New Online Investor Sentiment and Asset Returns

Zongwu Cai[a], Pixiong Chen[a,b]
[a]Department of Economics, University of Kansas
[b]Division of Quantitative Analytics, Wells Fargo Bank

November 23, 2022

Abstract

This paper proposes two data-driven econometric approaches to construct online investor sentiment indices based on internet search queries, which are built by the partial least squares and LASSO methods, respectively. By examining the relationship between investor sentiment and stock risk premium on overall market level, our empirical findings are that these sentiment indices have predictive power both in and out of sample, and the out-of-sample predictability of the online investor sentiment indices proposed by the paper is robust for different horizons. Moreover, our investor sentiment indices are also able to predict the returns of cross-sectional characteristics portfolios. This predictability based on investor sentiment has economic value since it improves portfolio performance, in terms of certainty equivalent return gain and Sharpe ratio, for investors who conduct the optimal asset allocation.

Keywords: Asset return; Data-driven method; Online investor sentiment; Partial least squares; Portfolio choice.

Classification JEL: C22, C53, G11, G17

1 Introduction

Understanding the time-varying movement of conditional market risk premium is one of the major challenges for academics and financial professionals as addressed by Spiegel (2008) and Cochrane (2011). The risk premium predictability helps testing market efficiency and improving asset allocation performance. Theoretically, the dynamic expected future stock returns are modeled as functions of state variables[1] that can detect real economic fluctuations, as argued in Rapach and Zhou (2013). As a consequence, the key is how to select the state variables from the numerous candidates that potentially contain relevant information. There is a long list of discovered state variables from preceding literature; see, for instance, Baker and Wurgler (2006, 2007), Welch and Goyal (2008), Green et al. (2013), and references therein. For example, Baker and Wurgler (2006), among others, claim that the stock return is influenced by investor sentiment, which drives the stock price to deviate from the fundamental level due to the irrational overreactions of market participants. Since investor sentiment is not directly observable, the main issue is how to measure and quantify its links to stock return. To provide a possible solution to these problems, there are some studies in literature to explore proxies from market-based, survey-based, and search-based data, as pointed out by Da et al. (2015), but have not reached a consensus on the measurement of investor sentiment.

In our paper, we explore the time-series relationship between online investor sentiment and risk premium in the stock market using the internet search queries, which are, commonly high-dimensional, the proxies of investor assessment from internet. For this purpose, we propose using two different methods, the partial least squared (PLS) and the LASSO, to directly and efficiently measure the online investor sentiment. Specifically, the information set of investor sentiment proxies is the internet search queries of households from Google Trends. These two varietal investor sentiment indices, as expected, are able to explain the time-varying movement of expected future excess returns of aggregate stock and cross-sectional stocks. The predictability of these two variants is examined under the linear univariate prediction model framework and evaluated by an out-of-sample R^2 statistic de-

[1] The state variables used in forecasting are also denoted as predictors.

fined by Campbell and Thompson (2008). In the previous literature, Da et al. (2015) select the sentiment proxies based on the t-statistics from a univariate regression of each change of search volume index (SVI), on contemporaneous stock return, and then aggregate such variables into sentiment index with equal weights. However, the approach proposed by Da et al. (2015) does not consider correlations among the search volume indices and is not in data-driven fashion. Instead of processing the selection and aggregation of the mass online information manually like the previous research, we modify the procedure of selection and aggregation by deploying data-driven variable selection approaches. We first propose the PLS method to extract the online investor sentiment, introduced first by Wold (1966, 1975) and extended by Kelly and Pruitt (2013, 2015) and Huang et al. (2015) with applications in economics and finance. This approach helps extracting the most relevant information from the online investor sentiment proxies and filtering out the noise as well as aggregating into a one-dimensional index. Meanwhile, another index is constructed by using the LASSO type approach proposed by Tibshirani (1996), which adds an L_1 norm penalty term to the criterion function of regression, to select the sentiment proxies and assign weights for combining a one-dimensional online investor sentiment index. It is well known in the literature that the LASSO type method leads to sparse estimation results and allocates different non-zero weights to each selected sentiment proxies. The weighted sum of them is our another one-dimensional online investor sentiment index. First, we examine the predictive power of online investor sentiment indices. Then, to evaluate the economic significance of new indices, we test the performance of portfolios which are based on investor sentiment indices under the Markowitz (1952) paradigm where a mean-variance investor optimally allocates the wealth across a risky asset and a risk-free asset. The economic value of predictability is measured by the gain of certainty equivalent return (CER) and Sharpe ratio. Besides the aggregated market, we also apply different characteristic portfolios to explore the cross-sectional impacts on online investor sentiment. Exploring the predictive ability of online investor sentiment indices on portfolios sorted by industries, momentum, book-to-market ratio, and size helps us to understand the economic sources of return predictability.

Our research is closely related to the literature on stock return predictability and investor sentiment measurements. First, empirical results provide evidence that the mean, variance and higher moments of stock returns are all time-varying as elaborated in Schwert (1989), Hansen (1994), and Ang and Bekaert (2007). Furthermore, there is an evidence of showing predictability of the mean and volatility of stork returns. Finally, Rapach and Zhou (2013) and Gu et al. (2020) review the literature on forecasting stock return based on investor sentiments.

Recognizing the effect of investor sentiment by researchers and market participants can be traced back as early as the book by Keynes (1936). Indeed, De Long at al. (1990) introduce an academic framework to explain the influence of irrational investors. Since the stock market participants are not as rational as theoretically assumed, their investment decisions are not always rationally based on the fundamental analyses. The irrational investment behaviors vary with time and circumstance. When the investor sentiment is high (low), investors tend to overbuy (oversell), which deviates the stock price from the fundamental level. However, the investor sentiment is unobservable so that one has to infer or extract it from observable proxies which ideally reflect the emotional changes of investors relating to the stock return expectation. For instance, Baker and Wurgler (2006, 2007) build an investor sentiment index from six financial market indices which include the closed-end fund discount, market turnover, number of IPOs, average first-day return on IPOs, equity share of new issuances, and the log difference in book-to-market ratios between dividend payers and dividend non-payers, and evidence the predictive power of their investor sentiment index. Further, Huang et al. (2015) propose an alternative index from these six similar sentiment proxies by applying the PLS method, and they argue that their index can improve the predictability of their sentiment index. Also, sentiment measurement has also been previously established by survey data as in Brown and Cliff (2004) and Lemmon and Portniaguina (2006). The investor sentiment surveys are not reliable since the survey takers tend to have good answers instead of true answers. Exploring not only the numeric market data, but also linguistic text data is necessary for constructing a measurement of investor sentiment. Therefore, Tetlock (2007) and Tetlock, at al. (2008) develop the sentiment index from traditional news media, such as Wall Street Journal, based on a word classification dictionary, while Jiang et al. (2019) construct a sentiment index based on the corporate financial disclosures. Recently, Zhou (2018) provides a comprehensive review to the investor sentiment literature.

Loughran and McDonald (2011) develop an updated word dictionary for capturing the sentiment in financial text. As internet has grown quickly in recent years, it has been a wildly-used data source on real-time economic activity. Online investor sentiment is a sub-area of investor sentiment research that extracts the sentiment from internet information. The work regarding abstracting, processing, and aggregating for a large volume of information is the main challenge. Studies have introduced online media-based and search-based sentiment indices. Actually, Antweiler and Frank (2004) find the posts on online message boards can predict well market return and volatility, while Siganos et al. (2014, 2017) examine the linkage between sentiment level and stock return by exploiting data from Facebook. Recently, Renault (2017) explores StockTwits, a social microblogging platform, and extracts the investor sentiment which has intraday predictive power. Since Choi and Varian (2012) introduce the Google search volume index to predict economic indicators, search-based data is used for financial empirical research on constructing investor sentiment; see, for example, Da et al. (2011, 2015), Andrei and Hasler (2015), and Chen et al. (2019). Compared with market-based or survey-based data, search-based data has advantages. A set of search-based data can be collected in a customized frequency with a small time lag, and the variables contained in this dataset can be large in number and highly customized. Da et al. (2015) find the link between the search behavior of households and the capital market. By quantifying the concerns about the economy with the search inquiry volume of economic terms, they build a measurement of online investor sentiment. Their aggregated index, which is termed as financial and economic attitudes revealed by search (FEARS) index, is a significant stock return predictor, and they also underscore the importance of researching data in the investor sentiment measurement application. In addition, the SVI from Google Trends is also used in pricing different assets like gold or crude oil studied by Balcilar et al. (2017) and Qadan and Nama (2018), respectively. We add to the expanding literature with extracting investor sentiment from search volume index from Google Trends to forecast market risk premium.

The main contribution in this paper is that we complement the empirical studies on investor sentiment by constructing two new online investor sentiment indices. We show the influence of these two new sentiment indices on financial markets and find a strong evidence of predictive power on both aggregate stock market return and cross-sectional stock returns. Our empirical results are in line with the previous literature on online investor sentiment. Although our research is closely related to Da et al. (2015), we exploit the similar online investor sentiment proxies, but we apply different data-driven methods to build online investor sentiment indices. The sentiment indices based on the method proposed by Da et al. (2015) is set as a benchmark for forecasting evaluation. In our research, however, we apply the partial least squares method to build investor sentiment indices as in Huang et al. (2015), but our investor sentiment indices are built from different sentiment proxies. In Huang et al.(2015), they use 6 Baker and Wurgler (2006) style sentiment proxies which are valid stock return predictors. Our online sentiment proxies have much higher dimension and are collected from the internet. More importantly, one single online sentiment proxy can not be a significant predictor of asset return, which requires multiple approaches to capture investor sentiment when using these proxies for forecasting.

The main motivations of this study come from academic research and financial applications. It is well documented that investor sentiment drives a portion of the time-varying market risk premium as a consequence of the irrational operations of market participants. However, investor sentiment can not be directly observed. Thus, the challenges are how to measure the investor sentiment and quantify its influence on the stock market, so that the measurements differ across studies. Da et al. (2015) illustrate the benefit of using search-based investor sentiment proxies, but we notice that the method they proposed may not necessarily capture the investor sentiment when a different data frequency is adopted. Therefore, we re-exploit the dataset and propose new sentiment indices for explaining the weekly dynamic of expected excess return on the aggregate stock market. Considering the weak predictive power of each single online investor sentiment proxy and the amount of proxies, we need to carefully apply data-driven methods in an efficient manner to construct new sentiment indices for the purpose of explaining the movement of market risk premium.

The rest of this paper is organized as follows. Section 2 describes in detail two data-driven econometric methods to construct sentiment indices from online sentiment proxies, together with the method as in Da et al. (2015). In Section 3, empirical results of predictability are reported. Finally, Section 4 concludes the paper.

2 Econometric Methods

2.1 Regression-Based Procedure

Da et al. (2015) introduce an idea of variable selection and aggregation for investor sentiment index. By running a simple linear regressions of each (information element variable) on stock returns, they decide the relationships among them and achieve those coefficient t-statistics as selection features. The aggregation is the average of those 30 investor sentiment proxies whose t-statistics are negative and have the largest absolute values. We can apply this method in two steps. At the first-step, we run an ordinary least squares (OLS) regression to estimate their coefficients,

$$r_{n,t} = \beta_{0,d} + \beta_{1,d} x_{d,t} + \mu_{n,t}, \tag{1}$$

to obtain $\hat{\beta}_{1,d}$ for $1 \leq d \leq D$, where $r_{n,t}$ denotes the contemporaneous return rate of asset n at time t and $x_{d,t}$ is the dth online investor sentiment proxy at time t. The regression (1) is used to identify the relationship of each information variable and stock return. At the next step, all $x_{d,t}$'s are sorted by the absolute values of t-statistics $\{|t_d|\}_{d=1}^D$ with $t_d = \hat{\beta}_{1,d}|/se(\hat{\beta}_{1,d})$ from the largest to the smallest. In the second-step, we construct the FEARS index as

$$\text{FEARS}_t^{reg} = \sum_{j=1}^{30} x_{(j),t}/\sqrt{30},$$

where $x_{(j),t}$ is the jth concordance of the jth order statistic of $\{|t_d|\}$, where orders run from the largest to the smallest, and FEARS_t^{reg} is the aggregated sentiment index by regression-based method. $1/\sqrt{30}$ is the normalized equal weight for index aggregation. We use FEARS^{reg} as a benchmark of sentiment index for comparing predicative efficiency. Clearly, one can see that FEARS^{reg} puts an equal weight to each information variable $x_{d,t}$ and it does not consider correlations among $\{x_{d,t}\}_{d=1}^D$ at time t. Finally, the selection based on the first sorted 30 online sentiment proxies is not data-driven so that it might not really present the true information.

2.2 LASSO Method

To extract the investor sentiment from the online investor sentiment proxies set in a data-driven fashion, we consider using a penalized method like LASSO type as in Tibshirani (1996) to accomplish the selection and aggregation task in an automatic manner. By comparing with the method which selects sentiment elements by t-statistics and aggregates into an equal-weighted index, the LASSO method uses all information variables simultaneously and picks up those correct among them, and allows more flexibility of weights of the selected sentiment elements to construct the investor sentiment index. For this purpose, we consider a contemporaneous return regression:

$$r_{n,t} = \beta_0 + \sum_{d=1}^{D} \beta_d x_{d,t} + u_{n,t}$$

with the penalized least squares error as

$$\text{PLSE}(\beta) = \sum_{t=1}^{T} \left[r_{n,t} - \beta_0 - \sum_{d=1}^{D} \beta_d \, x_{d,t} \right]^2 + \sum_{d=1}^{D} P_{\lambda_T}(|\beta_d|), \tag{2}$$

where $x_{d,t}$ is the dth information variable d at time t, $P_{\lambda_T}(|\beta_d|)$ is a penalty function, which, for the sake of simplicity, is taken to be the L_1 penalty function in our empirical study in Section 3[2], and λ_T is the hyper-parameter of the penalty. The sentiment index is combined as:

$$\text{FEARS}_t^{lasso} = \mathbf{x}_t^\top \tilde{\boldsymbol{\beta}}^{lasso},$$

where $\tilde{\boldsymbol{\beta}}^{lasso}$ is the normalized and shrunken coefficients $\boldsymbol{\beta}$ from penalized regression (2)[3], and $\mathbf{x}_t$ is the vector of all information variables observed at time t. When $\mathbf{x}_t$ is ultra dimensional, some screening and penalized procedures can be applied; see, for example, the review paper by Fan and Lv (2010).

2.3 PLS Approach

To extract the investor sentiment from the large (or very large) number of online investor sentiment proxies, we apply the PLS method, which is employed by Kelly and Pruitt (2013, 2015) to eliminate the common noise component and aggregate information from multiple variables, and is also applied by Huang et al. (2015) to create a sentiment index from several

[2] Of course, other types of penalty function can be used too, for example, the the smoothly clipped absolute deviation proposed in Fan and Li (2001).

[3] The normalization makes $\left\| \tilde{\boldsymbol{\beta}} \right\| \equiv 1$ and the shrinkage is for taking care of sparsity. For details, see Tibshirani (1996) or Fan and Li (2001).

traditional market-based sentiment proxies. The method is deployed in two steps of ordinary least squares (OLS) regressions. At the first-step, to decide the relation between sentiment proxies and stock return, for each information variable $x_{d,t}$, $d = 1, \ldots, D$, where $D = 170$ in our empirical studies, we run a time series regression of $x_{d,t}$ on contemporaneous realized return of asset n, $r_{n,t}$.

$$x_{d,t} = \theta_{0,d} + \theta_{1,d} r_{n,t} + \mu_{d,t}, \quad t = 1, \ldots, T,$$

which is the spirit of the inverse regression as studied in Li (1991) for dimension reduction to choose some *optimal* indices to characterize information variable $\{x_{d,t}\}_{d=1}^{D}$. At the second-step, at each time period t, we use a cross-sectional regression of information variable $x_{d,t}$ on estimated coefficients $\hat{\theta}_{1,d}$ passed by Step 1,

$$x_{d,t} = a_{0,t} + \alpha_{1,t}\, \hat{\theta}_{1,d} + \nu_{d,t}, \quad d = 1, \ldots, D,$$

which yields the estimate $\hat{\alpha}_{1,t}$ is the online investor sentiment index at time t, denoted by $\mathrm{FEARS}_t^{pls} = \hat{\alpha}_{1,t}$. Clearly, the PLS approach has an ability to deal with the case that D is very large.

3 Empirical Results

3.1 Data

The search query indices are the sentiment proxies that we use for constructing the sentiment index, and the stock market return and risk-free rate are applied in our analysis. In this section, we discuss the collecting and preparing procedure of data.

3.1.1 Search Query Index

Search query data series is provided by Google Trends which is powered by Google. A query data series, search volume index, represents the search volume of an input which can be a single word or a multi-word term. Such a SVI is rescaled by the historical maximum and ranges from 0 to 100. There are options to filter out results by a time period or a chosen region such as United States or worldwide. We use all 150 terms under the economic category including both positive and negative tones from General Inquirers Harvard IV-4

Dictionary, a widely used dictionary in the finance and textual analytics researches, like Tetlock (2007), Tetlock et al. (2008) and Da et al. (2015), Our basic term list includes words such as "rich," "savings," "subsidy," "gold," "crisis," "default" and "jobless." The query result returns the time series data of SVI at a specific frequency such as hourly, daily, weekly, or monthly. In addition, Google Trends also returns a list of terms related to the input. For example, when we search a word "contribution" in our term list, the top five related terms are "IRA contribution," "IRA," "401k," "401k contribution," and "Roth contribution." Table 1 shows the correlations of log difference of SVIs of these 6

Table 1: Correlations among the related terms.

	contribution	IRA contribution	IRA	401k	401k contribution	Roth contribution
contribution	1.00	0.83	0.80	0.44	0.62	0.81
IRA contribution	0.83	1.00	0.86	0.39	0.51	0.86
IRA	0.80	0.86	1.00	0.54	0.44	0.88
401k	0.44	0.39	0.54	1.00	0.51	0.54
401k contribution	0.62	0.51	0.44	0.51	1.00	0.54
Roth contribution	0.81	0.86	0.88	0.54	0.54	1.00

related terms in our sample period. We notice that the multiple-word terms usually have high correlations, which can be higher than 0.8, with the original input. If there are two terms that are highly correlated, these terms probably carry too similar information, and the collinearity may cause problems in a regression. Besides this high correlation problem, those multi-word terms frequently bring in noises, which include phrases or short sentences asking for words' definitions or synonyms. Since we think it is inappropriate to use those unrelated or duplicated information in our term list, we filter the multiple-word phrases out and only combine single-word related terms to our basic term list as complements. There are 20 related words added to the original term list, such as "IRA," "401K," "corruption" and "anxiety." Differing from the method proposed in Da et al. (2015) which selects related terms from the results manually, our cleaning procedure does not require personal judgment and manual discrimination for adding the financial terms to term list. The SVIs of all words in the full term list are collected under the region option of United States and in a weekly frequency. After collecting operation of the SVIs, we remove those words which do not have at least 80% of observations over our sample period. After the preparation procedure, there are 170 SVIs of corresponding words which are the elements of online investor sentiment information. Our data is sampled weekly from January 2004, which is the starting time for Google Trends, through November 2021 with 935 time series. We calculate the natural log differences of SVIs and denote them as ΔSVIs. To address the issue of outliers, seasonality and heteroskedasticity, we adjust the search query data in three steps. First, we winsorize each ΔSVI at the 5% level (2.5% at each tail). Then, we regress each ΔSVI on number of week dummies and month dummies and use the residuals. Finally, we standardize each ΔSVI by scaling each by the time-series mean and standard deviation. Each adjusted change of search volume index, ΔSVI, is stationary and comparable under the unified scale.

3.1.2 Stock Return Data

We collect weekly returns of the Standard and Poor's 500 (S&P500) index and six individual stocks from Yahoo Finance. The S&P500 index is the well-known capitalization-weighted stock price index and widely used for representing the aggregate stock market return. Weekly level asset returns are using the week-over-week log difference of adjusted closing price on Friday. The weekly risk free rate is obtained from Professor Kenneth French's data library.[4] Our sample period covers from January 2004 through November 2021. The weekly return of S&P500 index has 0.15% mean, 2.42% standard deviation, -1.10 skewness, and 13.10 kurtosis. The return ranges from -20.08% to 11.42%. The Sharpe ratio of S&P500 index is 0.05.

3.2 Selection Results and Sentiment Indices

In the building of FEARS^{lasso}, we select the hyper-parameter λ by using the cross-validation of time series split as in Hyndman and Athanasopoulos (2018). The naive index is the average of 170 ΔASVIs of information proxies. Figure 1 shows a portion of our 4 sentiment indices during the COVID-19 period (2019 to 2021). For the sentiment indices built by PLS and LASSO, we can observe low spikes in March 2020 when the COVID-19 pandemic started.

[4] The data library address is http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

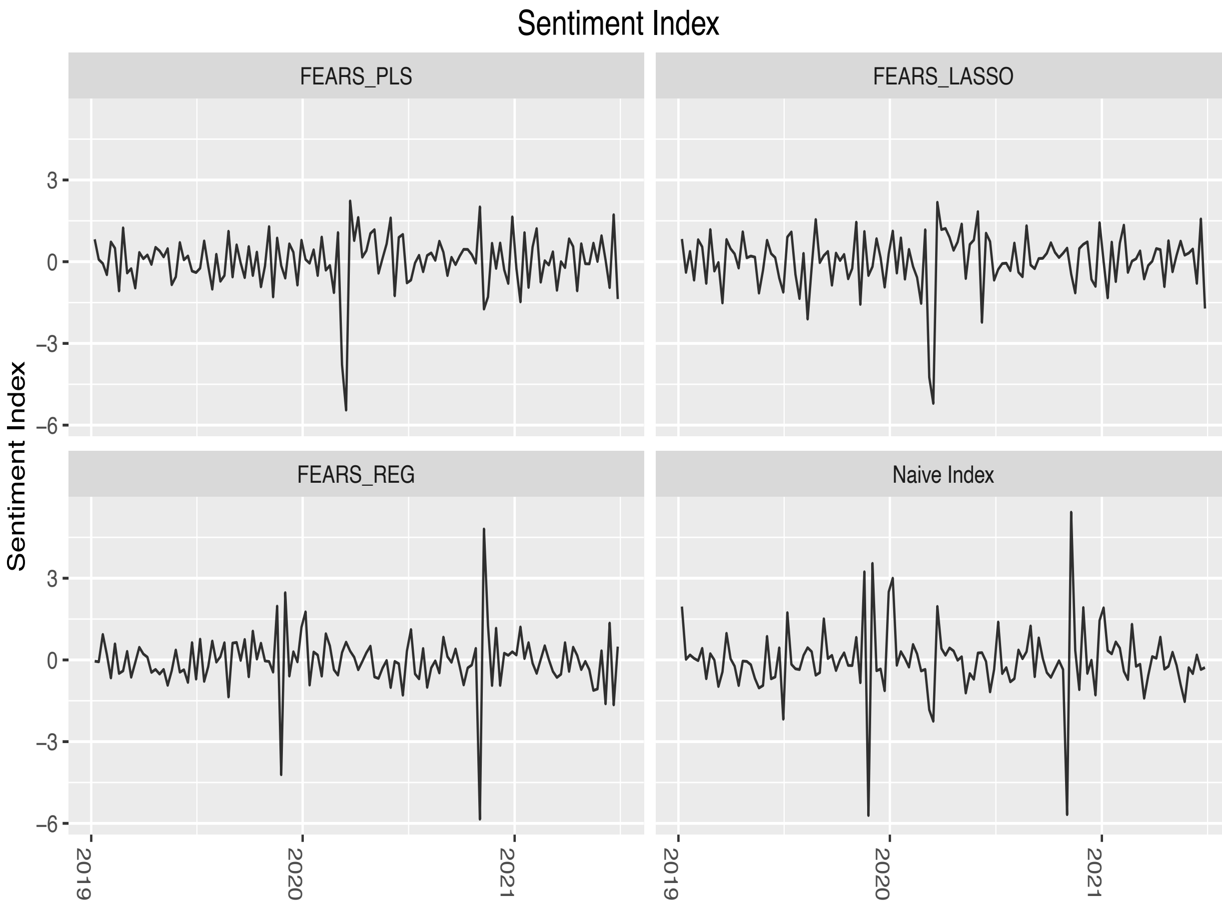

Figure 1: Investment indices.

Considering this NBER-dated start of recession was a time that market participants were extremely panicked indeed, a low level of our online investor sentiment indices imply that investors in the market concern about the economic situation. These two indices bounce back to positive after the trough. Meanwhile, in the case of sentiment index based on regression method, FEARS^{reg} does not show a similar pattern and is flatter in those recession time periods. Figure 2 shows the accumulated levels of each sentiment index. One can also see huge plunges in graphs of both FEARS^{pls} and FEARS^{lasso} as well. Table 2 reports the correlations of 4 sentiment indices including the naive index. From Table 2, it is clear that FEARS^{pls} and FEARS^{lasso} are highly correlated with a positive correlation of 0.87. FEARS^{reg} is negatively correlated with both FEARS^{pls} and FEARS^{lasso} with correlation of -0.52 and -0.32, respectively, which is consistent with the results that we observe in

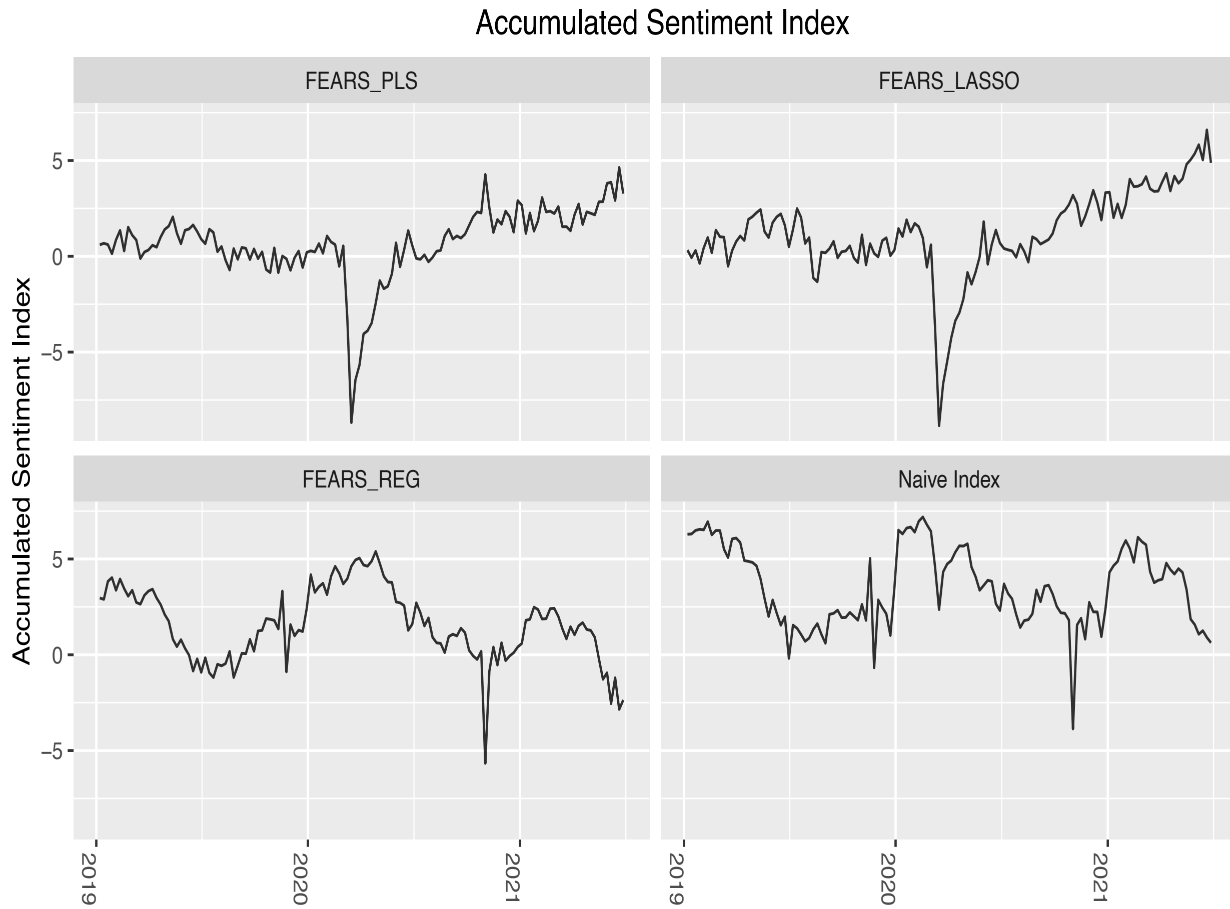

Figure 2: Accumulated investment indices.

Table 2: Correlations among the online investor sentiment indices.

	FEARS^{pls}	FEARS^{lasso}	FEARS^{reg}	Naive Index
FEARS^{pls}	1.00			
FEARS^{lasso}	0.87	1.00		
FEARS^{reg}	-0.53	-0.42	1.00	
Naive Index	-0.02	0.01	0.71	1.00

figures. It is implied by the low correlations that FEARS^{pls} and FEARS^{lasso} capture different fluctuations in sentiment from FEARS^{reg}. FEARS^{reg} has a correlation of 0.71 with the naive index, suggesting that FEARS^{reg} may capture the similar variations of the simple mean of raw sentiment proxies.

3.3 Forecasting Performances for Asset Returns

We consider a simple linear predictive regression model,

$$r_{n,t+1} = \beta_0 + \beta_1 \mathrm{SI}_t^k + \mu_{n,t+1} \tag{3}$$

where $r_{n,t+1}$ denotes the return rate of asset n in excess of the risk-free rate on day $t+1$, $k = 1$, 2, 3, and 4, denotes one of the investor sentiment indices in following order: FEARS^{pls}, FEARS^{lasso}, FEARS^{reg}, and the naive index at time t. Based on the predictive regression 3, we examine the return predictive power of online investor sentiment indices. The null hypotheses is that $\beta_1 = 0$, which means the input index has no predictive power. In this case, regression 3 reduces to $r_{n,t+1} = \beta_0 + \mu_{n,t}$ and the tested sentiment index does not contain information about future stock return. Under the alternative hypothesis, β_1 is statistically different with 0, so that the expected future stock return varies with the tested sentiment index. To alleviate the concern of autocorrelation and heteroskedasticity features of stock return, we also report the t-statistics adjusted by Newey-West standard error.

Table 3 (Panel A for weekly data and Panel B for monthly data) reports the empirical results of the predictive regression on excess market return. The investor sentiment index

Table 3: In-sample prediction performance results.

Index Type	β (%)	t-stat	NW-t-stat	R^2 (%)
Panel A: Weekly frequency				
FEARS^{pls}	-0.27***	-3.48	-3.20	1.28
FEARS^{lasso}	-0.22**	-2.82	-2.47	0.85
FEARS^{reg}	0.10	1.28	1.18	0.17
Naive Index	-0.07	-0.90	-0.92	0.09
Panel B: Monthly frequency				
FEARS^{pls}	-0.25	-0.88	-0.82	0.36
FEARS^{lasso}	-0.23	-0.80	-0.61	0.30
FEARS^{reg}	0.14	0.48	0.38	0.11
Naive Index	-0.20	-0.70	-0.50	0.23

The above table provides in-sample results of the predictive regression. The evaluation period is from January 2004 through November 2021. In the Newey-West test, the lag term is 4. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

constructed by PLS method, FEARS^{pls}, has a negative regression coefficient, β of -0.27%. It is consistent with the previous research that high investor sentiment implies a lower expected excess market return in the following week. The regression estimate of -0.27% implies that a one-standard-deviation increase in FEARS^{pls} is associated with a 0.27% decrease in expected excess market return for the next week. Also, FEARS^{pls} has a Newey-West adjusted t-statistics of -3.20 and an R^2 of 1.28%, which suggest that the predict power of FEARS^{pls} on excess market return is statistically significant. The t-statistic is higher than the threshold at 3.0 in absolute value which is suggested by Harvey et al. (2016) for an empirical significance level of asset pricing factor. Moreover, FEARS^{lasso}, the investor sentiment index constructed by LASSO method, like FEARS^{pls}, achieves a negative regression coefficient of -0.22% and it has a Newey-West adjusted t-statistics of -2.47, which is smaller than that for FEARS^{pls} in absolute value, with R^2 of 0.85% also lower than that for FEARS^{pls}. The coefficient of FEARS^{lasso} in predictive regression is still statistically significant. Such a sentiment index may have slightly weaker predictive power on excess market return than FEARS^{pls}. Unlike FEARS^{pls} or FEARS^{lasso}, FEARS^{reg} has a positive regression coefficient which is equal to 0.10%. An adjusted t-statistic of 1.18 and an R^2 of 0.17% are the lower than the results of two previous sentiment indices and suggest a statistically insignificant predictive power. The coefficient of predictive regression with naive index is not statistically significant with t-statistics and Newey-West adjusted t-statistics of -0.90 and -0.92, respectively. To do a comparison, we also examine the sentiment indices under a monthly frequency. From Panel B in Table 3 for monthly data, it is interesting to see that all 3 online investor sentiment indices as discussed in the previous sections do not show statistically significant predictive power in the sample. This implies that the monthly online sentiment should not provide any useful information to predict asset returns. Finally, we need to point out that we also analyze the daily data and the conclusion is similar to that for the monthly data, so that the results for daily data are not reported here, available upon request.

For both FEARS^{pls} and FEARS^{lasso}, a negative estimate indicates that the decrement (increment) of according online investor sentiment index predicts a higher (lower) expected return in the following week, which is caused by the overselling (overbuying) of irrational market participants the recent week. When the assets are mis-priced, rational investors should take actions to arbitrage, so that the pricing error caused by investor sentiment shock should eventually be corrected. The sentiment indices only predict in weekly frequency since within a longer time frame the arbitrage operations may be able to eliminate the effects of irrational overselling or overbuying. It is consistent with the previous literature; see, for example, McGurk et al. (2020), that the shorter a time horizon is, the larger impact of investor sentiment affects.

To make pairwise comparisons of forecasting results, we follow Diebold and Mariano (1995) and Diebold (2015), and apply the Diebold-Mariano (DM) test for differences in predictive accuracy between two models. The DM test makes assumptions directly on the forecast error loss differential. It is a model-free test that means it is intended for comparing forecasts, not for comparing models. Denote the loss associated with forecast error μ_t by $L(e_t)$, where $L(\cdot)$ is a loss function, such as quadratic loss or absolute loss. The time-t loss differential between forecasts 1 and 2 is then $d_{12,t} = L(\mu_{1,t}) - L(\mu_{2,t})$. Given assumptions of the DM test:

$$E(d_{12,t}) = \mu, \ \ \mathrm{Cov}(d_{12,t}, d_{12,(t-\tau)}) = \gamma(\tau), \ \ \text{and} \ \ \mathrm{Var}(d_{12,t}) = \sigma^2 < \infty.$$

Then, under the null hypothesis, $E[L(\mu_{1,t}) - L(\mu_{2,t})] = 0$, we have:

$$\mathrm{DM}_{12} = \bar{d}_{12}/\hat{\sigma}_{\bar{d}_{12}} \ \longrightarrow \ N(0,1),$$

where $\bar{d}_{12} = \frac{1}{T}\sum_{t=1}^{T} d_{12,t}$ is the sample mean loss differential and $\hat{\sigma}_{\bar{d}_{12}}$ is a consistent estimate of the standard deviation of $\bar{d}_{12}$. For our case, the loss function $L(\cdot)$ is taking a quadratic form, and $\hat{\sigma}_{\bar{d}_{12}}$ is estimated by Newey-West standard error over the in-sample period. Table 4 depicts the p-values of the DM test between two forecasts, in particular, one can see that the p-value for testing FEARS^{pls} versus FEARS^{reg} is 0.054, which means the forecasts based on such two sentiment indices are significantly different at 10% significance level, which means that FEARS^{pls} makes statistically significant improvements on forecasting over the FEARS^{reg}.

From the above, the empirical in-sample evidence shows strongly the predictive power of our new sentiment indices. However, Welch and Goyal (2008) claim that some stock return predictors are invalid because their predictions can not outperform the simple means of historical data. Thus, we examine the predictive power of the proposed sentiment indices by using the out-of-sample test. The key idea of the out-of-sample is that only the information

Table 4: Testing results for the Diebold-Mariano tests.

Forecast 1	Forecast 2	DM statistic	p-value
FEARS^{pls}	FEARS^{reg}	-1.929	0.054
FEARS^{lasso}	FEARS^{reg}	-1.170	0.242
FEARS^{pls}	FEARS^{lasso}	-1.572	0.116

The table above provides p-value of the Diebold-Mariano test. The evaluation period is from January 2004 through November 2021. Following the Newey-West method as in Newey and West (1994), the lag term applied is 4. The loss function is quadratic form.

available at time t can be used for predicting a variable, like excess market return, at time $t+1$. Following Welch and Goyal (2008) and Kelly and Pruitt (2013), we run the out-of-sample analysis by estimating the following predictive regression model recursively.

$$\hat{r}_{t+1} = \hat{\beta}_{0,t} + \hat{\beta}_{1,t}\text{SI}_{k,t:t},$$

where $\hat{\beta}_{0,t}$ and $\hat{\beta}_{1,t}$ are the estimates from regressing $\{r_{s+1}\}_{s=1}^{t-1}$ on a constant and a sentiment index $\{\text{SI}_{k,t:s}^{k}\}_{s=1}^{t-1}$ for each $k = 1$, 2, 3, and 4 denoting each sentiment index. We consider different recursively estimated sentiment indices, including $\text{FEARS}_{t:s}^{pls}$, $\text{FEARS}_{t:s}^{lasso}$, $\text{FEARS}_{t:s}^{reg}$, and Naive $\text{Index}_{t:s}$.

To conduct the out-of-sample test, we separate our original dataset into two subsets which are training set and testing set. Let q be a fix length chosen for the forecast evaluation period and the rest of $p = T - q$ observations are used for training the predictive model. The training set includes all the data available at time $t = 1, 2, \ldots, p$, and the future expected return is estimated at time $t = p+1, p+2, \ldots, T-1$. For the prediction at time $p+1$, we carry out operations in three steps. First, we construct a sentiment index $\{\text{SI}_{k,p:s}\}_{s=1}^{p}$ based on stock return and sentiment proxies at time $t = 1, 2, \ldots, p$. Second, we regress the expected future excess stock return, $\{r_{s+1}\}_{s=1}^{p-1}$, on a constant and the sentiment index, $\{\text{SI}_{k,p:s}\}_{s=1}^{p-1}$, to achieve the OLS estimates $\hat{\beta}_{0,p}$ and $\hat{\beta}_{1,p}$. Third, we predict the excess stock return at time $p+1$, $\hat{r}_{p+1}$, using the sentiment index at time p, $\text{SI}_{k,p:p}^{k}$. Meanwhile, we also calculate the average of historical data at time $t = 1, \ldots, p$ as a benchmark prediction of future excess stock return at time $p+1$. For the prediction of $p+2$, the three-step procedure is applied using available data till time $p+1$. Thus, by recursively predicting, there are q out-of-sample forecasts: $\{\hat{r}_{t+1}\}_{t=p}^{T-1}$. We explore the out-of-sample test in three different settings of forecast evaluation period $q = 12$, 26, and 52, respectively, which varies from a quarter to one year.

The out-of-sample forecasting performance is evaluated by out-of-sample R^2, denoted by R^2_{OS}, introduced by Campbell and Thompson (2008), which is the proportion of mean squared forecasting error of predictive regression and that of historical average.

$$R^2_{\mathrm{OS}} = 1 - \frac{\sum_{t=p}^{T-1} \left(r_{t+1} - \hat{r}_{t+1}\right)^2}{\sum_{t=p}^{T-1} \left(r_{t+1} - \bar{r}_{t+1}\right)^2}, \quad \text{where} \quad \bar{r}_{t+1} = \frac{1}{t}\sum_{s=1}^{t} r_s, \tag{4}$$

r_{t+1} is the excess asset return at time $t+1$, $\hat{r}_{t+1}$ is the forecast excess asset return based on a online investor sentiment at $t+1$, and $\bar{r}_{t+1}$ denotes the historical average benchmark corresponding to the constant expected return model ($r_{t+1} = \alpha + \epsilon_{t+1}$). R^2_{OS} is in a range of $(-\infty, 1]$. A positive R^2_{OS} means that the prediction from predictive regression outperforms the historical average in term of mean squared forecasting error.

Table 5 exhibits the evaluations for out-of-sample performance. We test the models in

Table 5: Out-of-sample performance.

Type	$R^2_{\mathrm{OS}}(\%)$		
Period	J = 12	J = 26	J = 52
FEARS^{pls}	3.37	2.08	1.85
FEARS^{lasso}	0.49	0.24	0.73
FEARS^{reg}	0.44	0.21	0.56
Naive Index	-1.66	-1.69	-0.69

J=12 is for the out-of-sample period over September 2021 to November 2021, J=26 is for the out-of-sample period over June 2021 to November 2021, and J=52 is for the out-of-sample period over December 2020 to November 2021.

three different period lengths which are a quarter (J=12), half a year (J=26), and a year (J=52). In this test, FEARS^{pls} generates positive R^2_{OS}s with 3.37%, 2.08%, and 1.85% for J=12, 26, and 52, respectively, and FEARS^{lasso} achieves positive R^2_{OS}s with 0.49%, 0.24%, and 0.73%, in such test time periods, which shows weaker out-of-sample predictive power. However, FEARS^{reg} has smaller R^2_{OS}s with 0.44%, 0.21%, and 0.56% in three out-of-sample periods of J=12, 26, and 52, respectively. The forecasting based on the naive index does not show a better out-of-sample predictive power than the simple historical mean, according to its all negative R^2_{OS}s. Even though all three constructed sentiment indices outperform the historical average in out-of-sample forecasting, FEARS^{pls} delivers substantially smaller mean squared forecast error than historical average, as R^2_{OS}s are more than double the second highest R^2_{OS}s in all three periods.

In summary, the newly proposed indices, FEARS^{pls}, FEARS^{lasso}, provide a strong evidence for in and out of sample predictive power. In particular, FEARS^{pls}, shows the robust predictive power on excess aggregate stock market return with a significant in-sample estimate and large R^2_{OS}s. While FEARS^{lasso} achieves a slightly less significant in-sample estimate and smaller R^2_{OS}s, which implies weaker predictive power. The in-sample estimate of FEARS^{reg} is insignificant, which suggests that the regression-based aggregation method is not an optimal choice in weekly frequency case. Aggregated from the same collection of online investor sentiment proxies with our new indices, naive index shows no predictive power, which emphasizes the necessity of choosing a proper aggregation method for online investor sentiment index.

3.4 Asset Allocation Implications

To examine the economic value of the stock return predictability of our online investor sentiment indices, we test the performance of dynamic asset allocation under the Markowitz paradigm. In this case, we consider a mean-variance investor who allocates wealth among one risky asset and one risk-free asset, and rebalances the portfolio at the end of each period based on the out-of-sample forecasting stock return for next period. Following Campbell and Thompson (2008) and Jiang et al. (2019), we use the certainty equivalent return gain and Sharpe ratio to measure the performance of portfolios and the economic value of predictors.

At the end of period t, the investor optimally allocates w_t in risky asset and $1 - w_t$ in risk-free asset for period $t + 1$

$$w_t = \frac{\hat{r}_{t+1}}{\gamma\, \hat{\sigma}^2_{t+1}},$$

where γ is the risk aversion parameter, $\hat{r}_{t+1}$ is the out-of-sample forecast of excess market return, and $\hat{\sigma}^2_{t+1}$ is the variance forecast of according stock return. Following Huang et al. (2015), we evaluate the portfolio performance with the risk aversion parameter of 1, 3, and 5. Following Campbell and Thompson (2008), we use a five-year moving window of past weekly returns to forecast the variance of excess market return. Considering the short-selling and leverage limitation, we constrain asset weights w_t to lie between 0 and 1.5 to exclude short sales and to allow for at most 50% leverage. The realized portfolio return in $t+1$ is

$$r^p_{t+1} = w_t r_{t+1} + r^f_{t+1},$$

where r^p_{t+1} is the portfolio return and r^f_{t+1} is the risk-free return. The CER of a portfolio is

$$\text{CER} = \hat{\mu} - 0.5\gamma\,\hat{\sigma}^2, \quad \text{and} \quad \text{CER}_{\text{gain}} = \text{CER} - \text{CER}^b,$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the sample mean and variance, respectively, of the portfolio return r^p_{t+1} over the out-of-sample forecasting evaluation periods, and CER^b is the CER level of benchmark portfolio. The CER gain, denoted by CER_{gain}, is the difference between the CER for the investor who uses the forecasts of market return based on an online investor sentiment k, where k is FEARS^{pls}, FEARS^{lasso}, FEARS^{reg} and Naive Index, and the CER of a benchmark portfolio, CER^b. In this case, the benchmark portfolio is built by using the historical average forecast in (4). We annualize the weekly CER gain by multiplying by 52 and use it to measure an annual portfolio management fee that investors are willing to pay for the advantage of the predictability of online investor sentiment index and the portfolio return that is higher than one based on historical average forecasts.

Table 6 shows the annualized CER gain and Sharpe ratio for each portfolio with risk aversion γ = 1, 3, and 5. The out-of-sample period is over 52 weeks from December 2020

Table 6: Asset allocation results.

	$\gamma = 1$		$\gamma = 3$		$\gamma = 5$	
Index Type	CER gain(%)	Sharpe ratio	CER gain(%)	Sharpe ratio	CER gain(%)	Sharpe ratio
FEARS^{pls}	0.56	0.30	0.19	0.30	0.11	0.30
FEARS^{lasso}	0.33	0.27	0.11	0.27	0.07	0.27
FEARS^{reg}	0.11	0.27	0.04	0.27	0.02	0.27
Naive Index	-0.04	0.25	-0.01	0.25	-0.01	0.25

The risk aversion γ = 1, 3, and 5. The out-of-sample period is 52 weeks.

through November 2021. The index type represents the portfolio that is constructed based on the out-of-sample forecasting of stock return depending on according online investor sentiment index. When the risk aversion is 1, the CER gain of FEARS^{pls} is 0.56% and ranking the top among all 4 online investor sentiment indices. The portfolio constructed with FEARS^{lasso} has CER gain of 0.33%. FEARS^{reg} has lower CER gain of 0.11%. The naive index has negative CER gain of -0.04. In this case, the Sharpe ratio of the portfolios built based on FEARS^{pls} is 0.30, which is higher than those of other three sentiment indices. When the risk aversion is 3, the FEARS^{pls} achieves the best performance with CER gain of 0.19% and Sharpe ratio of 0.30. Meanwhile, FEARS^{lasso} has CER gains of 0.11% and Sharpe ratio of 0.27, and FEARS^{reg} has lower CER gains of 0.04% and Sharpe ratio of 0.27. When risk aversion is 5, the FEARS^{pls} also outperforms other sentiment indices with the CER gain of 0.11%. FEARS^{lasso} and FEARS^{reg} has lower CER gains of 0.07% and 0.02%. Thus, FEARS^{pls} have the best over all performance in terms of CER gains and Sharpe ratio, and FEARS^{lasso} is ranked at second best. In every scenario, risk aversion of 1, 3 and 5, the CER gains for FEARS^{pls} are higher than those of other 3 sentiment indices. The consistently positive CER gains for FEARS^{pls} of 0.61%, 0.20%, and 0.12% can be explained as the maximum annual portfolio management fee that an investor with a risk aversion of 1, 3, or 5 is willing to pay, respectively, to obtain the benefit of predictive power of FEARS^{PLS} and have a higher return. Likewise in case of FEARS^{lasso}, investors are going to pay a premium for the predictive power of online investor sentiment. The Sharpe ratios of FEARS^{pls} and FEARS^{lasso}, which are around 0.30 and 0.27, respectively, exceed the market Sharpe ratio of 0.05%. The portfolios whose investment policy depends on our online investor sentiment indices outperform the aggregate market. In summary, FEARS^{pls} can generate economic value which is robust to common risk aversion levels, while FEARS^{reg} can not. Finally, to gauge the above observation, the accumulated return returns from allocation implication are computed and displayed in Figure 3, from which one see clearly that the accumulated return based on our sentiment indices are much better than the other three competitors (Reg, Naive and Benchmark).

3.5 Cross-sectional Characteristics Portfolios Forecasting

We have examined the relationship between investor sentiment and the dynamic risk premium of whole stock market in Sections 3.3 and 3.4. But, stocks with different features might have different movements. Therefore, we go one more step further to investigate the influence of online investor sentiment on cross-sectional stock returns. We examine four kinds of

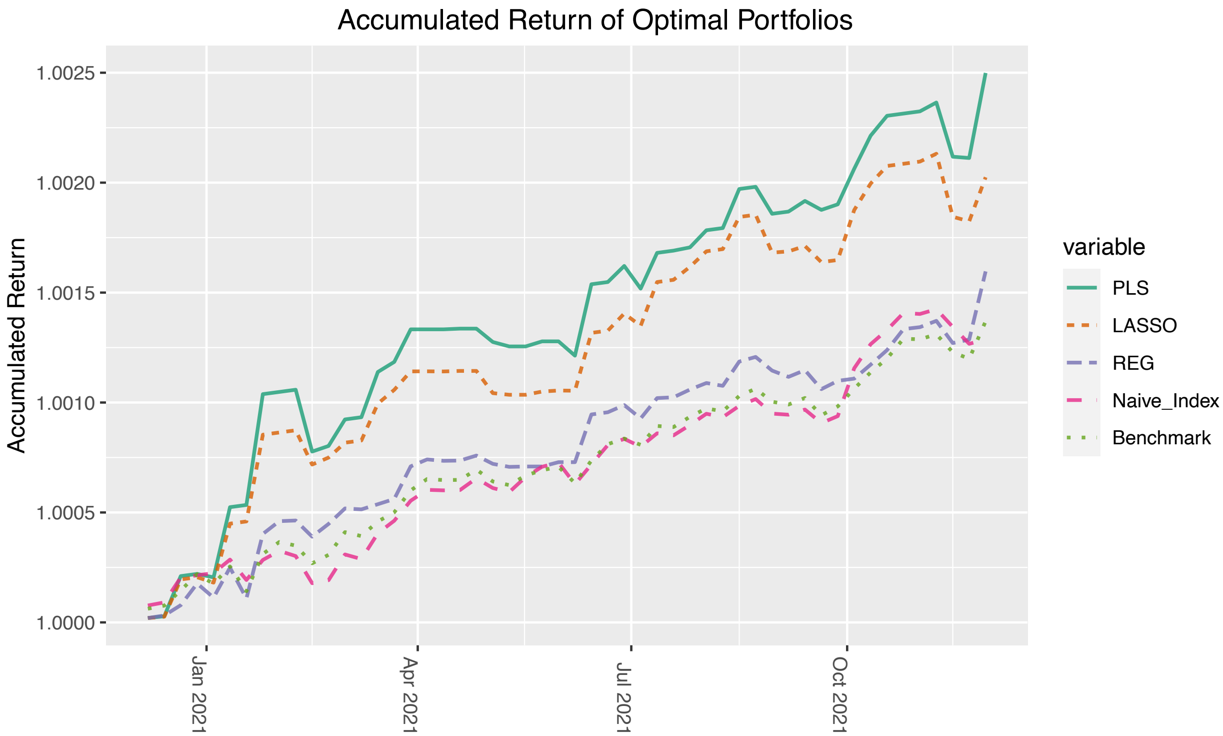

Figure 3: Allocation implication accumulated returns ($\gamma = 1$).

cross-sectional characteristics portfolios, which are sorted based on on industry, momentum, size, and book-to-market ratio.[5]

First, we consider the in-sample univariate predictive regression,

$$r_{t+1}^j = \beta_0^j + \beta_1^j \,\mathrm{SI}_\mathrm{t}^\mathrm{k} + \mu_{t+1}^j,$$

where r_{t+1}^j is each of the weekly excess returns for 10 industry, 10 momentum, 10 size, and 10 book-to-market ratio portfolios, SI_t^k is an online investor sentiment k, where SI^k is FEARS^{pls}, FEARS^{lasso}, FEARS^{reg}, and the naive index. In this predictive regression, the null hypotheses is $H_0 : \beta_1^j = 0$, and the alternative hypothesis is $H_A : \beta_1^j \neq 0$. If the alternative hypothesis is true, we believe that the expected future stock return of cross-sectional characteristics portfolios varies with the tested sentiment index. The regression results are shown in Tables 7 for industry, 8 for momentum, 9 for size, and 10 for book-to-

[5]The industry portfolios are based on the four-digit SIC code of stock. The portfolio of momentum are constructed on the prior return data. The portfolios of size are constructed on the market equity. The portfolios of book-to-market ratio are constructed on the ratio of book equity to market equity. All Portfolio returns are sorted and provided by Kenneth French.

market ratio, respectively.

In Panel A of Table 7, the estimation results of the predictive regression of FEARS^{pls} on 10 industry portfolios are reported. The predictive power is significant at 1% significance level in 6 out of 10 industries. The online investor sentiment have effects on industries like: nondurable, durable, manufacture, shop, utility, and other. In these industries, the coefficients, β, varies with industry and ranges from -0.44% to -0.19%, which implies the cross-sectional difference of industries in the exposures to investor sentiment. All of significant estimates of cross-sectional regressions are negative, which is consistent with the results we have in aggregate market. Meanwhile, estimates of FEARS^{lasso} are also negative and mostly statistically significant at 5% significance level across industries. The other online investor sentiment indices, FEARS^{reg} and the naive index, do not show significant estimation results in any industries. To be more specific, we also examine the detailed 49 industry portfolios. Results are reported in Table 11 in the appendix. FEARS^{pls} shows significant predictive power in all industries, and 18 out of 49 estimates have t-statistics than 3.0 in absolute value, the threshold of empirical significance. The coefficients range from -0.64% to -0.18% and vary with industries. Moreover, all negative estimates are consistent to our finding in the previous section. FEARS^{lasso} have significant estimates on 42 out 49 industries, and 5 estimates among them have t-statistics exceeding 3.0. Finally, FEARS^{reg} achieve 4 out of 49 significant estimates at 10% significance level, but naive index does not show any significant predictive power for 49 industries characteristic portfolios.

In Table 8, we report the estimation results of the predictive regression for 10 momentum portfolios. FEARS^{pls} achieves significant estimates in 5 out of 10 momentum portfolios. Excepting the most losing portfolio, FEARS^{pls} shows strong predictive power on each level of momentum portfolio. The estimates range from -0.39% to -0.27%. Every unit change of investor sentiment FEARS^{pls} tends to have smaller impact on those winner types portfolio according to the decline of estimated coefficients. FEARS^{lasso} has significant estimates in forecasting most of momentum portfolios. FEARS^{lasso} also achieve also have significant predictive power across different momentum portfolios under 5% significance level. The effects of FEARS^{lasso} on different momentum portfolios are varying in a small range. These results provide evidence that those winner portfolios tend to be overbuying under a positive

Table 7: Prediction performance for industry portfolios.

	β (%)	t-stat	NW-t-stat	R^2 (%)
Panel A: FEARS^{pls}				
Nondurable	-0.25 $***$	-3.83	-3.73	1.55
Durable	-0.44 $**$	-3.21	-3.09	1.09
Manufacture	-0.37 $***$	-4.10	-3.91	1.77
Energy	-0.25 $*$	-2.06	-1.96	0.45
Technology	-0.24 $**$	-2.70	-2.44	0.78
Telecom	-0.19 $*$	-2.36	-2.03	0.59
Shop	-0.33 $***$	-4.32	-4.00	1.96
Health	-0.18 $*$	-2.5	-2.53	0.67
Utility	-0.27 $***$	-3.36	-3.03	1.20
Other	-0.35 $***$	-3.37	-3.07	1.20
Panel B: FEARS^{lasso}				
Nondurable	-0.18 $**$	-2.75	-2.52	0.80
Durable	-0.34 $*$	-2.43	-2.33	0.63
Manufacture	-0.32 $***$	-3.57	-3.37	1.35
Energy	-0.18	-1.44	-1.31	0.22
Technology	-0.22 $*$	-2.47	-2.26	0.65
Telecom	-0.18 $*$	-2.17	-1.77	0.50
Shop	-0.32 $***$	-4.08	-3.64	1.76
Health	-0.17 $*$	-2.26	-2.19	0.54
Utility	-0.20 $*$	-2.56	-2.40	0.70
Other	-0.29 $**$	-2.74	-2.37	0.80
Panel C: FEARS^{reg}				
Nondurable	0.08	1.28	1.22	0.17
Durable	0.27	1.94	1.74	0.40
Manufacture	0.14	1.50	1.35	0.24
Energy	0.05	0.43	0.31	0.02
Technology	0.10	1.16	1.02	0.14
Telecom	0.10	1.23	1.03	0.16
Shop	0.13	1.62	1.68	0.28
Health	0.05	0.71	0.74	0.05
Utility	0.06	0.77	0.64	0.06
Other	0.11	1.04	0.92	0.12
Panel D: Naive Index				
Nondurable	-0.08	-1.30	-1.42	0.18
Durable	-0.01	-0.05	-0.05	0.00
Manufacture	-0.11	-1.19	-1.19	0.15
Energy	-0.16	-1.30	-0.99	0.18
Technology	-0.04	-0.47	-0.48	0.02
Telecom	-0.09	-1.07	-1.04	0.12
Shop	-0.05	-0.69	-0.81	0.05
Health	-0.05	-0.71	-0.81	0.05
Utility	-0.08	-1.02	-0.97	0.11
Other	-0.14	-1.31	-1.34	0.18

The table above provides in-sample results of the predictive regression of each online investor sentiment index on the characteristics portfolios. The sentiment indices considered are FEARS^{pls}, FEARS^{lasso}, FEARS^{reg} and the naive index. The evaluation period is from January 2004 through November 2021. In the Newey-West test, the lag term is set as 4. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 8: Prediction performance for momentum portfolios.

	β (%)	t-stat	NW-t-stat	R^2 (%)
Panel A: FEARS^{pls}				
Loser	-0.39 $*$	-2.31	-2.16	0.57
2	-0.37 $**$	-2.84	-2.43	0.86
3	-0.35 $**$	-3.13	-2.82	1.04
4	-0.34 $***$	-3.49	-3.06	1.29
5	-0.31 $***$	-3.46	-3.26	1.27
6	-0.27 $**$	-3.26	-2.97	1.12
7	-0.32 $***$	-4.05	-3.54	1.73
8	-0.28 $***$	-3.57	-3.24	1.35
9	-0.32 $***$	-3.77	-3.88	1.50
Winner	-0.32 $**$	-3.08	-3.15	1.00
Panel B: FEARS^{lasso}				
Loser	-0.28	-1.67	-1.48	0.30
2	-0.32 $*$	-2.50	-2.03	0.66
3	-0.32 $**$	-2.86	-2.41	0.87
4	-0.28 $**$	-2.88	-2.45	0.88
5	-0.26 $**$	-2.89	-2.66	0.89
6	-0.23 $**$	-2.72	-2.33	0.79
7	-0.28 $***$	-3.53	-3.14	1.32
8	-0.26 $**$	-3.21	-2.85	1.10
9	-0.28 $***$	-3.33	-3.37	1.17
Winner	-0.30 $**$	-2.86	-2.96	0.87
Panel C: FEARS^{reg}				
Loser	0.13	0.78	0.71	0.06
2	0.13	1.02	0.78	0.11
3	0.10	0.88	0.62	0.08
4	0.11	1.10	0.88	0.13
5	0.12	1.30	1.02	0.18
6	0.10	1.25	0.99	0.17
7	0.10	1.25	1.07	0.17
8	0.13	1.61	1.30	0.28
9	0.12	1.44	1.51	0.22
Winner	0.10	0.93	0.93	0.09
Panel D: Naive Index				
Loser	-0.17	-1.03	-1.08	0.11
2	-0.16	-1.26	-1.14	0.17
3	-0.15	-1.32	-1.06	0.19
4	-0.14	-1.39	-1.28	0.21
5	-0.09	-1.05	-0.95	0.12
6	-0.10	-1.14	-1.04	0.14
7	-0.09	-1.15	-1.20	0.14
8	-0.05	-0.64	-0.58	0.04
9	-0.04	-0.42	-0.49	0.02
Winner	-0.11	-1.01	-1.13	0.11

The table above provides in-sample results of the predictive regression of each online investor sentiment index on the characteristics portfolios. The sentiment indices considered are FEARS^{pls}, FEARS^{lasso}, FEARS^{reg} and the naive index. The evaluation period is from January 2004 through November 2021. In the Newey-West test, the lag term is set as 4. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

change of online investor sentiment level, and our new sentiment indices can indicate that negative movement of future expected return of such portfolios. FEARS^{reg} and the naive index do not show significant predictive power in this case.

In Panel A of Table 9, the results are generated from the predictive regressions of online investor sentiment FEARS^{pls} on 10 size sorted portfolio returns. The estimates are all negative, which is consistent with the literature and our estimation for the aggregate market. All estimates are statically significant. Especially, 6 out of 10 estimates are significant with the absolute values of t-statistics exceeding 3. The predictive power are more significant for portfolios consisted of companies larger than the median in market. The results in Panel A reveal the cross-sectional pattern between online investor sentiment and portfolio returns. The online investor sentiment FEARS^{pls} has strong effects on stocks of large companies. The regression results of FEARS^{lasso} are shown in Panel B. Most estimates are significant under 5% significance level. The coefficients, βs, are smaller than their counterparts from case of FEARS^{pls}. FEARS^{reg} and the naive index are in Panels C and D, respectively. The results are not statistically significant without any absolute values of t-statistics above 2.

Similarly, in Panel A of Table 10, the results are generated from the predictive regressions of online investor sentiment FEARS^{pls} on returns of 10 book-to-market sorted portfolios. All estimates are negative and significant. 5 out of 10 estimates are significant with the absolute values of t-statistics exceeding 3. The coefficient, β, ranges from -0.35% to -0.24%. FEARS^{pls} has better predictive power on growth types of portfolios in terms of t-statistics and R^2. Regression results of FEARS^{lasso} are shown in Panel B. Most estimates are significant under 5% significance level. It also exhibits predictive power on growth type portfolios. FEARS^{reg} and the naive index do not show significant results.

FEARS^{pls} has the most significant predictive power on different cross-sectional portfolios. FEARS^{lasso} also indicate some degrees of predictive power on such portfolios. The varying of coefficients with industry implies the cross-sectional difference in the exposures to investor sentiment. All R^2s of FEARS^{pls}, FEARS^{lasso} are larger than the corresponding R^2s of FEARS^{reg}. Our new online investor sentiment indices exhibit stronger predictive power than that of FEARS^{reg}.

Table 9: Prediction performance for size portfolios.

	β (%)	t-stat	NW-t-stat	R^2 (%)
Panel A: FEARS^{pls}				
Small	-0.25 **	-2.60	-2.73	0.72
2	-0.33 **	-2.94	-2.86	0.92
3	-0.35 **	-3.23	-3.17	1.11
4	-0.35 **	-3.28	-3.10	1.14
5	-0.37 ***	-3.61	-3.35	1.38
6	-0.34 ***	-3.46	-3.32	1.26
7	-0.34 ***	-3.54	-3.29	1.33
8	-0.35 ***	-3.86	-3.44	1.57
9	-0.31 ***	-3.63	-3.32	1.39
Large	-0.27 ***	-3.63	-3.37	1.39
Panel B: FEARS^{lasso}				
Small	-0.22 *	-2.27	-2.44	0.55
2	-0.30 **	-2.71	-2.60	0.78
3	-0.31 **	-2.83	-2.74	0.85
4	-0.31 **	-2.93	-2.72	0.91
5	-0.33 **	-3.20	-2.95	1.08
6	-0.29 **	-2.95	-2.77	0.92
7	-0.29 **	-3.07	-2.80	1.00
8	-0.30 ***	-3.30	-2.94	1.16
9	-0.26 **	-3.01	-2.67	0.96
Large	-0.22 **	-2.93	-2.59	0.91
Panel C: FEARS^{reg}				
Small	0.11	1.16	1.21	0.15
2	0.12	1.07	1.00	0.12
3	0.15	1.39	1.32	0.21
4	0.14	1.31	1.21	0.18
5	0.14	1.34	1.24	0.19
6	0.14	1.48	1.39	0.23
7	0.13	1.32	1.25	0.19
8	0.13	1.45	1.29	0.23
9	0.13	1.47	1.40	0.23
Large	0.10	1.33	1.22	0.19
Panel D: Naive Index				
Small	-0.01	-0.1	-0.11	0.00
2	-0.08	-0.67	-0.71	0.05
3	-0.04	-0.39	-0.42	0.02
4	-0.05	-0.50	-0.53	0.03
5	-0.07	-0.68	-0.73	0.05
6	-0.06	-0.60	-0.64	0.04
7	-0.07	-0.76	-0.82	0.06
8	-0.06	-0.67	-0.69	0.05
9	-0.04	-0.50	-0.55	0.03
Large	-0.08	-1.10	-1.14	0.13

The table above provides in-sample results of the predictive regression of each online investor sentiment index on the characteristics portfolios. The sentiment indices considered are FEARS^{pls}, FEARS^{lasso}, FEARS^{reg} and the naive index. The evaluation period is from January 2004 through November 2021. In the Newey-West test, the lag term is set as 4. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 10: Prediction performance for book-to-market ratio portfolios.

	β (%)	t-stat	NW-t-stat	R^2 (%)
Panel A: FEARS^{pls}				
Growth	-0.31 ***	-3.93	-3.74	1.62
2	-0.29 ***	-3.79	-3.47	1.51
3	-0.24 **	-3.26	-2.89	1.13
4	-0.30 ***	-3.53	-3.38	1.32
5	-0.32 ***	-3.67	-3.29	1.42
6	-0.29 **	-3.17	-2.90	1.06
7	-0.32 ***	-3.46	-3.20	1.27
8	-0.34 **	-3.20	-2.90	1.08
9	-0.30 **	-2.62	-2.55	0.73
Value	-0.35 **	-2.58	-2.35	0.71
Panel B: FEARS^{lasso}				
Growth	-0.27 ***	-3.46	-3.24	1.27
2	-0.24 **	-3.12	-2.77	1.03
3	-0.21 **	-2.85	-2.37	0.86
4	-0.24 **	-2.84	-2.59	0.86
5	-0.26 **	-3.03	-2.71	0.98
6	-0.23 *	-2.56	-2.38	0.70
7	-0.26 **	-2.74	-2.41	0.80
8	-0.27 *	-2.58	-2.24	0.71
9	-0.21	-1.87	-1.75	0.37
Value	-0.26	-1.92	-1.67	0.39
Panel C: FEARS^{reg}				
Growth	0.14	1.73	1.68	0.32
2	0.13	1.66	1.52	0.29
3	0.09	1.23	1.13	0.16
4	0.12	1.41	1.27	0.21
5	0.09	1.08	0.93	0.12
6	0.08	0.93	0.79	0.09
7	0.10	1.04	0.84	0.12
8	0.09	0.85	0.69	0.08
9	0.07	0.59	0.57	0.04
Value	0.10	0.76	0.71	0.06
Panel D: Naive Index				
Growth	-0.05	-0.69	-0.73	0.05
2	-0.04	-0.58	-0.59	0.04
3	-0.07	-0.89	-0.91	0.09
4	-0.07	-0.85	-0.85	0.08
5	-0.12	-1.33	-1.35	0.19
6	-0.10	-1.13	-1.15	0.14
7	-0.09	-1.00	-0.90	0.11
8	-0.14	-1.31	-1.20	0.18
9	-0.14	-1.22	-1.33	0.16
Value	-0.17	-1.26	-1.36	0.17

The table above provides in-sample results of the predictive regression of each online investor sentiment index on the characteristics portfolios. The sentiment indices considered are FEARS^{pls}, FEARS^{lasso}, FEARS^{reg} and the naive index. The evaluation period is from January 2004 through November 2021. In the Newey-West test, the lag term is set as 4. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

4 Conclusions

We show the procedure of using PLS and LASSO to efficiently extract new investor sentiment indices from a high dimensional online information set for aggregate stock return forecasting, and compare the new indices with those constructed by the previous method. We apply high dimensional online investor sentiment proxies of search volume data from Google Trends. The new sentiment indices have good predictive power for the aggregate stock return in weekly frequency, while the sentiment index built by the previous method is not significant. The predictive power of the new online investor sentiment indices, FEARS^{pls} and FEARS^{lasso}, is both statistically and economically significant. Moreover, our new online investor sentiments indices also show significant predictive power for cross-sectional stock portfolio returns. The empirical evidence concludes that the PLS and LASSO methods are consistent and efficient for exploiting high dimensional online investor sentiment proxies.

Some extensions related to this paper should be addressed in future research. For example, first, as mentioned in Section 3.1.1, ΔSVI is used instead of SVI since SVI might not be stationary. Therefore, one might use SVI as sentiment index rather than ΔSVI so that the econometric issues and testing predictability are totally different; see, for example, Campbell and Yogo (2006) and Liao et al. (2018) for more details on this aspect. This topic would be very interesting. Second, all sentiment indices used in this paper are constructed as a linear combination of online information variable proxies, which might not be appropriate if the truth is nonlinear. Third, when there are too many online information variable proxies (D is very large) and nonlinearity might exist, some machine learning methods might be suitable for building sentiment indices. Furthermore, similar to the manager sentiment index proposed in Jiang et al. (2019), it would be interesting to construct other type of new indices, such as combining the index from Google Trends with textual tone of corporate financial disclosure or media information; see, for example, the paper by Cai et al. (2023) for studying China economic policy uncertainty. Finally, one might apply other methods to construct sentiment indices such as the sliced inverse regression approach proposed by Li (1991).

References

Andrei, D. and Hasler, M. (2015). Investor attention and stock market volatility. *Review of Financial Studies*, 28(1), 33-72.

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3), 651-707.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Review of Financial Studies*, 59(3), 1259-1294.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Review of Financial Studies*, 61(4), 1645-1680.

Baker, M. and Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129-152.

Balcilar, M., Bonato, M., Demirer, R. and Gupta, R. (2017). The effect of investor sentiment on gold market return dynamics: Evidence from a nonparametric causality-in-quantiles approach. *Resources Policy*, 51(C), 77-84.

Brown, G.W. and Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 11(1), 1-27.

Cai, Z., Yuang, J. and Pan, Y. (2023). China economic policy uncertainty and its forecasting based on a new textual mining method. *China Journal of Econometrics*, 3(1), 1-21.

Campbell, J.Y. and Thompson, S.B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4), 1509-1531.

Campbell, Y.C. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1), 27-60.

Chen, J., Tang, G., Yao, J., and Zhou, G. (2019). Investor attention and stock returns. *Available at SSRN 3194387*.

Choi, H. and Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(Special Issue), 2-9.

Cochrane, J.H. (2011). Presidential address: Discount rates. *Journal of Finance*, 66(4), 1047-1108.

Da, Z., Engelberg, J. and Gao, P. (2011). In search of attention. *Journal of Finance*, 66(5), 1461-1499.

Da, Z., Engelberg, J. and Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies*, 28(1), 1-32.

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4), 703-738.

Diebold, F.X., and Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253-263.

Diebold, F.X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold - Mariano tests. *Journal of Business & Economic Statistics*, 33(1), 1-1.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348-1360.

Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1), 101-148.

Green, J., Hand, J.R., and Zhang, X.F. (2013). The supraview of return predictive signals. *Review of Accounting Studies*, 18(3), 692-730.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223-2273.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3), 705-730.

Harvey, C.R., Yan Liu, Y. and Zhu, H. (2016). ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5-68.

Huang, D., Jiang, F., Tu, J., and Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies*, 28(3), 791-837.

Jiang, F., Lee, J., Martin, X., and Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, 132(1), 126-149.

Kelly, B., and Pruitt, S. (2013). Market expectations in the crosssection of present values. *Journal of Finance*, 68(5), 1721-1756.

Kelly, B., and Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, 186(2), 294-316.

Keynes, J. (1936) *The General Theory of Employment, Interest and Money*. London: Macmillan.

Lemmon, M. and Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies*, 19(4), 1499-1529.

Li, K. (1991). Sliced inverse regression for dimension reduction. *Journal of The American Statistical Association*, 86(414), 316-327.

Liao, X., Cai, Z. and Chen, H. (2018). A perspective on recent models for testing predictability of asset returns. *Applied Mathematics – A Journal Of Chinese Universities, Series B*, 33(2), 127-144.

Loughran, T., and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10Ks. *Journal of Finance*, 66(1), 35-65.

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy*, 60(2), 151-158.

McGurk, Z., Nowak, A. and Hall, J.C. (2020). Stock returns and investor sentiment: Textual analysis and social media. *Journal of Economics and Finance*, 44(3), 458-485.

Newey, W.K. and West, K.D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61(4), 631-653.

Qadan, M. and Nama, H. (2018). Investor sentiment and the price of oil. *Energy Economics*, 69(1), 42-58.

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. In G. Elliott and A. Timmermann (Eds.), *Handbook of economic forecasting* (pp. 32883). Amsterdam, the Netherlands: Elsevier.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the US stock market. *Journal of Banking and Finance*, 84, 25-40.

Schwert, G.W. (1989). Why does stock market volatility change over time? *Review of Financial Studies*, 44(5), 1115-1153.

Siganos, A., Vagenas-Nanos, E. and Verwijmeren, P. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior and Organization*, 107 (Part B), 730-743.

Siganos, A., Vagenas-Nanos, E. and Verwijmeren, P. (2017). Divergence of sentiment and stock market trading. *Journal of Banking and Finance*, 78(1), 130-141.

Spiegel, M. (2008). Forecasting the equity premium: Where we stand today. *Review of Financial Studies*, 21(4), 1453-1454.

Tetlock, P.C. (2007) . Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62(3), 1139-1168.

Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance*, 63(3), 1437-1467.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B*. 58(1), 267288.

Welch, I. and Amit Goyal, A. (2008). A Comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455-1508.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiah (Ed.), *Multivariate Analysis* (pp. 391-420). New York: Academic Press.

Wold, H. (1975). Soft modeling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. *Journal of Applied Probability*, 12(S1), 117-142.

Zhou, G. (2018). Measuring investor sentiment. *Annual Review of Financial Economics*, 10, 239-259.

Appendix

Prediction Performance for 49 Industry Portfolios

Table 11: Prediction performance for 49 industry portfolios.

	β (%)	t-stat	NW-t-stat	R^2 (%)
Panel A: FEARS^{pls}				
Agric	-0.25*	-2.01	-1.95	0.43
Food	-0.23***	-3.67	-3.30	1.42
Soda	-0.21*	-2.27	-2.10	0.55
Beer	-0.24***	-3.50	-3.72	1.30
Smoke	-0.26**	-2.95	-2.62	0.93
Toys	-0.46***	-3.73	-3.16	1.47
Fun	-0.45**	-3.20	-2.44	1.09
Books	-0.28*	-2.50	-2.20	0.67
Hshld	-0.29***	-4.39	-3.91	2.02
Clths	-0.48***	-4.25	-3.66	1.90
Hlth	-0.20*	-2.05	-2.13	0.45
MedEq	-0.21*	-2.47	-2.47	0.65
Drugs	-0.18*	-2.47	-2.38	0.65
Chems	-0.38***	-3.57	-3.27	1.35
Rubbr	-0.29**	-2.98	-2.83	0.94
Txtls	-0.50**	-3.11	-3.06	1.02
BldMt	-0.52***	-4.29	-4.05	1.93
Cnstr	-0.64***	-4.37	-3.70	2.01
Steel	-0.60***	-3.79	-3.28	1.52
FabPr	-0.57***	-3.80	-3.26	1.53
Mach	-0.41***	-3.53	-3.04	1.32
ElcEq	-0.38***	-3.41	-3.21	1.23
Autos	-0.48**	-3.24	-2.99	1.11
Aero	-0.44***	-3.63	-2.95	1.39
Ships	-0.26*	-2.06	-2.28	0.45
Guns	-0.25**	-2.75	-1.98	0.80
Gold	-0.41*	-2.56	-2.03	0.70
Mines	-0.46**	-2.83	-2.35	0.85
Coal	-0.50*	-2.16	-1.62	0.50
Oil	-0.26*	-2.14	-2.02	0.49
Util	-0.28***	-3.46	-3.08	1.27
Telcm	-0.21*	-2.56	-2.14	0.70
PerSv	-0.31**	-2.91	-2.64	0.90
BusSv	-0.33***	-3.65	-3.18	1.41
Hardw	-0.26*	-2.47	-2.13	0.65
Softw	-0.22*	-2.54	-2.33	0.69
Chips	-0.28**	-2.73	-2.34	0.79
LabEq	-0.28**	-3.02	-2.77	0.97
Paper	-0.33***	-3.85	-3.90	1.56
Boxes	-0.49***	-4.88	-4.32	2.49
Trans	-0.40***	-3.92	-3.57	1.62
Whlsl	-0.28**	-3.18	-2.98	1.07

Rtail	-0.37***	-4.65	-4.10	2.27
Meals	-0.28**	-3.22	-3.06	1.10
Banks	-0.42**	-3.20	-2.78	1.08
Insur	-0.24*	-2.28	-2.36	0.56
RlEst	-0.35*	-2.54	-2.31	0.69
Fin	-0.46***	-3.82	-2.89	1.54
Other	-0.32***	-3.48	-3.34	1.28
Panel B: FEARS^{lasso}				
Agric	-0.13	-1.10	-1.00	0.13
Food	-0.18 **	-2.81	-2.57	0.84
Soda	-0.16	-1.79	-1.62	0.34
Beer	-0.16 *	-2.29	-2.32	0.56
Smoke	-0.19 *	-2.15	-1.80	0.49
Toys	-0.38 **	-3.03	-2.83	0.97
Fun	-0.41 **	-2.88	-2.15	0.88
Books	-0.24 *	-2.08	-1.70	0.46
Hshld	-0.25 ***	-3.75	-3.50	1.48
Clths	-0.37 **	-3.25	-2.74	1.12
Hlth	-0.15	-1.57	-1.60	0.26
MedEq	-0.19 *	-2.21	-2.25	0.52
Drugs	-0.17 *	-2.21	-2.06	0.52
Chems	-0.3 **	-2.86	-2.54	0.87
Rubbr	-0.25 **	-2.60	-2.37	0.72
Txtls	-0.30	-1.84	-1.81	0.36
BldMt	-0.43 ***	-3.55	-3.15	1.33
Cnstr	-0.54 ***	-3.68	-2.96	1.43
Steel	-0.52 ***	-3.31	-2.71	1.16
FabPr	-0.43 **	-2.85	-2.49	0.86
Mach	-0.37 **	-3.11	-2.58	1.02
ElcEq	-0.34 **	-2.99	-2.75	0.95
Autos	-0.35 *	-2.40	-2.19	0.62
Aero	-0.36 **	-2.96	-2.62	0.93
Ships	-0.18	-1.38	-1.64	0.20
Guns	-0.24 *	-2.56	-2.02	0.70
Gold	-0.36 *	-2.21	-1.67	0.52
Mines	-0.38 *	-2.34	-1.85	0.58
Coal	-0.47 *	-2.05	-1.48	0.45
Oil	-0.17	-1.40	-1.28	0.21
Util	-0.21 **	-2.61	-2.44	0.73
Telcm	-0.18 *	-2.22	-1.81	0.52
PerSv	-0.21 *	-1.99	-1.73	0.42
BusSv	-0.27 **	-3.00	-2.50	0.96
Hardw	-0.28 **	-2.68	-2.50	0.77
Softw	-0.18 *	-2.08	-1.97	0.46
Chips	-0.24 *	-2.32	-2.09	0.58
LabEq	-0.26 **	-2.82	-2.55	0.84
Paper	-0.27 **	-3.14	-3.06	1.04
Boxes	-0.40 ***	-4.01	-3.63	1.69
Trans	-0.37 ***	-3.63	-3.52	1.39
Whlsl	-0.24 **	-2.73	-2.61	0.79

Rtail	-0.35 ***	-4.36	-3.78	2.00
Meals	-0.23 **	-2.74	-2.51	0.80
Banks	-0.33 *	-2.48	-2.14	0.66
Insur	-0.14	-1.38	-1.35	0.20
RlEst	-0.29 *	-2.07	-1.78	0.46
Fin	-0.39 **	-3.20	-2.39	1.09
Other	-0.25 **	-2.71	-2.41	0.78

Panel C: FEARS^{reg}

Agric	0.04	0.36	0.35	0.01
Food	0.05	0.73	0.65	0.06
Soda	0.08	0.91	0.83	0.09
Beer	0.07	0.98	0.93	0.10
Smoke	0.13	1.47	1.25	0.23
Toys	0.36**	2.86	2.61	0.87
Fun	0.15	1.07	0.96	0.12
Books	0.11	0.93	0.84	0.09
Hshld	0.09	1.34	1.25	0.19
Clths	0.18	1.58	1.69	0.27
Hlth	0.03	0.34	0.38	0.01
MedEq	0.05	0.64	0.67	0.04
Drugs	0.05	0.73	0.74	0.06
Chems	0.16	1.55	1.41	0.26
Rubbr	0.13	1.37	1.24	0.20
Txtls	0.17	1.05	1.39	0.12
BldMt	0.26*	2.12	2.03	0.48
Cnstr	0.29	1.93	1.84	0.40
Steel	0.16	1.02	0.74	0.11
FabPr	0.25	1.68	1.45	0.30
Mach	0.17	1.40	1.06	0.21
ElcEq	0.19	1.72	1.40	0.32
Autos	0.31*	2.08	1.74	0.46
Aero	0.13	1.10	1.05	0.13
Ships	0.12	0.97	0.97	0.10
Guns	0.05	0.59	0.59	0.04
Gold	0.02	0.13	0.09	0.00
Mines	0.07	0.42	0.30	0.02
Coal	0.17	0.74	0.52	0.06
Oil	0.05	0.40	0.29	0.02
Util	0.06	0.77	0.64	0.06
Telcm	0.10	1.23	1.03	0.16
PerSv	0.16	1.45	1.33	0.23
BusSv	0.14	1.55	1.36	0.26
Hardw	0.06	0.60	0.52	0.04
Softw	0.10	1.16	1.08	0.14
Chips	0.10	1.00	0.82	0.11
LabEq	0.18*	1.97	1.71	0.42
Paper	0.10	1.17	1.05	0.15
Boxes	0.20	1.95	1.46	0.41
Trans	0.11	1.10	1.07	0.13
Whlsl	0.10	1.10	0.98	0.13

Rtail	0.15	1.80	1.83	0.35
Meals	0.07	0.80	0.87	0.07
Banks	0.10	0.79	0.66	0.07
Insur	0.05	0.47	0.48	0.02
RlEst	0.07	0.53	0.60	0.03
Fin	0.21	1.69	1.34	0.30
Other	0.09	1.04	0.93	0.12
Panel D: Naive Index				
Agric	-0.15	-1.27	-1.33	0.17
Food	-0.10	-1.51	-1.60	0.24
Soda	-0.08	-0.87	-0.88	0.08
Beer	-0.06	-0.92	-0.93	0.09
Smoke	-0.06	-0.63	-0.65	0.04
Toys	0.09	0.69	0.58	0.05
Fun	-0.10	-0.70	-0.79	0.05
Books	-0.10	-0.91	-0.97	0.09
Hshld	-0.11	-1.64	-1.80	0.29
Clths	-0.10	-0.87	-1.01	0.08
Hlth	-0.07	-0.76	-0.93	0.06
MedEq	-0.03	-0.36	-0.40	0.01
Drugs	-0.06	-0.75	-0.82	0.06
Chems	-0.07	-0.62	-0.64	0.04
Rubbr	-0.04	-0.36	-0.37	0.01
Txtls	-0.14	-0.89	-1.16	0.08
BldMt	-0.04	-0.36	-0.38	0.01
Cnstr	-0.05	-0.31	-0.35	0.01
Steel	-0.25	-1.61	-1.34	0.28
FabPr	0.01	0.08	0.07	0.00
Mach	-0.08	-0.70	-0.59	0.05
ElcEq	-0.01	-0.12	-0.10	0.00
Autos	0.02	0.14	0.13	0.00
Aero	-0.17	-1.40	-1.32	0.21
Ships	-0.18	-1.39	-1.48	0.21
Guns	-0.07	-0.73	-0.86	0.06
Gold	-0.23	-1.41	-1.04	0.21
Mines	-0.21	-1.29	-1.03	0.18
Coal	-0.13	-0.55	-0.41	0.03
Oil	-0.16	-1.29	-0.98	0.18
Util	-0.08	-1.02	-0.97	0.11
Telcm	-0.09	-1.07	-1.04	0.12
PerSv	-0.08	-0.74	-0.80	0.06
BusSv	-0.03	-0.37	-0.36	0.01
Hardw	-0.07	-0.72	-0.71	0.06
Softw	-0.04	-0.43	-0.48	0.02
Chips	-0.07	-0.69	-0.68	0.05
LabEq	0.06	0.63	0.63	0.04
Paper	-0.11	-1.22	-1.24	0.16
Boxes	-0.08	-0.79	-0.62	0.07
Trans	-0.12	-1.20	-1.22	0.15
Whlsl	-0.09	-1.03	-1.02	0.11

Rtail	-0.04	-0.50	-0.56	0.03
Meals	-0.07	-0.84	-1.03	0.08
Banks	-0.18	-1.39	-1.35	0.21
Insur	-0.13	-1.30	-1.48	0.18
RlEst	-0.06	-0.45	-0.57	0.02
Fin	-0.05	-0.45	-0.45	0.02
Other	-0.14	-1.58	-1.62	0.27

The table above provides in-sample results of the predictive regression of each online investor sentiment index on the characteristics portfolios. The sentiment indices considered are FEARS^{pls}. The evaluation period is from January 2004 through November 2021. In the Newey-West test, the lag term is set as 4. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.