Relative Sentiment and Machine Learning for Tactical Asset Allocation

Raymond C. Micaletti *

ABSTRACT

We examine Sentix sentiment indices for use in tactical asset allocation. In particular, we construct monthly relative sentiment factors for the U.S., Europe, Japan, and Asia ex-Japan by taking the difference in 6-month economic expectations between each region's institutional and individual investors. These factors (along with one-month forward equity returns) then serve as inputs to a wide array of machine learning algorithms. Employing combinatorial cross-validation and adjusting for data snooping, we find relative sentiment factors have robust and significant predictive power in all four regions; that they surpass both standalone sentiment and time-series momentum in terms of informational content; and that they demonstrate the ability to identify the subsequent best- and worst-performing global equity markets from along a cross-section. The results are consistent with previous findings on relative sentiment, discovered using unrelated datasets.

*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

Tactical asset allocation is the process of adjusting portfolio exposures based on short- or intermediate-term (e.g., weekly, monthly, quarterly) estimates of risk and return. Whereas a buy-and-hold benchmark allocation uses no additional information in its portfolio formation beyond benchmark weights, and whereas strategic asset allocation might only use estimates of long-term returns, tactical asset allocation often incorporates a variety of shorter-term measures to help contextualize the current state of the market and its likely intermediate-term evolution (Dahlquist and Harvey (2001)).

While the literature is full of tactical factors that have been put forward as being predictive of returns, often these factors do not survive out-of-sample testing (Welch and Goyal (2008)). The subsequent breakdowns are perhaps due to data snooping bias, i.e., the phenomenon whereby if enough factors are tested, some will appear to be predictive simply by random chance (White (2000)).

Not all factors fall by the wayside, however. Some have stood the test of time and have generally come to be recognized as legitimate market anomalies. In the context of tactical asset allocation, two such factors stand out—namely, value and momentum. These factors represent perhaps the two most popular ways of implementing tactical asset allocation strategies (e.g., Moskowitz et al. (2012), Asness et al. (2013), Asness et al. (2015), Faber (2007), Antonacci (2014), Zakamulin (2014)).

Despite their popularity and resilience, however, value and momentum each have certain weaknesses. Value, for example, often reduces equity allocations too early in uptrends and increases equity allocations too early in downtrends. Momentum tends to do the opposite (that is, reduce allocations too late after the market has peaked and increase allocations too late after the market has bottomed). This complementarity is one of the main reasons value and momentum factors are often traded together.

In recent years, an alternative tactical factor has been proposed that may help transcend the weaknesses of value and momentum. This factor, referred to as relative sentiment, is a measure of how bullish or bearish institutional investors are compared to individual investors in a given market.

In what was perhaps the first study to consider relative sentiment solely and explicitly, Edelen et al. (2010) measured the sentiment of individual investors relative to the sentiment of institutional investors using data from the Federal Reserve Z.1 reports. They found that when individual sentiment toward equities is high relative to the sentiment of institutions, equities tend to experience lower future returns, and vice versa.

More recently, Micaletti (2018)) used Commitments of Traders data to construct a composite indicator that measures how institutions are positioned relative to individuals in three areas of the market: equities, long-duration bonds, and along the yield curve. Adjusted for data snooping, this indicator demonstrated extreme levels of statistical and economic significance; contained more informational content than time-series momentum; aided in the forecasting of smart beta equity factors; and proved useful as a means of enacting a tactical asset allocation strategy—outperforming its value, momentum, and cross-asset strategy counterparts (Micaletti (2019)), despite having much lower time-averaged exposure to equities.

In addition to the aforementioned studies, a substantial body of research exists that, while not explicitly or solely employing relative sentiment, uses factors or frameworks closely related to relative sentiment. This is especially true when one considers the "adding-up constraint," whereby if institutions are acting in one manner, investors who are not part of the institutional class (e.g., hedge funds, individuals) must be acting in the opposite manner (Grinblatt and Keloharju (2000)). These studies have routinely shown that institutions tend to outperform individuals over intermediate time horizons (e.g., Schmeling (2007), Grinblatt and Keloharju (2000), Gibson and Safieddine (2003), Gibson et al. (2004)).

Taken as a whole, the foregoing studies strongly suggest institutional investors have superior forecasting abilities relative to individuals. This forecasting edge likely stems from the substantial and persistent structural advantages institutions possess, such as better information networks, deeper pools of human capital, and access to more advanced technologies. As these advantages are unlikely to disappear anytime soon, it is reasonable to believe relative sentiment might continue to be predictive of asset returns for the foreseeable future. Thus, it might make sense to examine relative sentiment factors wherever and whenever they appear.

Accordingly, in this study, we examine Sentix 6-month-forward economic expectations indices corresponding to institutional and individual investors. From these investor-class indices, we construct monthly relative sentiment factors for the U.S., Europe, Japan, and Asia ex-Japan. These factors, coupled with one-month forward regional equity returns, then serve as inputs to a wide selection of machine learning models (which produce directional forecasts of one-month-forward equity returns).

While the effectiveness of machine learning techniques for predicting financial markets has been called into question (e.g., Makridakis et al. (2018)), we find several of the machine learning techniques investigated here, in conjunction with various combinations of relative sentiment factors, produce statistically significant results, even after adjusting for data snooping.

As many studies have incorporated machine learning techniques for financial market forecasting (e.g., López de Prado (2018), Nalbantov et al. (2006), Oflus (2014), Chakravorty et al. (2018),), this paper breaks no new ground in that regard. Here, the machine learning algorithms are simply tools for unmasking any predictive power relative sentiment may have.

Nor is the present study the first to look at Sentix sentiment data, which has been investigated extensively (e.g., Schmeling (2007), Heiden et al. (2013), Hengelbrock et al. (2013), Zwergel and Klein (2006), Beaumont et al. (2006)).

But, as far as we can tell, Sentix data has never been used to construct explicit relative sentiment factors. Perhaps this is because researchers believe information would be lost by combining the standalone sentiments into an index that measures their differences.

Where this paper does contribute to the literature, however, is as follows:

1. It adds to the expanding list of explicit relative sentiment studies and confirms the results of prior studies regarding relative sentiment's statistical significance
2. It provides evidence relative sentiment surpasses standalone sentiment when it comes to forecasting equity returns
3. It demonstrates relative sentiment works not only for U.S. markets, but for global markets as well
4. It affirms the results in Micaletti (2018), which show relative sentiment contains more predictive power than time-series momentum.
5. And lastly, it corroborates the results in Micaletti (2019), which show relative sentiment has the potential to identify the subsequent best- and worst-performing markets from along a cross-section.

The rest of the paper is organized as follows: Section I describes the data and the methods used to analyze relative sentiment. Section II presents the results of our tests. Section III summarizes the main findings of the paper.

I. Data and Methodology

Relative Sentiment Factors

Our dataset consists of Sentix 6-month economic expectations indices for the U.S., Europe, Japan, and Asia ex-Japan. This dataset is available through platforms such as Bloomberg and Factset without any additional subscription fee.

The indices are the result of a monthly online survey conducted by Sentix, whereby participants—both institutional and private investors—are able to submit their six-month forward views on various regional economies. The survey results are mapped to index values and publicly released in the middle of the month. Conveniently, the time stamp of the data is set to the end of the month, which means no further lags of the data are necessary.

In addition to having an aggregate sentiment index for each region, Sentix breaks out the regional results into their component indices, namely, institutional investor sentiment and private (i.e., individual) investor sentiment. These component indices are the ones we makes use of in this study.

If we let IS_t^A and PS_t^A represent the institutional and private sentiment indices for region A at time t, respectively, then we can construct a corresponding relative sentiment factor, RS_t^A, for region A at time t by simply subtracting the private sentiment from the institutional sentiment. That is,

$$RS_t^A = IS_t^A - PS_t^A \tag{1}$$

We construct relative sentiment factors for the U.S., Europe, Japan, and Asia ex-Japan in this manner. These factors constitute the input to the machine learning models. Because not all machine learning models we consider are capable of using only a subset of the input factors to make predictions, we train and test the machine learning models on all factor subsets that contain at least two factors. Table I lists these possible subsets (there are 11 of them).

The sentiment indices available from Bloomberg start in August 2002 and run through September 2019 (with a three-month period from October 2002 to December 2002 missing from the dataset). This results in 204 monthly data points.

Regional Equity Returns

We use relative sentiment factors along with machine learning models to forecast one-month forward equity returns in the same regions for which we have relative sentiment data, namely, the U.S., Europe, Japan, and Asia ex-Japan.

We obtain daily total (U.S. dollar) returns for the four regions from Professor Kenneth R. French's data library (French (2017)) through May 31, 2017. To these returns we splice

Table I. Factor Subsets: This table enumerates all 2-, 3-, and 4-factor combinations of relative sentiment ($RS^{\star}$) for the U.S. (USA), Europe (EUR), Japan (JPN), and Asia ex-Japan (AEJ).

	Relative Sentiment Factor Combinations			
Index	RS^{USA}	RS^{EUR}	RS^{JPN}	RS^{AEJ}
1	★	★		
2	★		★	
3	★			★
4		★	★	
5		★		★
6			★	★
7	★	★	★	
8	★	★		★
9	★		★	★
10		★	★	★
11	★	★	★	★

daily total (U.S. dollar) returns (over the time period June 2017 to September 2019) from ETFs representing the four regions.

The ETFs we use are VTI (the Vanguard Total U.S. Market Index), IEUR (iShares Core MSCI Europe Index), EWJ (iShares MSCI Japan Index), and EPP (iShares MSCI Pacific ex-Japan Index).

We compound these daily returns into monthly returns and align the monthly returns with the prior month's relative sentiment factors. For each regional equity market there are 11 such input arrays, corresponding to the 11 factor subsets.

Combinatorial Cross-Validation

Once the input data has been configured, the next step is to divide the input data into training sets and test sets. We do this using combinatorial cross-validation (CCV) (López de Prado (2018)). Table II shows a schematic of the CCV concept, which we will use to briefly describe the process.

Consider an input time-series divided along the time axis into five roughly equal groups of data points. Further, let's assume we train a machine learning model using three of those groups and test the model using the other two. With five groups, there are 10 different

Table II. Combinatorial Cross-Validation: The following schematic illustrates the concept of combinatorial cross-validation (CCV) for an input time-series divided into five groups (G_i) where three groups are used to train the model and two are used to test it. There are 10 ways to divide five groups into three training and two test groups. The symbols ○ represent training groups and the symbols ★ represent test groups. Test groups of the same color are then concatenated to form prediction time-series (aka backtest paths).

	CCV Training Set / Test Set Combinations									
Group	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
G_1	★	★	★	★	○	○	○	○	○	○
G_2	★	○	○	○	★	★	★	○	○	○
G_3	○	★	○	○	★	○	○	★	★	○
G_4	○	○	★	○	○	★	○	★	○	★
G_5	○	○	○	★	○	○	★	○	★	★

ways we can parse them into three training and two test groups. Table II enumerates these splits, where the symbols ○ represent training groups and the symbols ★ represents test groups. In this example, with 10 combinations and two test groups per combination, there will be 20 different test groups for which the machine learning models will make predictions on unseen data.

We then take five of those test groups—one from each time-segment— and concatenate them to form a prediction time-series that spans the entire length of the original time-series. We repeat this process, creating additional prediction time-series, until all test groups have been used once and only once. Table II illustrates this process, where test groups of the same color are concatenated to form prediction time-series.

With five total groups and two test groups, we can construct four different prediction time-series. These prediction time-series (aka, backtest paths) give us an idea of how well the machine learning model can generalize from training data to test data.

In general, with N total groups and M test groups, there are $\frac{N!}{M!\,(N-M)!}$ possible training/test splits and $\frac{M}{N}\,\frac{N!}{M!(N-M)!}$ number of backtest paths (López de Prado (2018)).

With our actual dataset, we divide the 204 monthly data points into 14 groups of 15 points each (with the 14th group having only 9 points). We train the models on 12 of those 14 groups, and test them on the remaining two. (We will refer to this parsing as CCV{15,2}, where each group contains 15 points and there are two test groups.)

There are 91 different ways to parse 14 groups into 12 training and two test groups. We run the machine learning models on each of those 91 splits. From those 91 splits, we produce 13 different backtest paths. We then average the results across those 13 different backtest paths to arrive at one composite backtest for each machine learning model.

Machine Learning Algorithms

Given the skepticism surrounding machine learning, we were unsure at the outset whether machine learning algorithms would work at all, and thus had little conviction as to which ones might be effective. As a result, we tested several different algorithms.

Broadly speaking, we investigated linear and logistic regression, support vector machines (SVM), random forests, generalized boosted models (GBMs), and various implementations of classification and regression trees. All of these algorithms came from either the base R package or user-contributed packages.

We used bootstrap aggregation ("bagging") with 50 "bags" for all machine learning algorithms except random forests and GBMs, which underwent their own internal bagging procedures. For the SVM algorithms, we considered several different parameter combinations. For GBMs, we considered adaboost, bernoulli, and gaussian distributions. For the other algorithms, we used the default R-package parameters.

Henceforth, we will use the term "machine learning model" (or simply "model") to refer to a machine learning algorithm coupled with a particular set of parameters. Table III provides a list of the algorithms and their relevant parameter universes (if any).

Given the number of algorithms and combinations of parameters considered, there were a total of 90 machine learning models tested. With 11 factor subsets, this resulted in 990 backtests of relative sentiment for each regional equity market.

Machine Learning Output

The test-set predictions for all models (regardless of whether they were forecasting binary or real-valued outputs), were mapped to the values of -1 or 1, representing a negative or positive one-month-forward return forecast, respectively. These predictions were averaged across all bags and then further averaged across all sample backtest paths produced by the CCV procedure. This averaging yielded a composite forecast between -1 and 1. The composite forecasts were then mapped (linearly) to equity allocations between 0 and 1.

We used these equity allocations to construct time-series of strategy returns (i.e., backtests).

Table III. Machine Learning Algorithms: The following table lists the machine learning algorithms considered in this study, their R packages and functions, and their relevant parameter universes. (For example, certain support vector machines have a kernel parameter, a cost parameter, a ν or ϵ parameter, and a γ parameter.)

Machine Learning Algorithms Used in Study					
	R package/ function	kernel/ distribution	Cost	ν or ϵ	γ
linear regression	base/lm()				
logistic regression	base/glm()				
kernel logistic regression	base/glm() with kpca()	radial			
C-classification linear	e1071/svm()	linear	{1,10,100}		
C-classification nonlinear	e1071/svm()	radial	{1,10,100}		{0.5,1,2}
ν-classification linear	e1071/svm()	linear		{0.1,0.2,0.3}	
ν-classification nonlinear	e1071/svm()	radial		{0.1,0.2,0.3}	{0.5,1,2}
ν-regression linear	e1071/svm()	linear	{1,10,100}	{0.1,0.2,0.3}	
ν-regression nonlinear	e1071/svm()	radial	{1,10,100}	{0.1,0.2,0.3}	{0.5,1,2}
ϵ-regression linear	e1071/svm()	linear		{0.1,0.2,0.3,0.4}	
ϵ-regression nonlinear	e1071/svm()	radial		{0.1,0.2,0.3,0.4}	{0.5,1,2}
tree classification	rpart/rpart()				
tree regression	rpart/rpart()				
tree classification	tree/tree()				
tree regression	tree/tree()				
tree classification	party/ctree()				
tree regression	party/ctree()				
random forest classification	randomForest/randomForest()				
random forest regression	randomForest/randomForest()				
generalized boosted models	gbm/gbm()	{adaboost,bernoulli,gaussian}			

Two different tactical strategies were examined—one incorporating both equities and bonds, and the other using equities and cash. We looked at the former because of its seeming ubiquity amongst tactical strategies, and at the latter to get a clearer picture (unobscured by bond returns) of relative sentiment's ability to forecast equity markets.

If we let ψ_t represent the equity allocation for month t, R_t^e represent the monthly equity market return, and R_t^b represent the monthly return of an aggregate bond index, then the monthly strategy returns, R_t^s, for each strategy are given by:

$$\begin{aligned} &\text{Equity/bond strategy:} \quad R_t^s = \psi_t\, R_t^e + (1 - \psi_t)\, R_t^b \\ &\text{Equity/cash strategy:} \quad R_t^s = \psi_t\, R_t^e \end{aligned}$$

The benchmarks to these strategies are also taken to be equity/bond and equity/cash portfolios, respectively. The benchmark equity allocation is set to the average monthly equity allocation of the strategy, i.e.: $\bar{\psi}_t = \frac{1}{T} \sum_{s=1}^{T} \psi_s$, where T is the number of time points in the dataset. (Having the benchmark equity allocation reset to the same value each month implies monthly rebalancing—commensurate with monthly rebalancing of the strategies.)

Thus, the respective monthly benchmark returns are given by:

$$\begin{aligned} &\text{Equity/bond benchmark:} \quad R_t^s = \bar{\psi}_t\, R_t^e + (1 - \bar{\psi}_t)\, R_t^b \\ &\text{Equity/cash benchmark:} \quad R_t^s = \bar{\psi}_t\, R_t^e \end{aligned}$$

We obtain aggregate bond returns from the Bloomberg Barclays U.S. Aggregate Index. Further, no estimates of transaction costs or slippage were included in either the strategies or benchmarks—primarily because we are dealing with broad, liquid equity and bond markets, and have a two-week advance notice of any portfolio changes (which in turn only affect a fraction of the portfolio each month). Consequently, actual transaction costs, both explicit and implicit, should not be prohibitive.

Variations

In addition to using a CCV parsing with 15 data points per group and two test groups (i.e., CCV{15,2}), we also use a parsing with 24 points per group and two test groups (i.e., CCV{24,2}). We do this to see how the results change when using less data to train the models. CCV{15,2} uses about 85% of the data for training, while CCV{24,2} uses roughly 75%. We report results for both cases.

Further, we apply the exact same procedures described above to the components of the relative sentiment factors. For example, if a factor subset consists of RS_t^{USA} and RS_t^{EUR}, we decompose those factors into their component sentiments, namely, institutional and private investor sentiment in both the U.S. and Europe (i.e., IS_t^{USA}, PS_t^{USA}, IS_t^{EUR}, and PS_t^{EUR}), and run the machine learning models on those factors. We perform these decompositions on all 11 factor subsets shown in Table I.

We refer to this analysis as component sentiment analysis to distinguish it from relative sentiment analysis. We examine component sentiment in order to see whether (the explicit representation of) relative sentiment contains any additional predictive power beyond simply knowing the values of its standalone components.

Multiple Hypothesis Testing

As previously mentioned, the outputs of the forecasting procedure described above are 990 time-series of relative sentiment strategy returns (and their corresponding benchmarks). With that many strategies, some are bound to look statistically significant (relative to their benchmarks) just by chance. To determine whether a strategy may rightfully be considered significant, we have to control for data snooping. We do this using so-called "multiple hypothesis testing" (MHT) algorithms.

As a high-level overview, MHT algorithms take in a set of M strategy-return time-series, $X = \{X_{t,1}, X_{t,2}, \ldots, X_{t,M}\}$, and their benchmarks, $Y = \{Y_{t,1}, Y_{t,2}, \ldots, Y_{t,M}\}$, and create a new set of strategies $Z = X - Y = \{X_{t,1} - Y_{t,1}, X_{t,2} - Y_{t,2}, ..., X_{t,M} - Y_{t,M}\}$. They then perform bootstrap sampling (with replacement) on Z, using B samples, to create B sets of hypothetical strategies $\tilde{\mathbb{Z}} = \{\tilde{Z}^{(1)}, \tilde{Z}^{(2)}, ..., \tilde{Z}^{(B)}\}$.

Then, for each hypothetical set $\tilde{Z}^{(i)} = \{\tilde{Z}^{(i)}_{t,1}, \tilde{Z}^{(i)}_{t,2}, ..., \tilde{Z}^{(i)}_{t,M}\}$, the MHT algorithm computes the performance statistic of interest (e.g., the mean return, the Sharpe ratio) across all M strategies and finds its maximum value. Once this process has been completed, there will be a distribution of B maximum performance statistics corresponding to the B sets of hypothetical strategies (where each set contains M strategies).

If the performance statistic of an actual strategy in the original set Z compares favorably against this distribution of hypothetical maxima (i.e., if it resides in the upper tails of the distribution), we then have cause to say the strategy rejects the null hypothesis of no statistical significance.

We consider three MHT algorithms, namely, the Step-RC method (RC = "Reality Check") (Romano and Wolf (2007)), the Step-SPA method (SPA = "Superior Predictive Ability") (Hsu et al. (2014)), and a method found in the R package "multtest" (using the function "mhp") (Pollard et al. (2005)). Each approach is bootstrap-based and we use 10,000 bootstrap samples in our testing.

The aforementioned MHT algorithms are all designed to control the "k-family-wise error rate" (k-FWER), i.e., the probability of identifying at least k false positives amongst all the strategies identified as significant. The k-FWER is controlled via the relationship $\mathbb{P}\{\text{number of false positives} \geq k\} \leq \alpha$, where α is a user-specified probability.

We set k = 1 and test six different values of α, namely, $\alpha = \{0.01, 0.02, 0.05, 0.10, 0.15, 0.30\}$. López de Prado and Lewis (2018) show that for time series with a true Sharpe ratio of 1.0, the tradeoff between a test's confidence and its power is optimized at $\alpha \approx 0.30$. They also argue that unless a strategy has a true Sharpe ratio greater than 1 for a period of more than 10 years of daily data, α values less than 0.15 are likely to be excessively conservative (i.e., produce too many false negatives). Thus, the α values listed above would appear to err on the side of conservatism.

The output at each value of α is the number of strategies that reject the null hypothesis of no statistical significance. By construction, the probability more than one of those rejections is false is less than the corresponding α.

The three algorithms seem to produce a roughly similar number of rejections for a given set of strategies. Step-RC is broadly the most conservative, followed closely by Step-SPA, while the "multtest" function "mhp" is often less conservative at lower values of α and more conservative at higher values.

We run both Step-RC and Step-SPA over multiple different parameter combinations related to their bootstrap sampling processes. For example, we consider both stationary and circular bootstraps. We compute the variance of the bootstrap samples (for purposes of studentization) using both the sample variance and the variance given in Eq. (6) of Politis and Romano (1994) (which is independent of the bootstrap samples). With respect to bootstrap blocklengths, we first compute the optimal blocklengths of all 990 strategies using function "b.star" in R package "np" (**?**). We then alternately consider both the mean and the 95th percentile of those 990 optimal blocklengths as the blocklength of the overall sampling process.

Thus, there are eight different bootstrap-related parameter combinations for both Step-RC and Step-SPA. We augment those outputs with the output of one implementation of the "multtest" function "mhp" and report the average across all 17 results.

II. Results

Statistical Significance

Table IV presents the MHT results for the U.S. This table covers relative sentiment and component sentiment, tactical strategies with and without bonds, and both CCV cases. The table values are the number of strategies (out of 990) that reject the null hypothesis of no statistical signficance for each value of α.

As is apparent, relative sentiment far surpasses component sentiment in terms of significance. Several dozen relative sentiment strategies register as significant at the most extreme values of α, whereas no component sentiment strategies do. At $\alpha = 0.30$, relative sentiment registers approximately 300 strategies as signficant—about ten times as many as component sentiment.

As one might expect, the number of significant strategies declines slightly for the case of CCV{24,2}, as fewer points were used to train the models. Further, it is evident the with-bonds tactical strategies exhibit moderately more significance—likely because of the recurring performance boost from bonds during periods of poor equity market returns.

The corresponding MHT tables for Europe, Japan, and Asia ex-Japan can be found in Appendices A, B, and C, respectively. In those tables, we observe mostly similar patterns. Relative sentiment surpasses component sentiment in Europe and Japan at all values of α. In Asia ex-Japan, relative sentiment surpasses component sentiment at the most extreme values of α, but loses its edge at the less extreme values.

For Europe and Japan, we likewise observe fewer significant strategies in the CCV{24,2} case (in Europe the dropoff is notable). In contrast, Asia ex-Japan shows a *greater* number of significant strategies for the CCV{24,2} case, despite the models having been trained on less data.

Overall, relative sentiment produced both greater numbers of significant strategies and strategies that registered more extreme significance when compared to component sentiment. This suggests the difference in institutional and private investor sentiment might be more informative than their standalone values.

Table IV. Multiple Hypothesis Testing, Relative Sentiment vs. Component Sentiment, USA: The following tables present multiple hypothesis testing (MHT) results for both relative sentiment and component sentiment strategies in the U.S. The results correspond to different combinatorial cross-validation cases (CCV{15,2} and CCV{24,2}) as well as to different tactical strategies (equities + bonds and equities + cash). The reported values represent the numbers of strategies (from the pool of 990 tested) that reject the null hypothesis of no statistical significance (at various levels of the MHT parameter α). As MHT can be performed in numerous ways, the numbers presented here are averages, computed from the outputs of several different MHT implementations.

(a) USA, Relative Sentiment Strategies

		Average Number of Null-Hypothesis Rejections α					
Equities+	CCV	0.01	0.02	0.05	0.10	0.15	0.30
Bonds	{15,2}	34	69	133	191	236	359
Bonds	{24,2}	24	45	92	150	190	332
Cash	{15,2}	23	43	81	136	186	289
Cash	{24,2}	17	34	74	125	168	276

(b) USA, Component Sentiment Strategies

		Average Number of Null-Hypothesis Rejections α					
Equities+	CCV	0.01	0.02	0.05	0.10	0.15	0.30
Bonds	{15,2}	0	0	1	5	10	29
Bonds	{24,2}	0	0	0	1	4	14
Cash	{15,2}	0	0	2	8	13	33
Cash	{24,2}	0	0	2	6	13	35

Economic Significance

Figure 1 shows the performance of relative sentiment in the U.S. over the time period 2002-2019. It plots the "with-bonds" and "without-bonds" total return curves for both CCV cases. These total return curves are composites, averaged over the top 10 strategies (of the 990). The figure also plots their respective composite benchmarks. (Note: because the with-bonds benchmarks for CCV{15,2} and CCV{24,2} were virtually identical in performance, we combine them into one series. The same holds for the without-bonds benchmarks.) Table V presents the performance statistics.

The composite relative sentiment strategies in the U.S. outperformed their benchmarks by roughly 650-700 basis points (bps) per annum over the 17-year period—yielding a total return approximately three times the return of the benchmark. These relative sentiment

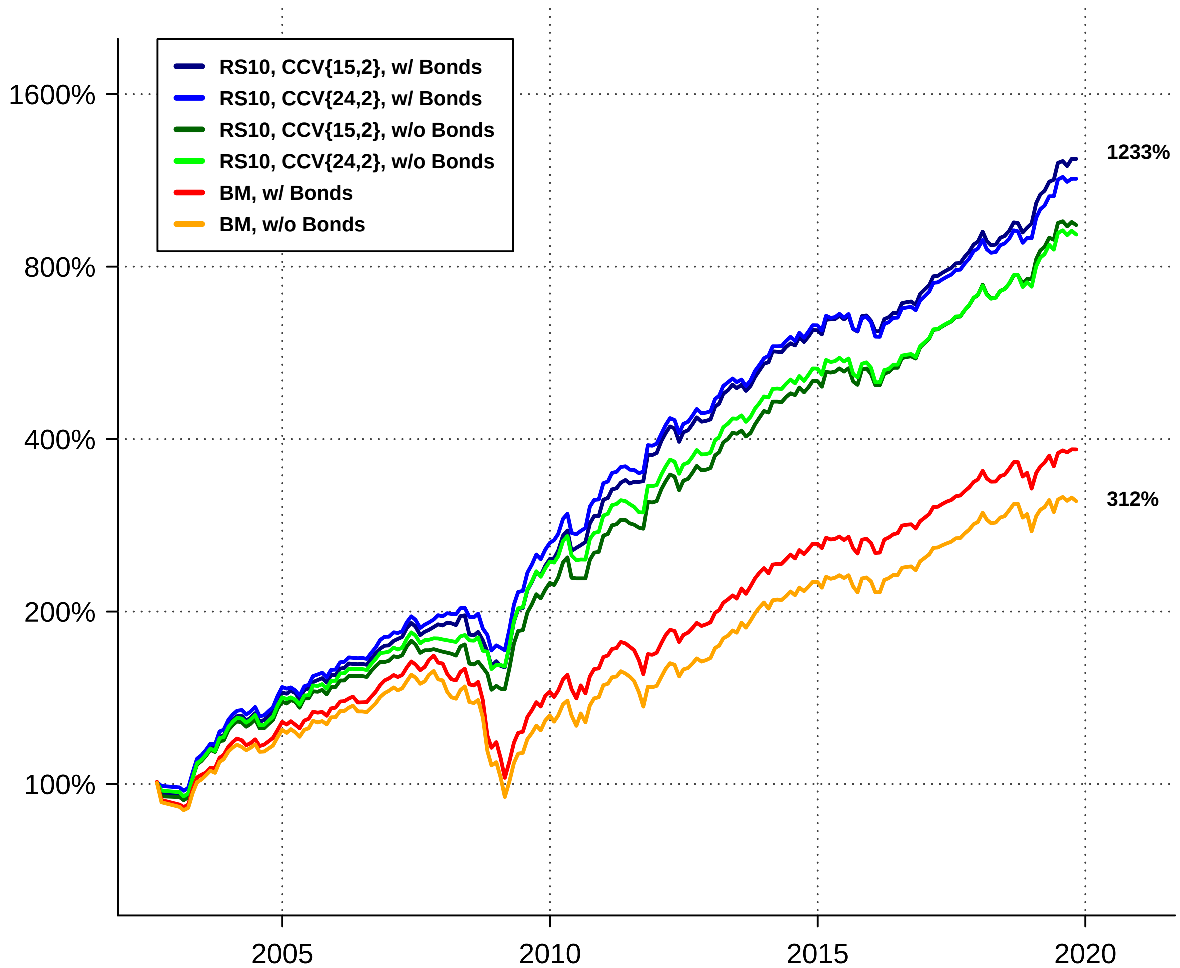

Figure 1. Relative Sentiment Performance, USA, 2002-2019: Composite total returns of the top 10 U.S. (USA) relative sentiment strategies (RS10) for CCV{15,2} and CCV{24,2}—both with and without bonds—plotted against their respective benchmarks (BM)

Table V. Relative Sentiment Performance Statistics, USA, 2002-2019: Performance statistics of the composite returns of the top 10 U.S. (USA) relative sentiment strategies (RS10) for CCV{15,2} and CCV{24,2}—both with and without bonds. Corresponding benchmark (BM) statistics included for comparison.

Strategy	Bonds	CCV	Total Return	CAGR	Sharpe Ratio	Maximum Drawdown	Average Equity Allocation	Average Monthly Turnover
RS10	With	{15,2}	1233%	15.9%	1.44	−18.9%	73.2%	29.8%
RS10	With	{24,2}	1139%	15.4%	1.41	−15.7%	74.1%	26.4%
RS10	Without	{15,2}	945%	14.1%	1.27	−17.9%	73.2%	29.8%
RS10	Without	{24,2}	910%	13.9%	1.26	−13.7%	73.9%	26.3%
BM	With	{15,2}	384%	8.2%	0.70	−38.8%	73.6%	
BM	Without	{24,2}	312%	6.9%	0.59	−39.7%	73.5%	

returns were achieved with higher Sharpe ratios and lower maximum drawdowns. The average monthly equity allocation of the top U.S. relative sentiment strategies was around 73%, while the average monthly turnover was between 26%-30%.

The corresponding figures and tables for Europe, Japan, and Asia ex-Japan are in Appendices A, B, and C, respectively. Again we observe similar patterns as in the U.S. In Europe, the relative sentiment strategies outperformed their benchmarks by roughly 400-550 bps per annum; in Japan, by 400-500 bps; and in Asia ex-Japan also by 400-500 bps. In all regions, the Sharpe ratios were higher than the benchmarks. In Europe and Japan, the maximum drawdowns were substantially less than their benchmarks, while in Asia ex-Japan the maximum drawdowns were similar to those of the benchmarks. Average equity allocations were in the 65%-68% range (slighly lower than in the U.S.) and average monthly turnover was between 26%-29%.

Relative Sentiment vs. Time-Series Momentum

The idea underpinning time-series momentum as a factor is that when momentum is positive (negative), subsequent returns are generally higher (lower). If momentum were a more powerful predictor than relative sentiment, given the state of momentum, the state of relative sentiment would make little difference.

Instead, what we observe is exactly the opposite. The state of relative sentiment appears to drive directional returns, regardless of the state of time-series momentum. Table VI shows the average one-month forward equity market returns conditioned on the four combinations of positive/negative momentum and positive/negative relative sentiment. We measure momentum based on whether the regional equity market is above or below its 10-month simple moving average. We measure relative sentiment based on whether a region's monthly equity allocation is above or below its average monthly equity allocation. (For example, if the average monthly equity allocation in the U.S. is 73%, then all months with equity allocations greater (less) than 73% would be deemed to have positive (negative) relative sentiment.)

The monthly equity allocations used to measure relative sentiment are composites averaged over the top 50 strategies in each region. Note also, the average one-month forward equity returns in Table VI are the regional equity markets' returns and *not* the allocation-weighted returns of the relative sentiment strategies.

From the table, we see when time-series momentum is negative, but relative sentiment is positive, the annualized average one-month forward return over all regions is approx-

Table VI. Relative Sentiment vs. Time-Series Momentum: This table presents the average one-month forward equity market return in all regions conditioned on the state of the regions' prior month-end time-series momentum (TSM) (whether positive or negative) and on the state of the prior month-end relative sentiment factor (whether bullish (RS+) or bearish (RS-)). The state of relative sentiment in each region is determined by aggregating the top 50 relative sentiment strategies in the region into one strategy and seeing whether a given month's equity allocation is above or below the average monthly equity allocation of the aggregated strategy. If above (below), relative sentiment is considered bullish (bearish). The table also shows the differences and annualized differences in these conditional returns. Results are reported for both combinatorial cross-validation cases (CCV{15,2} and CCV{24,2}).

		Avg Monthly Eqy Mkt Return when TSM Negative				Avg Monthly Eqy Mkt Return when TSM Positive			
Region	CCV	RS-	RS+	Difference	Annualized	RS-	RS+	Difference	Annualized
USA	{15,2}	-3.4%	3.0%	6.5%	77.5%	-0.3%	1.7%	2.0%	23.6%
USA	{24,2}	-3.4%	3.5%	6.9%	83.1%	-0.2%	1.5%	1.7%	20.0%
EUR	{15,2}	-2.9%	2.5%	5.3%	63.8%	0.0%	1.6%	1.6%	19.5%
EUR	{24,2}	-1.8%	1.8%	3.7%	44.3%	0.1%	1.5%	1.4%	16.8%
JPN	{15,2}	-0.9%	1.8%	2.7%	32.0%	-0.6%	1.1%	1.7%	20.6%
JPN	{24,2}	-1.0%	1.9%	2.9%	35.1%	-0.8%	1.3%	2.1%	25.5%
AEJ	{15,2}	-0.9%	1.8%	2.6%	31.3%	0.1%	1.8%	1.7%	20.7%
AEJ	{24,2}	-0.9%	1.7%	2.6%	30.8%	-0.5%	2.1%	2.6%	31.8%
				Average	**49.7%**			**Average**	**22.3%**

imately 27%. In contrast, when momentum is negative and relative sentiment is also negative, the annualized average one-month forward return over all regions is approximately -23%—a difference of 50 percentage points depending on the state of relative sentiment.

In the case of positive momentum, we observe the same basic pattern—returns are higher (lower) when relative sentiment is higher (lower)—although the spreads are not as large. The annualized average one-month forward return spread between positive and negative relative sentiment in this case is approximately 22%.

These results are consistent with Micaletti (2018), which found similar annualized spreads in U.S. equity market returns depending on the state of relative sentiment. Moreover, the results are robust to the number of relative sentiment strategies used to generate the composite equity allocations (from which we determine the polarity of relative sentiment). We tested composites ranging from 10 strategies to 200 strategies and the annualized spreads remained stable.

Relative Sentiment and Cross-sectional Returns

The foregoing results suggest relative sentiment might be useful for forecasting equity market performance within regions. A natural follow-up question is whether relative sentiment might be similarly useful for forecasting equity market performance *across* regions.

To address this question, we examine whether the rank order of relative sentiment across regions is predictive of one-month forward relative equity market returns. Each month we line up the regional equity allocations (which we use as proxies for relative sentiment) and rank them. We then collect the one-month forward returns corresponding to each rank position. We repeat this each month and concatenate the rank-position returns. (As before, the equity allocations for each region are composites computed from the region's top 50 relative sentiment strategies.)

Figure 2 plots the total returns corresponding to each relative sentiment rank position along the 4-region cross-section. The concatenated returns of the sequentially top-ranked regions produced a total return of 883% over the test period (the highest), while the lowest-ranked regions mustered a total return of just 248% (the lowest). The second- and third-ranked regions generated the second and third highest total returns, respectively (although the third-ranked regions had the lowest total return for much of the backtest history before eventually overtaking the lowest-ranked regions for third place).

Table VII shows the total returns at each relative sentiment rank position for all combinations of 3-region cross-sections. Again, we see the rank order of relative sentiment is predictive of the rank order of total returns. In each 3-region combination, the concatenated returns of the sequentially worst-ranked regions produced the lowest total return; the top-ranked regions produced the highest; and the middle-ranked regions fell somewhere in between.

We repeat the analysis for all 2-region cross-sections. The objective was to see whether when Region A had higher relative sentiment than Region B, did Region A's equity market have higher one-month forward returns than Region B's on average?

Table VIII shows the annualized average difference in one-month forward returns for every head-to-head matchup. For example, from the table we see that when Europe had higher relative sentiment than the U.S., Europe subsequently outperformed the U.S. in the following month (on average) by 3.9 percentage points on an annualized basis. Conversely, when the U.S. had higher relative sentiment than Europe, the U.S. subsequently

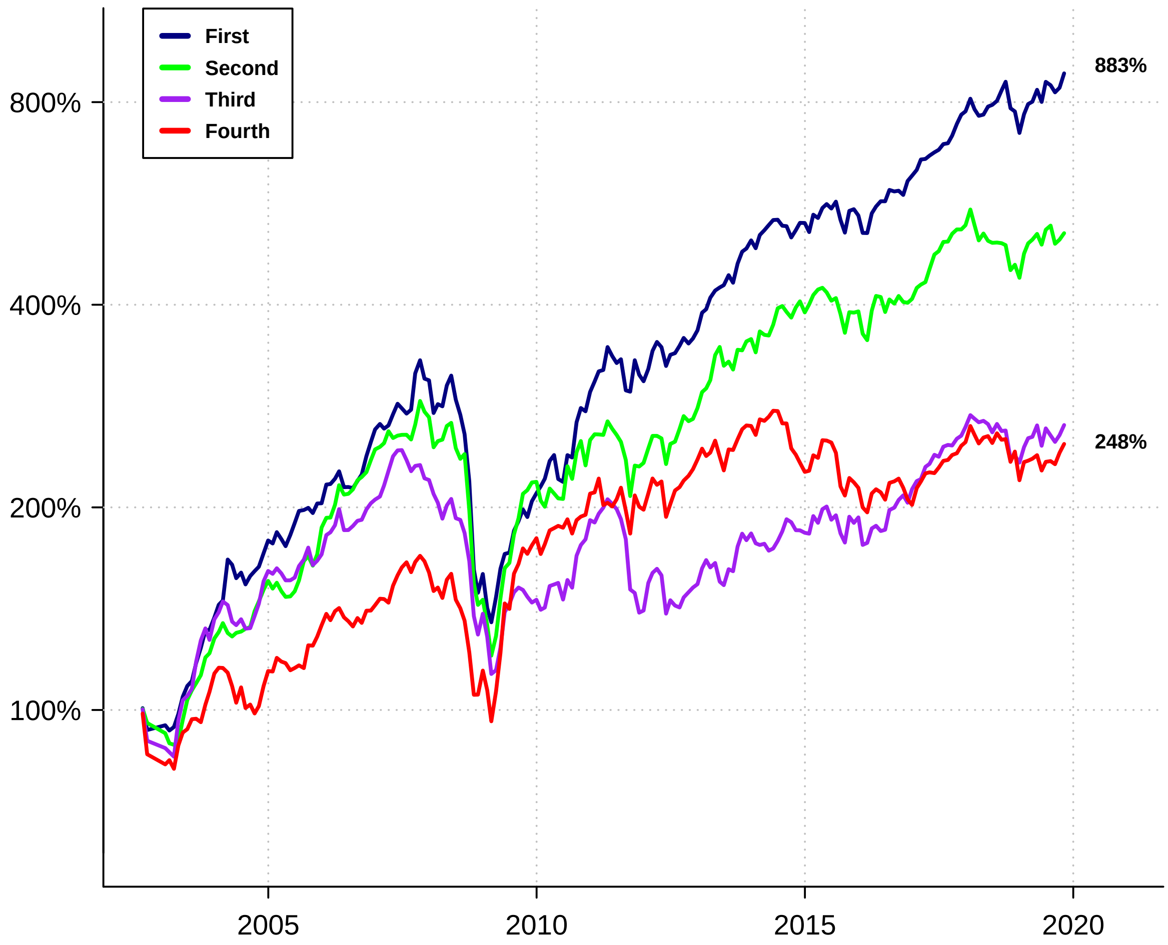

Figure 2. Relative Sentiment Cross-Sectional Returns, 2002-2019: This figure plots the cumulative total returns corresponding to the rank order of relative sentiment along the four-region cross-section. At each month-end, the composite equity allocations (our proxies for relative sentiment) of the top 50 strategies in each region were lined up and ranked. Then, the subsequent monthly returns were bucketed according to their prior month-end rank position and compounded to produce the chart above.

Table VII. Cumulative Total Returns Along Three-Region Cross-Section: This table shows the cumulative total return of each relative sentiment rank position along a three-region cross-section. At each month-end, the relative sentiment of the three regions is ranked and the corresponding forward monthly return of each region is assigned to a bucket—low, middle, or high—based on the region's prior month-end relative sentiment rank. Relative sentiment in each region for a given month is proxied by the region's equity allocation for that month, which itself is found by averaging the corresponding equity allocations of the region's top 50 relative sentiment strategies. The last column in the table indicates whether the cumulative total returns are a monotonically increasing function of the relative sentiment rank position. (Note: results correspond to combinatorial cross-validation case CCV{15,2}.)

3-Region Combinations			Total Return by Rank			
			Low	Middle	High	L < M < H?
USA	EUR	JPN	142%	335%	760%	TRUE
USA	EUR	AEJ	220%	561%	586%	TRUE
USA	JPN	AEJ	188%	288%	964%	TRUE
EUR	JPN	AEJ	169%	274%	807%	TRUE

Table VIII. Annualized Return Spreads Based on Head-to-Head Relative Sentiment Rankings: This table presents annualized average monthly return spreads between regional equity markets based on their head-to-head relative sentiment rankings. That is, for two regions, A and B, we subtract the monthly equity market return of B from the monthly equity market return of A, when A has higher relative sentiment (and vice versa when B has higher relative sentiment). We average over all such instances of that relative sentiment orientation and annualize the result. Relative sentiment in each region for a given month is proxied by the region's equity allocation for that month, which itself is found by averaging the corresponding equity allocations of the region's top 50 relative sentiment strategies. (Note: results correspond to combinatorial cross-validation case CCV{15,2}.)

	Higher-Ranked Region			
Lower-Ranked Region	USA	EUR	JPN	AEJ
USA		3.9%	9.3%	1.3%
EUR	4.1%		7.7%	2.4%
JPN	10.2%	5.6%		2.9%
AEJ	4.8%	7.4%	10.4%	

outperformed in the following month at a rate of 4.1 percentage points per annum. All values in the table are positive, meaning the region with higher-ranked relative sentiment tended to outperform the region with lower-ranked relative sentiment over a one-month horizon. This suggests relative sentiment might indeed have predictive power along the cross-section.

These results are consistent with Micaletti (2019), which reported similar effects using a positions-based relative sentiment indicator derived from the Commitments of Traders report.

Machine Learning Algorithms

One notable observation with respect to the machine learning algorithms is that the algorithms responsible for producing the best-performing strategies tended to be the same ones from region to region. For the U.S., Europe, and Japan, the best strategies arose from random forests and generalized boosted models, with the occasional appearance of $\nu-$SVM. For Asia ex-Japan, the best strategies arose from $\nu-$ and $C-$SVM, with the occasional random forest thrown in. This consistency suggests there might be some coherent structure to the data that these algorithms are uniquely suited to uncover.

Table IX and X list the top performing algorithms in the U.S. for CCV{15,2} and CCV{24,2}, respectively. Similar tables for Europe, Japan, and Asia ex-Japan may be found in Appendices A, B, and C, respectively.

Table IX. Best-performing machine learning algorithms, USA, CCV{15,2}: This table lists the best-performing machine learning algorithms in the U.S. (USA) for combinatorial cross-validation case CCV{15,2}, along with their factor subsets and relevant parameters.

						Additional Parameters			
Region	CCV	Rank	Method	Type	Factor Combo	Kernel/ Distribution	ν/ϵ	Cost	γ
USA	{15,2}	1	random forest	regression	7				
USA	{15,2}	2	random forest	regression	8				
USA	{15,2}	3	gbm	regression	4	gaussian			
USA	{15,2}	4	gbm	regression	1	gaussian			
USA	{15,2}	5	random forest	regression	2				
USA	{15,2}	6	gbm	regression	2	gaussian			
USA	{15,2}	7	gbm	regression	8	gaussian			
USA	{15,2}	8	random forest	regression	4				
USA	{15,2}	9	gbm	regression	3	gaussian			
USA	{15,2}	10	random forest	regression	1				

Table X. Best-performing machine learning algorithms, USA, CCV{24,2}: This table lists the best-performing machine learning algorithms in the U.S. (USA) for combinatorial cross-validation case CCV{24,2}, along with their factor subsets and relevant parameters.

						Additional Parameters			
Region	CCV	Rank	Method	Type	Factor Combo	Kernel/ Distribution	ν/ϵ	Cost	γ
USA	{24,2}	1	gbm	regression	4	gaussian			
USA	{24,2}	2	gbm	regression	2	gaussian			
USA	{24,2}	3	gbm	regression	5	gaussian			
USA	{24,2}	4	gbm	regression	1	gaussian			
USA	{24,2}	5	random forest	regression	7				
USA	{24,2}	6	gbm	classification	1	bernoulli			
USA	{24,2}	7	gbm	classification	1	adaboost			
USA	{24,2}	8	random forest	classification	2				
USA	{24,2}	9	random forest	regression	8				
USA	{24,2}	10	gbm	regression	7	gaussian			

III. Conclusion

Past work on relative sentiment, whether direct or indirect, has unearthed substantial evidence of its predictive power for equity markets over intermediate time horizons. The present study, which investigates relative sentiment factors derived from Sentix economic sentiment indices (for the U.S., Europe, Japan, and Asia ex-Japan), corroborates those findings and adds to that body of evidence.

Sentix-based relative sentiment factors, coupled with certain machine learning models, appear to generate statistically significant tactical asset allocation strategies for regional equity markets. Of the nearly 1000 relative sentiment strategies tested, dozens to several hundred (depending on the region) registered as significant after adjusting for data snooping. The levels of significance ranged from moderate in Asia-ex Japan and Europe, to strong in Japan, to extremely strong in the U.S.

In three of the four regions tested, relative sentiment markedly outperformed component sentiment (i.e., where the institutional and individual sentiments were considered as separate factors), and was roughly equivalent to component sentiment in the fourth region (although in this region, too, relative sentiment was stronger at more extreme levels of significance). This suggests the way in which different investor classes are situated *relative* to one another may contain more useful information than just their respective standalone orientations.

Relative sentiment also exhibited considerable economic significance. The composites of the top 10 relative sentiment strategies in each region outperformed their respective benchmarks by anywhere from 400 to 700 basis points per annum with higher Sharpe ratios and lower drawdowns. (Such levels would likely survive any realistic transaction-cost assumptions.)

Moreover, it again appears as though relative sentiment may provide more predictive power than time-series momentum, confirming the results in Micaletti (2018). Regional equity returns were higher (lower) when relative sentiment was higher (lower), regardless of the state of the region's time-series momentum. This result held for all regions.

Notably, relative sentiment seems particularly adept at identifying when to take equity market exposure during periods of negative momentum. The average annualized market return (across all regions) when momentum was negative and relative sentiment was positive was 27%, whereas when both momentum and relative sentiment were negative, the average annualized return was -23%—a spread of 50 percentage points depending on the polarity of relative sentiment. A similar (though not as dramatic) result was observed when momentum was positive.

Beyond demonstrating the ability to beneficially adjust equity allocations within regions, relative sentiment also demonstrated the ability to adjust equity allocations *across* regions (affirming the results in Micaletti (2019)). The rank ordering of the regions' monthly equity allocations—our proxy for relative sentiment—was a strong predictor of relative equity market returns one-month forward. This held for all 2-, 3-, and 4-region cross-sections.

Finally, we observe the best performing relative sentiment strategies from region to region tended to be produced by the same handful of machine learning algorithms (namely, generalized boosted models, random forests, and certain types of support vector machines). This consistency suggests there might be some underlying structure to the data that these algorithms are uniquely suited to uncover.

In light of the foregoing results, coupled with relative sentiment's existing body of evidence, it might be time to consider whether relative sentiment merits recognition as a legitimate market anomaly—one that is on par with, or perhaps even supersedes, value and momentum for purposes of tactical asset allocation.

Appendix A.

Table XI. Multiple Hypothesis Testing, Relative Sentiment vs. Component Sentiment, EUR: The following tables present multiple hypothesis testing (MHT) results for both relative sentiment and component sentiment strategies in Europe (EUR). The results correspond to different combinatorial cross-validation cases (CCV{15,2} and CCV{24,2}) as well as to different tactical strategies (equities + bonds and equities + cash). The reported values represent the numbers of strategies (from the pool of 990 tested) that reject the null hypothesis of no statistical significance (at various levels of the MHT parameter α). As MHT can be performed in numerous ways, the numbers presented here are averages, computed from the outputs of several different MHT implementations.

(a) EUR, Relative Sentiment Strategies

		Average Number of Null-Hypothesis Rejections α					
Equities+	CCV	0.01	0.02	0.05	0.10	0.15	0.30
Bonds	{15,2}	2	5	10	18	25	48
Bonds	{24,2}	0	0	0	2	3	6
Cash	{15,2}	2	3	8	16	21	39
Cash	{24,2}	0	0	0	1	2	5

(b) EUR, Component Sentiment Strategies

		Average Number of Null-Hypothesis Rejections α					
Equities+	CCV	0.01	0.02	0.05	0.10	0.15	0.30
Bonds	{15,2}	0	0	0	1	4	20
Bonds	{24,2}	0	0	0	0	0	0
Cash	{15,2}	0	0	3	5	11	34
Cash	{24,2}	0	0	0	0	0	2

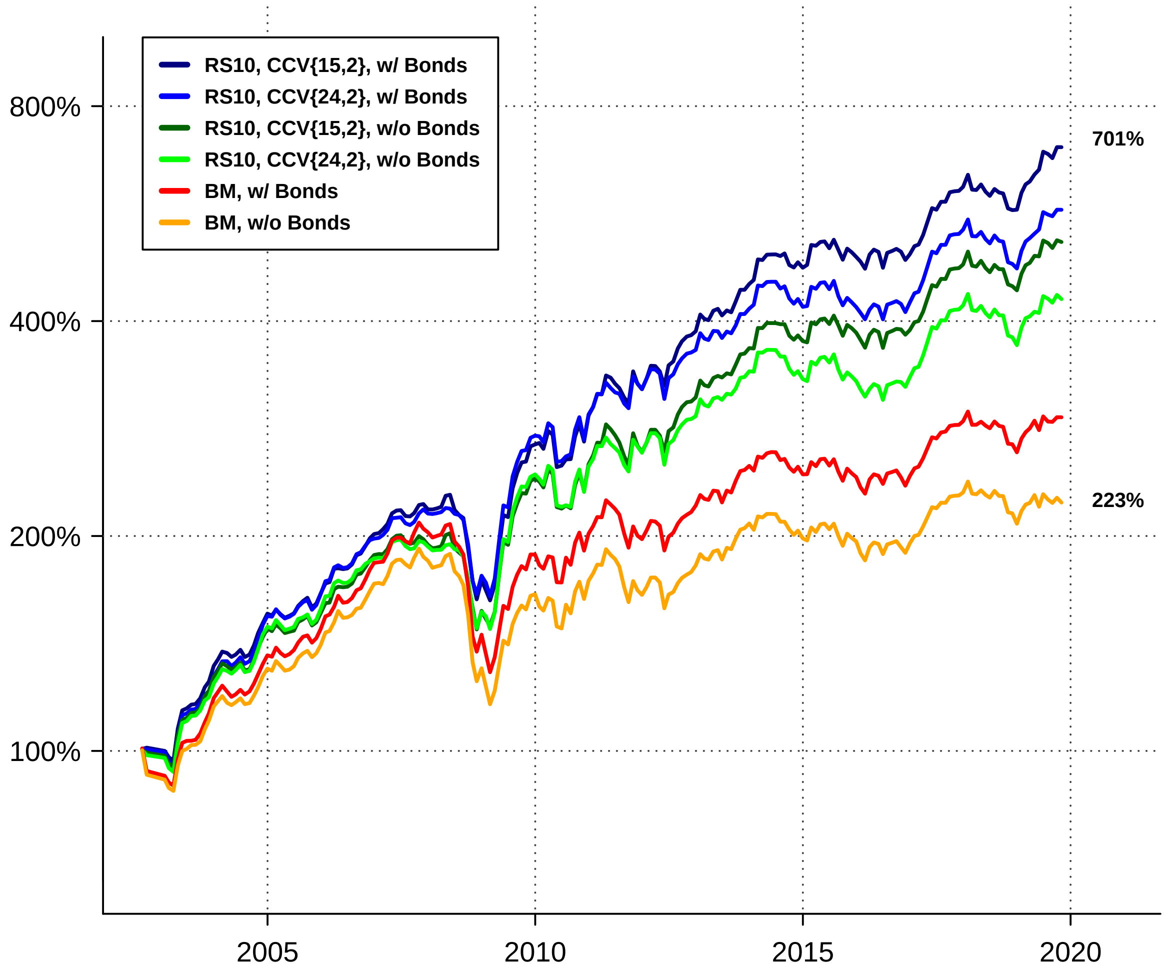

Figure 3. Relative Sentiment Performance, EUR, 2002-2019: Composite total returns of the top 10 Europe (EUR) relative sentiment strategies (RS10) for CCV{15,2} and CCV{24,2}—both with and without bonds—plotted against their respective benchmarks (BM)

Table XII. Relative Sentiment Performance Statistics, EUR, 2002-2019: Performance statistics of the composite returns of the top 10 Europe (EUR) relative sentiment strategies (RS10) for CCV{15,2} and CCV{24,2}—both with and without bonds. Corresponding benchmark (BM) statistics included for comparison.

Strategy	Bonds	CCV	Total Return	CAGR	Sharpe Ratio	Maximum Drawdown	Average Equity Allocation	Average Monthly Turnover
RS10	With	{15,2}	701%	12.1%	0.95	−28.9%	64.7%	27.5%
RS10	With	{24,2}	573%	10.8%	0.83	−24.7%	63.6%	26.2%
RS10	Without	{15,2}	516%	10.1%	0.80	−26.7%	64.5%	27.6%
RS10	Without	{24,2}	429%	8.9%	0.69	−25.1%	63.6%	26.2%
BM	With	{15,2}	293%	6.5%	0.51	−38.3%	64.2%	
BM	Without	{24,2}	223%	4.8%	0.37	−39.5%	64.1%	

Table XIII. Best-performing machine learning algorithms, EUR, CCV{15,2}: This table lists the best-performing machine learning algorithms in Europe (EUR) for combinatorial cross-validation case CCV{15,2}, along with their factor subsets and relevant parameters.

						Additional Parameters			
Region	CCV	Rank	Method	Type	Factor Combo	Kernel/ Distribution	ν/ϵ	Cost	γ
EUR	{15,2}	1	random forest	regression	7				
EUR	{15,2}	2	nu-SVM	regression	7	radial	0.3	100	1
EUR	{15,2}	3	nu-SVM	regression	7	radial	0.3	10	2
EUR	{15,2}	4	gbm	regression	4	gaussian			
EUR	{15,2}	5	nu-SVM	regression	7	radial	0.3	100	2
EUR	{15,2}	6	gbm	regression	1	gaussian			
EUR	{15,2}	7	gbm	regression	8	gaussian			
EUR	{15,2}	8	random forest	classification	8				
EUR	{15,2}	9	nu-SVM	regression	7	radial	0.2	100	2
EUR	{15,2}	10	C-SVM	classification	7	radial		10	2

Table XIV. Best-performing machine learning algorithms, EUR, CCV{24,2}: This table lists the best-performing machine learning algorithms in Europe (EUR) for combinatorial cross-validation case CCV{24,2}, along with their factor subsets and relevant parameters.

						Additional Parameters			
Region	CCV	Rank	Method	Type	Factor Combo	Kernel/ Distribution	ν/ϵ	Cost	γ
EUR	{24,2}	1	nu-SVM	regression	7	radial	0.3	100	1
EUR	{24,2}	2	nu-SVM	regression	7	radial	0.3	10	2
EUR	{24,2}	3	random forest	regression	7				
EUR	{24,2}	4	gbm	regression	4	gaussian			
EUR	{24,2}	5	gbm	regression	3	gaussian			
EUR	{24,2}	6	gbm	regression	11	gaussian			
EUR	{24,2}	7	gbm	regression	10	gaussian			
EUR	{24,2}	8	nu-SVM	regression	7	radial	0.3	100	2
EUR	{24,2}	9	gbm	regression	5	gaussian			
EUR	{24,2}	10	gbm	regression	1	gaussian			

Appendix B.

Table XV. Multiple Hypothesis Testing, Relative Sentiment vs. Component Sentiment, JPN: The following tables present multiple hypothesis testing (MHT) results for both relative sentiment and component sentiment strategies in Japan (JPN). The results correspond to different combinatorial cross-validation cases (CCV{15,2} and CCV{24,2}) as well as to different tactical strategies (equities + bonds and equities + cash). The reported values represent the numbers of strategies (from the pool of 990 tested) that reject the null hypothesis of no statistical significance (at various levels of the MHT parameter α). As MHT can be performed in numerous ways, the numbers presented here are averages, computed from the outputs of several different MHT implementations.

(a) JPN, Relative Sentiment Strategies

		Average Number of Null-Hypothesis Rejections					
		α					
Equities+	CCV	0.01	0.02	0.05	0.10	0.15	0.30
Bonds	{15,2}	4	9	21	38	56	91
Bonds	{24,2}	0	1	5	11	23	71
Cash	{15,2}	3	6	13	24	35	68
Cash	{24,2}	0	0	5	10	14	56

(b) JPN, Component Sentiment Strategies

		Average Number of Null-Hypothesis Rejections					
		α					
Equities+	CCV	0.01	0.02	0.05	0.10	0.15	0.30
Bonds	{15,2}	0	0	1	2	4	33
Bonds	{24,2}	0	0	0	0	1	10
Cash	{15,2}	0	0	0	1	2	11
Cash	{24,2}	0	0	0	0	0	1

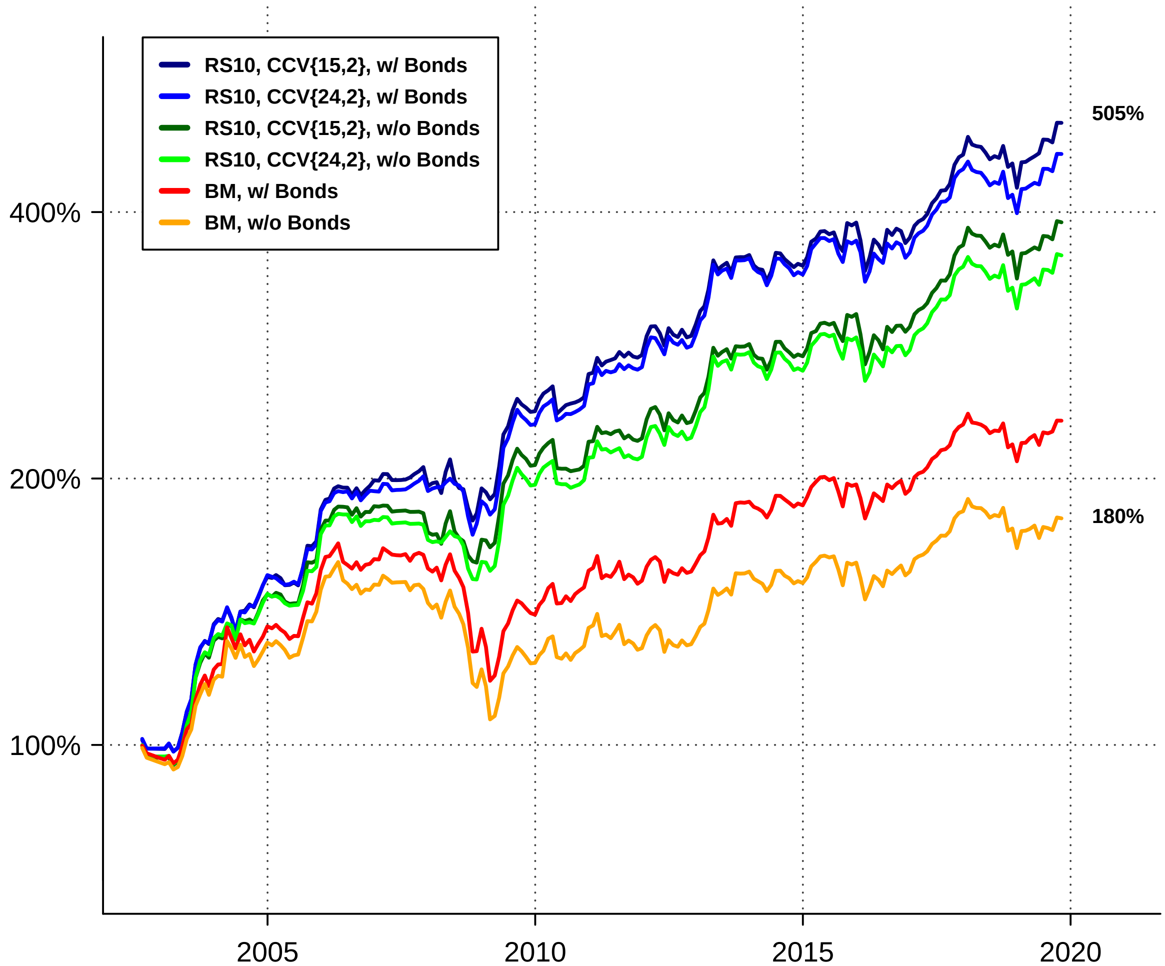

Figure 4. Relative Sentiment Performance, JPN, 2002-2019: Composite total returns of the top 10 Japan (JPN) relative sentiment strategies (RS10) for CCV{15,2} and CCV{24,2}—both with and without bonds—plotted against their respective benchmarks (BM)

Table XVI. Relative Sentiment Performance Statistics, JPN, 2002-2019: Performance statistics of the composite returns of the top 10 Japan (JPN) relative sentiment strategies (RS10) for CCV{15,2} and CCV{24,2}—both with and without bonds. Corresponding benchmark (BM) statistics included for comparison.

Strategy	Bonds	CCV	Total Return	CAGR	Sharpe Ratio	Maximum Drawdown	Average Equity Allocation	Average Monthly Turnover
RS10	With	{15,2}	505%	10.0%	0.90	−14.7%	65.7%	29.4%
RS10	With	{24,2}	465%	9.5%	0.87	−14.1%	64.6%	28.5%
RS10	Without	{15,2}	390%	8.3%	0.76	−13.8%	65.0%	29.3%
RS10	Without	{24,2}	357%	7.8%	0.72	−15.7%	64.6%	28.5%
BM	With	{15,2}	232%	5.1%	0.45	−30.1%	65.1%	
BM	Without	{24,2}	180%	3.5%	0.30	−33.6%	64.8%	

Table XVII. Best-performing machine learning algorithms, JPN, CCV{15,2}: This table lists the best-performing machine learning algorithms in Japan (JPN) for combinatorial cross-validation case CCV{15,2}, along with their factor subsets and relevant parameters.

						Additional Parameters			
Region	CCV	Rank	Method	Type	Factor Combo	Kernel/ Distribution	ν/ϵ	Cost	γ
JPN	{15,2}	1	random forest	regression	4				
JPN	{15,2}	2	random forest	regression	8				
JPN	{15,2}	3	random forest	regression	11				
JPN	{15,2}	4	gbm	regression	11	gaussian			
JPN	{15,2}	5	random forest	regression	1				
JPN	{15,2}	6	gbm	regression	8	gaussian			
JPN	{15,2}	7	gbm	regression	3	gaussian			
JPN	{15,2}	8	nu-SVM	regression	8	radial	0.2	100	2
JPN	{15,2}	9	gbm	regression	6	gaussian			
JPN	{15,2}	10	random forest	regression	3				

Table XVIII. Best-performing machine learning algorithms, JPN, CCV{24,2}: This table lists the best-performing machine learning algorithms in Japan (JPN) for combinatorial cross-validation case CCV{24,2}, along with their factor subsets and relevant parameters.

						Additional Parameters			
Region	CCV	Rank	Method	Type	Factor Combo	Kernel/ Distribution	ν/ϵ	Cost	γ
JPN	{24,2}	1	random forest	regression	4				
JPN	{24,2}	2	gbm	regression	7	gaussian			
JPN	{24,2}	3	gbm	regression	11	gaussian			
JPN	{24,2}	4	random forest	regression	1				
JPN	{24,2}	5	gbm	regression	1	gaussian			
JPN	{24,2}	6	random forest	regression	3				
JPN	{24,2}	7	gbm	classification	11	bernoulli			
JPN	{24,2}	8	random forest	regression	11				
JPN	{24,2}	9	random forest	regression	5				
JPN	{24,2}	10	random forest	regression	8				

Appendix C.

Table XIX. Multiple Hypothesis Testing, Relative Sentiment vs. Component Sentiment, AEJ: The following tables present multiple hypothesis testing (MHT) results for both relative sentiment and component sentiment strategies in Asia ex-Japan (AEJ). The results correspond to different combinatorial cross-validation cases (CCV{15,2} and CCV{24,2}) as well as to different tactical strategies (equities + bonds and equities + cash). The reported values represent the numbers of strategies (from the pool of 990 tested) that reject the null hypothesis of no statistical significance (at various levels of the MHT parameter α). As MHT can be performed in numerous ways, the numbers presented here are averages, computed from the outputs of several different MHT implementations.

(a) AEJ, Relative Sentiment Strategies

		Average Number of Null-Hypothesis Rejections α					
Equities+	CCV	0.01	0.02	0.05	0.10	0.15	0.30
Bonds	{15,2}	0	1	1	3	4	13
Bonds	{24,2}	4	6	9	15	19	33
Cash	{15,2}	0	1	1	2	3	8
Cash	{24,2}	4	5	9	13	17	31

(b) AEJ, Component Sentiment Strategies

		Average Number of Null-Hypothesis Rejections α					
Equities+	CCV	0.01	0.02	0.05	0.10	0.15	0.30
Bonds	{15,2}	0	0	0	0	1	14
Bonds	{24,2}	0	0	0	5	12	39
Cash	{15,2}	0	0	0	1	8	20
Cash	{24,2}	0	0	3	19	33	71

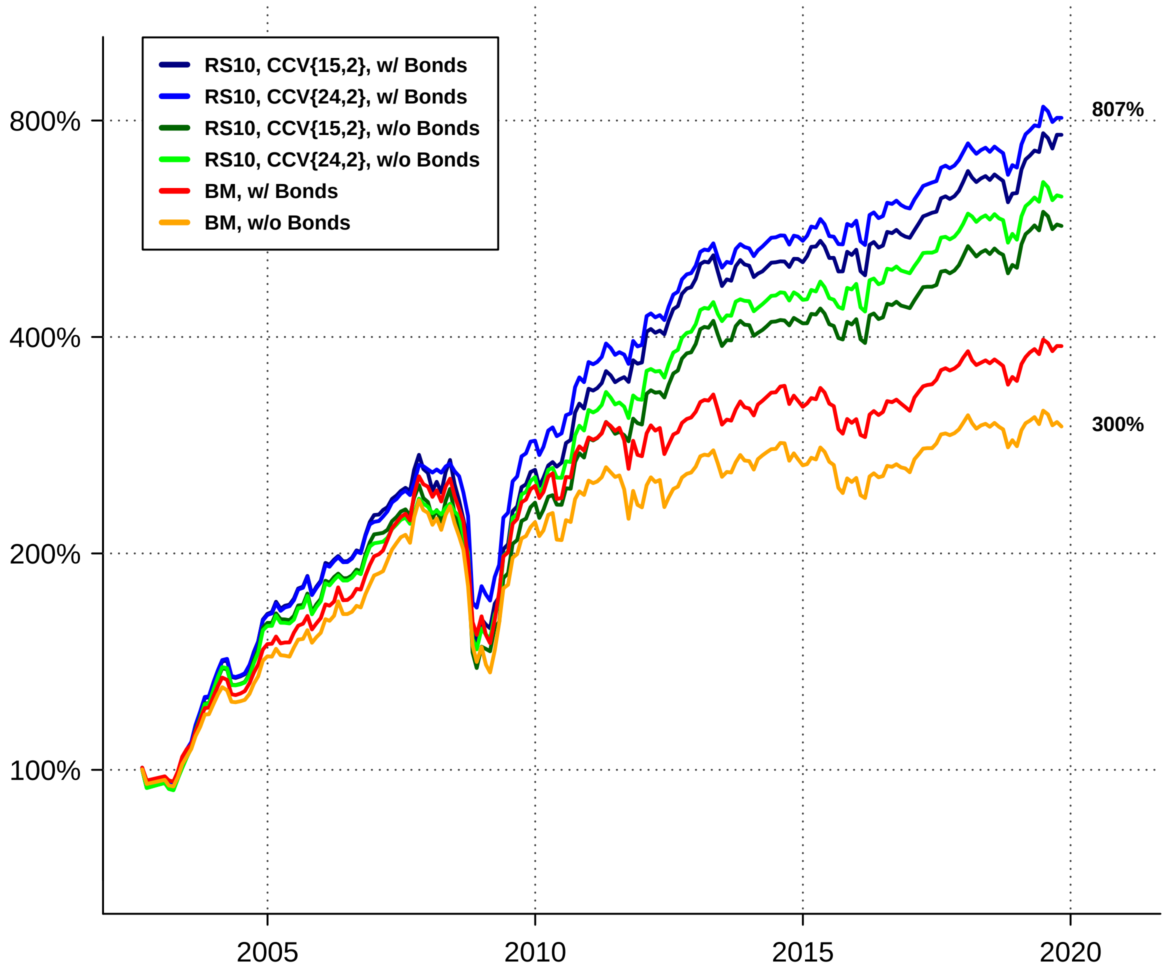

Figure 5. Relative Sentiment Performance, AEJ, 2002-2019: Composite total returns of the top 10 Asia ex-Japan (AEJ) relative sentiment strategies (RS10) for CCV{15,2} and CCV{24,2}—both with and without bonds—plotted against their respective benchmarks (BM)

Table XX. Relative Sentiment Performance Statistics, AEJ, 2002-2019: Performance statistics of the composite returns of the top 10 Asia ex-Japan (AEJ) relative sentiment strategies (RS10) for CCV{15,2} and CCV{24,2}—both with and without bonds. Corresponding benchmark (BM) statistics included for comparison.

Strategy	Bonds	CCV	Total Return	CAGR	Sharpe Ratio	Maximum Drawdown	Average Equity Allocation	Average Monthly Turnover
RS10	With	{15,2}	764%	12.7%	0.87	−44.9%	68.2%	25.5%
RS10	With	{24,2}	807%	13.1%	0.94	−36.9%	66.2%	24.7%
RS10	Without	{15,2}	571%	10.8%	0.77	−44.3%	67.9%	24.0%
RS10	Without	{24,2}	627%	11.4%	0.82	−38.3%	66.2%	25.3%
BM	With	{15,2}	389%	8.3%	0.60	−41.4%	67.2%	
BM	Without	{24,2}	300%	6.7%	0.49	−42.4%	67.1%	

Table XXI. Best-performing machine learning algorithms, AEJ, CCV{15,2}: This table lists the best-performing machine learning algorithms in Asia ex-Japan (AEJ) for combinatorial cross-validation case CCV{15,2}, along with their factor subsets and relevant parameters.

						Additional Parameters			
Region	CCV	Rank	Method	Type	Factor Combo	Kernel/ Distribution	ν/ϵ	Cost	γ
AEJ	{15,2}	1	C-SVM	classification	6	radial		10	2
AEJ	{15,2}	2	C-SVM	classification	6	radial		100	1
AEJ	{15,2}	3	nu-SVM	classification	6	radial	0.3		2
AEJ	{15,2}	4	C-SVM	classification	6	radial		100	2
AEJ	{15,2}	5	random forest	classification	6				
AEJ	{15,2}	6	nu-SVM	regression	6	radial	0.3	100	1
AEJ	{15,2}	7	nu-SVM	regression	6	radial	0.3	10	2
AEJ	{15,2}	8	nu-SVM	classification	3	radial	0.3		2
AEJ	{15,2}	9	nu-SVM	classification	6	radial	0.3		1
AEJ	{15,2}	10	C-SVM	classification	3	radial		10	2

Table XXII. Best-performing machine learning algorithms, AEJ, CCV{24,2}: This table lists the best-performing machine learning algorithms in Asia ex-Japan (AEJ) for combinatorial cross-validation case CCV{24,2}, along with their factor subsets and relevant parameters.

						Additional Parameters			
Region	CCV	Rank	Method	Type	Factor Combo	Kernel/ Distribution	ν/ϵ	Cost	γ
AEJ	{24,2}	1	C-SVM	classification	6	radial		10	2
AEJ	{24,2}	2	nu-SVM	classification	6	radial	0.3		2
AEJ	{24,2}	3	C-SVM	classification	6	radial		100	1
AEJ	{24,2}	4	C-SVM	classification	6	radial		100	2
AEJ	{24,2}	5	random forest	classification	6				
AEJ	{24,2}	6	nu-SVM	classification	6	radial	0.3		1
AEJ	{24,2}	7	nu-SVM	regression	6	radial	0.3	10	2
AEJ	{24,2}	8	eps-SVM	regression	6	radial	0.1		2
AEJ	{24,2}	9	nu-SVM	regression	6	radial	0.3	100	1
AEJ	{24,2}	10	random forest	regression	7				

REFERENCES

Antonacci, Gary, 2014, *Dual Momentum Investing: An Innovative Strategy for Higher Returns and Lower Risk* (McGraw-Hill Education).

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

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

Beaumont, Rob, Macro van Dael, Bart Frijns, and Thorsten Lehnert, 2006, On individual and institutional noise trading, Working Paper, Department of Finance Maastrict Universityl.

Chakravorty, Gaurav, Ankit Awasthi, and Brandon Da Silva, 2018, Deep learning for global tactical asset allocation, SSRN Electronic Journal, URL: https://ssrn.com/abstract=3242432.

Dahlquist, Magnus, and Campbell R. Harvey, 2001, Global tactical asset allocation, SSRN Electronic Journal, URL: https://ssrn.com/abstract=795376.

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

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

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.

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.

Heiden, Sebastian, Christian Klein, and Berhnhard Zwergel, 2013, Beyond fundamentals: Investor sentiment and exchange rate forecasting, *European Financial Management* 19, 558–578.

Hengelbrock, Jordis, Eric Theissen, and Christian Westheide, 2013, Market response to investor sentiment, *Journal of Business Finance & Accounting* 40, 901–917.

Hsu, Yu-Chin, Chung-Ming Kuan, and Meng-Feng Yen, 2014, A generalized stepwise procedure with improved power for multiple inequalities testing, *Journal of Financial Econometrics* 12, 730–755.

López de Prado, Marcos, 2018, *Advance in Financial Machine Learning* (John Wiley & Sons, Inc.).

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

Makridakis, Spyros, Evangelos Spiliotis, and Vassilios Assimakopoulos, 2018, Statistical and machine learning forecasting methods: Concerns and ways forward., *PLoS ONE* 13.

Micaletti, Raymond, 2018, Want smart beta? follow the smart money: Market and factor timing using relative sentiment, SSRN Electronic Journal, URL: https://ssrn.com/abstract=3164081.

Micaletti, Raymond, 2019, The smart money indicator: A new risk-management tool, Alpha Architect, URL: https://alphaarchitect.com/2019/02/08/relative-sentiment-a-unique-market-timing-tool-that-isnt-trend-following.

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.

Oflus, Kemal, 2014, Quantitative tactical asset allocation using ensemble machine learning methods, SSRN Electronic Journal, URL: https://ssrn.com/abstract=2438522.

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

Pollard, Katherine S., Sandrine Dudoit, and Mark J. van der Laan, 2005, *Multiple Testing Procedures: R multtest Package and Applications to Genomics, in Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (Springer).

Romano, Joseph P., and Michael Wolf, 2007, Control of generalized error rates in multiple testing, *Annals of Statistics* 35, 1378–1408.

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

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.

White, Halbert, 2000, A reality check for data snooping, *Econometrica* 68, 1097–1126.

Zakamulin, Valeriy, 2014, The real-life performance of market timing with moving average and time-series momentum rules, *Journal of Asset Management* 15, 261–278.

Zwergel, Bernhard, and Christian Klein, 2006, On the predictive power of sentiment - why institutional investors are worth their pay, SSRN Electronic Journal, URL: https://ssrn.com/abstract=965890.