Want Smart Beta? Follow the Smart Money: Market and Factor Timing Using Relative Sentiment

Raymond C. Micaletti *

ABSTRACT

We present a real-time, cross-asset, positions-based relative sentiment indicator to predict the U.S. equity market. Derived from the Commitments of Traders report, the indicator measures—in a novel way—the aggregate positioning in equities of institutional investors relative to individual investors. Applying a wide range of statistical tests and controlling for data snooping, we find this indicator to be highly significant, exceptionally robust, and substantially more powerful than both value and momentum in predicting U.S. equity returns. Beyond the broad market, the indicator also facilitates the timing of several "smart beta" equity factors—many of which were thought difficult or impossible to time. We propose a tactical asset allocation strategy based on the indicator and compare it to several value- and/or momentum-based alternatives—finding the proposed strategy produces higher returns (both absolute and risk-adjusted), while having considerably less time-averaged exposure to equities.

*Dr. Raymond C. Micaletti, is the co-Chief Investment Officer of Columbus Macro, LLC, an investment adviser registered with the Securities and Exchange Commission. The views expressed in this paper are those of the author and do not necessarily reflect the views or opinions of Columbus Macro, LLC or its employees. Please refer to Columbus Macro's Form ADV for additional information or visit their website at www.columbusmacro.com

Market timing has long been controversial. While generations of investors have aspired to skillful market timing, their collective lack of success has lent support to those who argue against its viability. Meanwhile, the relentless criticism of market timing by proponents of efficient markets has endowed the concept with indeliby negative connotations.

Nevertheless, the temptation to time the market is strong and various approaches have been proposed. Often these approaches take the form of tactical asset allocation strategies, which are generally less extreme variants of the binary, "risk-on/risk-off" approach many associate with market timing. Among these strategies, two of the more popular and durable approaches involve the use of value and (time-series) momentum (Asness et al. (2013), Moskowitz et al. (2012))—factors widely regarded as providing excess returns over time. Despite their durability, however, value and momentum both have notable weaknesses when it comes to tactical asset allocation.

Value, for example, is prone to reducing allocations (or exiting positions) too early during uptrends and prone to increasing allocations (or initiating positions) too early during downtrends. Momentum, in contrast, tends to have the opposite weaknesses. That is, momentum is vulnerable to staying in the market too long—while momentum indicators remain positive, but the market has topped and has begun to fall; and likewise, is vulnerable to staying out of the market too long—when momentum indicators remain negative, but the market has bottomed and has begun to rise. This latter weakness can be particularly vexing for momentum-based investors, who are susceptible to missing out on the often-substantive returns that tend occur right after significant market bottoms.

In this paper, we present what appears to be a more efficient and more powerful alternative to value and momentum—one that helps address the weaknesses listed above. The alternative we propose is a cross-asset, positions-based relative sentiment indicator derived from the Commitments of Traders (COT) report. This indicator measures the aggregate degree to which institutional investors are long or short equities relative to individual investors. The key innovation of the indicator is that it takes into account not only investors' positions in equities, but also their positions in closely related non-equity assets. For ease of reference, we will refer to this indicator as the Smart Money Indicator (SMI), as institutional investors are often thought of as the "smart money."

Applying a wide range of statistical tests—parametric, nonparametric, Bayesian, and classical—we find the SMI to have exhibited highly significant timing ability across the 25-year history of the weekly COT report. This timing ability is evident in both the first and second halves of the historical data, passes extremely stringent data snooping tests, and is exceptionally robust to different parameter combinations, lags in the data, and methods of constructing the indicator. Moreover, the SMI can be implemented in real-time, has holding periods amenable to tactical asset allocation, and its underlying composition lends insight into the behavioral forces that help drive markets.

To put it simply, the SMI, by virtue of its being positive (institutions relatively bullish equities) or negative (institutions relatively bearish equities), appears to partition weekly U.S. equity market returns into two distinct groups: One with substantially above average returns, and the other with substantially below average returns. The economic significance here is meaningful: When the SMI is positive, annualized equity market returns are approximately 20 percentage points higher—with lower volatility—than when the SMI is negative.

But beyond the broad market, the SMI also provides valuable timing information for:

1. **Time-series momentum**: The SMI essentially renders time-series momentum uninformative as a predictive factor for the U.S. equity market. That is, regardless of the state of the market's time-series momentum, returns are significantly higher (lower) when the SMI is positive (negative). This effect is especially pronounced when the market has negative momentum. In such cases, annualized market returns are approximately 30% (on average) when the SMI is positive and roughly -20% when the SMI is negative—a spread of 50 percentage points. In other words, the SMI appears to be adept at identifying when (and when not) to take market exposure during periods of negative momentum.

2. **"Smart beta" equity factors**: While conventional wisdom tends to hold that long/short equity factors are difficult, if not impossible, to time, we find the polarity of the SMI largely determines the return outcomes of several such factors, particularly ones based on fundamentals. Additionally, for many of these factors, the SMI's timing significance increases markedly when conditioned on the state of the factors' time-series momentum.

As market timing indicators often play a central role in tactical asset allocation (TAA), we develop a walk-forward TAA strategy based on the SMI and compare it to several benchmarks, both active and passive. We find the SMI-based strategy outperforms each benchmark—producing higher returns (both absolute and risk-adjusted), while having considerably less time-averaged exposure to equities than its active counterparts (all of which involve value, momentum, or a combination or the two).

As the preceding overview suggests, this paper lies at the intersection of several active debates in financial economics—debates that center on the identity of the Smart Money, the feasibilty of market and factor timing, and the predictive power of investor sentiment, among others. In light of those debates, the contributions of this paper to the literature include:

1. The introduction of a novel, robust, and behaviorally-justified relative sentiment indicator that:
   (a) exhibits extreme levels of market timing significance across more than 20 years of market history (even after controlling for data snooping)
   (b) appears to identify when (and when not) to take equity market exposure during periods of negative time-series momentum
   (c) helps facilitate the timing of several smart beta equity factors
   (d) seems to provide a more powerful and more efficient alternative to value and momentum indicators for TAA
   (e) lends insight into the underlying behavioral forces that drive equity markets

The remainder of this paper is organized as follows. In section I, we review the related literature. In section II, we describe the data used in the study, and in section III, we present the rationale behind the SMI, along with its computational details. In section IV, we establish the notation used throughout the rest of the paper. We outline the statistical tests used to evaluate the SMI's timing ability in section V and present the results of those tests in sections VI-VIII (for the broad equity market, time-series momentum, and smart beta equity factors, respectively). In section IX, we develop a TAA strategy based on the SMI and compare it to several active and passive benchmarks. We conclude in section X by summarizing the results.

I. Literature Review

The central theme of this paper is the development of a robust relative sentiment indicator to help discern up-markets from down-markets over intermediate time horizons—and in the process help address several of the weaknesses of value and (time-series) momentum in the context of tactical asset allocation. Whether any indicator can accomplish such a task, however, is a matter of contention. Welch and Goyal (2008) review a comprehensive list of putatively predictive factors and find most have lost statistical significance since publication. Similarly, Marshall et al. (2008) and Neuhierl and Schlusche (2011) systematically test a wide array of market timing rules and find few, if any, retain statistical significance after controlling for data snooping.

Thus, with respect to the viability of market timing, sharp divisions abound. Opponents say it requires high prediction accuracy, results in only modest outperformance at best, and has small probability of success (e.g., Sharpe (1975), Bauer and Dahlquist (2001), Estrada (2008)); while proponents claim it lowers volatility, does not require high prediction accuracy, and has ample margin for error (e.g., Shilling (1992), Clarke et al. (1989), Sy (1990)). Between these two extremes, Asness et al. (2015) acknowledge the oft-held belief that market timing is a "sin", but advocate investors should "sin a little" by combining value and momentum strategies for both equities and bonds—a prescription not without justification, as value and momentum have mostly proven effective over time (e.g., Asness et al. (2013), Moskowitz et al. (2012)).

Long-term effectiveness notwithstanding, both value and momentum, as tactical asset allocation tools, can suffer through unpleasant episodes of underperformance. Value, for example, tends to reduce allocations too soon in uptrends and increase allocations too soon in downtrends (as in 2008-2009); while momentum tends to reduce allocations too late after market tops and increase allocations too late after market bottoms. The SMI appears to transcend several of these weaknesses.

Further, because the SMI exhibits the ability to identify up-markets from down-markets, it naturally becomes a candidate to time smart beta equity factors—many of which do better in one market regime or the other. The prevailing sentiment regarding the viability of factor timing, however, is even less generous than the prevailing sentiment with respect to market timing. Contrarian factor timing is thought difficult, if not impossible (e.g., Asness et al. (2017a), Asness (2016)), while other, more complex, methods for timing factors—which mainly tend to focus on value-versus-growth or small-caps-versus-large-caps (e.g., Nalbantov et al. (2006), L'Her et al. (2007), Miller et al. (2013))—appear not to have gained any traction. Nevertheless, we find the polarity of the SMI largely explains the return outcomes for several fundamentally-based long/short equity factors—even after controlling for data snooping.

In light of its performance, it is natural to ask whether the SMI has any behavioral explanation. In fact, it does.

Studies repeatedly have shown institutions tend to outperform individuals over intermediate time horizons. These results are deep and robust, covering disparate metrics, markets, and time periods (e.g., Schmeling (2007), Gibson and Safieddine (2003), Gibson et al. (2004), Grinblatt and Keloharju (2000), among many others). The SMI's construction explicitly aligns with these empirical results by quantifying the market direction in which institutions are leaning relative to individuals.

Moreover, de Roon et al. (2000) show that futures risk premia depend not only on investors' positions in the futures under consideration, but also on their positions in closely related assets—so-called "cross-hedging pressure." The SMI explicitly takes into account certain cross-hedging pressures in non-equity assets when quantifying relative sentiment in equities. In fact, our tests show it is precisely the inclusion of these cross-hedging effects that drives the SMI's predictive power—underscored by the fact that when we more fully account for these cross-hedging pressures, the timing significance of the SMI increases.

Although a handful of prior studies have looked at relative sentiment (whether directly or indirectly) and some have even incorporated COT data and cross-hedging pressure, often these studies use data not available in real time (e.g., Edelen et al. (2010), Gibson and Safieddine (2003)) or use cross-hedging pressure unrelated to equities as part of a multi-factor, multi-asset analysis (e.g., Basu et al. (2006)) or cover only a small period of market history (e.g., Schmeling (2007), Basu et al. (2006)). Moreover, none of these prior studies controls for data snooping.

In contrast, the study presented here:

1. Introduces a single, binary, real-time relative sentiment indicator to predict the U.S. equity market
2. Includes only relevant (and justifiable) cross-hedging effects in the quantification of relative sentiment
3. Analyzes a much longer period of history than prior relative sentiment studies
4. Controls for data snooping

II. Data

We use data from the "legacy" Commitments of Traders report to construct the SMI. The legacy COT report separates traders into three classifications: "Commercials," "Non-commercials," and "Non-reportables." Commercials are defined as traders who use the derivatives markets primarily to hedge business risk; Non-commercials, as traders who use the derivatives market to speculate; and Non-reportables, as traders holding positions below the CFTC's reporting threshold (i.e., individuals).

The CFTC also releases a "disaggregated" report that breaks down the Commercial and Non-commercial categories even further. From this report, one observes significant overlap between Commercial traders and institutional asset managers in the financial futures used to construct the SMI. Because of this overlap, we use the terms "Commercials" and "institutions" interchangeably.

Prior to March 1995, the COT report monitored only futures positions. Since that time, the CFTC has released two reports: a futures-only report and one that aggregates both futures and options positions. For our primary results, we splice futures-only data until March 14, 1995 with futures-and-options data from March 21, 1995 onward (henceforth referring to this as "spliced COT data"). To demonstrate the SMI's robustness, however, we also present timing results using futures-only data across the entire time period.

Presently, the COT report is a snapshot of traders' positions taken every Tuesday and released every Friday at 3:30 p.m. Eastern on the CFTC's website[1]. The report's release and availability were not always this accommodating, however. To account for the evolving nature of the report's accessibility during a portion of the time period under consideration, we incorporate a one-week lag into the COT data for our primary results (but also show results for zero- and two-week lags as well).

Beyond COT data, we also require factor returns to administer the various statistical tests. We consider the following factors in our analysis:

- MRF: Market Minus the Risk-Free Rate (Fama and French (1993))
- SMB: Small Minus Big (Fama and French (1993))
- HML: High Minus Low (Fama and French (1993))
- UMD: Up Minus Down (Jegadeesh and Titman (1993))
- RMW: Robust Minus Weak Operating Profitability (Fama and French (2015))
- CMA: Conservative Minus Aggressive Investment (Fama and French (2015))
- LTR: Long-Term Reversal (Bondt and Thaler (1985))
- STR: Short-Term Reversal (Lehmann (1990), Jegadeesh (1990))
- BAB: Betting Against Beta (Frazzini and Pedersen (2014))
- QMJ: Quality Minus Junk (Asness et al. (2017b))

We obtain daily factor returns for BAB and QMJ from AQR's Data Library (AQR (2017)), and obtain all other daily factor returns from Professor Kenneth R. French's Data Library (French (2017)) .

[1] When Friday is a holiday, the report is released early the following week.

III. Smart Money Indicator

The SMI is the additive combination of three components, where each component measures "institutional-versus-individual" relative sentiment (IIRS). The three components include:

1. **Equities IIRS:** Relative sentiment in the S&P 500 Stock Index[2]
2. **Long-Duration IIRS:** Relative sentiment in the (30-Year) U.S. Treasury Bond
3. **Yield-Curve IIRS:** The *difference* between the relative sentiment in the 10-Year U.S. Treasury Note and the relative sentiment in the U.S. Treasury Bond

The first component is a direct measure of IIRS in equities. The second component is a direct measure of IIRS in long-duration bonds, and an indirect—and inverse—measure of IIRS in equities. It functions as an inverse measure because of the general negative correlation between equities and long-duration bonds. That is, if institutions are relatively bearish long-duration bonds, it may suggest institutions are relatively bullish equities.

The third component also functions as an indirect measure of IIRS in equities. The idea being if institutions are relatively longer shorter-duration assets, that might imply institutions are expecting longer-duration assets to fare relatively poorly. Given the general negative correlation between long-duration and equities, then, this might correspondingly imply institutions are expecting equities to fare relatively well. Hence, if institutions are more relatively bullish the 10-Year Note than the U.S. Treasury Bond, that too might imply relative institutional bullishness on equities.

The idea that investors' positioning in assets other than the S&P 500 (aka, "cross-hedging pressure") might affect the return of the S&P 500 itself is not unorthodox. de Roon et al. (2000) show that "both the futures own hedging pressure and cross-hedging pressures from related markets are important determinants of futures risk premia." Indeed, they find strong evidence of cross-hedging pressure between the S&P 500 Index and the U.S. Treasury bond.

We quantify IIRS using composite z-scores, i.e., weighted combinations of the respective z-scores of the institutions' and individuals' standalone net positioning in the applicable futures contracts. That is, if we let $\Delta_t^{i,j}$ represent the net percent of open interest in futures contract i for trader-class j at time t, and if we let $z(\Delta_t^{i,j}, N)$ represent the N-period z-score of $\Delta_t^{i,j}$, then the IIRS in futures contract i at time t is defined as:

[2] We use the large S&P 500 Stock Index contract rather than the S&P 500 Index E-Mini contract, which has less history. However, we also present results for a value-weighted combination of the two.

$$\hat{z}^{i}_{t,N} \equiv \frac{1}{2}\left(z(\Delta^{i,cm}_{t}, N) - z(\Delta^{i,nr}_{t}, N)\right) \tag{1}$$

where the superscripts cm and nr stand for "Commercials" (institutions) and "Non-reportables" (individuals), respectively.

The composite z-score is defined in such a way that it will be large in magnitude when institutions and individuals have strongly divergent positioning (from their respective averages), and small in magnitude when either their relative positioning is similar or when it does not deviate much from their respective averages.

As past studies have shown (e.g., Brown and Cliff (2005)), the time horizon over which one would expect institutions to outperform individuals is on the order of weeks and months. The composite z-score defined in Eq.(1), however, is prone to fluctuate from week to week. To slow these fluctuations and thereby prevent frequent reversals in the directional positions ultimately prescribed by the SMI, we instead look at the trailing M-period maxima or minima of the composite z-score. We denote these M-period extrema by $\hat{z}^{i,max}_{t,N,M}$ and $\hat{z}^{i,min}_{t,N,M}$, respectively.

We then define the three-component SMI composite z-score as:

$$\hat{z}^{\mathrm{SMI}}_{t,N,M} \equiv \hat{z}^{SP,max}_{t,N,M} - \hat{z}^{US,min}_{t,N,M} + (\hat{z}^{TY,max}_{t,N,M} - \hat{z}^{US,max}_{t,N,M}) \tag{2}$$

where the superscripts SP, US, and TY represent the S&P 500 Stock Index, the U.S. Treasury Bond, and the 10-Year U.S. Note, respectively.

The first component is the M-period maximum relative sentiment in the S&P 500. The higher (lower) this value, the more relatively bullish (bearish) institutions are on equities (or have been in the recent past).

The second component is the M-period minimum relative sentiment in the U.S. Treasury Bond. Our conjecture is that the lower (higher) this value, the more relatively bullish (bearish) institutions are on equities—due to the general inverse relationship between equities and long-duration bonds. Hence, we subtract this quantity in Eq. (2).

The last component, in parentheses, is the difference between the M-period maximum relative sentiment in the 10-Year Note and the M-period maximum relative sentiment in the U.S. Treasury Bond. Our conjecture here is that the more relatively bullish institutions are on the 10-Year Note compared to the U.S. Treasury Bond, the more relatively bullish they are on equities.

Finally, we define the SMI, parametrized by N and M, at time t as:

$$\mathrm{SMI}_{t,N,M} \;\equiv\; \hat{z}^{\mathrm{SMI}}_{t,N,M} \;-\; \underset{1\,\leq\, s\,\leq\, t}{\mathrm{median}}\left\{\,\hat{z}^{\mathrm{SMI}}_{s,N,M}\,\right\} \tag{3}$$

That is, we subtract off the cumulative median of the SMI composite z-score[3] and test whether the sign of the SMI is predictive.

As defined, the SMI depends on two parameters—the length of the z-score lookback, N, and the length of the extrema lookback, M. To test the robustness of the SMI's timing ability, we consider a range of values for both N and M. For N, we use the "round" values of 39, 52, 65, 78, 91, and 104 weeks (corresponding to 3, 4, 5, 6, 7, and 8 quarters, respectively). For M, we use all values ranging from 1 to 13 weeks—and we use the same value of M for each SMI component. That is, we make no effort to find either an optimal N value or an optimal combination of M values.

IV. Notation

With 13 M and 6 N values, there are 78 different pairwise SMI parameter combinations, generating 78 different SMI time-series. Given our focus on the SMI's sign, each SMI time-series generates two sets of conditional returns and two factor timing[4] strategies. These conditional returns and factor timing strategies serve as inputs to the statistical tests discussed in the following section.

For ease of exposition in subsequent sections, let $R(F)_t$ represent the time-series of weekly returns of factor F (time-stamped on the last business day of each week). Then, let $R(F, \mathrm{SMI}^+)_t$ (alternately, $R(F, \mathrm{SMI}^-)_t$) represent the conditional factor returns when the prior-week SMI is positive (negative). Further, let $S(F, \mathrm{SMI}^+)_t$ (alternately, $S(F, \mathrm{SMI}^-)_t$) represent the weekly returns of the "long-or-flat" factor timing strategy that goes long factor F when the prior-week SMI is positive (negative) and is flat otherwise.

In sections VI and VII, we condition the SMI's factor timing on the state of the factor's time-series momentum (TSM). Thus, we augment the notations above to indicate the state of time-series momentum as well. Let $R(F, \mathrm{SMI}^*, \mathrm{TSM}^*)_t$ represent the weekly factor returns conditioned on the state of both the prior-week SMI and the prior-week TSM—where the asterisks take on the relevant polarities describing the states of the respective

[3]We also present results using a rolling 3-year median and the cumulative 40*th* percentile.

[4]Throughout this section and the next, we use the term "factor timing" to include market timing as well.

quantities. Likewise, let $S(F, \text{SMI}^*, \text{TSM}^*)_t$ represent the weekly returns of the long-or-flat factor timing strategy that goes long factor F when the prior-week SMI and the prior-week TSM have the prescribed polarities (and is flat otherwise).

V. Statistical Tests for Factor Timing

We use the following statistical tests to measure the SMI's factor timing significance:

1. Treynor-Mazuy test: (Treynor and Mazuy (1966))
2. Henriksson-Merton Parametric test: (Henriksson and Merton (1981))
3. Henriksson-Merton Nonparametric test: (Henriksson and Merton (1981))
4. Granger-Pesaran test: (Granger and Pesaran (1999))
5. Cumby-Modest test: (Cumby and Modest (1987))
6. Markov Regression test: (Chu et al. (2009))
7. Jiang test: (Jiang (2003))
8. Welch t-test: (Welch (1947))
9. Bayesian t-test: (Bååth (2014))

Each test above takes in either the conditional returns or the long-or-flat strategy returns generated by an SMI time-series, and produces a p-value of factor timing significance (for that SMI time-series). By convention, we orient the tests to measure the timing significance of going long the factor when the SMI is positive, and take the resulting p-value to be the probability that one would observe a test statistic *less* than the one observed. Thus, p-values near 1 indicate positive factor timing skill (i.e., the factor does relatively better when the prior-week SMI is positive), while p-values near 0 indicate negative factor timing skill (i.e., the factor does relatively better when the prior-week SMI is negative).

For both the Treynor-Mazuy test and Henriksson-Merton Parametric test, in addition to specifying the tests using only the factor under consideration (the default specification), we also augment the tests using different asset pricing models, find the best fitting model (based on the Aikaike Information Criterion), and use it to determine the test's p-value (for each SMI parameter combination). The models we consider include the Capital Asset Pricing Model (Sharpe (1964)), the Fama-French 3-factor model (Fama and French (1993)), the Carhart 4-factor model (Carhart (1997)), and the Fama-French 5-factor model (Fama and French (2015)). To account for heteroskedaticity and autocorrelation, we use the technique of Andrews (1991). Further, we use three different variants of the Jiang test, each one accounting for autocorrelation of returns in a different way. Thus, for each SMI parameter combination, the 8 tests produce 11 separate p-values. (See the appendix for more implementation details.)

We find the median of those 11 p-values at each of the 78 SMI parameter combinations, resulting in a 13×6 matrix of median p-values. This matrix thus conveys how broadly significant each individual SMI parameter combination is, as well as how broadly significant the SMI is overall.

To measure the SMI's economic significance, we compute the difference in compounded annualized returns between $R(F, \text{SMI}^+)_t$ and $R(F, \text{SMI}^-)_t$ for each SMI parameter combination. We also compute the analogous differences in annualized volatilities.

In addition to the tests above, we also employ a test that controls for data snooping. The algorithm we use is the Multiple Hypothesis Testing (MHT) algorithm (Romano and Wolf (2005)). In short, this algorithm looks at the long/short strategies, LS_i, given by:

$$LS_i = \ln(1 + S_i(F, \text{SMI}^+)_t) - \ln(1 + S_i(F, \text{SMI}^-)_t), \quad i = 1, \ldots, 78, \tag{4}$$

and generates bootstrap samples (with replacement) from them. It then compares test statistics from the actual LS_i to the best test statistics across all bootstrap samples—resulting in a more stringent test of statistical significance.

The MHT algorithm takes as inputs an integer, k, representing a number of allowable false positives, and a probability, α, such that $\mathbb{P}\{\text{number of false positives} \geq k\} \leq \alpha$. We set $k = 1$ and consider α values of 0.10, 0.05, 0.01, 0.001. Romano et al. (2008) state that typical values for α are 0.10 and 0.05. However, de Prado and Lewis (2018) show that "unless a strategy has a true annualized Sharpe ratio above 1 over a period of more than 10 years of daily data," values of α below 0.15 are likely to be overly restrictive (i.e., likely to generate too many false negatives)[5]. In this context, our α values should be considered sufficiently, if not excessively, conservative.

We present two sets of MHT results. One set corresponds to using the raw bootstrap samples generated by the MHT algorithm. The other set corresponds to using bootstrap samples for which we constrain the potential oversampling of time-localized outliers (such as those produced during the Financial Crisis)[6].

[5]With 22 years of weekly data, we have roughly the equivalent of 5 years of daily data. If the SMI long-or-flat strategies had a true Sharpe ratio of 1, the optimal α based on the formula in de Prado and Lewis (2018) would be approximately 0.30. Further, if we were to use the market's Sharpe ratio as the true Sharpe ratio and if all the long-or-flat SMI strategies were essentially the same strategy due to high correlations (the worst-case scenario), the optimal α would still be approximately 0.20.

[6]We limit the number of plus or minus 3-standard-deviation-or-more outliers in each bootstrap sample to

(Note: The weekly COT data starts on October 6, 1992. With a maximum N value of 104 weeks, a maximum M value of 13 weeks, and a 52-week burn-in period to compute the expanding-window median, the earliest date by which all 78 SMI time-series have computable values is December 22, 1995. This results in just over 22 years of usable history up until December 29, 2017—a total of 1150 weekly data points.)

VI. Timing the Market Factor

To get a sense of the SMI's ability to time the market factor, MRF, we first plot in Figure 1 the cumulative returns of $S(MRF, \text{SMI}^+)_t$ versus $S(MRF, \text{SMI}^-)_t$ (where the former goes long MRF when the prior-week SMI is positive and is flat otherwise, and the latter goes long MRF when the prior-week SMI is negative and is flat otherwise). Here we use a one-week lag of spliced COT data with the parameters $M = 5$ and $N = 78$ (which correspond to the median-performing SMI parameter combination among a sizable contingent of the more significant ones).

Over the 22 year period beginning on December 22, 1995 and ending on December 29, 2017, \$1 invested in $S(MRF, \text{SMI}^+)_t$ would have grown to \$8.20 (net of the risk-free rate), while \$1 invested in $S(MRF, \text{SMI}^-)_t$ would have fallen to \$0.52 (i.e., one would have lost 48% net of the risk-free rate)—an outperformance multiple of 16. As a frame of reference, it took more than 62 years (from July 1, 1926 until November 11, 1988) before a total return index tracking HML's top quintile (value) was first worth 16 times the corresponding index tracking HML's bottom quintile (growth).

The substantial bifurcation in market returns exhibited in Figure 1 is not unique to the selected parameter combination nor is it the largest bifurcation. To demonstrate the SMI's robustness, we consider 16 different computational variants. These variants mostly differ based on the type of COT data used, the length (in weeks) of the SMI lag, and the portion of the dataset analyzed. But we also consider other variations, such as incorporating the S&P 500 E-mini futures contract into the SMI's construction, and alternately subtracting the rolling 3-year median and the cumulative 40^{th} percentile in Eq. (3). We also examine the timing ability of the three SMI components separately and in groups of two. Table I lists the specifications of each of the 16 cases.

Case 1 is our baseline and Tables III–V present its results in their entirety: median p-values and differences in annualized returns and volatilities across all SMI parameter

be no more than the number that occur in the raw factor return time-series, $R(F)_t$.

Case	COT Data	SMI Lag (Weeks)	Period Covered	Other Specification
1	Spliced	1	Full	
2	Spliced	0	Full	
3	Spliced	2	Full	
4	Futures-Only	1	Full	
5	Spliced	1	Full	Includes E-mini
6	Spliced	1	Full	3-year median
7	Spliced	1	Full	40th percentile
8	Spliced	1	Full	Equities IIRS
9	Spliced	1	Full	Long-Duration (LD) IIRS
10	Spliced	1	Full	Yield-Curve (YC) IIRS
11	Spliced	1	Full	Equities & LD IIRS
12	Spliced	1	Full	Equities & YC IIRS
13	Spliced	1	Full	LD & YC IIRS
14	Spliced	1	1st Half	
15	Futures-and-Options	1	2nd Half	
16	Futures-and-Options	1	2nd Half	Adds Ultra T-Bond

Table I. Computational specifications for the various SMI cases

combinations, along with results from the MHT algorithm. We are especially interested in the number of parameter combinations exhibiting high median statistical significance, the number exhibiting large positive (negative) differences in their annualized returns (volatilities), and the number deemed significant by the MHT algorithm as a function of α. Table II summarizes these aggregate numbers for each of the 16 computational cases.

From the summaries in Table II, one can see that each SMI case that analyzes the full dataset (i.e., Cases 1–7) has a solid, and often very large, contingent of parameter combinations exhibiting high statistical and economic significance. The only full-sample cases that have conspicuously less broad-based significance are the ones that investigate the individual components of the SMI separately and in groups of two (i.e., Cases 8–13). These results seem to suggest that the more-thorough accounting of relevant cross-hedging pressures found in the three-component SMI is the source of its superior predictive ability.

For the baseline, Case 1, 72 of the 78 SMI parameter combinations have median p-values greater than 0.975, 64 have median p-values greater than 0.99, and 38 have median p-values greater than 0.999. Economically, 75 parameter combinations have differences in their annualized returns greater than 10 percentage points, 62 have differences greater than 15 percentage points, and 49 have differences greater than 20 percentage points (which is also the average difference across all parameter combinations). We also see the higher returns when the SMI is positive come with lower volatility—contrary to what modern portfolio theory would predict.

When controlling for data snooping the statistical significance persists: 65 parameter combinations are deemed significant by the raw MHT algorithm with $\alpha = 0.05$, 50 retain significance with $\alpha = 0.01$, and 19 remain significant at the *unusually* stringent level of $\alpha = 0.001$—which strongly suggests that the SMI's timing results are not the product of data mining. We observe similar levels of significance and robustness for the other full-sample SMI variants (Cases 2–7).

Cases 14 and 15 cover the first and second halves of the time period, respectively. As a result, they each have half as many data points, and thus less power to quantify extreme significance. Nevertheless, they both exhibit a substantial number of parameter combinations with high significance. Case 14 has 55 combinations with median p-values greater than 0.99, 62 with annualized return differences greater than 20 percentage points, and 57 deemed significant by the outlier-adjusted MHT with $\alpha = 0.05$. The corresponding numbers for Case 15 are 15, 30, and 22.

The relatively fewer parameter combinations with exceptionally high significance in Case 15 is primarily because this case encompasses the Great Recession. During this time, certain SMI parameter combinations stayed positive well after the market top in late 2007, and—after eventually turning negative—did not turn positive again until several weeks after the initial surge off the bottom.

This slow-footedness of certain SMI parameter combinations was largely confined to ones having both higher N and M values. As the lengths of the z-score lookback, N, and extrema lookback, M, increase, the SMI becomes less responsive to weekly changes in the COT data. A cluster of these "slower-moving" SMI parameter combinations were particularly vulnerable to incorrect forecasts during the 2007–2009 period[7].

However, another development that affected the second half of the dataset was the 2010 launch of Ultra U.S. Treasury Bond futures, which have even longer duration than standard U.S. Treasury Bond futures. Ultra Treasury Bond futures therefore absorb some of the long-duration positions that would otherwise be allotted to the U.S. Treasury Bond contract. When we factor these additional long-duration positions into the analysis—Case 16—we find improved performance for the second half of the dataset—with 20 parameter cases having median p-values greater than 0.99, 39 having annualized return differences greater than 20%, and 27 registering as significant on the outlier-adjusted MHT algorithm at the $\alpha = 0.05$ level (compared to 15, 30, and 22 before). As before, we see the importance of more-thoroughly accounting for the relevant cross-hedging pressures when constructing

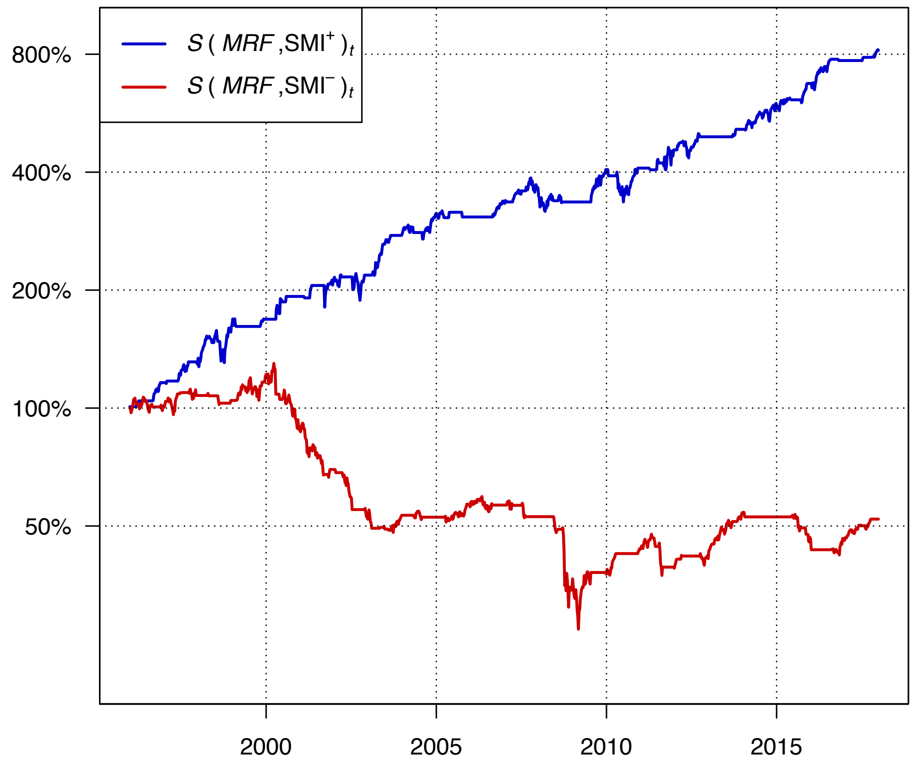
[7]The most highly significant SMI parameter combinations across cases tend to hold their long-or-flat positions an average of 16 weeks, while the slower-moving parameter combinations tend to hold their positions an average of 21 weeks.

the SMI.

When we take into consideration the totality of the foregoing results, the SMI appears to have exhibited highly significant and exceptionally robust market timing ability. In essence, the polarity of the SMI seems to be capturing:

1. the deep and persistent forecasting advantages institutions have over individuals
2. the general time horizons over which build-ups in (relative) sentiment unwind

Additionally, because institutions expend significant resources analyzing the ongoing state of corporate and economic fundamentals, how institutions (in the aggregate and at the margins) hedge their equity and bond portfolios in the futures markets ultimately represents their consensus opinion on the state of those fundamentals. Thus, the SMI might best be thought of as an implied fundamental indicator. This connection with fundamentals coupled with the SMI's behavioral explanations (outlined in Section I) may help contextualize its rather extreme economic and statistical significance.



Source: Columbus Macro, LLC

Figure 1. SMI Timing of MRF: Cumulative returns of $S(MRF, \mathrm{SMI}^{+})_t$ vs. $S(MRF, \mathrm{SMI}^{-})_t$ $(M = 5,\ \ N = 78)$

	Number of SMI Parameter Combinations with the Specified Characteristics												
	Median p-Values $\geq$			Differences in Annualized Returns $\geq$			Volatility $\leq$	Passes MHT at (k, α) Level: Unadjusted			Outlier-Adjusted		
Case	.975	.990	.999	10%	15%	20%	0	(1,.05)	(1,.01)	(1,.001)	(1,.05)	(1,.01)	(1,.001)
1	72	64	38	75	62	49	78	65	50	19	65	54	21
2	74	67	34	77	66	39	77	66	39	9	66	39	10
3	73	65	28	75	63	36	78	59	32	7	60	38	8
4	68	58	25	74	56	33	78	51	26	1	51	28	1
5	60	46	21	71	51	30	76	66	40	14	64	45	24
6	74	66	29	75	61	30	77	58	27	5	61	34	6
7	74	70	37	78	71	41	78	74	57	32	76	62	28
8	37	13	0	51	15	0	78	5	0	0	9	0	0
9	9	3	0	21	5	0	77	8	1	0	8	0	0
10	9	1	0	39	12	0	78	13	0	0	13	1	0
11	36	13	0	51	15	0	78	5	0	0	9	0	0
12	36	14	0	51	15	0	78	5	0	0	9	0	0
13	9	3	0	21	5	0	77	8	1	0	8	0	0
14	65	55	9	72	68	62	49	57	22	2	57	22	3
15	29	15	0	65	57	30	78	15	1	0	22	7	0
16	33	20	1	65	54	39	78	22	4	0	27	7	1

Table II. SMI Timing of MRF: Summary Results, Cases 1–16

	N					
M	39	52	65	78	91	104
1	0.9535	0.9449	0.8835	0.9384	0.9664	0.9417
2	0.9859	0.9951	0.9996	0.9974	0.9982	0.9993
3	0.9962	0.9998	0.9990	0.9992	0.9999	0.9999
4	0.9998	0.9999	0.9992	0.9981	0.9954	0.9988
5	0.9996	1.0000	1.0000	1.0000	0.9982	0.9987
6	0.9980	1.0000	1.0000	0.9995	0.9929	0.9962
7	0.9994	1.0000	0.9998	0.9999	0.9992	0.9991
8	0.9993	0.9999	0.9987	0.9988	0.9994	0.9994
9	0.9998	0.9999	0.9994	0.9972	0.9966	0.9994
10	0.9995	0.9986	0.9975	0.9854	0.9900	0.9853
11	0.9994	0.9990	0.9992	0.9977	0.9959	0.9888
12	0.9989	0.9992	0.9989	0.9946	0.9800	0.9811
13	0.9999	0.9993	0.9992	0.9976	0.9832	0.9762

Table III. SMI Timing of *MRF*, Case 1: The median of 11 timing test p-values, evaluated at each SMI parameter combination

	N					
M	39	52	65	78	91	104
1	9.6%	8.7%	7.8%	10.2%	13.2%	12.1%
2	13.4%	13.9%	20.6%	19.5%	18.1%	21.1%
3	16.4%	21.8%	19.5%	22.7%	23.1%	26.6%
4	22.2%	27.4%	22.8%	20.1%	17.4%	20.7%
5	24.6%	28.5%	27.2%	27.2%	19.4%	20.8%
6	23.0%	28.9%	28.1%	22.7%	16.9%	18.9%
7	26.1%	33.7%	27.6%	28.6%	22.6%	22.5%
8	24.7%	28.4%	22.7%	21.5%	20.9%	20.8%
9	27.7%	28.7%	24.9%	17.0%	17.2%	21.0%
10	24.6%	23.8%	22.2%	13.1%	15.0%	13.4%
11	24.1%	21.9%	23.5%	20.6%	16.9%	14.0%
12	22.7%	23.7%	23.5%	18.6%	12.0%	12.2%
13	25.4%	20.4%	21.9%	17.2%	11.3%	10.7%

(a) Differences in annualized returns

	N					
M	39	52	65	78	91	104
1	-0.9%	-0.8%	-1.2%	-2.4%	-2.0%	-2.4%
2	0.0%	-0.6%	-0.8%	-2.3%	-1.3%	-3.2%
3	-0.7%	-0.4%	-1.1%	-2.1%	-1.4%	-2.3%
4	-0.7%	-1.4%	-1.7%	-1.6%	-2.8%	-3.4%
5	-2.1%	-2.5%	-2.5%	-2.6%	-2.9%	-3.7%
6	-3.4%	-2.7%	-2.5%	-3.2%	-3.2%	-3.9%
7	-3.6%	-3.9%	-3.4%	-3.7%	-3.8%	-4.5%
8	-3.9%	-4.6%	-4.0%	-3.4%	-3.8%	-4.1%
9	-4.4%	-4.7%	-4.4%	-2.3%	-2.1%	-2.4%
10	-4.5%	-4.8%	-4.7%	-2.5%	-2.7%	-2.9%
11	-4.5%	-5.0%	-4.7%	-5.1%	-2.8%	-3.0%
12	-4.5%	-4.5%	-4.8%	-5.4%	-3.1%	-3.4%
13	-3.3%	-2.7%	-3.7%	-5.2%	-3.2%	-3.2%

(b) Differences in annualized volatilities

Table IV. SMI Timing of *MRF*, Case 1: Differences in annualized returns and volatilities between $R(MRF, \text{SMI}^+)_t$ and $R(MRF, \text{SMI}^-)_t$, evaluated at each SMI parameter combination

	α			
k	0.001	0.01	0.05	0.10
1	19	50	65	74

(a) MHT Results: Raw bootstrap

	α			
k	0.001	0.01	0.05	0.10
1	21	54	65	74

(b) MHT Results: Outlier-adjusted bootstrap

Table V. SMI Timing of *MRF*, Case 1: The number of SMI parameter combinations deemed significant by the Multiple Hypothesis Testing (MHT) algorithm as a function of k and α—using both raw bootstrap samples (left) and outlier-adjusted bootstrap samples (right)

VII. Timing Time-Series Momentum

Given the strength of the SMI's timing results in the previous section, it is natural to ask whether the SMI adds any incremental information beyond that provided by the state of the market's time-series momentum. When we investigate that question, however, what we find instead is that in the presence of the SMI[8], it is time-series momentum that adds virtually no incremental predictive power for the U.S. equity market.

To classify the state of time-series momentum, we use two separate indicators—an intermediate-term total return and a longer-term moving average. We say the market is in a state of positive time-series momentum when both momentum indicators are positive. Likewise, we say the market is in a state of negative time-series momentum when both indicators are negative. (These conventions ignore the state where one indicator is positive and the other negative, which encompasses about 13% of the weekly data points. However, this state can be lumped in with either of the other two states without materially affecting the results. For simplicity, then, we consider only the wholly-positive and wholly-negative cases.)

To test robustness, we consider three different time horizons for the intermediate-term total return (namely, 63, 84, and 105 days—approximately 3, 4, and 5 months, respectively) and three different lengths for the longer-term moving average (namely, 150, 200, and 250 days). This gives us 9 different pairwise parameter combinations to define the state of time-series momentum[9] (TSM).

To get a visual sense of the SMI's ability to time the market conditional on the state of time-series momentum, we begin by plotting in Figure 2 the cumulative returns of $S(MRF, \text{SMI}^+, \text{TSM}^+)_t$ versus $S(MRF, \text{SMI}^-, \text{TSM}^+)_t$. We plot the analogous returns for the state of negative time-series momentum in Figure 3.

These cumulative returns are composites—averaged across all 9 TSM parameter combinations. For both figures, we use the same SMI parameters used in Figure 1, namely $M = 5$ and $N = 78$. From the figures, we see that regardless of the state of time-series momentum, excess market returns are substantially higher when the SMI is positive.

For the state of positive momentum, \$1 invested in $S(MRF, \text{SMI}^+, \text{TSM}^+)_t$ would have

[8] The SMI we use here and in the remainder of the paper corresponds to Case 1 of Section V, i.e., one-week lag, spliced COT data.

[9] We lag these daily momentum indicators by one day to avoid look-ahead bias, and we extract the last-business-day-of-the-week values to align with the weekly factor returns and weekly SMI time-series.

appreciated to \$2.94, while \$1 invested in $S(MRF, \text{SMI}^{-}, \text{TSM}^{+})_t$ would have grown to only \$1.29. For the state of negative momentum, \$1 invested in $S(MRF, \text{SMI}^{+}, \text{TSM}^{-})_t$ would have risen to \$2.25, while \$1 invested in $S(MRF, \text{SMI}^{-}, \text{TSM}^{-})_t$ would have fallen to \$0.48—i.e., one would have lost 52%. (This latter result highlights the potential danger of being long the market when both TSM and the SMI are negative.)

Tables VI and VII present the summary SMI timing results for the states of positive and negative time-series momentum, respectively. (Note: In the tables, the notation in the "TSM Combo" column indicates which elements of the parameter vectors of the momentum indicators are being used in the analysis, along with whether we are conditioning on those indicators being positive or negative. For example, "32+" would indicate that the first momentum indicator (the intermediate total return) was computed using the third element of its parameter vector (105 days), while the second momentum indicator (the longer-term moving average) was computed using the second element of its parameter vector (namely, 200 days), and that we conditioned on both of those indicators being positive.)

Given the reduced number of data points for each momentum state relative to the full sample (positive momentum states are 63% of the sample, negative momentum states are 24% of the sample), *ceteris paribus*, we should expect to see less extreme statistical significance than in the previous section. This is indeed the case.

The SMI's ability to time the market conditional on the market having positive momentum, while respectable, is not as strong as its ability to time the market unconditionally—primarily because when the SMI is negative and TSM is positive, market returns are still relatively flat rather than negative. There is, however, a pocket of SMI parameter combinations exhibiting high median p-values, double digit differences in annualized returns, and negative differences in annualized volatilities. When we control for data snooping, about a dozen SMI parameter combinations, on average (per TSM parameter combination), register as significant at the (now even more restrictive due to the reduced sample size) $\alpha = 0.05$ level.

For the state of negative time-series momentum, the SMI timing results are considerably stronger. Here we see several dozen SMI parameter combinations with median p-values greater than 0.99, as well as several dozen that pass the outlier-adjusted MHT algorithm at the $\alpha = 0.01$ level—for each TSM parameter combination.

What is perhaps most noteworthy, however, is the SMI's economic significance when the market's time-series momentum is negative. The difference in compounded annualized market returns between $R(MRF, \text{SMI}^{+}, \text{TSM}^{-})_t$ and $R(MRF, \text{SMI}^{-}, \text{TSM}^{-})_t$, averaged over all ($78 \times 9 = 702$) SMI and TSM parameter combinations is just shy of 50 percentage points. When momentum is negative but the SMI is positive, the market's average annualized return is 30%. When both momentum and the SMI are negative, the market's average annualized return is approximately -20%. Further, the positive returns of the former have much lower volatility than the negative returns of the latter.

These results suggest the SMI is able to identify the times when it is advantageous to be long the market during states of negative momentum. In turn, this ability helps address one of the primary weaknesses of time-series momentum as a TAA tool—namely, its tendency to maintain reduced (or zero) equity allocations too long, well after the market has bottomed (while waiting for momentum indicators to turn positive). Thus, by monitoring the SMI when momentum is negative, momentum-based investors might be able to participate more fully in the often-substantive rebounds that tend to occur right after significant market troughs.

Using regression analysis, one can show more formally that momentum indicators have little or no predictive power in the presence of the SMI. We ran regressions of weekly excess market returns against (a combination of) the SMI and the two momentum indicators used above. We ran these regressions in three different forms (using both continuous and binary variables) across the 702 different SMI–TSM parameter cases and averaged the resulting p-values and t-stats of the SMI and momentum coefficients across all 2106 regressions. The SMI had an average t-stat of 2.83 with an average p-value 0.026. The momentum indicators had average t-stats of 0.56 (intermediate-term total return) and 0.18 (longer-term moving average), with corresponding p-values of 0.447 and 0.576, respectively. Thus, in the presence of the SMI, time-series momentum appears to add virtually no incremental information.

For completeness, Tables VIII–XIII present the full SMI timing results conditional on the state of time-series momentum. The results in each table are averaged across all 9 TSM parameter combinations.

TSM Combo	Median p-Values $\geq$ .975	.990	.999	Differences in Annualized Returns $\geq$ 10%	15%	20%	Volatility $\leq$ 0	Passes MHT at (k, α) Level, Unadjusted (1,.05)	(1,.01)	(1,.001)	Outlier-Adjusted (1,.05)	(1,.01)	(1,.001)
11+	39	30	7	66	35	15	65	32	12	6	33	11	9
12+	27	11	0	41	15	4	64	9	3	1	9	3	0
13+	18	7	0	34	11	2	64	9	2	0	9	1	0
21+	29	12	2	52	20	5	64	23	9	2	17	8	1
22+	21	7	1	42	11	1	63	14	2	1	10	2	1
23+	18	4	1	41	10	1	64	5	1	0	4	1	0
31+	30	14	4	49	17	7	64	20	9	2	18	9	1
32+	28	11	1	42	14	2	64	14	3	1	14	3	1
33+	16	4	0	38	9	1	64	4	1	0	4	1	0

Number of SMI Parameter Combinations with the Specified Characteristics

Table VI. SMI Timing of *MRF* Conditioned on Positive TSM: Summary results for all TSM parameter combinations

TSM Combo	Median p-Values $\geq$ .975	.990	.999	Differences in Annualized Returns $\geq$ 10%	15%	20%	Volatility $\leq$ 0	Passes MHT at (k, α) Level, Unadjusted (1,.05)	(1,.01)	(1,.001)	Outlier-Adjusted (1,.05)	(1,.01)	(1,.001)
11-	63	53	8	78	78	77	77	56	29	4	58	41	13
12-	46	25	2	78	76	71	75	51	15	1	53	26	8
13-	44	24	1	77	74	68	71	53	15	1	56	27	3
21-	58	42	7	78	77	76	71	54	20	6	55	38	8
22-	55	36	6	78	77	74	76	55	28	5	56	41	3
23-	55	38	7	78	76	73	77	55	25	6	57	44	11
31-	58	40	12	78	78	77	73	52	20	5	55	33	7
32-	59	46	8	78	78	77	76	56	29	1	59	47	6
33-	53	31	6	78	77	74	78	52	21	0	56	39	9

Number of SMI Parameter Combinations with the Specified Characteristics

Table VII. SMI Timing of *MRF* Conditioned on Negative TSM: Summary results for all TSM parameter combinations

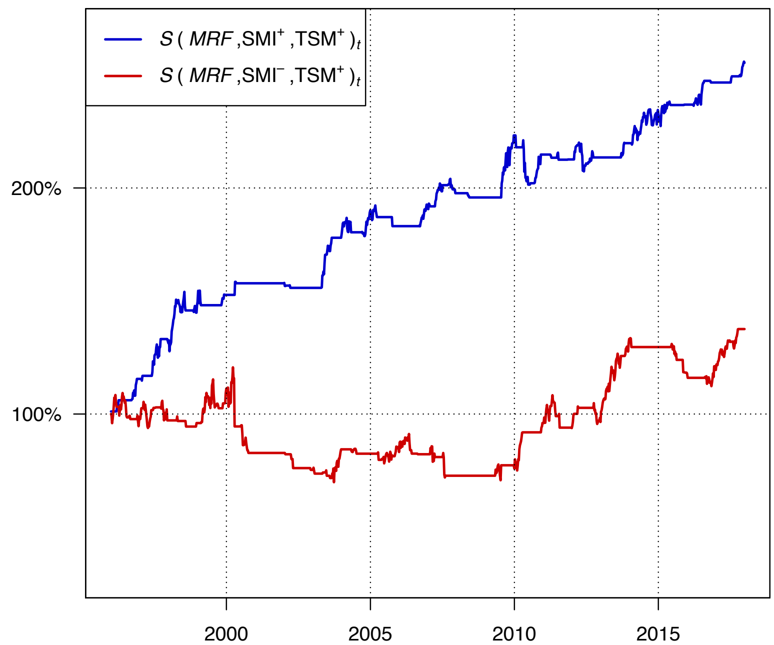


Figure 2. SMI Timing of *MRF* Conditioned on Positive TSM: Cumulative returns of $S(MRF, \mathrm{SMI}^{+}, \mathrm{TSM}^{+})_t$ vs. $S(MRF, \mathrm{SMI}^{-}, \mathrm{TSM}^{+})_t$ $(M = 5,\ \ N = 78$, averaged over all TSM parameter combinations)

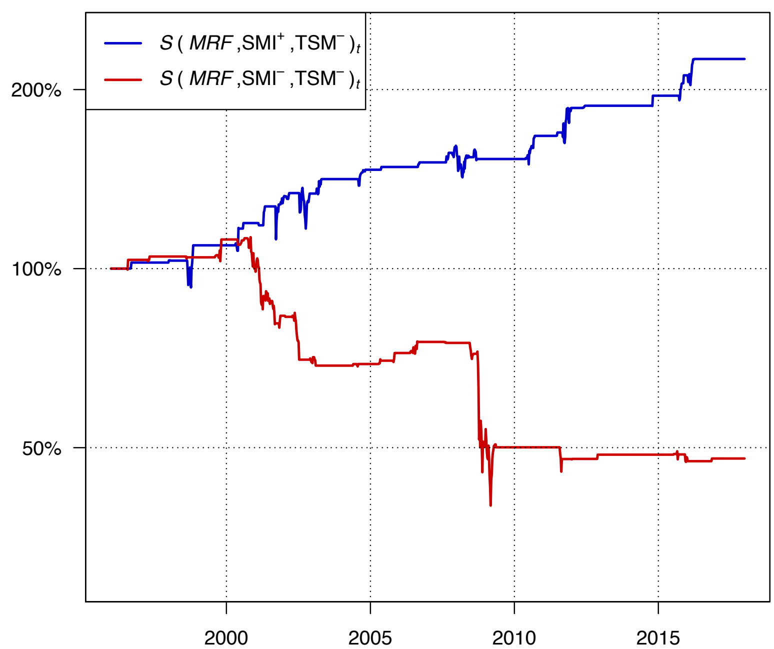


Figure 3. SMI Timing of *MRF* Conditioned on Negative TSM: Cumulative returns of $S(MRF, \mathrm{SMI}^{+}, \mathrm{TSM}^{-})_t$ vs. $S(MRF, \mathrm{SMI}^{-}, \mathrm{TSM}^{-})_t$ $(M = 5,\ \ N = 78$, averaged over all TSM parameter combinations)

	N					
M	39	52	65	78	91	104
1	0.9129	0.7789	0.6202	0.8434	0.9368	0.8848
2	0.9464	0.9493	0.9713	0.8723	0.8513	0.9616
3	0.9687	0.9937	0.9706	0.9117	0.9748	0.9840
4	0.9825	0.9882	0.9584	0.8561	0.7373	0.8877
5	0.9777	0.9674	0.9842	0.9794	0.8528	0.8508
6	0.9694	0.9711	0.9785	0.9451	0.8385	0.8569
7	0.9430	0.9962	0.9575	0.9587	0.9110	0.9352
8	0.9723	0.9923	0.9169	0.8933	0.9300	0.9796
9	0.9947	0.9979	0.9837	0.9696	0.9481	0.9803
10	0.9933	0.9839	0.9295	0.8233	0.8615	0.8620
11	0.9896	0.9876	0.9640	0.9130	0.8735	0.8751
12	0.9822	0.9811	0.9429	0.8309	0.7926	0.7465
13	0.9991	0.9768	0.9481	0.8372	0.7881	0.7316

Table VIII. SMI Timing of *MRF* Conditioned on Positive TSM: Median of 11 timing test p-values, evaluated at each SMI parameter combination (averaged over all TSM parameter combinations)

	N					
M	39	52	65	78	91	104
1	8.1%	3.3%	1.2%	7.7%	11.9%	10.2%
2	9.4%	8.8%	10.3%	8.1%	3.5%	12.7%
3	11.7%	14.3%	10.5%	10.2%	9.9%	14.5%
4	12.2%	16.1%	12.4%	8.0%	4.0%	9.2%
5	14.8%	13.9%	14.8%	14.3%	7.4%	7.4%
6	13.9%	14.0%	15.5%	10.4%	7.1%	7.7%
7	12.1%	18.8%	13.2%	11.2%	8.9%	10.2%
8	14.4%	17.9%	11.1%	8.5%	10.1%	14.3%
9	20.0%	21.8%	15.6%	11.8%	10.9%	14.1%
10	18.6%	16.8%	10.7%	6.2%	7.5%	7.8%
11	17.9%	17.3%	13.3%	8.8%	8.3%	7.4%
12	15.5%	14.5%	11.4%	5.5%	6.7%	6.1%
13	23.7%	11.4%	7.4%	4.7%	5.7%	4.3%

(a) Differences in annualized returns

	N					
M	39	52	65	78	91	104
1	0.6%	0.8%	1.0%	-1.2%	-0.9%	-1.2%
2	1.9%	1.3%	1.5%	-1.3%	1.0%	-1.8%
3	1.3%	1.3%	1.2%	-0.9%	0.7%	-1.2%
4	1.5%	-0.7%	-0.7%	-0.4%	-1.2%	-1.6%
5	-1.0%	-1.1%	-1.5%	-1.4%	-1.5%	-1.7%
6	-1.4%	-1.4%	-1.4%	-1.9%	-1.8%	-2.2%
7	-1.5%	-1.4%	-1.2%	-1.5%	-1.6%	-2.2%
8	-1.9%	-1.9%	-1.5%	-1.9%	-1.9%	-2.2%
9	-2.7%	-2.3%	-1.4%	-1.8%	-2.0%	-2.3%
10	-2.6%	-2.6%	-1.4%	-1.9%	-2.2%	-2.6%
11	-2.7%	-2.7%	-1.3%	-2.0%	-2.2%	-2.3%
12	-2.3%	-1.7%	-1.1%	-1.9%	-2.7%	-2.8%
13	-2.1%	0.6%	1.2%	-1.3%	-2.2%	-2.1%

(b) Differences in annualized volatilities

Table IX. SMI Timing of *MRF* Conditioned on Positive TSM: Differences in annualized returns and volatilities between $R(MRF, \text{SMI}^+, \text{TSM}^+)_t$ and $R(MRF, \text{SMI}^-, \text{TSM}^+)_t$ (averaged over all TSM parameter combinations)

	α			
k	0.001	0.01	0.05	0.10
1	1	5	14	24

(a) MHT Results: Raw data

	α			
k	0.001	0.01	0.05	0.10
1	1	4	13	22

(b) MHT Results: Outlier-adjusted data

Table X. SMI Timing of *MRF* Conditioned on Positive TSM: Number of SMI parameter combinations deemed significant by the Multiple Hypothesis Testing (MHT) algorithm as a function of k and α—using both raw bootstrap samples (left) and outlier-adjusted bootstrap samples (right) (averaged over all TSM parameter combinations)

M	N: 39	52	65	78	91	104
1	0.7334	0.8752	0.8564	0.8596	0.7856	0.7859
2	0.8667	0.9180	0.9930	0.9958	0.9966	0.9829
3	0.9445	0.9843	0.9823	0.9957	0.9971	0.9978
4	0.9962	0.9990	0.9956	0.9931	0.9870	0.9836
5	0.9950	0.9994	0.9982	0.9976	0.9827	0.9907
6	0.9923	0.9996	0.9977	0.9939	0.9652	0.9760
7	0.9991	0.9995	0.9963	0.9992	0.9901	0.9895
8	0.9946	0.9945	0.9846	0.9869	0.9739	0.9119
9	0.9919	0.9909	0.9863	0.9560	0.9444	0.9611
10	0.9846	0.9844	0.9911	0.9438	0.9513	0.9198
11	0.9840	0.9717	0.9939	0.9914	0.9836	0.9623
12	0.9857	0.9906	0.9933	0.9881	0.9611	0.9776
13	0.9850	0.9877	0.9949	0.9903	0.9488	0.9594

Table XI. SMI Timing of *MRF* Conditioned on Negative TSM: Median of 11 timing test p-values, evaluated at each SMI parameter combination (averaged over all TSM parameter combinations)

M	N: 39	52	65	78	91	104
1	16.1%	25.4%	24.0%	24.6%	18.7%	19.0%
2	27.3%	31.3%	50.0%	54.2%	55.3%	44.7%
3	34.9%	45.4%	45.2%	55.0%	56.4%	58.0%
4	58.2%	67.5%	55.0%	51.5%	45.6%	43.6%
5	57.2%	70.8%	61.3%	59.5%	43.0%	48.6%
6	59.5%	74.6%	60.2%	55.1%	36.9%	40.7%
7	74.8%	82.2%	62.4%	73.3%	51.9%	51.9%
8	63.2%	64.2%	51.6%	49.1%	41.8%	30.0%
9	61.8%	60.8%	54.0%	28.2%	27.6%	30.4%
10	52.8%	53.7%	58.3%	25.0%	27.6%	20.3%
11	51.6%	47.1%	61.8%	56.6%	36.7%	29.8%
12	51.7%	58.6%	61.9%	54.8%	27.1%	30.9%
13	46.7%	50.3%	65.4%	52.9%	23.8%	26.3%

(a) Differences in annualized returns

M	N: 39	52	65	78	91	104
1	-5.1%	-4.7%	-6.4%	-6.6%	-6.7%	-6.8%
2	-4.9%	-5.0%	-5.9%	-6.3%	-6.1%	-7.2%
3	-5.3%	-4.9%	-5.9%	-6.2%	-5.8%	-5.8%
4	-5.0%	-4.2%	-5.0%	-5.1%	-6.9%	-7.0%
5	-5.1%	-5.4%	-4.9%	-5.8%	-6.6%	-7.2%
6	-7.4%	-4.7%	-4.4%	-6.4%	-6.1%	-6.9%
7	-7.3%	-6.6%	-6.2%	-7.7%	-7.9%	-7.9%
8	-6.9%	-7.2%	-6.6%	-5.3%	-6.8%	-6.6%
9	-6.7%	-6.6%	-6.5%	-0.7%	-0.9%	-0.7%
10	-6.8%	-6.4%	-7.0%	0.0%	-0.6%	-0.5%
11	-6.6%	-5.9%	-6.4%	-6.2%	-0.3%	-0.7%
12	-6.7%	-5.8%	-6.8%	-6.8%	0.4%	-0.7%
13	-3.1%	-3.5%	-6.3%	-7.5%	0.2%	-1.0%

(b) Differences in annualized volatilities

Table XII. SMI Timing of *MRF* Conditioned on Negative TSM: Differences in annualized returns and volatilities between $R(MRF, \mathrm{SMI}^+, \mathrm{TSM}^-)_t$ and $R(MRF, \mathrm{SMI}^-, \mathrm{TSM}^-)_t$ (averaged over all TSM parameter combinations)

k	α: 0.001	0.01	0.05	0.10
1	3	22	54	59

(a) MHT Results: Raw data

k	α: 0.001	0.01	0.05	0.10
1	8	37	56	62

(b) MHT Results: Outlier-adjusted data

Table XIII. SMI Timing of *MRF* Conditioned on Negative TSM: Number of SMI parameter combinations deemed significant by the Multiple Hypothesis Testing (MHT) algorithm as a function of k and α—using both raw bootstrap samples (left) and outlier-adjusted bootstrap samples (right) (averaged over all TSM parameter combinations)

VIII. Timing Smart Beta Factors

We now examine the SMI's ability to time smart beta equity factors, where our focus is on the nine factors (other than MRF) listed in Section II. In general, we observe three types of factors:

Group 1. Those that do better when the SMI is positive (SMB, STR)
Group 2. Those that do better when the SMI is negative (HML, RMW, CMA, LTR, BAB, QMJ)
Group 3. Those that exhibit no evidence of SMI timability (UMD)

To get a sense of the economic significance of SMI-based factor timing, we plot in Figures 4–7 the cumulative returns of $S(F, \mathrm{SMI}^+)_t$ versus $S(F, \mathrm{SMI}^-)_t$ for a selection of Group 2 factors. (For each factor, we use the SMI parameter combination with the *median* median p-value from among a selection of the more highly significant parameter combinations.) As is evident from the figures, the polarity of the SMI seems to largely determine the factor return outcomes.

Tables XIV and XV summarize the factor timing and data snooping results for Groups 1 and 2, respectively. Factor symbols followed by the "+" sign indicate the SMI's timing was conditioned on the factor having positive time-series momentum (with the results averaged across all TSM parameter combinations). (We will refer to such cases as "conditional" factor timing.)

Notably, the SMI's timing significance for Group 1 factors is virtually nonexistent when timing using the SMI only and only weakly and narrowly significant when the Group 1 factors have positive momentum. In contrast, the results for Group 2 are much stronger and broader. Here, the SMI exhibits moderately-strong statistical significance in the unconditional case, and strong-to-exceptionally-strong statistical significance when the factors have positive time-series momentum.

For example, when timing HML unconditionally, there are 6 SMI parameter combinations with p-values less than 0.01, none with (absolute) differences in annualized returns greater than 15 percentage points, and only 3 deemed significant by the outlier-adjusted MHT with $\alpha = 0.01$. When HML has positive momentum, however, 33 SMI parameter combinations have timing significance at the 0.01 level, 27 have (absolute) differences in annualized returns greater than 15 percentage points, and 25 register as significant on the outlier-adjusted MHT with $\alpha = 0.01$. We see similar improvements moving from the unconditional to the conditional case for RMW (which particularly stands out in this regard), CMA, LTR, and QMJ.

The Group 2 results appear to agree with observed market behavior. As Group 2 factors are based on fundamentals or proxies for fundamentals, their short portfolios tend to consist of expensive, low quality "glamour" stocks. When the SMI is positive and the broad market is generally rising, investors tend to engage in more speculative behavior. During such times, glamour stocks tend to do relatively well, likely causing long/short factor performance to suffer.

In contrast, when the SMI is negative and the overall market is generally stagnant or falling, investors tend to seek refuge in higher quality stocks. This flight to quality—and away from glamour—would appear to have positive repercussions for long/short factor performance.

	Number of SMI Parameter Combinations with the Specified Characteristics												
				Differences in Annualized				Passes MHT at (k,α) Level					
	Median p-Values $\geq$			Returns $\geq$			Volatility $\leq$	Unadjusted			Outlier-Adjusted		
Case	.975	.990	.999	10%	15%	20%	0	(1,.05)	(1,.01)	(1,.001)	(1,.05)	(1,.01)	(1,.001)
SMB	0	0	0	0	0	0	78	0	0	0	0	0	0
SMB+	1	0	0	10	2	0	77	4	1	0	4	0	0
STR	0	0	0	13	0	0	78	0	0	0	0	0	0
STR+	0	0	0	12	0	0	78	1	0	0	5	0	0

Table XIV. Smart Beta Factors, Group 1: Summary Results

	Number of SMI Parameter Combinations with the Specified Characteristics												
				Differences in Annualized				Passes MHT at (k,α) Level					
	Median p-Values $\leq$			Returns $\leq$			Volatility $\leq$	Unadjusted			Outlier-Adjusted		
Case	.025	.010	.001	-10%	-15%	-20%	0	(1,.05)	(1,.01)	(1,.001)	(1,.05)	(1,.01)	(1,.001)
HML	13	6	0	5	0	0	78	18	2	0	27	3	0
HML+	50	33	0	60	27	0	78	47	13	0	53	25	4
RMW	6	1	0	3	0	0	74	10	1	0	19	1	0
RMW+	70	59	16	75	45	1	78	56	23	2	74	54	23
CMA	22	2	0	0	0	0	74	13	0	0	14	0	0
CMA+	46	25	0	60	5	0	72	43	12	0	47	15	1
LTR	15	5	0	1	0	0	78	3	0	0	4	0	0
LTR+	14	9	0	25	8	0	78	9	1	0	10	2	0
BAB	20	7	1	25	3	0	75	22	5	1	23	5	1
BAB+	30	18	1	40	13	0	59	10	1	0	15	4	0
QMJ	18	11	0	12	0	0	72	16	4	0	18	4	0
QMJ+	64	38	2	76	41	7	78	36	12	0	45	19	2

Table XV. Smart Beta Factors, Group 2: Summary Results

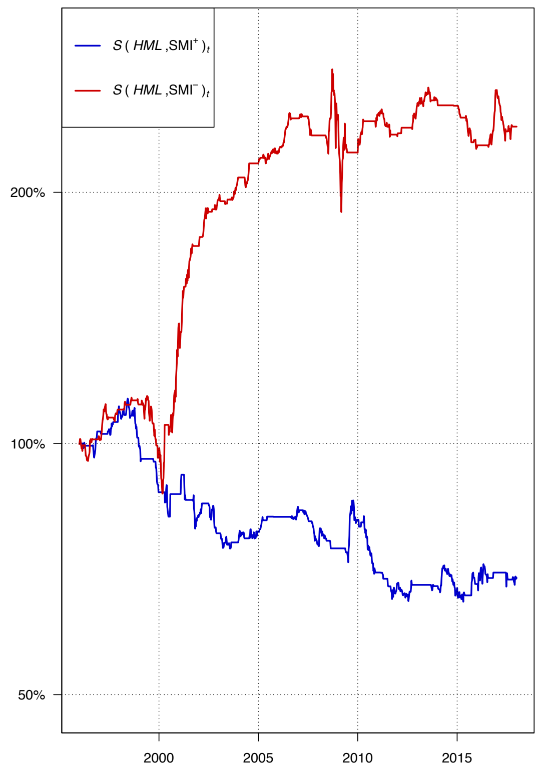


Figure 4. Cumulative $S(HML, \mathrm{SMI}^+)_t$ vs. $S(HML, \mathrm{SMI}^-)_t$ $(M = 4, \quad N = 65)$

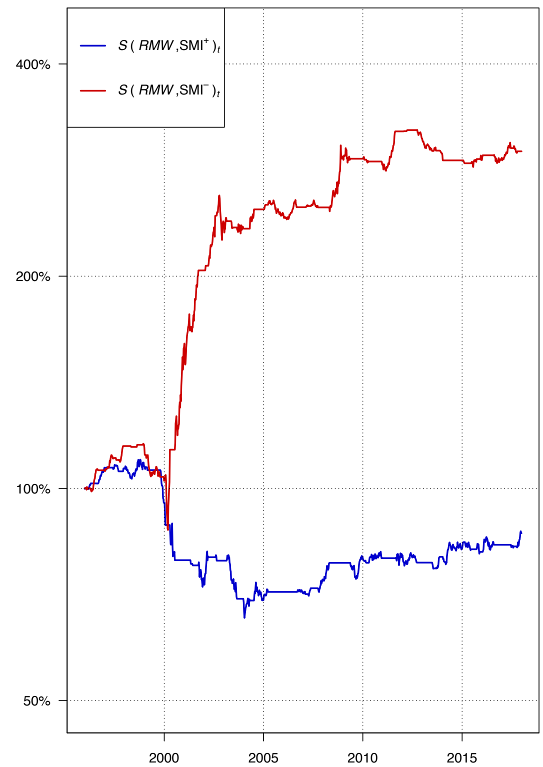


Figure 5. Cumulative $S(RMW, \mathrm{SMI}^+)_t$ vs. $S(RMW, \mathrm{SMI}^-)_t$ $(M = 7, \quad N = 52)$

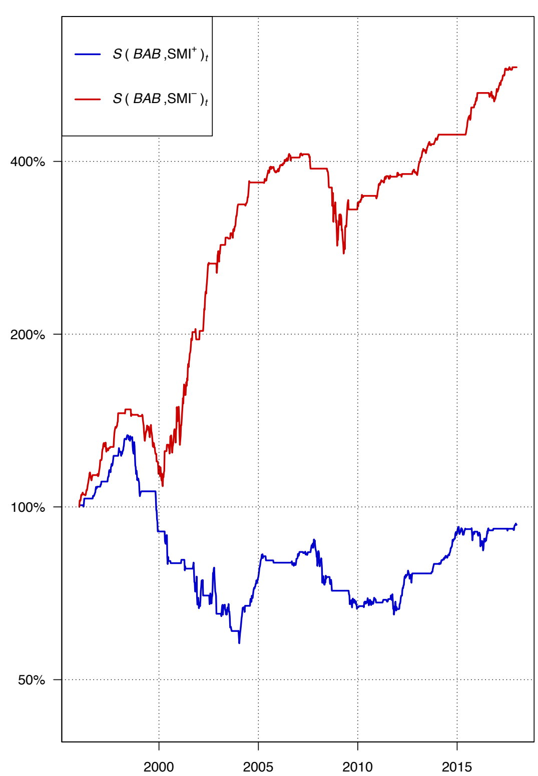


Figure 6. Cumulative $S(BAB, \mathrm{SMI}^+)_t$ vs. $S(BAB, \mathrm{SMI}^-)_t$ $(M = 5, \quad N = 78)$

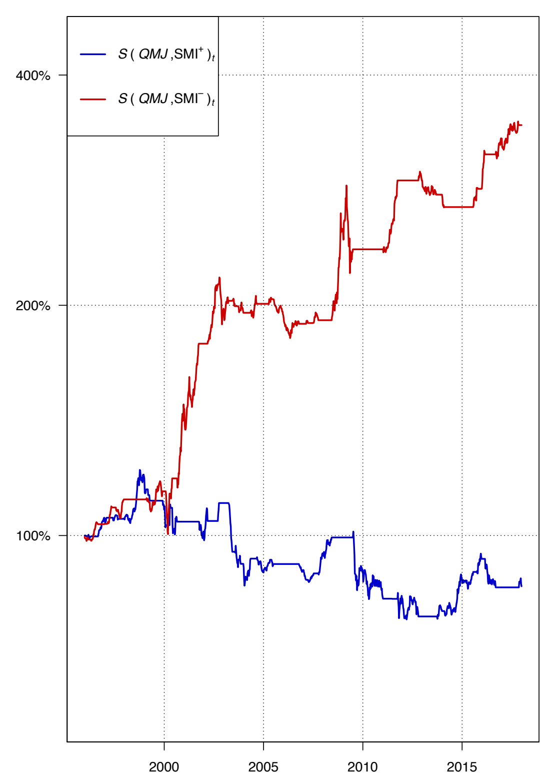


Figure 7. Cumulative $S(QMJ, \mathrm{SMI}^+)_t$ vs. $S(QMJ, \mathrm{SMI}^-)_t$ $(M = 13, \quad N = 65)$

IX. Tactical Asset Allocation

In this section, we attempt to simulate a TAA strategy involving the SMI that could have been exploited in real time. To do this, we begin at the common starting date of the 78 SMI time-series and walk forward, assessing at every point in time the statistical significance of each of the 78 SMI-generated long-or-flat market timing strategies, $S_i(MRF, \text{SMI}^+)_t$. For the remainder of this section, we will refer to these timing strategies as substrategies, to avoid confusion with the overall strategy we are developing.

To measure the timing significance of the SMI substrategies, we use only statistical tests known at the time of the first assessment (December 22, 1995). Based on dates of publication, those tests include the Treynor-Mazuy, Henriksson-Merton Parametric, Henriksson-Merton Nonparametric, Cumby-Modest, and Welch t tests.

We then walk forward, week by week, finding the median p-value of those five tests for each SMI substrategy. If the median p-value for any substrategy is greater than a miniminally acceptable p-value, we consider that substrategy "eligible" for trading. If a substrategy is eligible and its corresponding SMI is positive, we consider that substrategy "triggered."

If the ratio of triggered substrategies to eligible substrategies (weighted by their respective p-values) exceeds a "critical-mass" threshold, we invest 100% in (broad-market U.S.) equities. If not, we invest 100% in bonds. We determine this critical-mass threshold dynamically, by again walking forward (using a two-year burn-in period) from the beginning of the SMI dataset and finding—at each point in time—the threshold that would have produced the highest cumulative Sharpe ratio as of the prior week[10]. We then use this threshold to make our investment decision for the current week. We repeat this procedure weekly.

We consider three different p-value thresholds—0.95, 0.975, and 0.99—resulting in three separate equity allocation time-series, where each time-series is invested either 0% or 100% in equities. We average these three time-series to get a composite equity allocation. If the composite equity allocation is less than 100%, we allocate the remaining capacity to bonds.

We compare the SMI-based strategy to the following benchmark strategies:

1. **60/40**: A passive portfolio consisting of 60% U.S. equities and 40% U.S. bonds, rebal-

[10]If there is a tie for the best performance between discretized thresholds, we take the maximum threshold. However, the results are robust to breaking the tie using the minimum threshold, as well as to breaking the tie randomly.

anced quarterly

2. **SMA10M**: The 10-month moving average strategy (Faber (2007)), which is 100% in equities, if equities are above their 10-month moving average, or 100% in bonds otherwise.
3. **SMA10M+TR4M**: A variant of the 10-month moving average strategy that looks at both the 10-month moving average and the 4-month total return (Keller and van Putten (2012)). If both are positive, the strategy is 100% in equities. If only one is positive, the strategy is 50% in equities and 50% in bonds. Otherwise the strategy is 100% in bonds.
4. **CAPE+TREND**: The value-and-momentum strategy presented in Asness et al. (2015). This strategy combines the Cyclically-Adjusted Price to Earnings (CAPE) ratio[11] (Campbell and Shiller (1998)) with a 12-month trend indicator. See Asness et al. (2015) for details. (Note: unlike the other strategies, whose equity allocations range between 0%–100%, the equity allocations in this strategy range between 50%–150%. Whenever the equity allocation is less than 100%, we invest the difference in bonds.) We also present results for CAPE and TREND standalone.

Our backtests incorporate the following assumptions, procedures, and data:

1. For each strategy, we average the performance across three different execution variants. For example, for the SMI strategy, one variant executes on Thursday, one executes on Friday, and one executes on the following Monday. (Recall that the information necessary to make the allocation decisions for the SMI strategy is available on the previous Friday due to our lagging of the data.) We do this for the sake of robustness and to better simulate real-world trading procedures.
2. For the month-end and quarter-end strategies (i.e., SMA10M, SMA10M+TR4M, CAPE+TREND, 60/40), one variant executes on the second-to-last day of the month, one on the last day of the month, and one on the first day of the following month. The allocation decisions for these variants are made using data from the day prior to execution.(That is, the variant that executes on the first day of the following month uses data from the last day of the previous month, and so on.)
3. All strategies assume entry at closing prices on the day of execution plus four basis points of (detrimental) slippage.
4. For the CAPE strategy, we lag the Shiller earnings by one quarter to account for the lag in reported earnings.
5. All trades are charged a $4.95 commission and all active strategies incorporate a

[11] In actuality, it inverts this ratio and looks at the Cyclically-Adjusted Earnings to Price ratio.

15-basis-point annual management fee.

6. We source daily equity market returns and the daily risk-free rate (for Sharpe ratio calculations) from Professor Kenneth R. French's Data Library and use daily bond returns derived from the Bloomberg Barclays U.S. Aggregate Bond Total Return Index. (Correspondingly, all performance metrics are computed using daily returns.)

Figure 8 shows the equity curves of each strategy, starting from the point in time where the SMI strategy becomes operational. Table XVI presents the performance metrics, from which we see the SMI strategy has both a higher absolute return and a higher risk-adjusted return (i.e., Sharpe ratio) than any of the benchmarks. The SMI strategy also has the highest capture ratio and a maximum drawdown on par with the smallest. It achieves this while having a time-averaged equity exposure of only 54%. In contrast, the five active benchmarks have time-averaged equity exposures between 70% and 95%.

It thus appears the SMI is both more powerful and more efficient than either value or momentum (or their combination) in identifying higher-yielding equity market opportunities. This power and efficiency, however, do come at the expense of higher average monthly turnover[12].

Figures 9 and 10 show the comparative performance starting both one-third and two-thirds of the way through the 20-year backtest period. Tables XVII and XVIII report the corresponding performance metrics. The results over both subperiods are consistent with the full period results. That is, the SMI strategy has the highest (absolute and risk-adjusted) returns, the highest capture ratios, and generally lower drawdowns than the alternatives—while still having much lower time-averaged equity exposure than its active counterparts.

Notably, from 2011–2017, a time period over which the U.S. equity market rose strongly—thus making it potentially detrimental *not* to have been fully invested in equities—the SMI strategy outpaced its nearest competitor, TREND, by approximately 75 basis points per year while having a time-averaged equity exposure of only 61% compared to 101% for TREND. Moreover, the maximum drawdown of the SMI strategy over this time period, -10.7%, was the lowest drawdown of all the strategies tested, and half the magnitude of the TREND strategy's -21.2% drawdown.

[12]The turnover for the SMI strategy decreased as time went on, likely due to the stabilization of the dynamically-determined parameters as more data points became available.

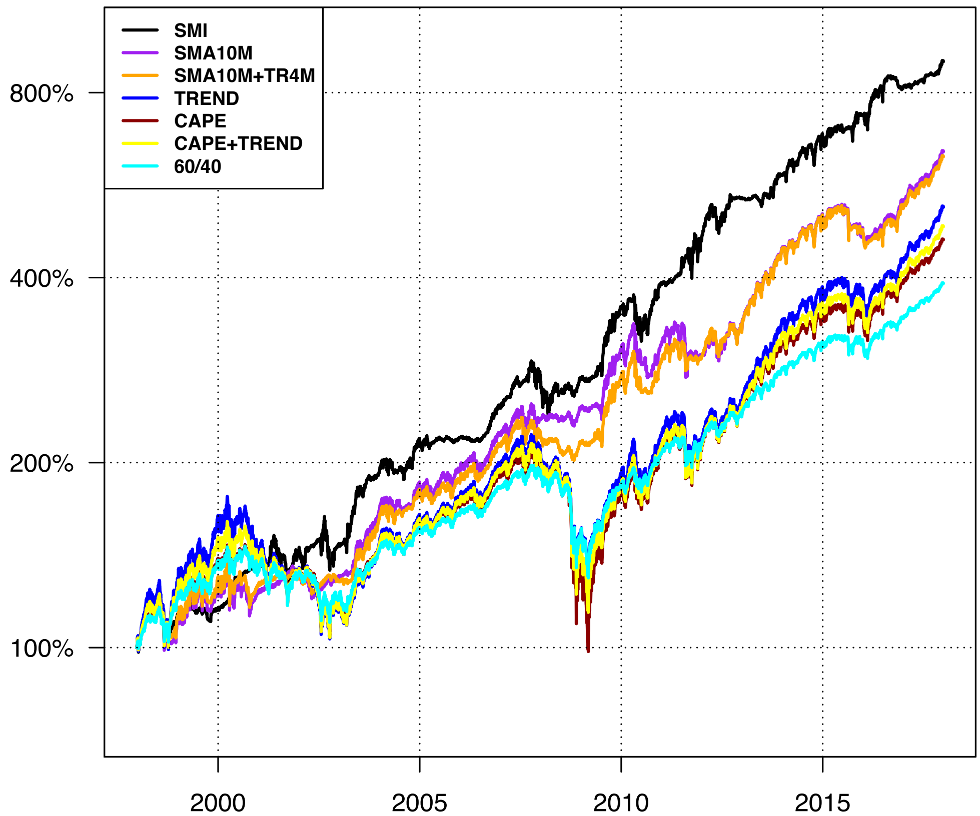


Figure 8. TAA Strategies: Comparative performance (start date: December 26, 1997)

	Performance Metrics					
Strategy	CAGR	Sharpe Ratio	Capture Ratio	Maximum Drawdown	Average Equity Allocation	Average Monthly Turnover
SMI	11.62%	0.81	1.11	-22.43%	53.88%	33.62%
SMA10M	9.75%	0.65	1.05	-20.97%	72.15%	12.48%
SMA10M+TR4M	9.65%	0.68	1.05	-18.42%	69.66%	14.56%
TREND	8.61%	0.46	0.98	-41.40%	95.73%	3.65%
CAPE	7.95%	0.44	0.98	-52.91%	78.85%	1.80%
CAPE+TREND	8.22%	0.45	0.98	-46.93%	87.30%	2.17%
60/40	7.07%	0.50	1.00	-34.42%	60.07%	0.79%

Table XVI. TAA Strategies: Performance Metrics (start date: December 26, 1997)

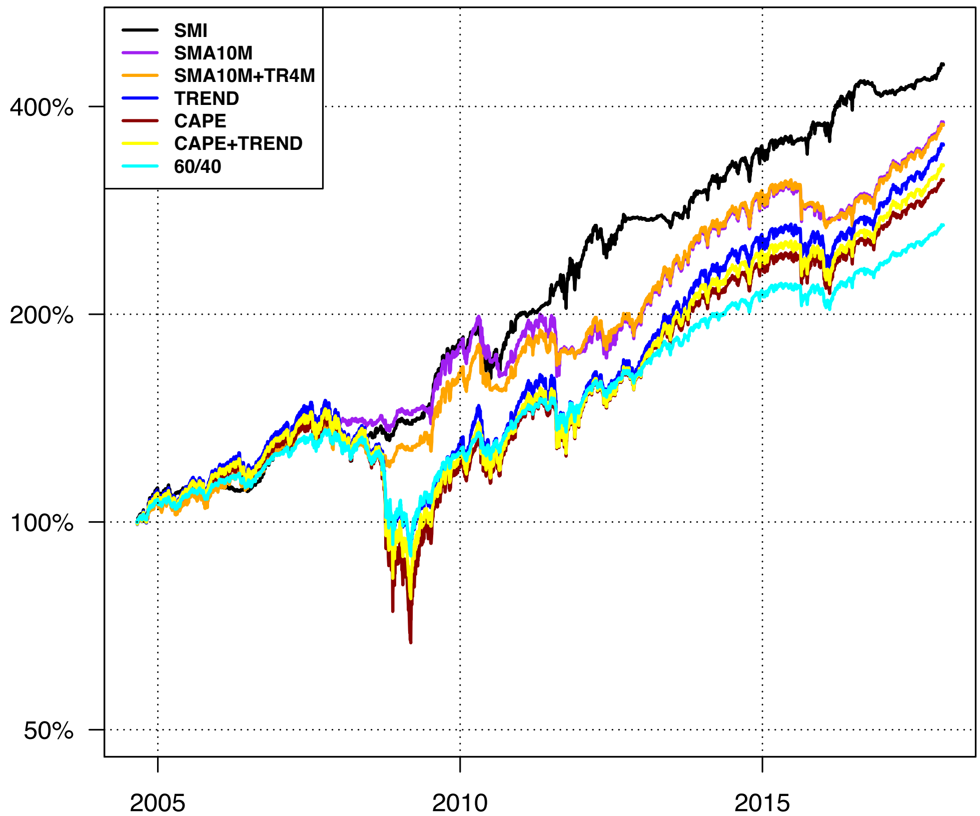


Figure 9. TAA Strategies: Comparative performance (start date: August 27, 2004)

Strategy	CAGR	Sharpe Ratio	Capture Ratio	Maximum Drawdown	Average Equity Allocation	Average Monthly Turnover
	Performance Metrics					
SMI	12.11%	0.94	1.11	-18.22%	55.99%	26.91%
SMA10M	10.51%	0.78	1.04	-18.77%	79.82%	10.15%
SMA10M+TR4M	10.44%	0.83	1.04	-14.99%	76.73%	12.31%
TREND	9.90%	0.60	0.98	-40.23%	96.75%	3.39%
CAPE	8.93%	0.51	0.97	-52.91%	85.09%	1.83%
CAPE+TREND	9.33%	0.56	0.97	-46.93%	90.92%	2.06%
60/40	7.71%	0.64	1.00	-34.42%	60.07%	0.70%

Table XVII. TAA Strategies: Performance Metrics (start date: August 27, 2004)

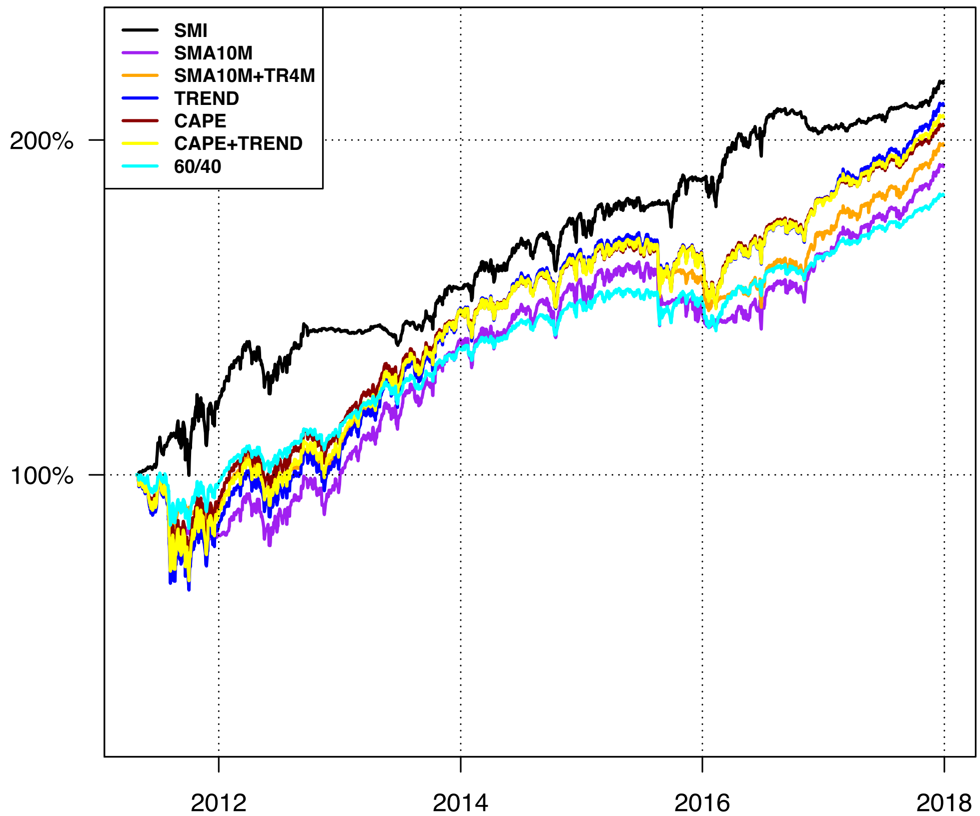


Figure 10. TAA Strategies: Comparative performance (start date: April 29, 2011)

	Performance Metrics					
Strategy	CAGR	Sharpe Ratio	Capture Ratio	Maximum Drawdown	Average Equity Allocation	Average Monthly Turnover
SMI	12.91%	1.11	1.07	-10.67%	61.12%	25.00%
SMA10M	10.04%	0.81	0.97	-18.80%	88.77%	9.00%
SMA10M+TR4M	10.76%	0.96	0.99	-14.85%	85.08%	10.54%
TREND	12.15%	0.85	0.96	-21.18%	100.58%	3.53%
CAPE	11.45%	0.91	0.97	-17.83%	85.37%	1.84%
CAPE+TREND	11.75%	0.87	0.97	-19.65%	92.97%	2.41%
60/40	9.06%	1.06	1.00	-10.84%	60.25%	0.89%

Table XVIII. TAA Strategies: Performance Metrics (start date: April 29, 2011)

X. Conclusion

Although the concept of market timing is often derided, studies have repeatedly shown institutional investors tend to outperform individual investors over intermediate time horizons. Thus, should one attempt to time the market, mimicking the actions of institutions seems to be a better bet than siding with individuals. Accordingly, we present a relative sentiment indicator—the SMI—that measures the aggregate positioning in equities of institutions relative to individuals. The key innovation of the SMI's construction is its incorporation of cross-hedging pressures from long-duration bonds and from points along the yield curve.

Subjecting the SMI to a battery of statistical tests, we find it possesses highly significant and exceptionally robust market timing ability. The significance is both economic and statistical: Annualized market returns are (on average) 20 percentage points higher (with lower volatility) when the SMI is positive compared to when it is negative; moreover, the SMI exhibits statistical significance beyond the 0.001 level for a large number of parameter combinations. When controlling for data snooping, these results persist at highly restrictive significance levels, and are robust to changes in the underlying data, construction methods, and time periods analyzed.

We further find the SMI to be substantially more powerful than time-series momentum in predicting equity market returns. That is, returns are significantly higher when the SMI is positive, regardless of the state of the market's time-series momentum. This effect is particularly striking when momentum is negative—where the state of negative momentum and positive SMI outperforms the state of negative momentum and negative SMI (on average) by 50 percentage points on an annualized basis. This characteristic of the SMI may help it address one of the weaknesses of time-series momentum as a tactical allocation tool—namely, momentum's tendency to maintain reduced (or zero) equity allocations well after the market has bottomed (while waiting for momentum indicators to turn positive).

Beyond the broad market, the SMI also helps time several smart beta equity factors—primarily ones based on fundamentals, where we find such factors tend to do much better when the SMI is negative (i.e., "risk-off" markets reward fundamentals). For several factors, the statistical and economic significance becomes quite strong when the factors also have positive time-series momentum.

We further develop a walk-forward TAA strategy based on the SMI and compare it to several active and passive benchmarks. Over the 20-year period analyzed (1997–2017), the SMI strategy produced higher (absolute and risk-adjusted) returns, with generally lower drawdowns, while spending much less time invested in equities than its value-or-momentum counterparts—suggesting the SMI may be both more powerful and more efficient as a tactical allocation tool than either value or momentum.

In short, the polarity of the SMI seems to be capturing the deep and persistent forecasting advantages that institutions have over individuals, and the time horizons over which those advantages play out. Hence, in the quest to take smart beta, following the Smart Money appears to be a worthwhile strategy.

Appendix A

A. Statistical Tests of Factor Timing

Here we describe the tests used to assess factor timing significance—from the perspective of “unconditional” factor timing (i.e., factor timing using the SMI only). For the case of “conditional” factor timing (i.e., conditional on the state of the factor’s time-series momentum), simply substitute the analagous returns and strategies defined in Section IV into the formulas below.

1. **Treynor-Mazuy test**: For our purposes, the TM test can be formulated as:

$$S(F, \text{SMI}^{+})_t \;=\; \beta_0 \;+\; \beta_1\, R(F)_t \;+\; \beta_2\, R^2(F)_t \;+\; \epsilon_t \tag{5}$$

   The term $\beta_2\, R^2(F)_t$ arises from TM’s argument that a fund manager who increases factor exposure when the factor is rising and decreases factor exposure when the factor is falling will have a convex characteristic line.

   Since the publication of the TM test, various other asset pricing models have been proposed. Among these are the Fama-French three-factor model (FF3, Fama and French (1993)), the Carhart four-factor model (C4, Carhart (1997)), and the Fama-French five-factor model (FF5, Fama and French (2015)). To allow for the possibility that one of these other pricing models better explains the SMI-generated returns, we fit the following five models for each SMI parameter combination:

$$S(F,\mathrm{SMI}^+)_t \;=\; \beta_0 \;+\; \beta_1\, R(F)_t \;+\; \beta_2\, R^2(F)_t \;+\; \epsilon_t \tag{6}$$

$$S(F,\mathrm{SMI}^+)_t \;=\; \beta_0 \;+\; \beta_1\, R(F)_t \;+\; \beta_2\, R^2(F)_t \;+\; \beta_3\, R(MRF)_t \;+\; \epsilon_t \tag{7}$$

$$\begin{aligned} S(F,\mathrm{SMI}^+)_t \;=\; & \beta_0 \;+\; \beta_1\, R(F)_t \;+\; \beta_2\, R^2(F)_t \;+\; \beta_3\, R(MRF)_t \;+\; \beta_4\, R(SMB)_t \\ & +\; \beta_5\, R(HML)_t \;+\; \epsilon_t \end{aligned} \tag{8}$$

$$\begin{aligned} S(F,\mathrm{SMI}^+)_t \;=\; & \beta_0 \;+\; \beta_1\, R(F)_t \;+\; \beta_2\, R^2(F)_t \;+\; \beta_3\, R(MRF)_t \;+\; \beta_4\, R(SMB)_t \\ & +\; \beta_5\, R(HML)_t \;+\; \beta_6\, R(UMD)_t \;+\; \epsilon_t \end{aligned} \tag{9}$$

$$\begin{aligned} S(F,\mathrm{SMI}^+)_t \;=\; & \beta_0 \;+\; \beta_1\, R(F)_t \;+\; \beta_2\, R^2(F)_t \;+\; \beta_3\, R(MRF)_t \;+\; \beta_4\, R(SMB)_t \\ & +\; \beta_5\, R(HML)_t \;+\; \beta_6\, R(RMW)_t \;+\; \beta_7\, R(CMA)_t \;+\; \epsilon_t \end{aligned} \tag{10}$$

The equations above correspond to TM tests using the factor only, the factor plus the CAPM, the factor plus the FF3, the factor plus the C4, and the factor plus the FF5, respectively. If the factor, F, itself is part of the CAPM, FF3, C4, or FF5, we adjust the equations accordingly.

From these models, we select the one with the lowest Aikaike Information Criterion (AIC) and take the p-value of the β_2 coefficient as the measure of timing significance (Treynor and Mazuy (1966), Henriksson and Merton (1981)). We use robust regression to estimate the models, and the technique of Andrews (1991) to compute heteroscedasticity-and-autocorrelation-consistent (HAC) standard errors.

2. **Henriksson-Merton Parametric test**: The HM-P test is virtually identical to the TM test described above. It simply uses a different term—one meant to proxy the payoff to a protective put option strategy—to identify the convexity of portfolio returns:

$$S(F,\mathrm{SMI}^+)_t \;=\; \beta_0 \;+\; \beta_1\, R(F)_t \;+\; \beta_2\, \max\{0,-R(F)_t\} \;+\; \epsilon_t \tag{11}$$

We use the same estimation, selection, and evaluation procedures described above (in the case of the TM test) to determine the p-value of timing significance for the HM-P test.

3. **Henriksson-Merton Nonparametric test**: Unlike the TM and HM-P tests, which seek to measure the convexity of returns, the HM-NP test focuses on the frequency of correct predictions. It requires knowledge of the factor's directional forecasts

(which we have, courtesy of whether the SMI is positive or negative) and assumes those forecasts are independent of return magnitudes. The HM-NP test looks at the following four sets of returns to construct the test statistic:

$$\text{Forecast positive, return positive:} \quad \{R(F)_t \mid \text{SMI}_{t-1} > 0 \;\; \& \;\; R(F)_t > 0\} \tag{12}$$

$$\text{Forecast positive, return negative:} \quad \{R(F)_t \mid \text{SMI}_{t-1} > 0 \;\; \& \;\; R(F)_t < 0\} \tag{13}$$

$$\text{Forecast negative, return positive:} \quad \{R(F)_t \mid \text{SMI}_{t-1} \leq 0 \;\; \& \;\; R(F)_t > 0\} \tag{14}$$

$$\text{Forecast negative, return negative:} \quad \{R(F)_t \mid \text{SMI}_{t-1} \leq 0 \;\; \& \;\; R(F)_t < 0\} \tag{15}$$

As the HM-NP test seeks to quantify factor timing skill based on the frequency of successful forecasts, the test statistic ends up taking the form of a binomial random variable, which converges to a standard normal variable as the number of trials increases.

4. **Granger-Pesaran test**: The Granger-Pesaran (GP) test is similar to the HM-NP test in that it measures the frequency of correct predictions, requires knowledge of the directional forecasts, and does not take into account the magnitude of returns. In particular, GP test looks at the probability of forecasting a bad outcome relative to the probability of a bad outcome. It uses the same four sets of returns as the HM-NP test and its test statistic likewise converges to a standard normal variable.

5. **Cumby-Modest test**: The Cumby-Modest (CM) test is a generalization of the HM-NP test. It is formulated as:

$$S(F, \text{SMI}^+)_t \; = \; \beta_0 \; + \; \beta_1 \, I_{\{\text{SMI}_{t-1} > 0\}} \; + \; \epsilon_t \tag{16}$$

The CM test posits that if skillful factor timing is present, the β_1 coefficient must be positive and significant, otherwise the indicator function is not adding any incremental information beyond that provided by the mean of the factor returns, β_0.

6. **Markov Regression test**: The Markov Regression (MR) test takes the same general form as the CM test, however, it adds lagged factor returns as predictor variables in order to deal with potential autocorrelation in the factor returns:

$$S(F, \text{SMI}^+)_t \; = \; \beta_0 \; + \; \beta_1 \, I_{\{\text{SMI}_{t-1} > 0\}} \; + \; \beta_2 \, R(F)_{t-1} + \ldots + \beta_{k+1} \, R(F)_{t-k} + \epsilon_t \tag{17}$$

In applying this test, we fit (for each SMI parameter combination) four different models, each with an increasing number of lagged factor returns, out to a lag of $t-4$ (weeks). As with the other regression-based tests, we use robust regression with HAC covariance adjustments and select the model with the lowest AIC. We then take the p-value of the β_1 coefficient as the measure of factor timing significance.

7. **Jiang test**: The Jiang test looks at what is known as a "U-statistic," which seeks to measure the frequency at which the concavity of returns is greater than 0. Positive concavity (i.e., convexity) indicates that the portfolio manager was increasing factor exposure during up-markets and decreasing factor exposure during down-markets. Unlike the parametric tests above, this test makes no assumptions about the distribution of returns. And unlike the aforementioned nonparametric tests, which assume return forecasts are independent of return magnitudes, this test explicitly takes the magnitudes of returns into account.

   The Jiang test works by sorting factor returns ($R(F)_t$) from smallest to largest alongside the corresponding SMI-based factor timing strategies ($S(F, \text{SMI}^+)_t$). It then looks at all triplets of factor returns $\{R(F)_i,\ R(F)_j,\ R(F)_k\}$ where $R(F)_i \leq R(F)_j \leq R(F)_k$ and computes the concavity of the triplet:

$$\text{concavity} = \frac{S(F,\text{SMI}^+)_k - S(F,\text{SMI}^+)_j}{R(F)_k - R_(F)_j} - \frac{S(F,\text{SMI}^+)_j - S(F,\text{SMI}^+)_i}{R(F)_j - R(F)_i} \tag{18}$$

   In order to deal with potential distortions caused by the autocorrelation of factor returns, we make the adjustment suggested by Jiang (2003), whereby we look only at triplets for which the sorted factor returns are at least K-periods apart in the original chronological order. In our tests, we use three different values for K, namely, 2, 4, and 8 weeks (resulting in three separate p-values for the Jiang test).

8. **Welch t-test**: The unpaired t-test (with unequal volatilities) looks at two sets of conditional returns—namely, $R(F,\text{SMI}^+)_t$ and $R(F,\text{SMI}^-)_t$—to test the null hypothesis that the difference in their means is 0 (i.e., that the SMI has no predictive power). We test the null hypothesis against the alternative hypothesis that $R(F,\text{SMI}^+)_t$ has a higher mean than $R(F,\text{SMI}^-)_t$.

9. **Bayesian t-test**: The Bayesian t-test takes as input the same two sets of returns as the Welch t-test. However, whereas the Welch t-test tells us the probability of finding a test statistic greater than the one observed *if* the true difference in the means were

0, the Bayesian t-test estimates the actual probability that one set of returns has a higher mean than the other—a much stronger result (Bååth (2014)). The Bayesian t-test accomplishes this estimation using a Markov Chain Monte Carlo simulation. The resulting "p-value," then, is the probability that $R(F, \text{SMI}^+)_t$ has a higher mean than $R(F, \text{SMI}^-)_t$.

B. Multiple Hypothesis Testing

Our implementation of the Multiple Hypothesis Testing (MHT) algorithm (Romano and Wolf (2005)) seeks to quantify the statistical significance of the long/short strategies LS_i given by:

$$LS_i \;=\; \ln(1 \;+\; S_i(F, \text{SMI}^+)_t \;-\; \ln(1 \;+\; S_i(F, \text{SMI}^-)_t, \qquad i = 1, 2, \ldots, 78$$

The MHT operates by generating hypothetical strategy returns from the actual LS_i using bootstrap sampling with replacement. It then compares a test statistic (e.g., the mean) of each actual LS_i against a distribution of the best (corresponding) test statistics from across all the hypothetical strategy outcomes.

To generate the bootstrap samples, we need to specify the bootstrap's blocklength. To make this specification, we first find the optimal blocklength of each of the 78 LS_i using the algorithm in Patton et al. (2009). We then find the average of these 78 optimal blocklengths and use this average to generate 5000 stationary bootstrap samples[13].

For the test statistic, we use the studentized mean. As prescribed in Romano et al. (2008), we studentize the LS_i means using an HAC-consistent estimator and we studentize the bootstrap means using the method given in Politis and Romano (1994).

REFERENCES

Andrews, Donald W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.

AQR, 2017, Data library, URL: https://www.aqr.com/library/data-sets.

[13] We also ran the algorithm using various fixed blocklengths and with optimal blocklengths for each of the 78 different LS_i, with no material difference in the results.

Asness, Cliff, Antti Ilmanen, and Thomas Maloney, 2015, Back in the hunt, Institutional Investor.

Asness, Clifford, 2016, The siren song of factor timing aka "smart beta timing" aka "style timing", *The Journal of Portfolio Management* 42, 1–6.

Asness, Clifford, Swati Chandra, Antti Ilmanen, and Ronen Israel, 2017a, Contrarian factor timing is deceptively difficult, *The Journal of Portfolio Management* 43, 72–87.

Asness, Clifford S., Andrea Frazzini, and Lasse Heje Pedersen, 2017b, Quality minus junk, SSRN Electronic Journal, URL: https://ssrn.com/abstract=2312432.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen, 2013, Value and momentum everywhere, *The Journal of Finance* 68, 929–985.

Bååth, Rasmus, 2014, Bayesian first aid: A package that implements bayesian alternatives to the classical *.test functions in r, in *UseR! 2014 - the International R User Conference*.

Basu, Devraj, Roel Oomen, and Alexander Stremme, 2006, Exploiting the informational content of the linkages between spot and derivatives markets, https://www2.warwick.ac.uk/fac/soc/wbs/subjects/finance/research/wpaperseries/wf06-249.pdf.

Bauer, Richard J., and Julie R. Dahlquist, 2001, Market timing and roulette wheels, *Financial Analysts Journal* 57, 28–40.

Bondt, Werner F.M. De, and Richard Thaler, 1985, Does the stock market overreact, *The Journal of Finance* 40, 793–805.

Brown, Gregory W., and Michael T. Cliff, 2005, Investor sentiment and asset valuation, *The Journal of Business* 78, 405–440.

Campbell, John Y., and Robert J. Shiller, 1998, Valuation ratios and the long-run stock market outlook, *The Journal of Portfolio Management* 24, 11–26.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chu, Chia-Shang, Liping Lu, and Zhentao Shi, 2009, Pitfalls in market timing test, *Economic Letters* 103, 123–126.

Clarke, Roger G., Michael T.W. Fitzgerald, Philip Berent, and Meir Statman, 1989, Market timing with imperfect information, *Financial Analysts Journal* 45, 27–36.

Cumby, Robert E., and David M. Modest, 1987, Testing for market timing abiility: A framework for forecast evaluation, *Journal of Financial Economics* 19, 169–189.

de Prado, Marcos López, and Michael J. Lewis, 2018, What is the optimal significance level for investment strategies?, SSRN Electronic Journal, URL: https://ssrn.com/abstract=3193697.

de Roon, Frans A., Theo E. Nijman, and Chris Veld, 2000, Hedging pressure effects in futures markets, *The Journal of Finance* 55, 1437–1456.

Edelen, Roger M., Alan J. Marcus, and Hassan Tehranian, 2010, Relative sentiment and stock returns, *Financial Analysts Journal* 66, 20–32.

Estrada, Javier, 2008, Black swans and market timing: How not to generate alpha, *Journal of Investing* 17, 20–34.

Faber, Mebane, 2007, A quantitative approach to tactical asset allocation, *Journal of Wealth Management* 9, 69–79.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

French, Kenneth R., 2017, Data library, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Gibson, Scott, and Assem Safieddine, 2003, Does smart money move markets?, *The Journal of Portfolio Management* 29, 66–77.

Gibson, Scott, Assem Safieddine, and Ramana Sonti, 2004, Smart investments by smart money: Evidence from seasoned equity offerings, *Journal of Financial Economics* 72, 581–604.

Granger, Clive W.J., and M. Hashem Pesaran, 1999, Economic and statistical measures of forecast accuracy, *Journal of Forecasting* 19, 537–560.

Grinblatt, Mark, and Matti Keloharju, 2000, The investment behavior and performance of various investor types: a study of finland's unique data set, *Journal of Financial Economics* 55, 43–67.

Henriksson, Roy D., and Robert C. Merton, 1981, On market timing and investment performance. ii. statistical procedures for evaluating forecasting skills, *The Journal of Business* 54, 513–533.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jegadeesh, Nrasimhan, 1990, Evidence of predictable behavior of security returns, *The Journal of Finance* 45, 881–898.

Jiang, Wei, 2003, A nonparametric test of market timing, *Journal of Empirical Finance* 10, 339–425.

Keller, Wouter J., and Hugo S. van Putten, 2012, Generalized momentum and flexible asset allocation (faa): An heuristic approach, SSRN Electronic Journal, URL: https://ssrn.com/abstract=2193735.

Lehmann, Bruce N., 1990, Fads, martingales, and market efficiency, *Quarterly Journal of Economics* 105, 1–28.

L'Her, Jean-Francois, Tammam Mouakhar, and Mathieu Roberge, 2007, Timing small versus large stocks, *The Journal of Portfolio Management* 34, 41–50.

Marshall, Ben R., Rochester H. Cahan, and Jared M. Cahan, 2008, Can commodity futures be profitably traded with quantitative market timing strategies, *Journal of Banking & Finance* 32, 1810–1819.

Miller, Keith L., Chee Ooi, Hong Li, and Daniel Giamouridis, 2013, Size rotation in the u.s. equity market, *The Journal of Portfolio Management* 39, 116–127.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of Financial Economics* 104, 228–250.

Nalbantov, Georgi, Rob Bauer, and I. G. Sprinkhuizen-Kuyper, 2006, Equity style timing using support vector regressions, *Applied Financial Economics* 16, 1095–1111.

Neuhierl, Andreas, and Bernd Schlusche, 2011, Data snooping and market-timing rule performance, *Journal of Financial Econometrics* 9, 550–587.

Patton, Andrew, Dimitris N. Politis, and Halbert White, 2009, Correction to "automatic block-length selection for the dependent bootstrap" by d.politis and h. white, *Econometric Reviews* 28, 372–375.

Politis, Dimitris N., and Joseph P. Romano, 1994, The stationary bootstrap, *Journal of the American Statistical Association* 89, 1303–1313.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf, 2008, Formalized data snooping based on generalized error rates, *Econometric Theory* 24, 404–447.

Romano, Joseph P., and Michael Wolf, 2005, Stepwise multiple testing as formalized data snooping, *Econometrica* 73, 1237–1282.

Schmeling, Maik, 2007, Institutional and individual sentiment: Smart money and noise trader risk?, *International Journal of Forecasting* 23, 127–145.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *The Journal of Finance* 19, 425–442.

Sharpe, William F., 1975, Likely gains from market timing, *Financial Analysts Journal* 31, 60–69.

Shilling, A. Gary, 1992, Market timing: Better than a buy-and-hold strategy, *Financial Analysts Journal* 48, 46–50.

Sy, Wilson, 1990, Market timing: Is it a folly?, *The Journal of Portfolio Management* 16, 11–16.

Treynor, Jack L., and Kay K. Mazuy, 1966, Can mutual funds outguess the market?, *Harvard Business Review* 44, 131–136.

Welch, B.L., 1947, The generalization of student's problem when several different population variances are involved, *Biometrika* 34, 28–35.

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of the equity premium prediction, *Review of Financial Studies* 21, 1455–1508.