioned previously how the Fama French way of defining the size factor is dependent on which value metric we use.)



Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

Looking at the statistical significance of each factor through its t-statistic, we can see that Price-to-Earnings and Price-to-Cash Flow yielded higher significance for the HML factor than Price-to-Book. And those two along with Dividend Yield all eclipsed the Price-to-Book construction of the SMB factor.

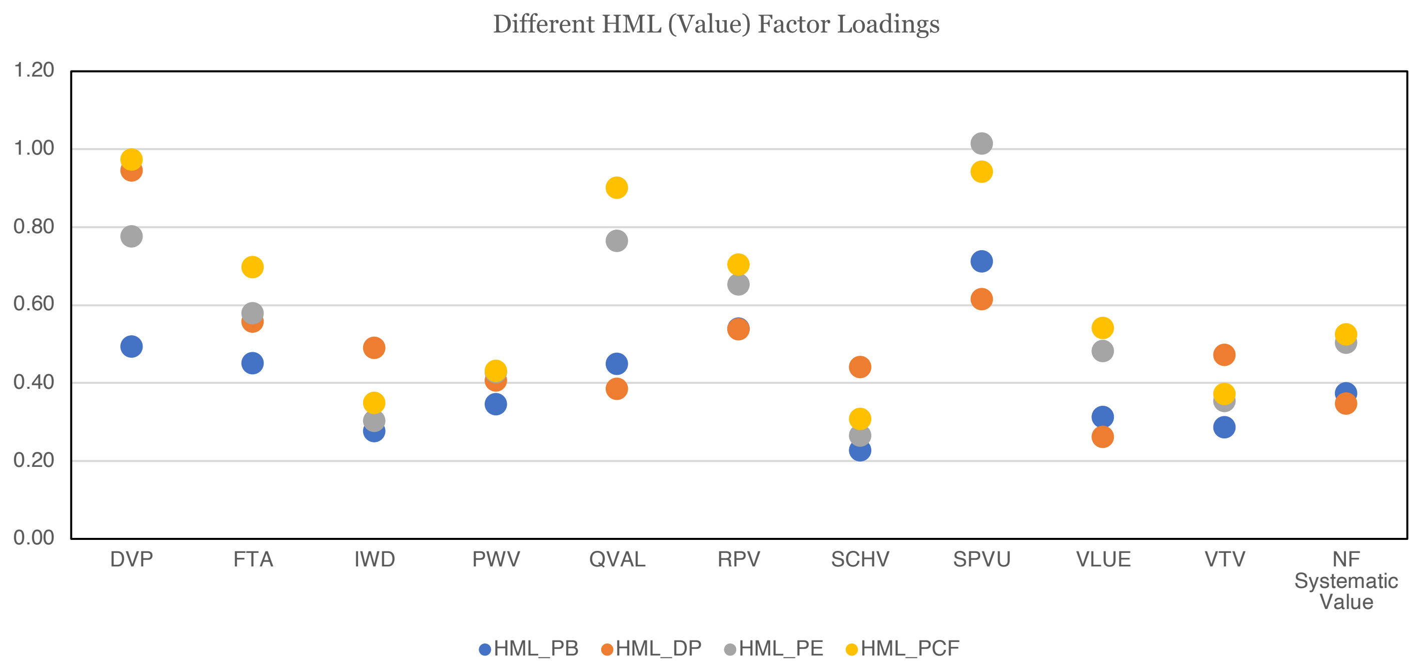
T-Statistics for HML and SMB Using Various Value Metrics

	Price-to-Book	Dividend Yield	Price-to-Earnings	Price-to-Cash Flow
HML	2.9	0.0	3.7	3.4
SMB	1.0	2.4	1.6	1.9

Assuming that we do consider all metrics to be appropriate ways to assess the value of companies, even if possibly under different circumstances, how do different variants of the Fama French three-factor model change for each scenario with regression analysis?

The Impact on Factor Regressions

Using a sample of U.S. value ETFs and our Systematic Value strategy, we plot the loadings for the different versions of HML. The regressions are carried out using the trailing three years of monthly data ending on October 2019.



Source: Tiingo, Kenneth French Data Library. Calculations by Newfound Research. Past performance is not an indicator of future results. Returns represent live strategy results. Returns for the Newfound Systematic Value strategy are gross of all management fees and taxes, but net of execution fees. Returns for ETFs included in study are gross of any management fees, but net of underlying ETF expense ratios. Returns assume the reinvestment of all distributions.

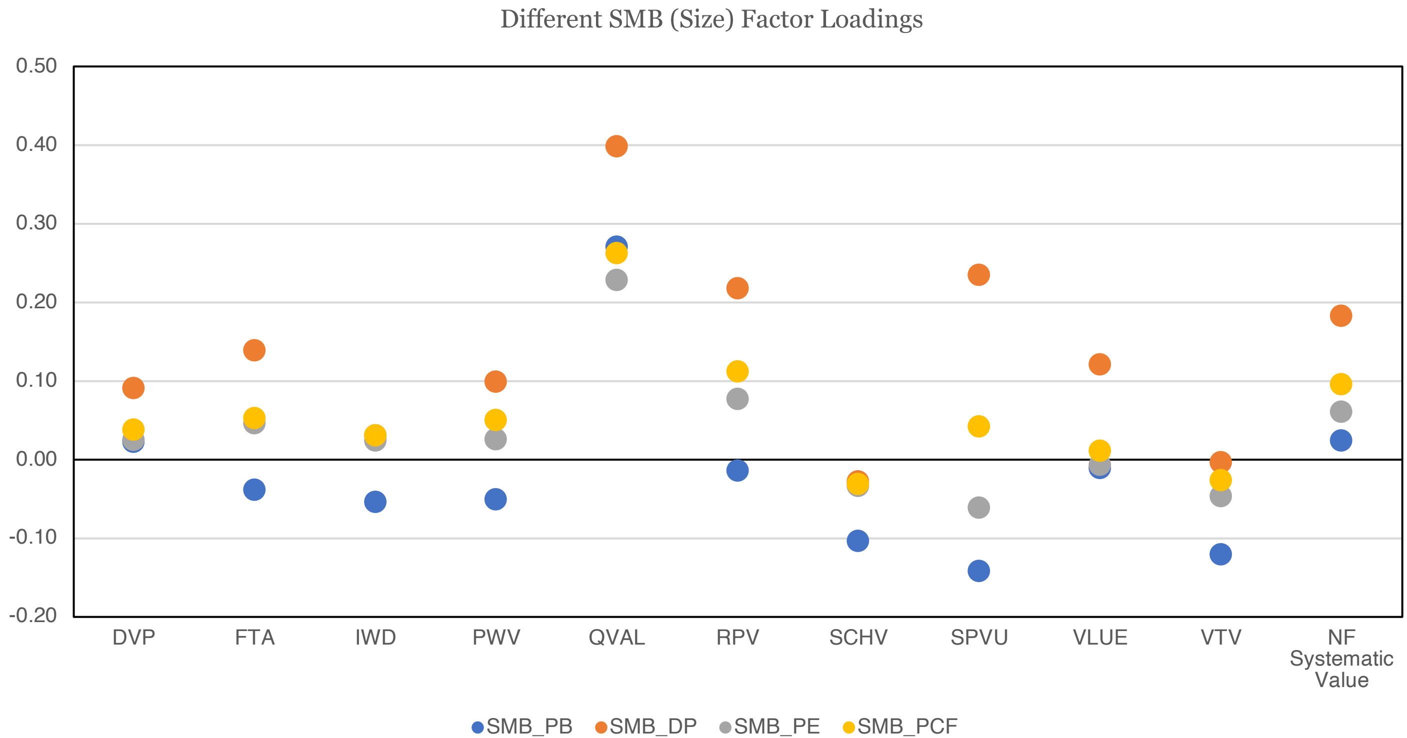
For each different specification of HML, the differences in the loading between investments is generally directionally consistent. For instance, DVP has higher loadings than FTA for all forms of HML.

However, sometimes this is not the case.

VLUE looks more attractive than VTV based on price-to-cash flow but not dividend yield. FTA is roughly equivalent to QVAL in terms of loading when price-to-book is used for HML, but it varies wildly when other metrics are used.

The tightest range for the four models for any of the investments is 0.09 (PWV) and the widest is 0.52 (QVAL). When we factor in that these estimates each have their own uncertainty, distinguishing which investment has the better value characteristic is tough. Decisions are commonly made on much smaller differences.

We see similar dispersion in the SMB loadings for the various constructions.



Source: Tiingo, Kenneth French Data Library. Calculations by Newfound Research. Past performance is not an indicator of future results. Returns represent live strategy results. Returns for the Newfound Systematic Value strategy are gross of all management fees and taxes, but net of execution fees. Returns for ETFs included in study are gross of any management fees, but net of underlying ETF expense ratios. Returns assume the reinvestment of all distributions.

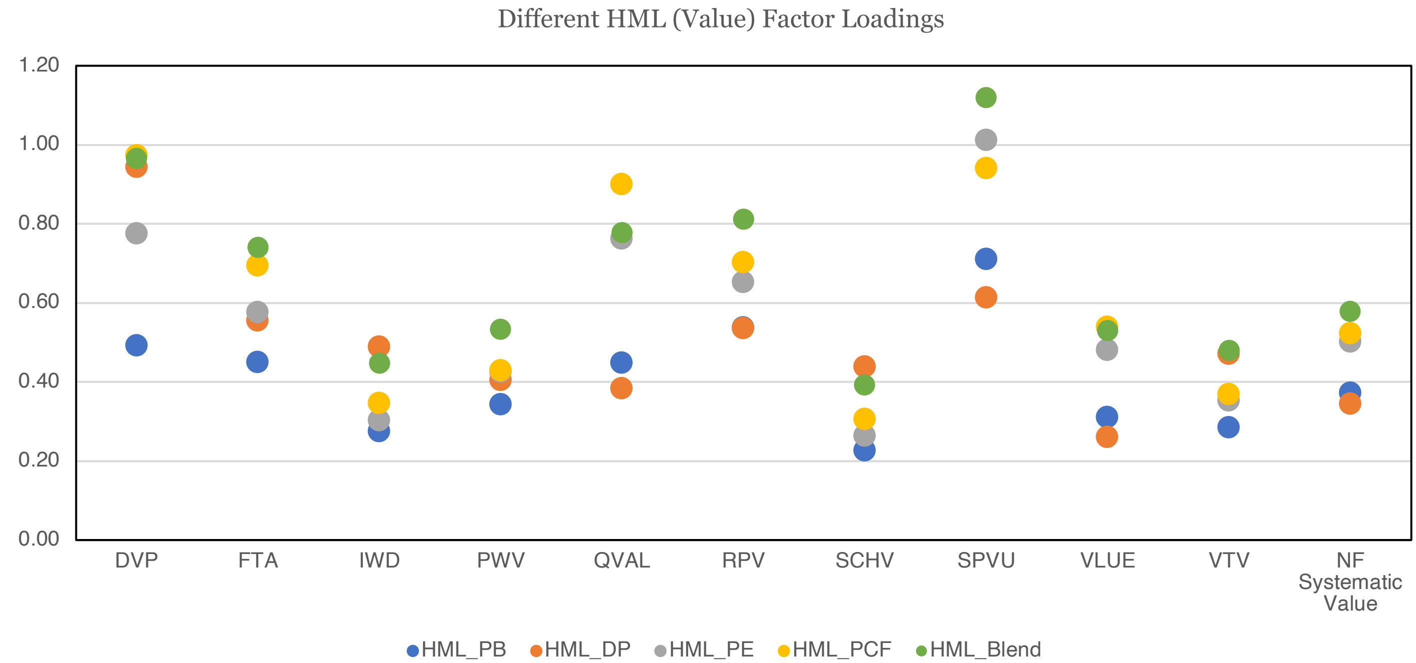
Many of these values are not statistically significant from zero, so someone who has a thorough understanding of uncertainty in regression would likely not draw a strict comparison between most of these investments.

However, one implication of this is that if a metric is chosen that does ascribe significant size exposure to one of these investments, an investor may make a decision based on not wanting to bear that risk in what they desire to be a large-cap investment.

Can We Blend Our Way Out?

One way we often mitigate model specification risk is by blending a number of models together into one.

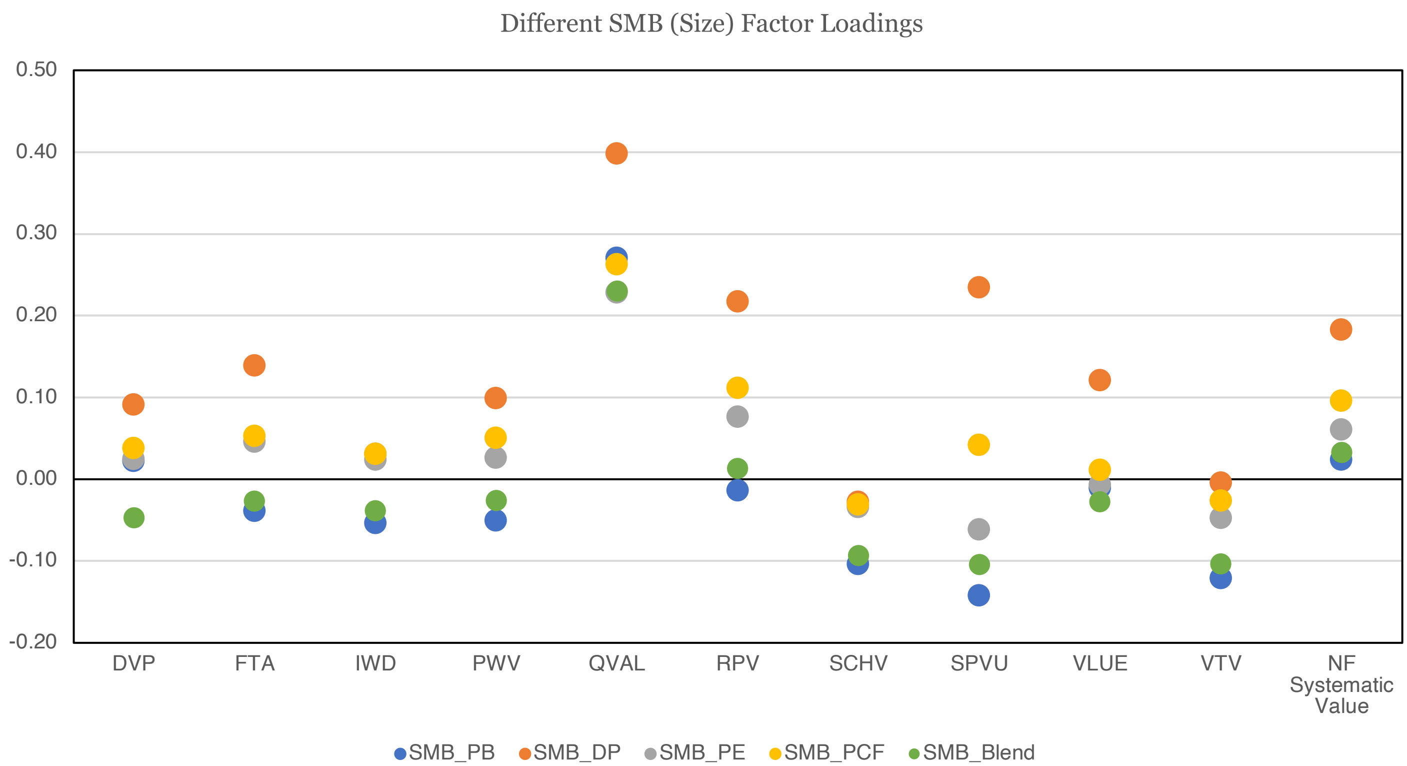
By averaging all of our HML and SMB factors, respectively, we arrive at blended factors for the three-factor model.



Source: Tiingo, Kenneth French Data Library. Calculations by Newfound Research. Past performance is not an indicator of future results. Returns represent live strategy results. Returns for the Newfound Systematic Value strategy are gross of all management fees and taxes, but net of execution fees. Returns for ETFs included in study are gross of any management fees, but net of underlying ETF expense ratios. Returns assume the reinvestment of all distributions.

All of the investments now have HML loadings in the top of their range of the individual model loadings, and many (FTA, PWV, RPV, SPVU, VTV, and the Systematic Value strategy) have loadings to the blended HML factor that exceed the loadings for all of the individual models.

The opposite is the case for the blended SMB factor: the loadings are in the low-end of the range of the individual model loadings.



Source: Tiingo, Kenneth French Data Library. Calculations by Newfound Research. Past performance is not an indicator of future results. Returns represent live strategy results. Returns for the Newfound Systematic Value strategy are gross of all management fees and taxes, but net of execution fees. Returns for ETFs included in study are gross of any management fees, but net of underlying ETF expense ratios. Returns assume the reinvestment of all distributions.

So which is the correct method?

That's a good question.

For some investments, it is situation-specific. If a strategy only uses price-to-earnings as its value metric, then putting it up against a three-factor model using the P/E ratio to construct the factors is appropriate for judging the efficacy of harvesting that factor.

However, if we are concerned more generally about the abstract concept of "value", then the blended model may be the best way to go.

Conclusion

In this study, we have explored the impact of model specification for the value and size factor in the Fama French three-factor model.

We empirically tested this impact by designing a variety of HML and SMB factors based on three additional value metrics (price-to-earnings, price-to-cash flow, and dividend yield). These factors were constructed using the same rules as for the standard method using price-to-book ratios.

Each factor, with the possible exceptions of the dividend yield-based HML, has performance that could make it a legitimate specification for the three-factor model over the time that common data is available.

Running factor regressions using these alternate specifications on a suite of value ETFs and Newfound's Systematic Value strategy, led to a wide array of results, both numerically and directionally.

While many investors consider the uncertainty of the parameter estimates from the regression using the three-factor model, most do not consider the uncertainty that comes from the assumption of how you construct the equity factors in the first place.

Understanding the additional uncertainty is crucial for decision-making. Managers and investors alike must consider what risks they are trying to measure and control by using tools like factor regression and make sure their assumptions align with their goals.

"Value" is in the eye of the beholder, and blind applications of two different value factors may lead to seeing double conclusions.

Timing Trend Model Specification with Momentum

December 23, 2019

Summary

- Over the last several years, we have written several research notes demonstrating the potential benefits of diversifying "specification risk."
- Specification risk occurs when an investment strategy is overly sensitive to the outcome of a single investment process or parameter choice.
- Adopting an ensemble approach is akin to creating a virtual fund-of-funds of stylistically similar managers, exhibiting many of the same advantages of traditional multi-manager diversification.
- In this piece, we briefly explore whether model specification choices can be timed using momentum within the context of a naïve trend strategy.
- We find little evidence that momentum-based parameter specification leads to meaningful or consistent improvements beyond a naively diversified approach.

Over the last several years, we've advocated on numerous occasions for a more holistic view of diversification: one that goes beyond just *what* we invest in, but also considers *how* those decisions are made and *when* they are made.

We believe that this style of thinking can be applied "all the way down" our process. For example, how-based diversification would advocate for the inclusion of both value and momentum processes, as well as for different approaches to capturing value and momentum.

Unlike correlation-based *what* diversification, *how*-based diversification often does little for traditional portfolio risk metrics. For example, in *Is Multi-Manager Diversification Worth It?* we demonstrated that within most equity categories, allocating across multiple managers does almost nothing to reduce portfolio volatility. It does, however, have a profound impact on the dispersion of terminal wealth that is achieved, often by avoiding manager-specific tail-risks. In other words, our certainty of achieving a given outcome may be dramatically improved by taking a multi-manager approach.

Ensemble techniques to portfolio construction can be thought of as adopting this same multi-manager approach by creating a set of virtual managers to allocate across.

In late 2018, we wrote two notes that touched upon this: *When Simplicity Met Fragility* and *What Do Portfolios and Teacups Have in Common?* In both studies we injected a bit of randomness into asset returns to measure the stability of trend-following strategies. We found that highly simplistic models tended to exhibit significant deviations in results with just slightly modified inputs, suggesting that they are highly fragile. Increasing diversification across *what, how*, and *when* axes led to a significant improvement in outcome stability.

As empirical evidence, we studied the real-time results of the popular Dual Momentum GEM strategy in our piece *Fragility Case Study: Dual Momentum GEM*, finding that slight deviations in model specification lead to significantly different allocation conclusions and therefore meaningfully different performance results. This was particularly pronounced over short horizons.

Tying trend-following to option theory, we then demonstrated how an ensemble of trend following models and specifications could be used to increase outcome certainty in *Tightening the Uncertain Payout of Trend-Following.*

Yet while more diversification appears to make portfolios more consistent in the outcomes they achieve, empirical evidence also suggests that certain specifications can lead to superior results for prolonged periods of time. For example, slower trend following signals appear to have performed much, much better than fast trend following signals over the last two decades.

One of the benefits of being a quant is that it is easy to create thousands of virtual managers, all of whom may follow the same style (e.g. "trend") but implement with a different model (e.g. prior total return, price-minus-moving-average, etc) and specification (e.g. 10 month, 200 day, 13 week / 34 week cross, etc). An ancillary benefit is that it is also easy to re-allocate capital among these virtual managers.

Given this ease, and knowing that certain specifications can go through prolonged periods of out-performance, we might ask: can we time specification choices with momentum?

Timing Trend Specification

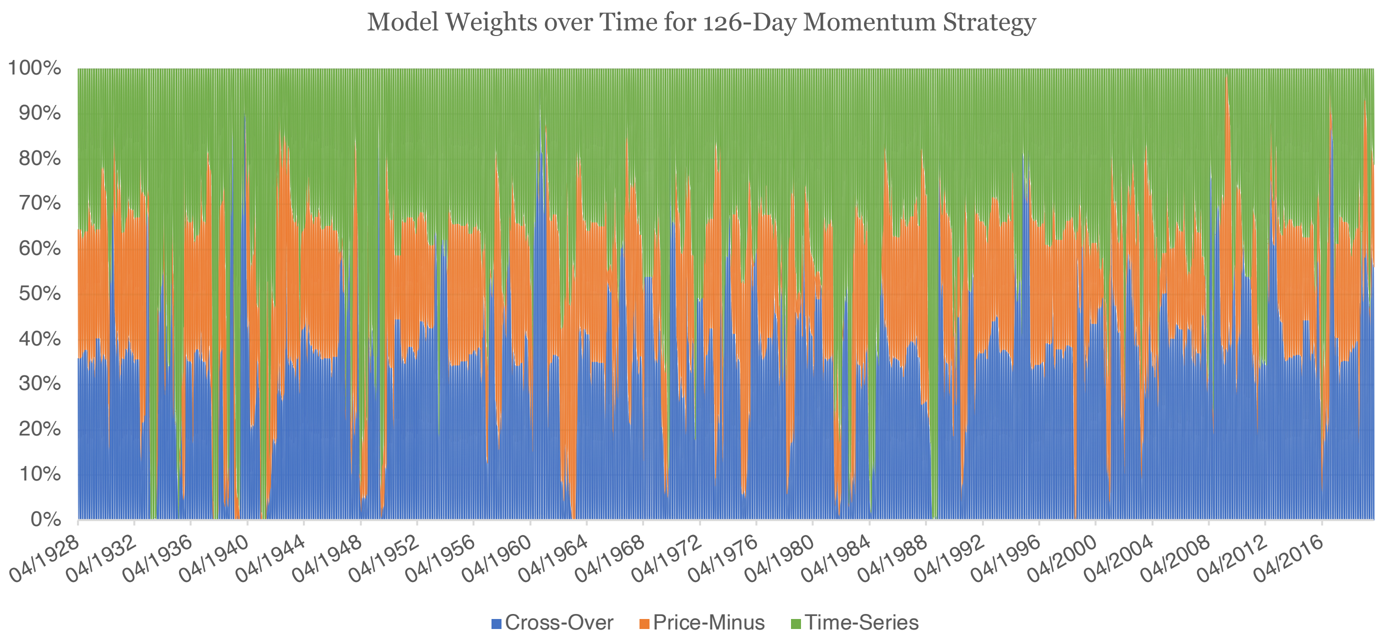
In this research note, we will explore whether momentum signals can help us time out specification choices as it relates to a simple long/flat U.S. trend equity strategy.

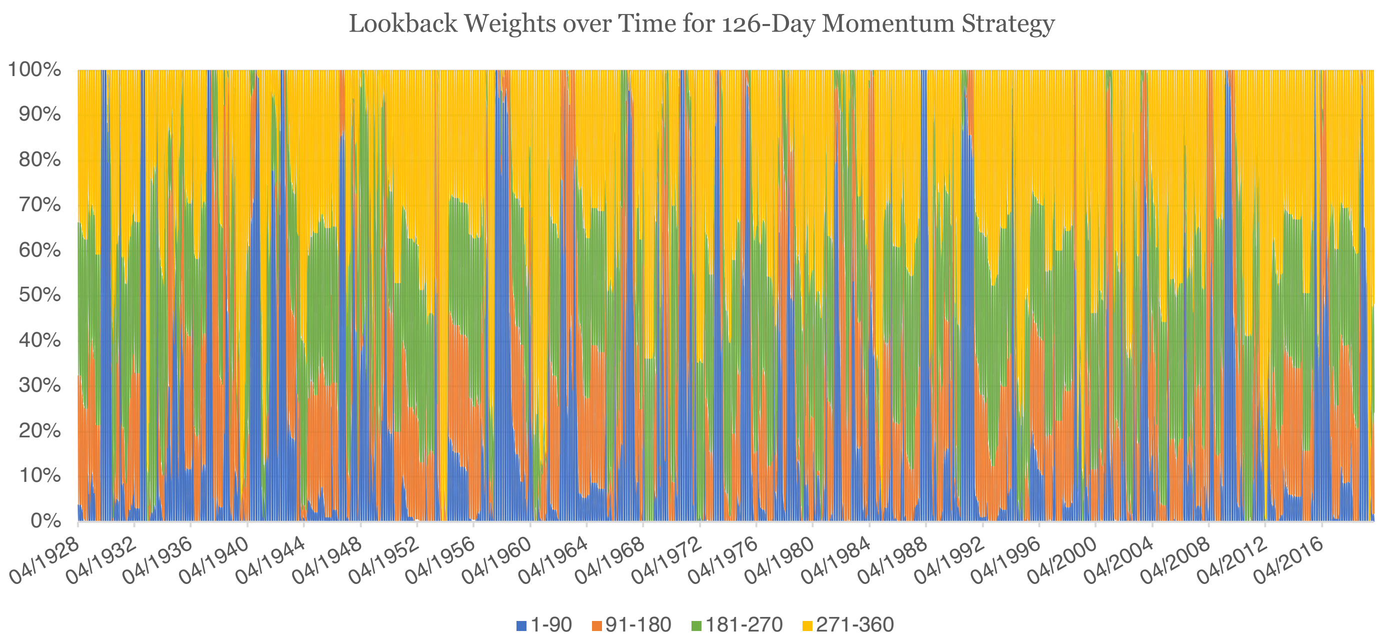
Using data from the Kenneth French library, our strategy will hold broad U.S. equities when the trend signal is positive and shift to the risk-free asset when trends are negative. We will develop 1023 different strategies by employing three different models – prior total return, price-minus-moving-average, and dual-moving-average-cross-over – with lookback choices spanning from 20-to-360 days in length.

After constructing the 1023 different strategies, we will then apply a momentum model that ranks the models based upon prior returns and equally-weights our portfolio across the top 10%. These choices are made daily and implemented with 21 overlapping portfolios to reduce the impact of rebalance timing luck.

It should be noted that because the underlying strategies are only allocating between U.S. equities and a risk-free asset, they can go through prolonged periods where they have identical returns or where more than 10% of models share the highest prior return. In these cases, we select all models that have returns equal-to-or-greater-than the model identified at the 10th percentile.

Before comparing performance results, we think it is worthwhile to take a quick look under the hood to see whether the momentum-based approach is actually creating meaningful tilts in specification selection. Below we plot both aggregate model and lookback weights for the 126-day momentum strategy.



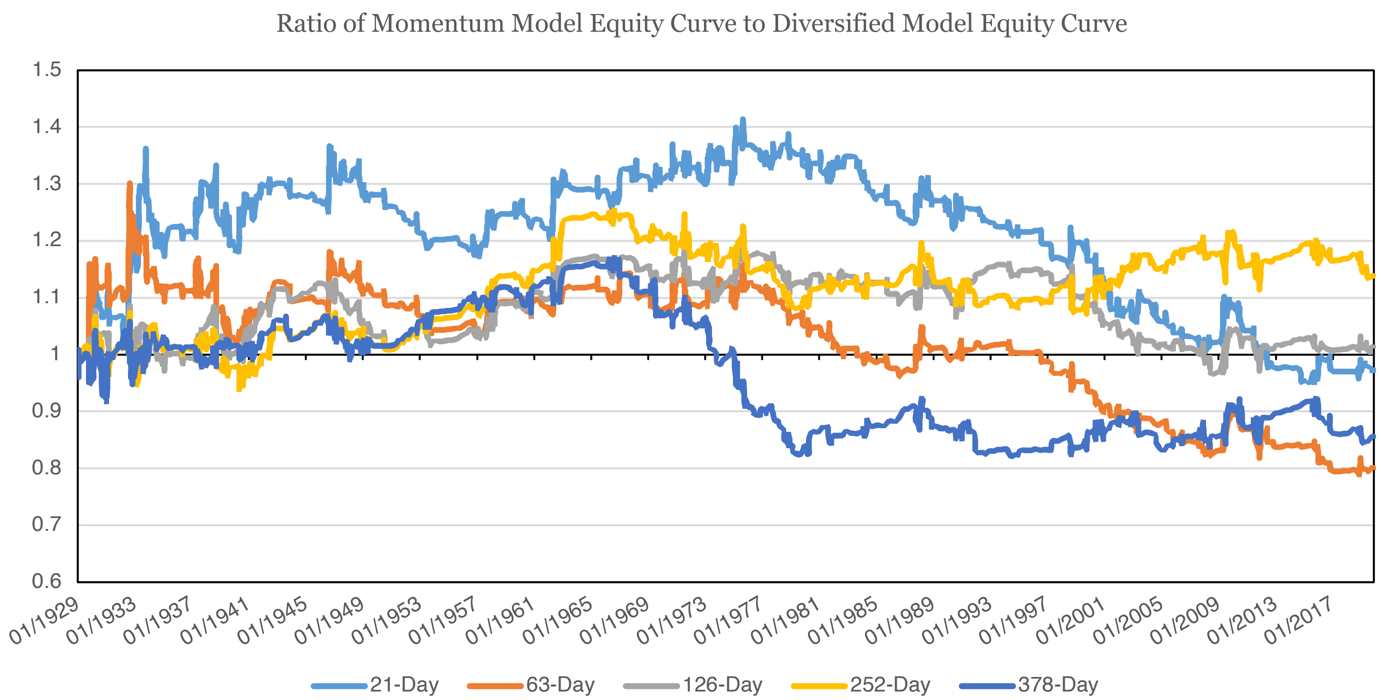


Source: Kenneth French Data Library. Calculations by Newfound Research.

We can see that while the model selection remains largely balanced, with the exception of a few periods, the lookback horizon selection is far more volatile. On average, the strategy preferred intermediate-to-long-term signals (i.e. 181-to-360 day), but we can see intermittent periods where short-term models carried favor.

Did this extra effort generate value, though? Below we plot the ratio of the momentum strategies' equity curves versus the naïve diversified approach.

We see little consistency in relative performance and four of the five strategies end up flat-to-worse. Only the 252-day momentum strategy out-performs by the end of the testing period and this is only due to a stretch of performance from 1950-1964. In fact, since 1965 the relative performance of the 252-day momentum model has been negative versus the naively diversified approach.



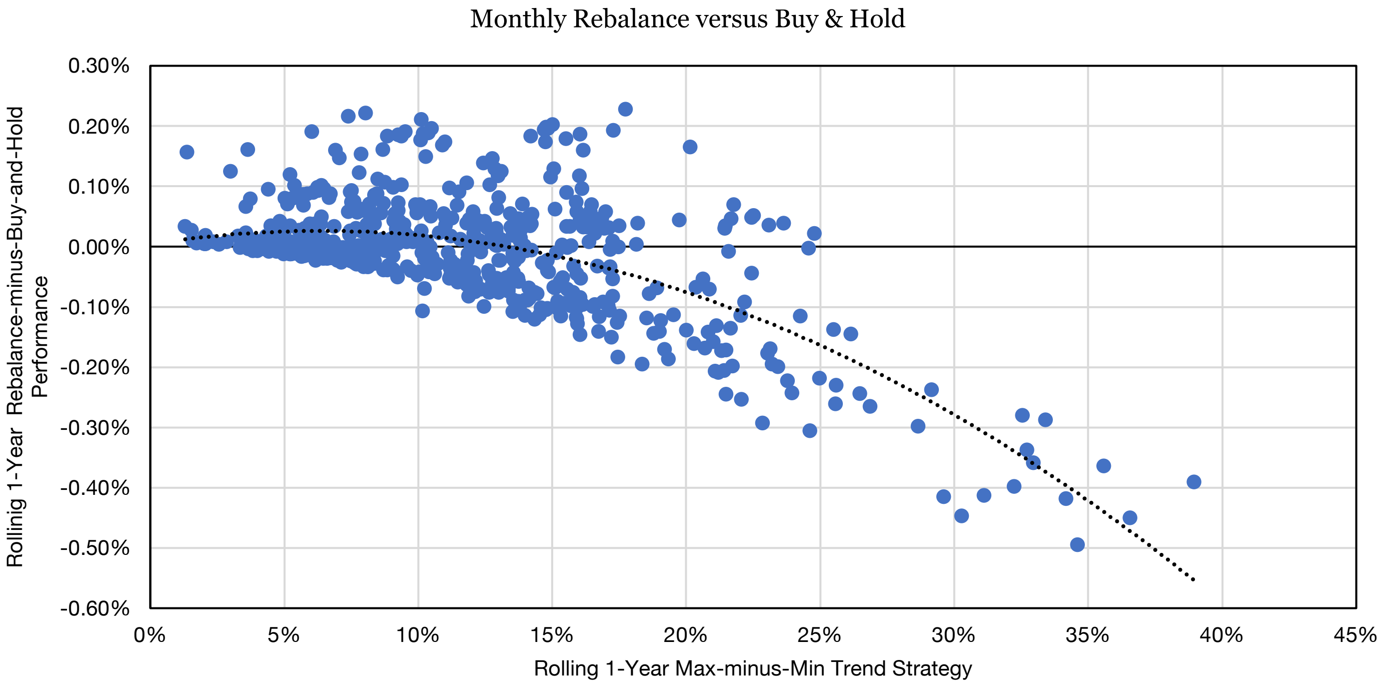
Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

This analysis suggests that naïve, momentum-based specification selection does not appear to have much merit against a diversified approach for our simple trend equity strategy.

The Potential Benefits of Virtual Rebalancing

One potential benefit of an ensemble approach is that rebalancing across virtual managers can generate growth under certain market conditions. Similar to a strategically rebalanced portfolio, we find that when returns across virtual managers are expected to be similar, consistent rebalancing can harvest excess returns above a buy-and-hold approach.

The trade-off, of course, is that when there is autocorrelation in specification performance, rebalancing creates a drag. However, given that the evidence above suggests that relative performance between specifications is not persistent, we might expect that continuously rebalancing across our ensemble of virtual managers may actually allow us to harvest returns above and beyond what might be possible with just selecting an individual manager.



Source: Kenneth French Data Library. Calculations by Newfound Research. Past performance is not an indicator of future results. Performance is backtested and hypothetical. Performance figures are gross of all fees, including, but not limited to, manager fees, transaction costs, and taxes. Performance assumes the reinvestment of all distributions.

Conclusion

In this study, we explored whether we could time model specification choices in a simple trend equity strategy using momentum signals.

Testing different lookback horizons of 21-through-378 days, we found little evidence of meaningful persistence in the returns of different model specifications. In fact, four of the five momentum models we studied actually under-performed a naïve, diversified. The one model that did out-perform only seemed to do so due to strong performance realized over the 1950-1964 period, actually relatively under-performing ever since.

While this evidence suggests that timing specification with momentum may not be a fruitful approach, it does suggest that the *lack* of return persistence may benefit diversification for a second reason: rebalancing. Indeed, barring any belief that one specification would necessarily do better than another, consistently re-pooling and distributing resources through rebalancing may actually lead to the growth-optimal solution.[66] This potentially implies an even higher hurdle rate for specification-timers to overcome.

[66] See section 4.1 Cooperation in the lecture notes provided by Ole Peters (https://ergodicityeconomics.files.wordpress.com/2018/06/ergodicity_economics.pdf)

To read other commentaries or to subscribe to future posts, please visit blog.thinknewfound.com.