Economics Working Paper Series

2019/013

Investor Sentiment as a Predictor of Market Returns

Kim Kaivanto and Peng Zhang

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

LUMS home page: http://www.lancaster.ac.uk/lums/

Investor Sentiment as a Predictor of Market Returns*

Kim Kaivanto[§] and Peng Zhang[‡]
[§]Lancaster University, Lancaster LA1 4YX, UK
[‡]Guizhou Minzu University, Guiyang City, PRC 550025

this version: June 26, 2019

Abstract

Investor sentiment's effect on asset prices has been studied extensively — to date, without delivering consistent results across samples and datasets. We investigate the asset-pricing impacts of eight widely cited investor-sentiment indicators (one direct, six indirect, one composite), within a unified long-horizon regression framework, predicting real NYSE-index returns over horizon lengths of 1, 3, 12, 24, 36, and 48 months. Results reveal that three of the non-composite indicators have consistent predictive power: the Michigan Index of Consumer Sentiment ($MICS$), IPO volume ($NIPO$), and the dividend premium ($PDND$). This finding has implications for the widely cited Baker-Wurgler first principal component ($SFPC$) composite indicator, which extracts information from the full set of six indirect indicators. As the diffusion-index literature shows, this type of wide-net approach is likely to impound idiosyncratic noise into the composite summary indicator, exacerbating forecasting errors. Therefore we create a new 'targeted' composite indicator from the first principal component of the three indicators that perform well in long-horizon regressions, i.e. $MICS$, $NIPO$, and $PDND$. The resulting targeted composite indicator out-performs $SFPC$ in a market-returns prediction horse race. Whereas $SFPC$ primarily predicts Equally Weighted Returns (EWR) rather than Value Weighted Returns (VWR), our new sentiment indicator performs better than $SFPC$ in predicting both VWR and EWR. This improved performance is due in part to a reduction in overfitting, and in part to incorporation of the direct sentiment indicator $MICS$.

Keywords: investor sentiment, market return, predictability, long-horizon regression, bootstrap diffusion index, composite index, overfitting
JEL classification: G12, G17

 Authors are joint equal contributors, listed in alphabetical order.

1 Introduction

Market-return predictability is a well-established area of empirical inquiry. Moreover, it informs cognate questions on the drivers of market risk, the pricing of market risk, and the quantitative modeling of asset prices in general equilibrium. In this paper we undertake a comprehensive investigation of the market-level asset-pricing impacts of eight widely cited investor-sentiment indicators, employing a unified long-horizon-regression framework and an identical sample period for all indicators.

Market-return predictability studies have discovered statistically significant correlations between market returns and various indicators, including short-term interest rates (Fama and Schwert 1977), interest-rate term structure (Keim and Stambaugh 1986, Campbell 1987), past returns (Lo and MacKinlay 1988, Poterba and Summers 1988, Fama and French 1988a, Jegadeesh 1991), and price-related ratios (Fama and French 1988b, Fama and French 1989, Campbell and Shiller 1988, Hodrick 1992).

A further branch of this empirical literature grew from De Long et al.'s (1990) (DSSW) finding that noise traders can prevent well-informed rational traders from arbitraging away disparities between market price and fundamental value. DSSW showed that rational arbitrageurs face systematic risk resulting from uncertainty regarding noise traders' beliefs. By affecting noise traders' beliefs, investor sentiment can affect asset mispricing marketwide. Accordingly, a substantial post-DSSW literature incorporates proxies for investor sentiment into tests of market-return predictability. These fall in five clusters: survey-based proxy studies (Schmeling 2009 and Lee, Jiang and Indro 2002); tests of the Law of One Price (Baker, Wurgler and Yuan 2011); market-timing studies (Baker and Wurgler 2000); textual analysis of news and social media feeds (Tetlock 2007, Da, Engelberg and Gao 2007, Garcia 2013, and Sun, Najan and Shen 2016); and distillation of common components from multiple indicators (Baker and Wurgler 2006, Baker and Wurgler 2007, Kim Ryu and Seo 2014, Huang et al. 2015, and Han and Li 2017 for international evidence). Empirically, strong positive investor sentiment generally predicts lower returns in the near future, while several studies find that the correlation can turn positive over a longer horizon.

A range of different sentiment indicators have been studied in the cross-sectional asset pricing literature. In this literature, multiple investor-sentiment indicators show strong predictive power in explaining portfolio returns, particularly for small-firm stocks and hard-to-arbitrage stocks. Such predictability persists even after introducing macroeconomic control variables and multiple risk factors. For a review, see Baker and Wurgler (2007).

Several indicators share common factor structure, at least in part, despite different approaches to operationalization. Existing empirical studies typically examine the effects of up to three indicators on market returns or portfolio returns, without paying much heed to ensuring comparability with other studies in terms of target return structure, data frequency, sample periods, or model specification.[1] Consequently, it is not straightforward to synthesize across studies for drawing general or summative conclusions. Moreover, with a rich list of sentiment indicators successfully predicting cross-sectional returns, it is natural to ask whether such predictability extends to the aggregate market level. When the effect of sentiment is tested, many indicators can be used to explain cross-sectional returns as long as the influence can be rationalized within a specific sub-market. In contrast, candidates for influential market-level sentiment are more limited. Any candidate indicator must capture latent investor sentiment that (i) affects a significant population of market participants, (ii) becomes reflected in the final price through a clear pricing mechanism, and (iii) has a non-transient effect on aggregate asset prices that is not immediately eliminated by arbitrage.

Consequently, caution must be exercised in choosing sentiment indicators. We undertake a comprehensive investigation of the asset-pricing impacts of seven widely cited non-composite investor-sentiment indicators, employing a unified framework and an identical sample period for all indicators. Results from long-horizon single-factor regressions suggest that three indicators ($MICS$, $NIPO$, $PDND$) predict market returns well, while the other indicators display little or inconsistent predictive power. This pattern continues to hold even when the economic-fundamentals-reflecting component of the investor-sentiment indicator is removed through orthogonalization with respect to a bank of 12 fundamental variables.

Although investor sentiment's posited effect on asset prices has been studied extensively, the time lag with which investor sentiment is expected to become impounded into asset prices is neither theoretically nor empirically well understood.[2] It has also been found that the effects of certain investor-sentiment proxies tend to reverse over longer horizons. Yet literature on the

[1] Existing literature on the effect of investor sentiment in market return predictability consists of studies that (i) target various market index returns, including value-weighted and equal-weighted NYSE, S&P500, DJIA, AMEX, NASDAQ, RUSSEL3000, and different combinations of these indices; (ii) analyze samples of frequencies from half-hour to annual data; (iii) focus on a variety of sample lengths between four years and a century; and (iv) adopt multiple models for the estimation. The story on investor sentiment and cross-sectional returns is even more diverse.

[2] Baker and Wurgler (2006) argue that certain indicators will reflect sentiment with lag, however they do not discuss how to search for such time lag empirically. For the influence of sentiment on returns over mixed time lengths, see Fisher and Statman (2000), Brown and Cliff (2004) and Chung, Hung and Yeh (2017) for short-term results, and Neal and Wheatley (1998) and Brown and Cliff (2005) for long-term evidence.

time-lag dimension remains limited. We implement long-horizon double-factor regressions to investigate the time dimension of investor-sentiment's effect on market returns. For each of the horizons $T \in \{1, 3, 12, 24, 36, 48\}$ we estimate a specification that controls for average returns realized up to time $T - 1$, thereby revealing whether the sentiment indicator retains predictive power on month-T returns that is independent of its effect on returns in the $T - 1$ preceding months. The three indicators that show strong predictive power in single-factor regressions ($MICS$, $NIPO$, $PDND$) all demonstrate consistent, long-lasting predictive power in double-factor regressions. But we observe little evidence of eventual reversal, unlike some previous studies. This is our second contribution.

Lastly, the results of the predictive-power tests may be refined further within the 'diffusion index' framework. The common tendencies in a large collection of time series may be summarized by a limited number of 'diffusion indices', which may be constructed using a variety of different techniques, including e.g. principal component analysis, dynamic factor analysis, partial least squares, and least-angle regression. Diffusion indices may be used to improve the performance of forecasting models, or to mitigate overidentification problems. In the area of investor sentiment, Baker and Wurgler's (2006, 2007) widely cited approach constructs a diffusion index from the first principal component (SFPC) of all six indirect investor-sentiment indicators. However, as Bai and Ng (2008) argue, indiscriminate inclusion of predictors in the diffusion index may introduce too much idiosyncratic noise, thereby leading to potential overfitting and exacerbating forecasting errors.[3] We follow Bai and Ng (2008) in selecting a subset of targeted predictors, using significant predictive power as a marker for inclusion in the subset, from which we construct a new diffusion index ($T3$-$SFPC$) that is the first principal component of this subset. In a market-returns prediction horse race, our new index consistently beats the $SFPC$ index of Baker and Wurgler (2006, 2007). This improved control of overfitting is our third contribution.

The sequel is structured as follows: Section 2 introduces the data. Methodology is summarized in Section 3. Sections 4 and 5 report results from single-factor and double-factor regressions over multiple horizons, respectively. We extract new 'targeted' diffusion indices in Section 6 and test their predictive power against the benchmarks of Baker and Wurgler (2006, 2007). Section 7 concludes.

[3]Bai and Ng (2008) share with a rich literature the same spirit of prerequisite test for diffusion index with a large pool of predictors, e.g. Bair et al. (2006), Boivin and Ng (2006), and Forni et al. (2005).

2 Data

We use real NYSE index returns to represent the market return. Both equal-weighted and value-weighted monthly NYSE index returns are obtained from the Center for Research in Security Prices (CRSP), which are then adjusted with the Consumer Price Index (CPI) into real returns.[4] For investor-sentiment indicators, our choice is mainly constrained by data availability. Most of the indicators are only available over relatively limited time intervals and in frequencies that are not uniformly comparable. Striking a balance between the number of indicators and achievable sample size, we elect to focus on eight monthly indicators from January 1978 through to December 2007. These indicators are introduced in the following subsections, including direct, indirect and composite sentiment measures.

2.1 Direct sentiment measures

The most straightforward, direct indicators of sentiment are provided by survey data. Shiller (1999) suggests that the Yale School of Management Stock Market Confidence Indices can reflect the attitudes of institutional investors. Qiu and Welch (2006) show that data from the UBS/Gallup surveys can explain returns, particularly small-stock returns and returns of stocks held disproportionately by retail investors. Similar findings have also been obtained by Lemmon and Portniaguina (2006) with data from both the Index of Consumer Confidence and the University of Michigan Consumer Confidence Index. Brown and Cliff (2005) find significant long-horizon explanatory power in the Investors Intelligence survey.

Since survey-data availability is the main constraint on sample size, we give data availability the highest priority when choosing among different surveys. As a result, we use the Index of Consumer Sentiment from the University of Michigan Consumer Confidence Index. Although the survey was not originally designed to reflect investor sentiment in asset markets — but rather, to gauge general consumer confidence — it serves the sample-size-maximization objective in that it has been compiled consistently, without interruption, for longer than any other comparable survey. At an annual frequency the survey is available as far back as 1952, while at a monthly frequency the survey is available from January 1978 onward.

The Michigan Index of Consumer Sentiment ($MICS$) indicator is calculated as a linear

[4]Real returns tend to have clearer economic meaning in theoretical asset pricing models, especially in terms of consumption-based interpretations . However, real returns, nominal returns, and excess returns tend to be highly correlated and the choice generally would not affect empirical conclusions. See e.g. Fama and French (1988a) for usage of real return in empirical asset pricing study, Fama and French (1989) for excess return, Fama and French (1988b) for both nominal and real returns, and Poterba and Summers (1988) for both real and excess returns.

transformation of the percentages of positive and negative responses on five telephone-survey questions. The five questions cover (i) change in perceived household financial situation over the last year, (ii) expected year-ahead change in household financial situation, (iii) expected year-ahead national financial business conditions, (iv) expected national business conditions (continuous good times vs. periods of widespread unemployment or depression) over the coming 5 years, and (v) current purchasing conditions for major household durable items.

2.2 Indirect sentiment measures

We also adopt the following six indirect indicators from Jeffrey Wurgler's online data library, including closed-end-fund discount ($CEFD$), NYSE turnover ($TURN$), IPO volume ($NIPO$), IPO first day return ($RIPO$), net equity issuance fraction in total issuance ($NEIF$), and dividend premium ($PDND$). The indicators are constructed as follows.

Closed-end fund discount: Zweig (1973) uses the discount to verify that prices are likely to deviate from fundamental values when 'noise' is present. DeLong et al. (1990) attribute the discount to the fact that closed-end funds are mainly held by individual investors and that the noise brought by these investors will lead to an extra risk premium. Lee et al. (1991), Neal and Wheatley (1998) and Swaminathan (1996) find evidence that the discount is a measure of investor sentiment and can help explain market returns. The closed-end-fund discount ($CEFD$) indicator is calculated as the average percentage difference between the market-based Net Asset Value of the shares held by the closed-end funds and the prices at which closed-end funds' shares change hands.

Market liquidity: Empirical studies have long found coexistence of higher liquidity and lower future returns[5]. Baker and Stein (2004) argue that liquidity provides an indicator of the presence or absence of irrational investors who face short-sale constraints and are active only in optimism. Scheinkman and Xiong's work (2003) also points out the link between sentiment and market liquidity. NYSE turnover ($TURN$) is obtained as the natural logarithm of the ratio of reported share volume over average number of shares listed on NYSE.

IPO related data: Several studies link IPO activity with investor sentiment. Specifically, both IPO volume and IPO first-day return can be viewed as indicators of investor sentiment. For instance, Lee et al. (1991) find evidence that more IPOs happen when investor sentiment is high. Ljungqvist et al. (2006) show that sentiment can lead to IPO underpricing and hence

[5]See, for instance, Amihud and Mendelson (1986), Brennan and Subrahmanyam (1996), and Brennan et al. (1998).

cause high post-IPO returns. IPO volume ($NIPO$) records the number of IPOs in the month. IPO first-day return ($RIPO$) records the average first-day percentage return of all IPOs in the month.

New equity issuance: Given that IPOs are just one indicator of equity financing, a more general indicator of investor sentiment (in stock markets) can be gauged by the fraction of equity issuance to total asset issuance. Baker and Wurgler (2000) find a negative relationship between equity issuance and stock market returns, and attribute this relationship to issuers shifting between equity and debt to minimize the cost of financing. New-equity-issuance fraction ($NEIF$) is the proportion of new equity issued out of total issuance of equity and debt.

Dividend premium: Baker and Wurgler (2004a,b) argue that since dividend-paying equities have characteristics like coupon bonds, they represent 'safety' compared to dividend-nonpaying equities. As a result Baker and Wurgler (2006, 2007) argue that when investors perceive high risk level and look for safety, investor demand for dividends will drive prices of the dividend-paying and dividend-non-paying stocks apart. The dividend premium ($PDND$) is calculated as the log difference of the value-weighted average market-to-book ratios of dividend payers and dividend non-payers.

2.3 Composite sentiment measure

The principal component sentiment index ($SFPC$), which is based on the first principal components of the six standardized[6] indirect indicators, as calculated in Baker and Wurgler (2007).

Note that $SFPC$ may be understood as a 'diffusion index': i.e. a linear summary of the common tendencies in a large collection of time series. In many-predictor forecasting problems the diffusion-index method is widely accepted for its ability to reduce dimensionality, overcome overfitting, and improve forecasting accuracy (see e.g. Stock and Watson 2002, Boivin and Ng 2006, and Ludvigson and Ng 2007, among others).[7]

2.4 Orthogonalized sentiment indicators

Raw sentiment indicators may contain not only information reflecting investor sentiment but also macroeconomic fundamentals (Baker and Wurgler 2006, 2007, Brown and Cliff 2005, and Neal and Wheatley 1998). In order to exclude the component reflecting fundamentals, indicators

[6] Here standardization means subtracting the mean and then dividing by the standard deviation.

[7] Huang et al. (2015) introduce a composite sentiment indicator based on partial least squares. However, we do not include this indicator in the present analysis because its computation is horizon-specific, and we are specifically interested in evaluating indicators' predictive performance across a range of horizon lengths.

are standardized and then orthogonalized with respect to fundamental variables prior to being used in further analysis. We employ two sets of fundamental variables. The first set follows Baker and Wurgler's (2006, 2007) consumption-based asset-pricing approach in incorporating: growth in industrial production; real growth in durable, non-durable, services and total consumption; growth rates in employment, CPI; and an NBER recession dummy variable. The second set follows Brown and Cliff's (2005) and Neal and Wheatly's (1998) conditional asset pricing approach in incorporating: 1-month real US Treasury bill return; the difference between 3-month and 1-month real US treasury bill returns; the difference between 10-year and 3-month real US treasury bill returns; and the default spread between yields on Moody's Baa- and Aaa-rated corporate bonds. Data series for the first set have been obtained from Jeffrey Wurgler's online data library, whilst data series for the second set have been obtained from the US Federal Reserve. F tests show that the two sets of variables are jointly significant in explaining all sentiment indicators. Criteria including AIC and BIC suggest retaining both sets.

We use two parallel approaches for generating orthogonalized sentiment indicators. In the first approach all twelve fundamental variables are retained and therefore each orthogonalized sentiment indicator excludes the same fundamental information. In the second approach each sentiment indicator is orthogonalized with respect to only those fundamental variables that are statistically significant explanators of the indicator in question. Both approaches lead to qualitatively identical results in further analysis.

2.5 Sample characteristics

Table 1 summarizes the sample characteristics of the seven non-composite sentiment indicators (one direct and six indirect). All sentiment indicators except $MICS$ have positive excess kurtosis values. The skewness values of all sentiment indicators except $MICS$ and $CEFD$ show that investor sentiment is right skewed,[8] suggesting fat tails for bullish investor sentiment.[9] Intuitively, since previous studies find a negative correlation between investor sentiment and market returns, *right-skewed investor sentiment* is consistent with the *negative skewness in asset returns* (e.g. Cont 2001, Jackwerth and Rubinstein 1996). Indeed a new literature specifically tests the

[8] $CEFD$ and $PDND$ are supposedly negatively correlated to investor sentiment whilst all other variables are positively correlated to sentiment. If the correlations were perfect, the skewness for distribution of investor sentiment would be opposite to those of $CEFD$ and $PDND$ and be the same as those of other indicators.

[9] One may argue that several indicators including $TURN$, $NIPO$, $RIPO$ and $NEIF$ are bounded above zero so the evidence here may not indeed imply fat tail for bullish investor sentiment. However the same conclusion of fat tail for bullish investor sentiment can be drawn from Table 2, where the orthogonalized indicators are not bounded above zero. We consider the consistent results through Table 1 and Table 2 as evidence that the implication is not an artifact of truncated data.

relationships between investor sentiment and return skewness. For instance, Han (2008) shows that the level of investor sentiment affects the skewness of S&P 500 return and the slope of index option volatility smile.

Table 2 summarizes the sample characteristics of the orthogonalized sentiment indicators. Again, kurtosis values are inflated. Because the orthogonalized indicators come from the residuals of orthogonalization, the means are all extremely close to 0. Compared to Table 1, the signs on skewness and excess kurtosis statistics remain unchanged after orthogonalisation for every sentiment indicator except $CEFD$. However the values of the third and fourth central moments are often different from those in Table 1, showing that the distributional features of sentiment indicators are only partly preserved after excluding the impact of macroeconomic fundamentals.

Table 1: Summary statistics of original indicators

	Mean	Median	S.D.	Skewness	Kurtosis
$MICS$	88.0033	90.9000	12.0700	-0.5570	-0.1308
$CEFD$	8.6801	8.4950	5.6863	0.7743	0.4117
$TURN$	0.6816	0.5970	0.3518	1.8182	4.5544
$NIPO$	32.1667	26.0000	24.5374	0.9375	0.4307
$RIPO$	19.2269	14.1000	19.8890	2.4491	6.9768
$NEIF$	0.1609	0.1379	0.1097	1.4860	2.1447
$PDND$	-13.3101	-12.8300	10.2970	-0.9999	3.5731

This table shows summary statistics for the data of original sentiment indicators used in the analysis. The full monthly sample contains 360 observations from Januaray 1978 through December 2007.

Table 2: Summary statistics of orthogonalized indicators

	Mean	Median	S.D.	Skewness	Kurtosis
$MICS^{\perp}$	0.0000	0.0119	0.5092	-0.1215	-0.0780
$CEFD^{\perp}$	0.0000	0.0810	0.5830	-0.5657	-0.1578
$TURN^{\perp}$	0.0000	-0.0326	0.4194	1.1033	5.4288
$NIPO^{\perp}$	0.0000	-0.1316	0.7371	1.1396	1.8480
$RIPO^{\perp}$	0.0000	-0.0765	0.8699	1.3997	3.5062
$NEIF^{\perp}$	0.0000	-0.0634	0.5779	0.6605	1.8671
$PDND^{\perp}$	0.0000	0.0215	0.6966	-1.0584	3.4518

This table shows summary statistics for the data of sentiment indicators orthogonalized with twelve fundamental variables. The full monthly sample contains 360 observations from Januaray 1978 through December 2007.

Figure 1 provides the histogram distributions of all sentiment indicators. Panel A contains original indicators; Panel B contains orthogonalized indicators. As discussed above, the signs on skewness and kurtosis remain unchanged after orthogonalisation in all cases except $CEFD$. However the distinctions between Panels A and B suggest that excluding the influence of fundamental variables clearly changes the distributions of sentiment indicators.

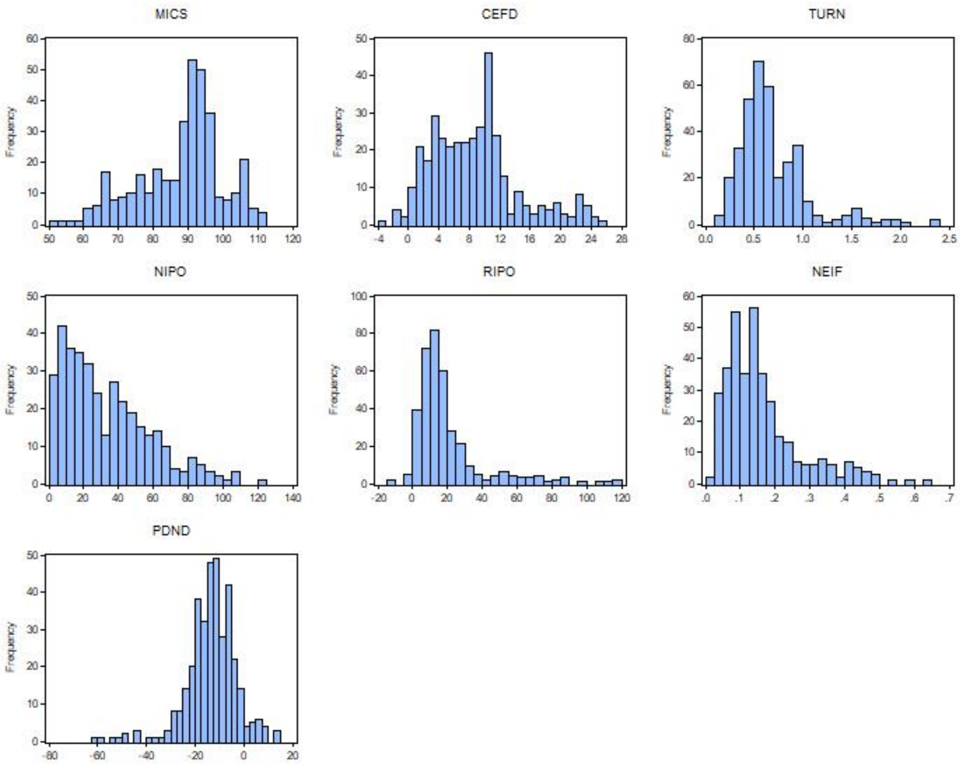
Tables 3 and 4 show the correlation structure between sentiment indicators. As found in the literature (e.g. Brown and Cliff 2004 and Baker and Wurgler 2006), correlations between different indicators are usually small in magnitude. The pairwise correlations are all below 0.5. This is consistent with the notion that indicators are noisy proxies for investor sentiment and reflect investor sentiment in partial and disparate ways. As discussed by Baker and Wurgler (2006), each indicator contains not only a 'sentiment component' but also'idiosyncratic, non-sentiment-related components' (Baker and Wurgler 2006, p.1656). For instance, the survey indicator ($MICS$) reflects consumer confidence in general, and thus contains information on e.g. employment conditions, which may be correlated with the broader economy in addition to the stock market. Indirect sentiment indicators reflect different mechanisms, including market-timing behaviour, liquidity-provision activity, safety seeking, and the noise-trader risk premium. Thus each indicator contains idiosyncratic noise components which are not directly related to investor sentiment and hence lead to small pairwise correlations. This has an important implication for our empirical tests: we expect to find nonuniform results across indicators due to different, possibly stochastic levels of noise across indicators.

Similar results are obtained with original indicators and orthogonalized indicators, suggesting a robust correlation structure that is not merely a reflection of common fundamentals.

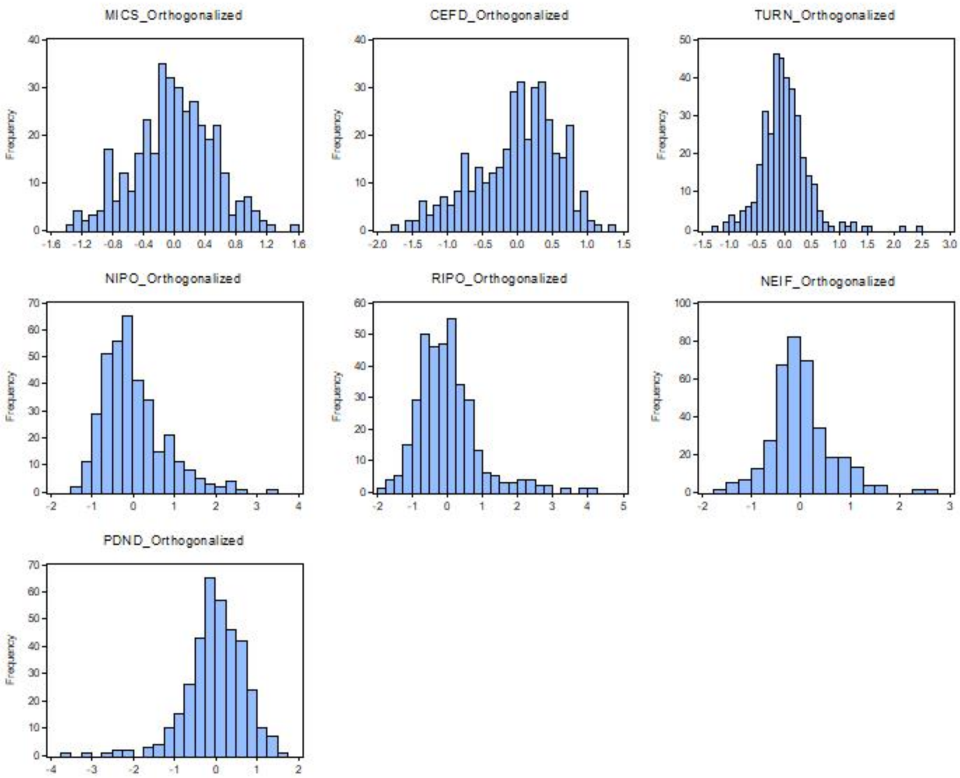
Table 3: Correlation of original indicators

	$MICS$	$CEFD$	$TURN$	$NIPO$	$RIPO$	$NEIF$	$PDND$
$MICS$	1						
$CEFD$	-0.3376	1					
$TURN$	0.2923	-0.3472	1				
$NIPO$	0.2941	-0.2044	-0.1656	1			
$RIPO$	0.1224	0.2978	-0.0365	-0.0024	1		
$NEIF$	-0.3530	0.3282	-0.4663	0.3353	0.1728	1	
$PDND$	-0.1410	-0.1385	-0.0094	-0.3765	-0.4560	-0.4326	1

This table shows the correlation coefficients of original indicators.



(a) Panel A: Histograms of original sentiment indicators



(b) Panel B: Histograms of orthogonalized sentiment indicators

Figure 1: Histogram distributions of all sentiment indicators.

Table 4: Correlation of orthogonalized indicators

	$MICS^{\perp}$	$CEFD^{\perp}$	$TURN^{\perp}$	$NIPO^{\perp}$	$RIPO^{\perp}$	$NEIF^{\perp}$	$PDND^{\perp}$
$MICS^{\perp}$	1						
$CEFD^{\perp}$	-0.2642	1					
$TURN^{\perp}$	0.2529	-0.0530	1				
$NIPO^{\perp}$	0.1969	0.0224	0.2269	1			
$RIPO^{\perp}$	-0.0023	0.1230	0.0518	0.0713	1		
$NEIF^{\perp}$	-0.0957	0.2302	0.0152	0.3603	0.1816	1	
$PDND^{\perp}$	-0.1058	-0.1438	-0.0990	-0.3571	-0.4507	-0.3569	1

This table shows the correlation coefficients of orthogonalized indicators.

3 Methodology

3.1 Model specification

Although investor sentiment's effect on asset prices has been studied extensively, giving rise to a rich empirical literature, theoretical models nevertheless offer no guidance as to the length of the time horizon over which sentiment becomes impounded into asset prices. Evidence has been found in both short-horizon and long-horizon analysis. We first follow practice in the market-return predictability literature and conduct our analysis over multiple horizons. Existing studies of market-return predictability mainly employ three long-horizon model specifications:

(i) Campbell and Shiller (1988) adopt a VAR model as follows

$$z_t = Az_{t-1} + \upsilon_t$$

where z_t is a matrix consisting of return and dividend-price ratio. Long-horizon analysis is implicit in the model.

(ii) Fama and French (1988a, 1988b) use single-factor regression to explain future multiple-period returns with past return or current dividend-price ratio, as in

$$r(t, t+T) = \alpha(T) + \beta(T)Y(t) + \varepsilon(t+T)$$

where $r(t, t+T)$ is future return (with T set to various horizon lengths), $Y(t)$ represents past return or the dividend-price ratio, and $\varepsilon(t+T)$ is the error term.

(iii) Jegadeesh (1991) also employs single-factor regression. He predicts single-period-ahead returns with the sum of lagged returns as a single regressor, as in

$$R_t = a_k + b_k \sum_{i=1}^{K} R_{t-i} + u_{K,t}$$

where R_t is the next-period return, $\sum_{i=1}^{K} R_{t-i}$ is the sum of lagged returns, and $u_{K,t}$ is the error term.

Further discussion of the similarities across models and the advantages of each specification can be found in Hodrick (1992) and Campbell (2001).

In this paper we first follow the Fama and French (1988b) model by estimating single-factor regressions over multiple horizons. We examine the null hypothesis that investor-sentiment indicators have no predictive power for market returns. We follow Fama and French (1988b, 1989) in setting the horizon lengths to 1, 3, 12, 24, 36, and 48 months. In Section 4, we run the single-factor regressions by regressing future k-month average returns on a constant and an indicator of investor sentiment.

$$\frac{1}{k}\sum_{i=1}^{k} r_{t+i} = c^{(k)} + \beta^{(k)} S_t + \epsilon_t^{(k)} \tag{1}$$

where:

(i) r can refer to equal-weighted return (EWR) or value-weighted return (VWR) of the NYSE index;

(ii) S represents one of the sentiment indicators and can refer to $MICS$, $CEFD$, $TURN$, $NIPO$, $RIPO$, $NEIF$, or $PDND$;

(iii) k represents the horizon length and can take the values 1, 3, 12. 24, 36, or 48.

(iv) the coefficient $\beta^{(k)}$ represents how sensitive the future return is to investor sentiment, given the horizon length k. If $\beta^{(k)}$ is statistically significant then evidence of predictive power in the investor-sentiment indicator is present.

In Section 5 we incorporate first-order lagged (future) returns as an additional predictor, testing the weaker null hypothesis that investor-sentiment indicators have no incremental predictive power over lagged returns. This specification addresses the self-predictive power of returns found by Poterba and Summers (1988) and Fama and French (1988a), and the fact that sentiment may be impounded into market returns over time, from month 1 through to month $k-1$, in addition to that component that becomes impounded into market returns in the final month, between month $k-1$ and month k. Such self-predictivity may result from either a time-decaying component in the price process or aggregation biases.

$$\frac{1}{k}\sum_{i=1}^{k} r_{t+i} = c^{(k)} + \alpha^{(k)}\left(\frac{1}{k}\sum_{i=1}^{k} r_{t-1+i}\right) + \beta^{(k)} S_t + \epsilon_t^{(k)} \tag{2}$$

The introduction of first-order lagged (future) returns means that estimating equation 2 for different horizons $k \in \{1, 3, 12, 24, 36, 48\}$ allows the analysis to search for the time period over which investor sentiment becomes impounded into market returns. For example, when k is set to 3, the coefficient $\beta^{(k)}$ will capture the effect of S_t on $\frac{1}{3}\sum_{i=1}^{3} r_{t+i}$ that is not already incorporated into $\frac{1}{3}\sum_{i=1}^{3} r_{t-1+i}$, i.e. r_t, r_{t+1}, and r_{t+2}. **[KK: I think the original notation's subscript was incorrect (I have commented it out above). Please verify that my change here is correct. PZ: I was trying to point out the new component in $\frac{1}{3}\sum_{i=1}^{3} r_{t+i}$ compared to $\frac{1}{3}\sum_{i=1}^{3} r_{t-1+i}$.]** If $\beta^{(k)}$ is statistically significant, then it can be claimed that the S_t retains predictive power on 3-month-ahead returns independently of any effect via current-period, 1-month-ahead, and 2-months-ahead returns. By assembling estimation results for multiple horizon lengths, we map the predictive power of different sentiment indicators along the time line.

The horizon lengths we use here contain both monthly and long-horizon frequencies. As is well known in the literature on long-horizon regression, the overlapping dependent variables will introduce strong autocorrelation within the residuals and therefore lead to biased and in most cases inconsistent estimates for least square coefficients (see e.g. Valkanov 2003). Furthermore, the distributions of the estimated coefficients are often not normal, together with the calculated standard errors being incorrect. As a result standard hypothesis tests do not provide reliable results. We use bootstrap methods to correct for the bias, as detailed in the online appendix.

4 Single-factor regressions

Tables 5 and 6 present the sentiment-indicator coefficients from regression Equation 1. Table 5 is based on original investor-sentiment indicators, whereas Table 6 is based on orthogonalized sentiment indicators. In both tables coefficient estimates $\widehat{\beta^{(k)}}$ are reported, with the adjusted p-values from bootstrap distributions in parentheses. Each p-value below 5% is denoted by an asterix (*).

As shown in Valkanov (2003), the R^2 in long-horizon regressions may have undesirable characteristics such as failure to converge to one in probability, and is likely to become inflated as the horizon length increases. Consequently, little may be gained by dwelling on long-horizon regression R^2 values. Due to this fact, we follow Fama and French's (1988a) practice of reporting only coefficient estimates and significance levels, omitting R^2 values from Tables 5 and 6. Instead, we discuss the R^2 at 1-month horizon briefly.

In line with the evidence from single-factor models of return predictability (e.g. Fama and French 1988b), the magnitudes of R^2 at 1-month horizon are very small. R^2 values remain below 5% for all seven indicators, and this is true both before and after orthogonalization. We leave R^2 values' market-efficiency implications to readers.[10]

4.1 Signs

Although theory (e.g. Shiller et al. 1984, Summers 1986, and Scheinkman and Xiong 2003) places no restrictions on the horizon over which investor sentiment may affect asset prices, it is understood that when investor sentiment is high, current asset prices will be driven up and therefore reduce expected future returns. When investor sentiment is low the opposite is generally true. An empirical question therefore is: How long does it take for the market to demonstrate such a negative correlation between investor sentiment and future returns? The long-horizon structure in our model provides an opportunity to answer this question.

Theory predicts positive $\widehat{\beta^{(k)}}$ for all indicators that reflect sentiment negatively ($CEFD$ and $PDND$), and negative $\widehat{\beta^{(k)}}$ for all indicators that reflect sentiment positively ($MICS$, $TURN$, $NIPO$, $RIPO$ and $NEIF$). In the rest of this section we discuss whether the signs of $\widehat{\beta^{(k)}}$ stay as expected over various horizons.

Most coefficient estimates' signs are consistent with those predicted by and as found in the existing literature (74 out of 84 coefficients across Tables 5 and 6).

All but one of the coefficients with the unexpected sign are statistically insignificant. There is weak evidence of a long-term reversal pattern for predicting VWR with $NIPO$ and $NEIF$, although few of these negative coefficients are statistically significant.

Perhaps the most interesting finding regarding the signs of coefficients comes from $RIPO$. The correlation often stays positive for short horizons (1 month and 3 months) — contrary to theoretical predictions. The correlation turns negative for longer horizons. While behavioral asset pricing theories and anecdotal evidence generally agree that firms and investment banks

[10]After all, as discussed by Fama (1991, p.1576), "academics largely agree on the facts that emerge from the tests, even when they disagree about their implications for efficiency."

Table 5: Coefficients of original sentiment indicators and p-values

This table records the estimated coefficient and p-values in the regression equation
$\frac{1}{k}\sum_{i=1}^{k} r_{t+i} = c^{(k)} + \beta^{(k)} S_t + \epsilon_t^{(k)}$ where
(i) r can refer to either equal-weighted return (EWR) or value-weighted return (VWR);
(ii) S represents a single-factor regressor, i.e. one of seven sentiment indicators and can refer to $MICS$, $CEFD$, $TURN$, $NIPO$, $RIPO$, $NEIF$, or $PDND$;
(iii) k represents the horizon length and can take the values 1, 3, 12. 24, 36, or 48.
(iv) the coefficient $\beta^{(k)}$ represents how sensitive the future return is to investor sentiment, giving the horizon length k. If $\beta^{(k)}$ is statistically significant then evidence of predictive power in the investor-sentiment indicator is present.

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
$MICS$	-0.000321	-0.000684	-0.000334	-0.000746	-0.000214	-0.000520	-0.000155	-0.000322	-0.000185	-0.000257	-0.000228	-0.000214
	(0.047*)	(0.002*)	(0.027*)	(0.000*)	(0.055)	(0.000*)	(0.059)	(0.000*)	(0.013*)	(0.000*)	(0.000*)	(0.000*)
$CEFD$	0.000102	0.000163	0.000121	0.000092	0.000199	0.000273	0.000128	0.000245	-0.000011	0.000170	0.000015	0.000263
	(0.402)	(0.381)	(0.372)	(0.421)	(0.238)	(0.218)	(0.287)	(0.121)	(0.468)	(0.146)	(0.473)	(0.020*)
$TURN$	-0.007972	-0.009969	-0.007157	-0.008858	-0.009255	-0.006883	-0.013667	-0.008347	-0.013691	-0.006644	-0.013285	-0.004340
	(0.100)	(0.108)	(0.106)	(0.119)	(0.041*)	(0.143)	(0.004*)	(0.026*)	(0.002*)	(0.030*)	(0.001*)	(0.072)
$NIPO$	-0.000078	-0.000300	-0.000105	-0.000297	-0.000042	-0.000239	0.000014	-0.000147	0.000023	-0.000142	0.000018	-0.000125
	(0.207)	(0.006*)	(0.092)	(0.003*)	(0.241)	(0.000*)	(0.382)	(0.000*)	(0.293)	(0.000*)	(0.310)	(0.000*)
$RIPO$	-0.000014	0.000102	-0.000009	0.000066	-0.000114	-0.000046	-0.000171	-0.000065	-0.000149	-0.000038	-0.000121	-0.000007
	(0.451)	(0.246)	(0.467)	(0.296)	(0.058)	(0.290)	(0.000*)	(0.085)	(0.000*)	(0.153)	(0.000*)	(0.407)
$NEIF$	-0.016613	-0.017561	-0.010525	-0.010381	-0.002942	-0.010815	0.007607	0.000695	0.008825	-0.003956	0.012378	-0.002609
	(0.216)	(0.262)	(0.293)	(0.336)	(0.417)	(0.267)	(0.245)	(0.471)	(0.165)	(0.322)	(0.049*)	(0.352)
$PDND$	0.000267	0.000426	0.000293	0.000547	0.000330	0.000479	0.000245	0.000314	0.000190	0.000233	0.000100	0.000122
	(0.107)	(0.074)	(0.072)	(0.024*)	(0.018*)	(0.005*)	(0.015*)	(0.002*)	(0.024*)	(0.003*)	(0.124)	(0.039*)

This table shows the coefficients of original sentiment indicators in regressions at six horizon lengths. Each indicator is used as the only regressor to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, using moving-block resampling of residuals with block length 10.

Table 6: Coefficients of orthogonalized sentiment indicators and p-values

This table records the estimated coefficient and p-values in the regression equation
$\frac{1}{k}\sum_{i=1}^{k} r_{t+i} = c^{(k)} + \beta^{(k)} S_t + \epsilon_t^{(k)}$ where
(i) r can refer to equal-weighted return (EWR) or value-weighted return (VWR);
(ii) S represents a single-factor regressor, i.e. one of seven orthogonalized sentiment indicators and can refer to $MICS^{\perp}$, $CEFD^{\perp}$, $TURN^{\perp}$, $NIPO^{\perp}$, $RIPO^{\perp}$, $NEIF^{\perp}$, or $PDND^{\perp}$;
(iii) k represents the horizon length and can take the values 1, 3, 12. 24, 36, or 48.
(iv) the coefficient $\beta^{(k)}$ represents how sensitive the future return is to investor sentiment, giving the horizon length k. If $\beta^{(k)}$ is statistically significant then evidence of predictive power in the investor-sentiment indicator is present.

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
$MICS^{\perp}$	-0.005213	-0.010357	-0.004975	-0.012194	-0.001638	-0.007810	-0.000830	-0.005523	0.000038	-0.002801	-0.002258	-0.004669
	(0.123)	(0.040*)	(0.087)	(0.003*)	(0.257)	(0.000*)	(0.318)	(0.000*)	(0.517)	(0.012*)	(0.035*)	(0.000*)
$CEFD^{\perp}$	0.002513	0.001684	0.001515	-0.000097	0.003315	0.002354	0.002941	0.000547	0.001805	-0.000614	0.002328	0.000866
	(0.249)	(0.356)	(0.325)	(0.493)	(0.096)	(0.219)	(0.058)	(0.380)	(0.106)	(0.329)	(0.041*)	(0.204)
$TURN^{\perp}$	-0.002542	-0.007756	0.000822	-0.005488	0.000309	-0.003911	-0.000016	-0.004989	-0.000793	-0.006240	-0.002238	-0.006894
	(0.322)	(0.130)	(0.430)	(0.171)	(0.455)	(0.156)	(0.501)	(0.013*)	(0.363)	(0.000*)	(0.153)	(0.000*)
$NIPO^{\perp}$	-0.003380	-0.007320	-0.005218	-0.007468	-0.003204	-0.005897	-0.002051	-0.003628	-0.001046	-0.002933	-0.000968	-0.002647
	(0.143)	(0.036*)	(0.023*)	(0.016*)	(0.040*)	(0.002*)	(0.057)	(0.000*)	(0.165)	(0.000*)	(0.140)	(0.000*)
$RIPO^{\perp}$	0.001689	0.005125	0.001273	0.003283	-0.001668	-0.000489	-0.002154	-0.001186	-0.001648	-0.000683	-0.000932	-0.000276
	(0.256)	(0.059)	(0.267)	(0.110)	(0.106)	(0.382)	(0.015*)	(0.096)	(0.018*)	(0.162)	(0.085)	(0.299)
$NEIF^{\perp}$	-0.007254	-0.005342	-0.007578	-0.005178	-0.006748	-0.007159	-0.003456	-0.002254	-0.002310	-0.001885	-0.000620	-0.000206
	(0.035*)	(0.141)	(0.007*)	(0.095)	(0.000*)	(0.002*)	(0.005*)	(0.057)	(0.030*)	(0.038*)	(0.026*)	(0.393)
$PDND^{\perp}$	0.001435	-0.002506	0.004834	0.004255	0.006748	0.006417	0.004640	0.004093	0.002656	0.002612	0.000638	0.001020
	(0.331)	(0.275)	(0.047*)	(0.135)	(0.000*)	(0.002*)	(0.000*)	(0.001*)	(0.009*)	(0.004*)	(0.265)	(0.112)

This table shows the coefficients of orthogonalized sentiment indicators in regressions at six horizon lengths. Each orthogonalized indicator is calculated from the corresponding original indicator orthogonalized with 12 fundamental control variables. Each new orthogonalized indicator is then used as the only regressor to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients ofsentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, using moving-block resampling of residuals with block length 10.

are 'timing' the market by launching IPOs when investor sentiment is high, therefore entailing that high $RIPO$ will be followed by low future returns, our empirical findings suggest that the prediction is true only for longer horizons. In other words, in the short term there must exist more complicated dynamics between $RIPO$ and investor sentiment or other sentiment indicators. First, $RIPO$ is also part of market return and may be driven up by low sentiment in previous periods. Second, it is widely agreed that IPO pricing is extremely difficult and that even seasoned professionals — including IPO-underwriting investment banks — can make pricing mistakes, and that underwriter's pricing decision is conditioned by an asymmetric loss function. Therefore high $RIPO$ may simply be a consequence of IPO undervaluation instead of resulting from high demand for IPO equities driven by high investor sentiment. Last but not least, it is possible that $RIPO$ affects returns with a lag. As there is a long lead time for preparing an IPO, high $RIPO$ will make initial public offerings attractive but can only lead to a wave of new IPOs with several months' delay. In this way IPO volume ($NIPO$) will lag $RIPO$. If $NIPO$ constitutes a good proxy of investor sentiment, then $RIPO$ will also affect future returns, but only in a lagged way. These relationships are borne out in the next subsection.

4.2 Significance

First, the seven indicators show different levels of predictive power. The average numbers of significant coefficients across Tables 5 and 6 are 8.5 for $MICS$, 1 for $CEFD$, 4.5 for $TURN$, 7 for $NIPO$, 2.5 for $RIPO$, 4 for $NEIF$, and 7.5 for $PDND$.

As the only direct sentiment indicator, survey data ($MICS$) shows strong predictive power across the full range of horizons. This finding is particularly noteworthy given that $MICS$ achieves such predictive performance despite being designed to capture general consumer sentiment rather than any factors specific to stock-market returns. Since comparatively few studies have focused on predicting market returns with survey data, our results validate survey-based sentiment indicators for use in future studies. It is entirely possible that surveys focusing specifically on investor sentiment can further improve predictive performance.

Number of IPOs ($NIPO$) and dividend premium ($PDND$) also show strong predictive power. $NIPO$ significantly predicts EWR across horizons, whist it also offers some weak evidence of VWR predictivity. $PDND$ performs well over horizons from 3 months up to 3 years.

Market turnover ($TURN$) shows predictivity over horizons in excess of 1 year. Return on IPOs ($RIPO$) only predicts future returns at horizons of 2 years and above.

Net equity issuance ($NEIF$) fails to predict market returns in its original form, but its performance improves after orthogonalization.

The closed-end fund discount ($CEFD$) shows little predictive power, suggesting that its performance in cross-sectional studies (Zweig 1973, DeLong et al. 1990) does not generalize to aggregate market returns.

Compared to $MICS$, $NIPO$ and $PDND$, the predictive performance of $TURN$, $NEIF$ and $RIPO$ appears distinctly less systematic and more variable. We argue that the liquidity-based $TURN$ may reflect heterogeneity in investor beliefs, which does not necessarily lead to bullish or bearish investor sentiment at the aggregate-market level since bullish belief and bearish belief may well cancel out. $NEIF$ may capture firm-level decisions — the choice between debt financing and equity financing is influenced not only by investor sentiment, but notably by tax policy (e.g. tax-deductible interest cost) and capital structure (e.g. debt ceiling). Moreover $NEIF$ may reflect investor sentiment only to the extent that the equity market is more sensitive to sentiment than the bond market. As discussed in last subsection, the dynamics between $RIPO$ and market returns may go beyond a simple form of return predictability, involving complex leading-following order, IPO pricing-error, and lagged influence.

Second, several indicators predict EWR better than VWR. $NIPO$ consistently predicts EWR over all horizons, but for VWR its predictive power is very limited. $MICS$ performs better in explaining future EWR than VWR, particularly after orthogonalization. Previous emipirical studies show that new stocks and small stocks are generally more affected by investor sentiment. Theoretical work argues that these stocks are harder to value and more difficult to arbitrage and hence are more likely to be subject to mispricing.[11] Empirical work on this question has successfully made use of a variety of sentiment indicators.[12] As new stocks and small stocks have lower capitalization levels, they contribute more to equal-weighted index returns than to value-weighted index returns. Therefore compared to VWR, EWR is better suited to capturing the effect of investor sentiment on new and small stocks. Overall our results support the conjecture that small stocks are more prone to being affected by sentiment than large stocks.

Third, considerable predictive power is found in $PDND$, primarily at horizons of 1 year or longer. $RIPO$ becomes significant in explaining VWR for horizons over 2 years. Market turnover ($TURN$) predicts returns at horizons in excess of 1 year. We interpret these findings

[11] A good review on this literature can be found in Baker and Wurgler (2007).

[12] See e.g. Lee, Shleifer and Thaler (1991), Neal and Wheatley (1998), Kamstra et al. (2003) and Edmans et al. (2007).

as evidence that these indicators predict market returns with a lag.[13] When investors seek to switch to dividend-paying firms they are not only searching for 'safety' in the immediate future but also 'safety' in the more distant future. Therefore $PDND$ reflects investor attitude toward the more distant future, and consequently the dividend premium affects future returns with a lag. With regard to $RIPO$, several studies point out that it leads IPO volume ($NIPO$). As $NIPO$ predicts future returns, it is natural for $RIPO$ to predict returns with a lag.

4.3 Robustness

Above we showed that these empirical results are robust to two independent implementations of orthogonalization (with all 12 macroeconomic variables and with the significant subset of the 12 variables for each indicator). In this section we report additional robustness checks, which confirm that our findings are not an artifact of any particular choice made in implementing the bootstrap. We explore the robustness of the results appearing in Tables 5 and 6 using three different approaches: (i) by varying the bootstrap's moving block length, (ii) by employing a paired moving block re-sampling technique inspired by Freedman (1981, 1984), and (iii) by combining (i) and (ii). We first introduce each approach and then briefly discuss the associated results, especially how they compare with those appearing in Tables 5 and 6.

First, given the moving-average structure of overlapping returns, it is arguably more appropriate to choose the block lengths according to the horizon lengths. For instance, at 3-months horizon length the return in Equation 1 or 2 becomes $\frac{1}{3}\sum_{i=1}^{3} r_{t+i}$ and therefore is expected to have the characteristics of a $MA(2)$ process. As a result it is likely that the residuals also follow the $MA(2)$ process. In this case choosing a block length of 3 in the moving block bootstrap will better capture the structure of the original data. By setting the block length equal to the horizon length (1, 3, 12, 24, 36, and 48 months respectively) we obtain the first set of robustness-test results.

Second, since model misspecification in the single-factor regression (Equation 1) is almost certainly present, the influence of any omitted predictor will likely be captured in the residuals. Unless all the possibly omitted predictors are independent of the sentiment indicator, there will be dependence between the regressor and the residuals. As discussed by Freedman (1981, 1984), in such cases it is important to calibrate such dependence within the data generating process

[13] A related argument about the influence of sentiment indicators on returns in a lagged way can be found in Baker and Wurgler (2006). They argue that generally indicators that involve firm supply responses should lag behind indicators based on investor demand.

(DGP) of any bootstrap implementation in order to achieve satisfactory asymptotic results. In fact, Freedman (1984) proves that assuming a joint distribution between the regressors and residuals and bootstrapping them in pairs is at least as sound as the conventional asymptotic methods.

This article is by no means the first study to pair regressors and residuals in the bootstrap or to resample the pair from blocks. Li and Maddala (1997) implicitly follow this approach and combine it with a parametric DGP for the regressors. MacKinnon (2006) suggests the use of a similar approach for all multivariate models. Our second robustness check is constructed by pairing regressor and residuals in the moving-block resampling.

In the third approach, we make both changes mentioned above to the bootstrap DGP.

Since robustness tests only manipulate the bootstrapping process, coefficient estimates from Equation 1 remain unchanged and the level of robustness is reflected solely in variations of bootstrapped standard deviations and the resulting p-values. In what follows we briefly discuss the headlines of the robustness evidence to show that the originally observed patterns persist across all three robustness-check variations.

Compared with the results reported in Table 5, the robustness-check variations have at most a marginal impact upon the number of significant coefficients. For $MICS$ there is a slight decrease from 10 to 9, 10 and 10. For $CEFD$ the change is ambiguous, from 1 to 0, 2 and 3. For $TURN$ the number changes from 6 to 6, 5 and 5. For $NIPO$ the number stays at 6. For $RIPO$ there is an increase from 3 to 3, 5 and 4. For $NEIF$ the change is ambiguous, from 1 to 0, 2 and 2. For $PDND$ there is a marginal decrease from 8 to 7, 8 and 8.

A similar pattern emerges from the robustness checks applied to orthogonalized data in Table 6. For $MICS^{\perp}$ there is a slight decrease from 7 to 6, 7 and 8. For $CEFD^{\perp}$ there is a slight decrease from 1 to 0, 1 and 1. For $TURN^{\perp}$ from 3 to 3, 4 and 4. For $NIPO^{\perp}$ the number increases from 8 to 8, 9 and 9. For $RIPO^{\perp}$ there is a change from 2 to 3, 1 and 1. For $NEIF^{\perp}$ the number changes from 8 to 6, 8 and 8. For $PDND^{\perp}$ there is a marginal decrease from 7 to 7, 6 and 7.

Across four robustness-check variations, $MICS$, $NIPO$, and $PDND$ keep showing significant predictive power, both before and after orthogonalization. Stronger predictability is found in $MICS$, $NIPO$ and $PDND$ for EWR than VWR, aligning with the theoretical prediction that small stocks are more affected by investor sentiment. $CEFD$ remains insignificant. Evidence from $TURN$, $RIPO$ and $NEIF$ fall in a less consistent pattern. The time-lag effect in

$PDND$, $TURN$ and $RIPO$ remains in the robustness-check variations.

Table 7 reports a further robustness summary measure: the count, across the four bootstrap implementations, of the number of times the coefficient is significant in a regression of market returns on sentiment. This count is reported for each indicator, in both original and orthogonalized forms, for horizon lengths of 1, 3, 12, 24, 36, and 48 months, separately for equally-weighted returns (EWR) and value-weighted returns (VWR). Perfectly consistent (perfectly robust) results generate counts of 4 (all significant) or 0 (all non-significant). Least-consistent (least-robust) results generate a count of 2. The significant-coefficient counts reported in Table 7 show that in most cases the four bootstrap implementations lead to the same conclusion: '4' or '0' appears in 143 out of 168 (85.1%) indicator-horizon-return-type combinations. Meanwhile the value '2' appears in only 11 out of 168 (6.5%) combinations. At the level of individual indicators, $NIPO/NIPO^{\perp}$ displays the most consistency, then $PDND/PDND^{\perp}$, after which $MICS/MICS^{\perp}$ and $RIPO/RIPO^{\perp}$ fall in third place. But even this third-place category musters perfect consistency for 87.5% of the models. We infer that the globally best-performing indicators — $MICS$, $NIPO$, and $PDND$ — achieve robust predictive success.

Table 7: Number of significant coefficients across four bootstrap implementations (single-factor regression)

	EWR						*VWR*					
	1-m	3-m	12-m	24-m	36-m	48-m	1-m	3-m	12-m	24-m	36-m	48-m
$MICS$	4	4	4	4	4	4	0	4	0	0	3	4
$MICS^{\perp}$	4	4	4	4	4	4	0	1	0	0	0	3
$CEFD$	0	0	0	1	2	3	0	0	0	0	0	0
$CEFD^{\perp}$	0	0	0	0	0	0	0	0	0	0	0	3
$TURN$	2	1	0	2	2	0	0	1	2	4	4	4
$TURN^{\perp}$	0	0	0	4	4	4	0	0	0	0	0	2
$NIPO$	4	4	4	4	4	4	0	0	0	0	0	0
$NIPO^{\perp}$	4	4	4	4	4	4	0	4	4	3	0	0
$RIPO$	0	0	0	0	0	0	0	0	2	4	4	4
$RIPO^{\perp}$	0	0	0	1	0	0	0	0	0	4	2	0
$NEIF$	0	0	0	0	0	0	0	0	0	0	2	3
$NEIF^{\perp}$	0	0	4	2	3	0	4	4	4	4	4	1
$PDND$	0	4	4	4	4	3	0	0	4	4	4	0
$PDND^{\perp}$	0	0	4	4	4	0	0	2	4	4	4	0

This table reports the significant-coefficient count across four bootstrap implementations of regressing EWR and VWR, respectively, on each sentiment indicator in turn, at horizon lenghts of 1, 3, 12, 24, 36, and 48 months. For each indicator-return-horizon combination the four bootstrap implementations are: one primary and three robustness-check variants. Perfect robustness is indicated by a significant-coefficient count of either 4 or 0. A count of 2 indicates the least-robust case. Results are reported for both original and orthogonalized variants of each indicator.

5 Impounding horizon: double-factor regressions

Tables 8 and 9 present the sentiment-indicator coefficient estimates from Equation 2, which includes lagged returns as an additional predictor. Table 8 is based on original investor-sentiment indicators, whereas Table 9 is based on orthogonalized sentiment indicators. Both tables report $\widehat{\beta^{(k)}}$-coefficient estimates, with bootstrapped p-values in parentheses. Each $< 5\%$ p-value is marked with an asterix (*).

We interpret results in Tables 8 and 9 through (i) contrasting the results with those obtained in Section 4 (single-factor regression), and (ii) pointing out that by controlling for average returns realized up to time T-1 (T representing horizon length), the specification of double-factor regression reveals whether the sentiment indicator retains predictive power on month-T returns that is independent of its effect on returns in all the T-1 proceding months. The latter approach offers additional and explicit searches for (a) the time lag with which investor sentiment is expected to become impounded into market prices and (b) the duration of investor sentiment's influence on market prices before such influence gradually decays to zero.

For reasons discussed in Section 4, we follow Fama and French's (1988a) practice of reporting only coefficient estimates and significance levels, omitting R^2 values from Tables 8 and 9. Due to the additional predictor, the R^2 at 1-month horizon is generally higher than that in Section 4. However the typical values of R^2 still fall below 10%.

5.1 Signs

The coefficient signs reported in Tables 8 and 9 generally support the hypothesis that investor sentiment is negatively correlated with future returns (70 out of 84 coefficients in Table 8 and 67 out of 84 in Table 9). However slightly greater inconsistency is present, compared to single-factor regression results. The inconsistency is particularly prominent in $CEFD$ and $NEIF$.

All but one of the coefficients with the unexpected sign are statistically insignificant. There exists weak evidence of a long-term reversal pattern from $CEFD$ and $NEIF$, although few reverse-signed coefficients are statistically significant.

The direct indicator $MICS$ has negative coefficients in all 24 regressions across Tables 8 and 9.

The six indirect indicators also have expected signs in most cases. The fraction of expected signs is 58/72 (81%) in Table 8, and 55/72 (76%) in Table 9. Compared to the fractions from single-factor regressions, the decreases mainly come from two indicators: $CEFD$ and $NEIF$. The former has only 5 (Table 8) and 3 (Table 9) coefficients with the expected positive signs, while the latter has only 6 (Table 8) and 8 (Table 9).

The positive coefficients on $RIPO$ found at shorter horizons in single-factor analysis (Section

Table 8: Coefficients of original sentiment indicators and p-values

This table records the estimated coefficient and p-values in the regression equation
$\frac{1}{k}\sum_{i=1}^{k} r_{t+i} = c^{(k)} + \alpha^{(k)}(\frac{1}{k}\sum_{i=1}^{k} r_{t-1+i}) + \beta^{(k)}S_t + \epsilon_t^{(k)}$ where
(i) r can refer to equal-weighted return (EWR) or value-weighted return (VWR);
(ii) S represents a single-factor regressor, i.e. one of the seven sentiment indicators and can refer to $MICS$, $CEFD$, $TURN$, $NIPO$, $RIPO$, $NEIF$ or $PDND$;
(iii) k represents the horizon length and can take the values 1, 3, 12. 24, 36, or 48.
(iv) the coefficient $\beta^{(k)}$ represents how sensitive the future return is to investor sentiment, giving the horizon length k. If $\beta^{(k)}$ is statistically significant then evidence of predictive power in the investor-sentiment indicator is present.

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
$MICS$	-0.000319	-0.000628	-0.000167	-0.000348	-0.000030	-0.000066	-0.000008	-0.000013	-0.000015	-0.000017	-0.000015	-0.000008
	(0.037*)	(0.000*)	(0.014*)	(0.000*)	(0.127)	(0.040*)	(0.235)	(0.212)	(0.038*)	(0.099)	(0.011*)	(0.188)
$CEFD$	0.000095	0.000058	-0.000047	-0.000069	0.000003	0.000015	-0.000023	-0.000001	-0.000020	-0.000001	-0.000006	0.000018
	(0.406)	(0.438)	(0.377)	(0.340)	(0.469)	(0.416)	(0.171)	(0.483)	(0.120)	(0.470)	(0.322)	(0.135)
$TURN$	-0.008007	-0.008202	-0.002672	-0.003032	-0.001308	-0.001317	-0.000404	-0.000291	-0.000594	-0.000439	-0.000419	-0.000180
	(0.100)	(0.137)	(0.178)	(0.196)	(0.120)	(0.181)	(0.284)	(0.367)	(0.136)	(0.251)	(0.152)	(0.348)
$NIPO$	-0.000083	-0.000271	-0.000066	-0.000118	-0.000009	-0.000035	0.000003	-0.000010	-0.000001	-0.000016	-0.000002	-0.000013
	(0.183)	(0.005*)	(0.056)	(0.009*)	(0.245)	(0.019*)	(0.296)	(0.108)	(0.428)	(0.006*)	(0.198)	(0.002*)
$RIPO$	-0.000025	-0.000041	-0.000078	-0.000135	-0.000063	-0.000068	-0.000027	-0.000021	-0.000025	-0.000021	-0.000013	-0.000010
	(0.390)	(0.377)	(0.061)	(0.020*)	(0.000*)	(0.000*)	(0.000*)	(0.004*)	(0.000*)	(0.000*)	(0.001*)	(0.007*)
$NEIF$	-0.016088	-0.017546	-0.004367	-0.004884	0.001494	-0.000745	0.001683	0.000730	0.001489	0.000169	0.000660	-0.000447
	(0.214)	(0.215)	(0.296)	(0.314)	(0.300)	(0.436)	(0.094)	(0.330)	(0.059)	(0.448)	(0.179)	(0.287)
$PDND$	0.000298	0.000556	0.000233	0.000421	0.000097	0.000130	0.000038	0.000054	0.000015	0.000028	0.000011	0.000020
	(0.086)	(0.018*)	(0.008*)	(0.001*)	(0.001*)	(0.001*)	(0.003*)	(0.001*)	(0.074)	(0.023*)	(0.057)	(0.009*)

This table shows the coefficients of original sentiment indicators in regressions at six horizon lengths. Each indicator is used with first-order lag of the dependent variable as the regressors to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, using moving-block resampling of residuals with block length 10.

Table 9: Coefficients of orthogonalized sentiment indicators and p-values

This table records the estimated coefficient and p-values in the regression equation
$\frac{1}{k}\sum_{i=1}^{k} r_{t+i} = c^{(k)} + \alpha^{(k)}(\frac{1}{k}\sum_{i=1}^{k} r_{t-1+i}) + \beta^{(k)} S_t + \epsilon_t^{(k)}$ where
(i) r can refer to equal-weighted return (EWR) or value-weighted return (VWR);
(ii) S represents a single-factor regressor, i.e. one of seven orthogonalized sentiment indicators and can refer to $MICS^{\perp}$, $CEFD^{\perp}$, $TURN^{\perp}$, $NIPO^{\perp}$, $RIPO^{\perp}$, $NEIF^{\perp}$ or $PDND^{\perp}$;
(iii) k represents the horizon length and can take the values 1, 3, 12. 24, 36, or 48.
(iv) the coefficient $\beta^{(k)}$ represents how sensitive the future return is to investor sentiment, giving the horizon length k. If $\beta^{(k)}$ is statistically significant then evidence of predictive power in the investor-sentiment indicator is present.

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
$MICS^{\perp}$	-0.005383	-0.011855	-0.004981	-0.008929	-0.001393	-0.002117	-0.000198	-0.000314	-0.000407	-0.000448	-0.000353	-0.000517
	(0.117)	(0.011*)	(0.007*)	(0.000*)	(0.008*)	(0.003*)	(0.238)	(0.201)	(0.025*)	(0.044*)	(0.010*)	(0.006*)
$CEFD^{\perp}$	0.002072	-0.000010	-0.000831	-0.001865	0.000011	-0.000473	-0.000238	-0.000469	-0.000177	-0.000367	-0.000102	-0.000091
	(0.295)	(0.488)	(0.304)	(0.194)	(0.484)	(0.245)	(0.157)	(0.052)	(0.153)	(0.044*)	(0.207)	(0.268)
$TURN^{\perp}$	-0.002510	-0.006790	0.000284	-0.001896	-0.000559	-0.001067	0.000188	-0.000080	-0.000300	-0.000613	-0.000227	-0.000398
	(0.323)	(0.140)	(0.464)	(0.274)	(0.246)	(0.144)	(0.317)	(0.441)	(0.164)	(0.056)	(0.167)	(0.112)
$NIPO^{\perp}$	-0.003538	-0.007015	-0.003925	-0.003789	-0.000882	-0.001220	-0.000187	-0.000348	-0.000176	-0.000404	-0.000157	-0.000296
	(0.126)	(0.027*)	(0.004*)	(0.015*)	(0.017*)	(0.011*)	(0.161)	(0.085)	(0.102)	(0.014*)	(0.057)	(0.011*)
$RIPO^{\perp}$	0.001434	0.001283	-0.001724	-0.003401	-0.001555	-0.001904	-0.000498	-0.000569	-0.000484	-0.000640	-0.000247	-0.000358
	(0.292)	(0.335)	(0.074)	(0.014*)	(0.000*)	(0.000*)	(0.001*)	(0.001*)	(0.000*)	(0.000*)	(0.001*)	(0.000*)
$NEIF^{\perp}$	-0.007169	-0.005309	-0.004363	-0.003440	-0.000616	-0.001070	0.000076	0.000031	0.000145	0.000020	-0.000002	-0.000086
	(0.028*)	(0.130)	(0.007*)	(0.056)	(0.118)	(0.051)	(0.378)	(0.453)	(0.203)	(0.459)	(0.490)	(0.278)
$PDND^{\perp}$	0.002079	0.002847	0.005804	0.008877	0.002568	0.003105	0.000920	0.001252	0.000375	0.000749	0.000238	0.000543
	(0.251)	(0.226)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.008*)	(0.000*)	(0.011*)	(0.000*)

This table shows the coefficients of orthogonalized sentiment indicators in regressions at six horizon lengths. Each orthogonalized indicator is calculated from the corresponding original indicator orthogonalized with 12 fundamental control variables. Each new orthogonalized indicator is then used with first-order lag of the dependent variable as the regressors to explain both value-weighted and equal-weighted future NYSE returns at 1 months, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, using moving-block resampling of residuals with block length 10.

4) persist at 1-month horizon in Table 9.

5.2 Significance

As in the single-factor regressions, there are marked differences between indicators. The average number of significant coefficients across Tables 8 and 9 is 8 for $MICS$, 0.5 for $CEFD$, 0 for $TURN$, 6 for $NIPO$, 9 for $RIPO$, 1 for $NEIF$, and 9.5 for $PDND$. These results clearly show that the indicators are far from equally informative over longer horizons.

The indicators that perform well in single-factor regressions — $MICS$, $NIPO$ and $PDND$ — retain strong predictive performance over and above market return's self-explanatory power. The performance of $RIPO$ is better in Tables 8 and 9 than in Tables 5 and 6. As in single-factor regressions, $CEFD$ fails to predict future returns. Relative to single-factor regressions, the predictive power of $TURN$ and $NEIF$ fades once the autocorrelation in market return is taken into account.

EWR is still marginally better explained by sentiment indicators than VWR, supporting the hypothesis that small stocks are more affected by investor sentiment. However compared to Section 4 the finding becomes less clear-cut, being strongly evident for only $NIPO$ and arguably present for $MICS$ and $PDND$.

The dividend premium ($PDND$) maintains strong predictive power at 3-month and longer horizons, both before and after orthogonalization. Similar results emerge for first-day IPO return ($RIPO$), which also becomes significant over 3-month and longer horizons, both before and after orthogonalization. These estimates provide stronger evidence that $PDND$ and $RIPO$ have a persistent effect on market returns across a range of horizons.

$MICS$, $NIPO$ and $PDND$ all display sustained predictive power across multiple horizons. $MICS$' and $NIPO$'s predictivity come into effect from 1-month and last up to 4 years. $PDND$'s predictivity comes online at the 3-month horizon, and also lasts up to 4 years.

$CEFD$ and $TURN$ show little predictive power in double-factor regressions. The predictive power that $TURN$ displayed in single-factor regressions may be explained as a short-horizon effect that was then picked up at longer horizons due to the slowly decaying autocorrelation of market returns.

$NEIF$ shows little predictive power, and orthogonalized $NEIF$ shows predictive power for VWR at short horizons (1 month to 3 months), suggesting that $NEIF$'s multi-horizon predictive power in single-factor regressions may result from predicting a highly persistent component of market returns.

$RIPO$ predicts VWR at long horizons of 12 months and above in double-factor regressions, verifying its predictive power found in single-factor regressions for VWR at 2 years and above.

5.3 Robustness

The robustness tests implemented here emulate the template developed in Section 4.3: setting the block lengths equal to the horizon lengths (1, 3, 12, 24, 36 and 48 months respectively), by adopting a paired moving block resampling technique (pairing the sentiment indicator and residuals), and by making both changes simultaneously. In this subsection we briefly discuss the robustness-check results as compared to Tables 8 and 9.

Compared with the results reported in Table 8, the robustness-check variations have at most a marginal impact upon the number of significant coefficients. For $MICS$ there is a slight decrease from 7 to 7, 6 and 6. For $CEFD$ the count stays as 0 across four variants. For $TURN$ the number changes from 0 to 0, 1 and 0. For $NIPO$ the number stays as 5. For $RIPO$ there is a decrease from 9 to 9, 8 and 8. For $NEIF$ the count changes from 0 to 0, 0 and 2. For $PDND$ there is a change from 9 to 9, 8 and 10.

A similar pattern emerges from the robustness checks applied to orthogonalized data in Table 9. For $MICS^{\perp}$ the count stays as 9. For $CEFD^{\perp}$ the count stays as 1. For $TURN^{\perp}$ the number changes from 0 to 1, 0 and 0. For $NIPO^{\perp}$ the number increases from 7 to 7, 9 and 9. For $RIPO^{\perp}$ there is a change from 9 to 9, 9 and 10. For $NEIF^{\perp}$ the count stays as 2. For $PDND^{\perp}$ there is a marginal decrease from 10 to 10, 10 and 9.

Table 10 reports the same robustness summary measure as in Table 7 — the count of the number of times for a significant coefficient across robustness-check measures. Perfectly consistent (perfectly robust) results generate counts of 4 (all significant) or 0 (all non-significant). Least-consistent (least-robust) results generate a count of 2. The significant-coefficient counts reported in Table 10 show that in most cases the four bootstrap implementations lead to the same conclusion: ‘4’ or ‘0’ appears in 153 out of 168 (91.1%) indicator-horizon-return-type combinations. Meanwhile the value ‘2’ appears in only 5 out of 168 (3.0%) combinations.

Table 10: Number of significant coefficients across four bootstrap implementations (double-factor regression)

	EWR						*VWR*					
	1-m	3-m	12-m	24-m	36-m	48-m	1-m	3-m	12-m	24-m	36-m	48-m
$MICS$	4	4	2	0	1	0	4	4	0	0	3	4
$MICS^{\perp}$	4	4	4	0	4	4	0	4	4	0	4	4
$CEFD$	0	0	0	0	0	0	0	0	0	0	0	0
$CEFD^{\perp}$	0	0	0	0	4	0	0	0	0	0	0	0
$TURN$	1	0	0	0	0	0	0	0	0	0	0	0
$TURN^{\perp}$	0	0	0	0	1	0	0	0	0	0	0	0
$NIPO$	4	4	4	0	4	4	0	0	0	0	0	0
$NIPO^{\perp}$	4	4	4	2	4	4	0	4	4	0	0	2
$RIPO$	0	2	4	4	4	4	0	0	4	4	4	4
$RIPO^{\perp}$	0	4	4	4	4	4	0	1	4	4	4	4
$NEIF$	0	0	0	0	0	0	0	0	0	1	1	0
$NEIF^{\perp}$	0	0	0	0	0	0	4	4	0	0	0	0
$PDND$	4	4	4	4	3	4	0	4	4	4	2	1
$PDND^{\perp}$	0	4	4	4	4	4	0	3	4	4	4	4

This table reports the significant-coefficient count across four bootstrap implementations of regressing EWR and VWR, respectively, on each sentiment indicator in turn, at horizon lenghts of 1, 3, 12, 24, 36, and 48 months. For each indicator-return-horizon combination the four bootstrap implementations are: one primary and three robustness-check variants. Perfect robustness is indicated by a significant-coefficient count of either 4 or 0. A count of 2 indicates the least-robust case. Results are reported for both original-form and orthogonalized variants of each indicator.

6 Composite indicators

One approach to dealing with the multitude of sentiment indicators is to adopt the 'diffusion index' method of extracting a composite summary indicator, which is then utilized in further analysis.

In their widely cited work, Baker and Wurgler (2006, 2007) construct a diffusion index — hereafter referred to as the sentiment index $SFPC$ — of the first principal component of the six indirect investor-sentiment indicators. However, Baker and Wurgler (2006, 2007) proceed without subjecting the set of indirect indicators to prescreening, which potentially leads to overfitting in predictive analysis. As a substantial literature argues,[14] such a lack of prescreening is likely to impound an undesirable level of idiosyncratic noise into the composite summary indicator, thereby leading to potential overfitting and exacerbating forecasting errors. According to Subramanian and Simon (2013), such overfitting and consequential forecasting errors are likely to be present in any predictive analysis, even for low-dimention regressions as in the present setting.

In order to mitigate such risk of overfitting, we propose a 'targeted diffusion index' as a further refinement of $SFPC$. We follow Bai and Ng's (2008) approach of selecting a subset of targeted predictors, using consistent and significant predictive power as a marker for inclusion in the subset, from which we construct a new diffusion index that is the first principal component of this subset. Since $CEFD$, $TURN$ and $NEIF$ fail to predict market returns consistently in Sections

[14]e.g. Forni et al. (2005), Bair et al. (2006), Boivin and Ng (2006), and Bai and Ng (2008).

4 and 5, inclusion of these indicators in a diffusion index could potentially introduce more noise than signal. $RIPO$ fails to demonstrate predictive power in Section 4, but its performance improves in Section 5 over long horizons. We adopt a cautious approach to interpreting these results, and classify $RIPO$ as an inconsistent predictor. This leaves $MICS$, $NIPO$ and $PDND$. Additionally, the time-lag effect for $PDND$ requires appropriate treatment for accurate time-series structure.

Thus we select $MICS$, $NIPO$, $PDND$ as our 'targeted' subset of indicators, and we apply a 3-month lag structure to $PDND$.[15] Each indicator is standardized first, and then orthogonalized with respect to 12 macroeconomic variables. The new diffusion index is obtained as the first principal component of the three targeted, orthogonalized time series, as set out in the following equation:

$$T3\text{-}SFPC_t = 0.4153MICS_t + 0.6867NIPO_t - 0.5967PDND_{t-3} \tag{3}$$

Whereas $MICS$ is a direct sentiment indicator that is not included in Baker and Wurgler's (2007) $SFPC$, both $NIPO$ and $PDND$ are indirect sentiment indicators that are present in $SFPC$. In order to test the overfitting hypothesis, we introduce the following targeted index for comparison with the non-targeted $SFPC$:

$$T2\text{-}SFPC_t = 0.7616NIPO_t - 0.6480PDND_{t-3} \tag{4}$$

If $T2$-$SFPC$ out-performs $SFPC$, then we may conclude that including all six indirect indicators in $SFPC$ results in overfitting that may be eliminated by dropping the four indicators that impound more noise than signal.

We include the new indices in a market-returns-prediction horse race, alongside the widely cited $SFPC$. Results are reported in Tables 11 and 12, for single-factor and double-factor regressions respectively.

In Table 11, both $T2$-$SFPC$ and $T3$-$SFPC$ display consistently better predictive performance than $SFPC$, and this is particularly clear for VWR. The fact that $T2$-$SFPC$ performs better than $SFPC$ confirms that the wide-net approach to constructing $SFPC$ impounds noise and results in overfitting, validating our recourse to targeting. Comparison between $T2$-$SFPC$

[15] Baker and Wurgler (2007) choose a lag of 12 months, following cross-sectional results. However their sentiment index is later used in a market-level analysis, making the 12-month lag possibly sub-optimal. Our results in Sections 4 and 5 suggest that the influence of $PDND$ on market returns comes into effect over 3-month and longer horizons.

and $T3$-$SFPC$ in turn confirms that inclusion of the direct sentiment indicator ($MICS$) further improves the targeted diffusion index. Thus $T3$-$SFPC$'s performance improvement over $SFPC$ is due in part to (i) control of overfitting, and in part to (ii) incorporation of the direct sentiment indicator $MICS$, a powerful predictor of market-level returns.

Comparing $T3$-$SFPC$ in Table 11 with $MICS$ in Table 5, it is evident that these indicators have similar performance: significant coefficients for 4 VWR horizons and 6 EWR horizons. However, $T3$-$SFPC$ is based on orthogonalized data, so the relevant comparison is with $MICS^{\perp}$, for which Table 6 records significant coefficients on 1 VWR horizon and 6 EWR horizons. Thus the gain in moving from the direct sentiment indicator $MICS$ to the composite sentiment indicator $T3$-$SFPC$ is primarily found in VWR predictivity and in non-redundancy with fundamentals information.

From Tables 5 and 6,[16] Tables 8 and 9,[17] and Table 11,[18] no clear answer is forthcoming to the impounding-horizon question. Indicators have predictive power across a range of different horizons, often with gaps over sub-intervals, with no clear impounding-horizon cutoff. This lack of a cutoff is present even in the double-factor regressions reported in Tables 8 and 9. These results — especially the occurrence of significant coefficients following an interval of non-significant coefficients — do not admit to straightforward interpretation.

However in the double-factor regressions of Table 12, $T3$-$SFPC$[19] sustains predictive power over 1-month, 3-month, and 12-month horizons, while $SFPC$'s predictive power is limited to the 1-month horizon. In other words, sentiment as gauged by $SFPC$ becomes impounded into market returns over one month, while sentiment as gauged by $T3$-$SFPC$ becomes impounded into market returns over 12 months. In the case of $T3$-$SFPC$, this impounding horizon holds for both VWR and EWR.

In contrast to 'raw' direct and indirect indicators, both $SFPC$ and $T3$-$SFPC$ have a clear and interpretable impounding-horizon cutoff. Among these composite indicators, $T3$-$SFPC$ distinguishes itself in incorporating less noise by construction, and by having both a longer predictive horizon (1 to 12 months versus 1 month) as well as a wider predictive scope (VWR and EWR versus EWR alone).

[16] single-factor regressions, indicators and orthogonalized indicators respectively

[17] double-factor regressions, indicators and orthogonalized indicators respectively

[18] single-factor regressions, composite indicators

[19] disregarding the marginally significant ($p = 0.045$) result for $T3$-$SFPC$ in predicting EWR at the 48-month horizon

Table 11: Single-factor-regression coefficients and p-values of original and new diffussion indices

This table records the estimated coefficient and p-values in the regression equation
$\frac{1}{k}\sum_{i=1}^{k} r_{t+i} = c^{(k)} + \beta^{(k)} S_t + \epsilon_t^{(k)}$ where
(i) r can refer to either equal-weighted return (EWR) or value-weighted return (VWR);
(ii) S represents a composite sentiment index, either SFPC, T2-SFPC or T3-SFPC;
(iii) k represents the horizon length and can take the values 1, 3, 12. 24, 36, or 48.
(iv) the coefficient $\beta^{(k)}$ represents how sensitive the future return is to investor sentiment, giving the horizon length k. If $\beta^{(k)}$ is statistically significant then evidence of predictive power in the investor-sentiment indicator is present.

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
$SFPC$	-0.005059	-0.008363	-0.004461	-0.007123	-0.002917	-0.004279	-0.001987	-0.003268	-0.000797	-0.002240	-0.000908	-0.002401
	(0.054)	(0.020*)	(0.061)	(0.030*)	(0.111)	(0.066)	(0.122)	(0.024*)	(0.302)	(0.045*)	(0.240)	(0.009*)
$T2-SFPC$	-0.004143	-0.006398	-0.005648	-0.007073	-0.005391	-0.006813	-0.003263	-0.003987	-0.001358	-0.002488	-0.000649	-0.001890
	(0.068)	(0.042*)	(0.008*)	(0.014*)	(0.001*)	(0.000*)	(0.004*)	(0.000*)	(0.096)	(0.002*)	(0.230)	(0.002*)
$T3-SFPC$	-0.004767	-0.007995	-0.006254	-0.008955	-0.005303	-0.007835	-0.003230	-0.004782	-0.001329	-0.002967	-0.001006	-0.002644
	(0.047*)	(0.014*)	(0.005*)	(0.002*)	(0.001*)	(0.000*)	(0.005*)	(0.000*)	(0.100)	(0.000*)	(0.148)	(0.000*)

This table shows the coefficients of original sentiment indicators in regressions at six horizon lengths. Each indicator is used as the only regressor to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, using moving-block resampling of residuals with block length 10.

Table 12: Double-factor-regression coefficients and p-values of original and new diffusion index

This table records the estimated coefficient and p-values in the regression equation
$\frac{1}{k}\sum_{i=1}^{k} r_{t+i} = c^{(k)} + \alpha^{(k)}(\frac{1}{k}\sum_{i=1}^{k} r_{t-1+i}) + \beta^{(k)} S_t + \epsilon_t^{(k)}$ where
(i) r can refer to equal-weighted return (EWR) or value-weighted return (VWR);
(ii) S represents a composite sentiment index, either SFPC, T2-SFPC or T3-SFPC;
(iii) k represents the horizon length and can take the values 1, 3, 12. 24, 36, or 48.
(iv) the coefficient $\beta^{(k)}$ represents how sensitive the future return is to investor sentiment, giving the horizon length k. If $\beta^{(k)}$ is statistically significant then evidence of predictive power in the investor-sentiment indicator is present.

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
$SFPC$	-0.005129	-0.006325	-0.000726	-0.001010	0.000101	0.000277	0.000236	0.000198	0.000216	0.000176	0.000091	0.000033
	(0.043*)	(0.031*)	(0.286)	(0.272)	(0.400)	(0.333)	(0.114)	(0.222)	(0.065)	(0.175)	(0.192)	(0.415)
$T2-SFPC$	-0.004099	-0.005382	-0.003279	-0.002926	-0.000790	-0.000888	-0.000086	-0.000212	-0.000089	-0.000051	-0.000019	-0.000116
	(0.061)	(0.043*)	(0.003*)	(0.028*)	(0.021*)	(0.039*)	(0.307)	(0.179)	(0.237)	(0.393)	(0.430)	(0.153)
$T3-SFPC$	-0.004785	-0.007260	-0.004009	-0.004390	-0.001009	-0.001292	-0.000123	-0.000274	-0.000001	-0.000161	-0.000079	-0.000207
	(0.047*)	(0.011*)	(0.000*)	(0.001*)	(0.004*)	(0.006*)	(0.246)	(0.126)	(0.490)	(0.182)	(0.208)	(0.045*)

This table shows the coefficients of original sentiment indicators in regressions at six horizon lengths. Each indicator is used with first-order lag of the dependent variable as the regressors to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, using moving-block resampling of residuals with block length 10.

7 Conclusion

Long-horizon regression has revealed considerable heterogeneity among direct and indirect sentiment indicators. Heterogneneity is manifest across the horizon of returns as well as across the weighting of returns (equal versus value weighting). A subset of indicators — the Michigan Index of Consumer Sentiment ($MICS$), the number of IPOs ($NIPO$), and the dividend premium ($PDND$) — are more consistent predictors of market returns than the remaining indirect indicators. This result is robust to (i) orthogonalisation of each indicator by 12 variables capturing macroeconomic fundamentals, (ii) the inclusion of one-period-lagged market returns as a predictor, to filter out the effect that sentiment may have had on market returns in previous periods, and (iii) 4 different variations of implementing the bootstrapping procedure.

These results on direct and indirect sentiment indicators can be used to refine the construction of composite indicators, such as Baker and Wurgler's (2006, 2007) first principal component sentiment indicator ($SFPC$). Here we do so by constructing a targeted composite indicator as the first principal component of the subset of indicators that perform consistently well in long-horizon regressions. This has the effect of reducing the amount of noise present in the composite indicator. The resulting targeted composite indicator ($T3$-$SFPC$) performs better than $SFPC$. The targeted composite indicator $T3$-$SFPC$ performs well in predicting not only equally weighted returns (EWR), but also value-weighted returns (VWR). The non-targeted composite indicator $SFPC$ predicts only the former. Controlling for lagged market returns, the targeted composite indicator $T3$-$SFPC$ also performs well in predicting not only 1-month-ahead market returns, but also market returns at 3-month and 12-month horizons. This is in contrast to the non-targeted composite indicator $SFPC$, which predicts only 1-month-ahead market returns when controlling for lagged market returns. By construction, the targeted composite indicator is orthogonalized with respect to 12 macroeconomic variables, and thus represents sentiment that is distinct from the component of expectations that reflects fundamentals.

References

Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17, 223-249.

Bai, J., Ng, S., 2008. Forecasting economic time series using targeted predictors. *Journal of Econometrics* 146, 304-317.

Bair, E., Hastie, T., Paul, D., Tibshirani, R., 2006. Prediction by supervised principal components. *Journal of the American Statistical Association* 101, 119-137.

Baker, M., Stein, J.C., 2004. Market liquidity as a sentiment indicator. *Journal of Financial Markets* 7, 271-299.

Baker, M., Wurgler, J., 2000. The equity share in new issues and aggregate stock returns. *Journal of Finance* 55, 2219-2257.

Baker, M., Wurgler, J., 2004a. A Catering theory of dividends. *Journal of Finance* 59, 1125-1165.

Baker, M., Wurgler, J., 2004b. Appearing and disappearing dividends: the link to catering incentives. *Journal of Financial Economics* 73, 271-288.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. *Journal of Finance* 61, 1645-1690.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market, *Journal of Economic Perspectives* 21, 129-151.

Baker, M., Wurgler, J., Yuan, Y., 2012. Global, local, and contagious investor sentiment. *Journal of Financial Economics* 104, 272-287.

Boivin, J., Ng, S., 2006. Are more data always better for factor analysis?. *Journal of Econometrics* 132, 1679-194.

Brennan, M., Chordia, T., Subrahmanyam, A., 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics* 49, 345-373.

Brennan, M., Subrahmanyam, A., 1996. Market microstructure and asset pricing: on the compensation for illiquidity in stock returns. *Journal of Financial Economics* 41, 441-464.

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

Brown, G.W., Cliff, M.T., 2005. Investor sentiment and asset valuation. *Journal of Business* 78, 405-440.

Campbell, J.Y., 1987. Stock returns and the term structure. *Journal of Financial Economics* 18(2), 373-399.

Campbell, J.Y., 2001. Why long horizons? A study of power against persistent alternatives. *Journal of Empirical Finance* 8, 459-491.

Campbell, J.Y., Shiller, R.J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1, 195-228.

Chung, S., Huang, C., Yeh, C., When does investor sentiment predict stock returns?. *Journal of Empirical Finance* 19, 217-240.

Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1(2): 223–236.

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

DeLong, J.B., Shleifer, A., Summers, H., Waldmann, R., 1990. Noise trading risk in financial markets. *Journal of Political Economy* 98, 703-738.

Edmans, A., Garcia, D., Norli, O., 2007. Sports sentiment and stock returns. *Journal of Finance* 62(4), 1967-1998.

Fama, E.F., 1991. Efficient capital markets: II. *Journal of Finance* 46(5), 1575-1617.

Fama, E.F., French, K.R., 1988a. Permanent and temporary components of stock prices. *Journal of Political Economy* 96(2), 246-273.

Fama, E.F., French, K.R., 1988b. Dividend yields and expected stock returns. *Journal of Financial Economics* 22, 3-25.

Fama, E.F., French, K.R., 1989. Business conditions and expected stock returns. *Journal of Financial Economics* 25, 23-49.

Fama, E.F., Schwert, G.W., 1977. Asset returns and inflation. *Journal of Financial Economics* 5, 115-146.

Fisher, K.L., Statman, M., 2000. Investor sentiment and stock returns. *Financial Analysts Journal* 56, 16-23.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association* 100, 830-840.

Freedman, D.A., 1981. Bootstrapping regression models, *The Annals of Statistics* 9(6), 1218-1228.

Freedman, D.A., 1984. On bootstrapping two-stage least-square estimates in stationary linear models. *The Annals of Statistics* 12(3), 827-842.

Garcia, D., 2013. Sentiemnt during recessions. *Journal of Finance* 3, 1267-1300.

Goetzmann, W., Jorion, P., 1993. Testing the predictive power of dividend yields. *Journal of Finance* 48, 663-679.

Han, B., 2008. Investor sentiment and option prices. *Review of Financial Studies* 21(1), 387-414.

Han, X., Li, Y., 2017. Can investor sentiment be a momentum time-series predictor? Evidence from China. *Journal of Empirical Finance* 42, 212-239.

Hansen, L.P., Hodrick, R.J., 1980. Forward exchange rates as optimal predictors of future spot rates: an empirical analysis. *Journal of Political Economy* 88, 829-853.

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

Hodrick, R.J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. *Review of Financial Studies* 5(3), 357-386.

Jackwerth, J. C., Rubinstein, M., 1996. Recovering probability distributions from option prices. *Journal of Finance* 51, 1611-1631.

Jegadeesh, N., 1991. Seasonality in stock price mean reversion: evidence from the US and the UK. *Journal of Finance* 46(4), 1427-1444.

Kamstra, M.J., Kramer, L.A., Levi, M.D., 2003. Winter blues: a sad stock market value. *American Economic Review* 93(1), 324-343.

Keim, D.B, Stambaugh, R.F., 1986. Predicting returns in the stock and bond markets. *Journal of Financial Economics* 17, 357-390.

Kim, J.S., Ryu, D., Seo, S.W., 2014. Investor sentiment and return predictability of disagreement. *Journal of Banking and Finance* 42, 166-178.

Lahiri, S.N., 1999. Theoretical comparisons of block bootstraps methods. *Annals of Statistics* 27, 386-404.

Lee, W.Y., Jiang, C.X., Indro, D.C., 2002. Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking and Finance* 26, 2277-2299.

Lee, C., Shleifer, A., Thaler, R.H., 1991. Investor sentiment and the closed-end fund puzzle. *Journal of Finance* 46(1), 75-109.

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

Li, H., Maddala, G.S., 1997. Bootstrapping cointegrating regressions. *Journal of Econometrics* 80, 297-318.

Lo, A.W., MacKinlay, A.C., 1988. Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies* 1(1), 41-66.

Ljungqvist, A., Nanda, V., Singh, R., 2006. Hot Markets, investor sentiment, and IPO pricing. *Journal of Business* 79(4), 1667-1702.

Ludvigson, S.C., Ng, S., 2007. The empirical risk-return relation: a factor analysis approach. *Journal of Financial Economics* 83, 171-222.

Mackinnon, J., 2006. Bootstrap methods in econometrics. *Queen's Economics Department Working Paper No.1028.*

Mishkin, F., 1992. Is the fisher effect for real?: A reexamination of the relationship between inflation and interest rates. *Journal of Monetary Economics* 30(2), 195-215.

Neal, R., Wheatley, S.M., 1998. Do measures of investor sentiment predict returns?. *Journal of Financial and Quantitative Analysis* 33, 523-547.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation cnsistent covariance matrix. *Econometrica* 55, 703-708.

Politis, D.N., Romano, J.P., 1992. General resampling scheme for triangular arrays of mixing random variables with application to the problem of spectral density estimation. *Annals of Statistics* 20, 1985–2007.

Poterba, J.M., Summers, L.H., 1988. Mean reversion in stock prices. *Journal of Financial Economics* 22(1), 27-59.

Qiu, L., Welch, I., 2006. Investor sentiment measures. *NBER Working Paper*, 10794

Scheinkman, J.A., Xiong, W., 2003. Overconfidence and speculative bubbles. *Journal of Political Economy* 111(6), 1183-1219.

Schmeling, M., 2009. Investor sentiment and stock returns: some international evidence. *Journal of Empirical Finance* 16, 394-408.

Shiller, R.J., 1999. Measuring bubble expectations and investor confidence. *NBER Working Paper*, 7008

Shiller, R.J., Fischer, S., Friedman, B.M., 1984. Stock prices and social dynamics,.*Brookings Papers on Economic Activity* 2, 457-510.

Stambaugh, R., 1999. Predictive regressions. *Journal of Financial Economics* 54, 375-421.

Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20(2), 147-162.

Subramanian, J., Simon, R., 2013. Overfitting in prediction models Is it a problem only in high dimensions?. *Contemporary Clinical Trials* 36(2), 636-641.

Summers, L.H., 1986. Does the stock market rationally reflect fundamental values?. *Journal of Finance* 41(3), 591-601.

Sun, L., Najand, M., Shen, J., 2016. Stock return predictability and investor sentiment: a high-frequency perspective. *Journal of Banking and Finance* 73, 147-164.

Swaminathan, B., 1996. Time-varying expected small firm returns and closed-end fund discounts. *Review of Financial Studies* 9(3), 845-887.

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

Valkanov, R., 2003. Long-horizon regressions: theoretical results and applicatiosn. *Journal of Financial Economics* 68, 201-232.

Wang, Y., Keswani, A., Taylor, S., 2006. The relationships between sentiment, returns and volatility. *International Journal of Forecasting* 22, 109-123.

Wu, C.F.J., 1986. Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* 14, 1261–1295.

Zweig, M.E., 1973. An investor expectation stock price predictive model using closed-end fund premiums. *Journal of Finance* 28(1), 67-78.

A Online Appendix: Bootstrap Procedure

Several methods have been proposed for obtaining unbiased estimates and standard errors. For instance, Hansen and Hodrick (1980) and Newey and West (1987) propose the use of adjusted standard errors in calculating the t statistic. Valkanov (2003) proposes substitution of $t/\sqrt{T}$ for the standard t-test statistic,[20] as the latter does not converge to any well-defined distribution whilst the former does. Our proposed approach is to use bootstrap simulations to generate an empirical sampling distribution for the t-test statistic under the null hypothesis. This approach has several advantages in implementation. Firstly, it is based on less strict asymptotic assumptions than the alternatives and therefore will offer better performance in finite (and particularly small) samples or when the degree of overlapping is relatively large.[21] Secondly, it can deal with not only autocorrelation but also possible heteroskedasticity in the residuals (with the right bootstrap method),[22] while for instance Hansen-Horick standard errors do not correct for heteroskedasticity. Thirdly, bootstrapping is relatively flexible, with a range of different approaches suitable for particular circumstances. Bootstrap can even overcome the initial small-sample problem through careful choice of the most suitable data-generating process to increase the sample size.

We use the moving-block bootstrap approach to deal with both autocorrelation and possible heteroskedasticity in the residuals. In order to take account of possible autocorrelation and heteroskedasticity even at short horizon lengths,[23] the bootstrap is implemented at all horizon lengths. We describe the bootstrap steps below.

(1.) For each pair of (averaged) future return (EWR or VWR) and sentiment indicator (one of the seven non-composite indicators for S_t) as well as a given horizon length (k), regression Equation 1 or Equation 2 is estimated using OLS. Store the coefficient estimate $\widehat{\beta^{(k)}}$, t-test statistic $\widehat{t}$ and the residual series $\widehat{\epsilon_t^{(k)}}$.

(2.) Overlapping moving blocks of 10 residuals are generated from $\widehat{\epsilon_t^{(k)}}$. Moving blocks of fixed length tend to work better (see e.g. Lahiri (1999)). We relax this setting and match block lengths to horizon lengths later in the robustness tests.

(3.) Moving blocks are drawn with replacement until the sample size (given the horizon length)

[20] t is the standard t–test statistic, and T is the sample size.

[21] See e.g. Mishkin (1992) and Goetzmann and Jorion (1993) for evidence on limited performance of the adjusted standard error approach.

[22] See e.g. Wu (1986) for wild bootstrap, Freedman (1984) for pairs bootstrap and Politis and Romano (1992) for block bootstrap which can deal with both autocorrelation and heteroskedasticity.

[23] which may come as a result of, e.g. small-sample biases as discussed in Stambaugh (1999).

is reached. The last moving block is truncated to fit the sample size if needed. The bootstrapped series of residuals are recorded as $\overline{\epsilon_t^{(k)}}$. We allow residuals and investor sentiment indicator to be resampled in pairs later in the robustness tests.

(4.) A pseudo series of the dependent variable (average future returns) is generated under the null hypothesis and recorded as $\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}}$.

The series in single-factor analysis is generated according to the following equation:

$$\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}} = \widehat{c^{(k)}} + \overline{\epsilon_t^{(k)}} \tag{5}$$

where $\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}}$ is the generated dependent variable; $\widehat{c^{(k)}}$ is the estimate of $c^{(k)}$ from the regression Equation 1; $\overline{\epsilon_t^{(k)}}$ is the bootstrapped series of the residuals from step 3.

The series in double-factor analysis is generated recursively according to the following equation:

$$\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}} = \widehat{c^{(k)}} + \widehat{\alpha^{(k)}}(\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t-1+i}}) + \overline{\epsilon_t^{(k)}} \tag{6}$$

where $\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}}$ is the generated dependent variable; $\widehat{c^{(k)}}$ and $\widehat{\alpha^{(k)}}$ are the estimates of $c^{(k)}$ and $\alpha^{(k)}$ from the regression Equation 2; $\overline{\epsilon_t^{(k)}}$ is the bootstrapped series of the residuals from step 3. Following Mackinnon's (2006) recommendations, the pre-sample value of $\frac{1}{k}\sum_{i=1}^{k} r_{t+i}$ is used to start the recursive process.

(5.) Regress the generated pseudo dependent variable $\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}}$ on the estimates of constant and the regressor(s) according to

$$\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}} = \widehat{c^{(k)}} + \beta^{(k)} S_t + \varepsilon_t^{(k)} \tag{7}$$

and

$$\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}} = \widehat{c^{(k)}} + +\widehat{\alpha^{(k)}}(\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t-1+i}}) + \beta^{(k)} S_t + \varepsilon_t^{(k)} \tag{8}$$

in Section 4 and Section 5 respectively. Record the coefficient estimate as $\widetilde{\beta^{(k)}}$ and the t statistic of $\widetilde{\beta^{(k)}}$ from the regression Equation 5 or 6.

(6.) Steps 1 through 5 are repeated 4999 times. The bootstrap sample size is chosen so that $\alpha(1+B)$ becomes an integer, making the simulation closer to be exact, where α is the significance level and B is the bootstrap sample size (MacKinnon, 2006).

(7.) The empirical sampling distribution of the t statistic under the null hypothesis is then obtained by pooling together the 4999 values from step 6.

(8.) Reject the null hypothesis at the 5% level, if the test statistic $\widehat{t}$ locates outside the 95% quantile of the empirical sampling distribution.[24]

[24] As suggested in MacKinnon (2006, p. 21), for hypothesis tests based on signed statistics, we may or may not wish to assume symmetry when calculating p-values. In present study we do not assume symmetry and therefore calculate the p-value under the null as in a single-tail test. This choice is validated by the fact that the empirical distribution generated from data is often heavily skewed in our sample.