Rethinking Measures of Sentiment: A New Approach

H. Christopher Kazemi*

Job Market Paper

This Draft: December 5, 2025

Updated Regularly: Latest Draft

Abstract

I argue that valid measures of investor sentiment must satisfy additional conditions beyond conventional return predictability tests to imply sentiment-induced misvaluation. Specifically, both positive and negative sentiments should explain returns and volatility contemporaneously, *and* forecast returns. I show that several well-known sentiment indexes fail to fully meet these necessary conditions and introduce three new empirical indexes that perform better. These proposed measures demonstrate superior predictive power for returns both in-sample and out-of-sample, particularly over longer horizons; and survive the inclusion of non-sentiment variables known to predict returns. Further evidence shows that their robust forecasting ability extends to broader financial outcomes, including changes in flows to actively-managed equity mutual funds, the VIX, and credit spreads.

Keywords: Investor sentiment; Sentiment-induced misvaluation; Return predictability; Long-run return predictability; Behavioral finance; Efficient market deviations; Empirical asset pricing.

JEL codes: G12, G14, G41, E44, C53.

*Arizona State University, chris.kazemi@asu.edu. I am grateful to my supervisor, Seth Pruitt, for his invaluable mentorship and continued support. I am also thankful to my other advisors, Hendrik Bessembinder, Sunil Wahal, and Oliver Boguth, for their helpful advice and constant guidance. I especially thank Denis Sosyura for his insightful conversations and suggestions. I would also like to thank Michael Barnett, George Aragon, Maziar Kazemi, Yuri Tserlukevich, Laura Lindsey, and David Schreindorfer for their constructive feedback.

1 Introduction

Behavioral models show that psychological biases, such as extrapolation, diagnostic expectations, and overconfidence, can lead to belief-driven asset misvaluation. Relatedly, behavioral empirical research focuses on identifying latent investor sentiment, which distorts beliefs and induces misvaluation, and examines its effects on returns and other financial outcomes. Given the growing influence of social networks, investors can now express their sentiment instantaneously to a wide audience and seek confirmatory information, thereby shaping financial activities (Cookson et al., 2023a,b). The recent surge in retail investor participation, fueled in part by the rise of commission-free platforms like Robinhood, can further amplify the role of sentiment in market dynamics. Consequently, identifying investor sentiment and understanding its effects on financial markets has become more important than ever.

Because sentiment is unobservable, researchers use proxies derived from surveys, composite indexes, text-based analysis, and more recently, AI and large language models (LLMs).[1] With recent advances in technology and data availability, constructing empirical measures of investor sentiment using such proxies is now more feasible than ever before. However, despite extensive work, a clear framework for evaluating these empirical measures remains underdeveloped. The literature often relies on return predictability to validate whether a measure reflects sentiment, but belief-based theories of misvaluation imply additional relationships that are largely overlooked.

In this paper, I propose testable conditions implied by sentiment-induced misvaluation, derived from a parsimonious theoretical framework. I use these joint conditions to evaluate empirical measures of investor sentiment and show that existing measures perform poorly when tested against them. I then introduce new measures that better satisfy the proposed conditions and exhibit stronger predictive power across different asset markets.

I begin by adopting a signaling framework inspired by leading belief-based models in behavioral finance. Treating investor sentiment as an exogenous signal, my model captures the core intuition of this literature: investor sentiment drives prices away from intrinsic value, creating misvaluation that persists until corrected by fundamentals. Based on this setup, I derive three joint conditions for sentiment-induced misvaluation. The first, the *Contemporaneity* condition, relates changes in investor sentiment and its volatility to contemporaneous market returns and volatility. The second,

[1] Common sentiment proxies include survey-based measures such as the University of Michigan Consumer Sentiment Index, the AAII sentiment index, and the Conference Board survey indexes; market-based composites (e.g., Baker and Wurgler, 2006; Huang et al., 2015); textual sentiment derived from news or regulatory filings (e.g., Tetlock, 2007; Jiang et al., 2019); and LLM-based tools (e.g., Bybee, 2023).

the *Predictability* condition, states that investor sentiment levels negatively forecast returns, with stronger effects over longer horizons. The third, the *Consistency* condition, requires that these patterns hold for both positive and negative sentiments.

I test this framework on 11 measures of investor sentiment, including eight widely used survey-based measures and three sentiment indexes from prior literature. I then construct two *sentiment statistics* that quantify each measure's alignment with the theoretical constraints implied by the three conditions and capture their joint statistical significance. I find that commonly used sentiment measures in the literature *do not* fully satisfy the joint conditions from sentiment-induced misvaluation. In particular, the three empirical measures from the literature—the widely used Baker and Wurgler (2006) sentiment index (BW), the Huang et al. (2015) sentiment index which reconstructs BW, and the Jiang et al. (2019) manager sentiment index—generally fail to relate to contemporaneous returns. The survey-based measures often relate to contemporaneous returns but exhibit weak to no return predictability, particularly over short horizons.

I present three new sentiment measures that better satisfy the proposed conditions. The first measure, MPsy, is constructed by aggregating daily firm-level and macro sentiments from LSEG MarketPsych analytics, which provide textual sentiment scores derived from media. The second measure, BW_{adj}, adjusts the timing of the Baker and Wurgler (2006) index. I reduce the lags of the original index's raw components and construct it as the cumulative sum of the first principal component (PC) of their changes. The third measure, CBW_{adj}, is the cumulative sum of the first PC of changes in BW_{adj} and the Conference Board's Consumer Confidence survey. I construct the latter two indexes by ensuring that their contemporaneous changes are positively correlated with 12-month market excess returns. This approach is consistent with the intuition behind the Contemporaneity condition that a measure capable of reflecting the correction of misvaluation should also capture its buildup. BW_{adj} and CBW_{adj} demonstrate that simple adjustments using readily available data can improve existing measures, enabling them to better capture sentiment-induced misvaluation. Among the three, MPsy performs the best—exhibiting both high contemporaneous and predictive relationships with market returns—followed by CBW_{adj} and BW_{adj}.

Next, I extend the analysis to forecast S&P 500 excess returns over longer horizons and two sample periods: 1980–2022 and 1998–2022. I also use out-of-sample and bootstrap resampling procedures, and control for other established return predictors. I find that the newly proposed sentiment measures outperform existing ones in return predictability, as reflected by both their statistical significance and estimated R^2 values, across all tests. While the new indexes perform similarly to the three measures from the literature over short horizons (1-3 months), they consistently explain more return variation beyond the 12-month horizon. Notably, MPsy, which records the highest sentiment statistics, is the only measure with statistically significant forecasting power across all horizons in the bootstrap analysis, reflecting high consistency over the sample period. The high sentiment statistics of MPsy are driven by its strong ability to capture sentiment contemporaneously, which in turn increases its forecasting power. Even after controlling for established return predictors, the new indexes continue to outperform other measures.

The relatively strong short-term (1–3 months) return predictability of sentiment measures from the literature, despite their failure to satisfy the contemporaneous relationships, could initially be viewed as puzzling. I show that this apparent predictability may stem from mistiming in the construction of these indexes, causing them to capture lagged rather than contemporaneous sentiment. Supporting this view, contemporaneous regressions show that lagged versions of the new sentiment measures perform significantly worse, while predictive regressions indicate that such mistimed measures can still forecast returns, particularly over short horizons. This finding explains why sentiment measures exhibit short-term predictive power despite their limited alignment with the joint conditions. By contrast, the three new indexes better satisfy the second part of the Predictability condition, which requires forecasting power to strengthen over longer horizons, and their predicted returns align more closely with actual returns as the horizon increases. Taken together, these findings highlight that the forecasting power of sentiment measures improves markedly when they are also contemporaneously related to returns.

The top-performing measures from return predictability tests are MPsy, CBW_{adj}, BW_{adj}, and BW (only in the post-1998 period), in that order, consistent with the ranking of their sentiment statistics. Averaged across analyses and horizons, MPsy explains more than double the return variations relative to the BW (38.6%, vs 18.3%) over the 1998-2022 period. Similarly, during the 1980-2022 period, CBW_{adj} and BW_{adj} significantly outperform other measures. I show that these overall forecasting R^2 values exhibit a clear positive relationship with the two sentiment statistics. This suggests that the joint conditions provide an effective framework for evaluating empirical sentiment measures by helping identify when a measure is mistimed or improperly scaled.

Lastly, I extend my predictability analysis beyond market returns to other financial outcomes. Specifically, I examine the ability of sentiment measures to forecast changes in active equity fund flows, market-implied volatility (VIX), and Moody's 10-year corporate credit spread.[2] The pre-

[2]The literature documents that sentiment can affect these three financial outcomes. See, for example, fund flows: Brown et al. (2003); Brown and Cliff (2004); Ben-Rephael et al. (2012); Greenwood and Shleifer (2014); Da et al. (2015); market volatility: Da et al. (2015); Chen et al. (2021); Ding et al. (2021); and credit spreads: Tang and Yan (2010); López-Salido et al. (2017).

dictive power of sentiment measures for these outcomes is generally weaker than for return predictability, and the R^2 values are particularly smaller for fund flows and credit spreads. However, MPsy and $\mathrm{CBW}_{\mathrm{adj}}$ still outperform other measures by exhibiting greater statistical significance in forecasting the three outcomes—especially MPsy, which retains predictive strength across all horizons. $\mathrm{BW}_{\mathrm{adj}}$ and BW also display some predictive ability, but only over limited horizons. In contrast, the remaining measures show little to no predictive power across most outcomes.

I also examine the contemporaneous relationships between changes in investor sentiment and these financial outcomes. The results show that sentiment measures with stronger predictive performance also exhibit meaningful contemporaneous relationships, reinforcing prior findings on the alignment between contemporaneous and predictive effects. This alignment supports the hypotheses that (1) investor sentiment negatively predicts future changes in active mutual fund flows, as sentiment-driven inflows reverse when misvaluation corrects; (2) sentiment positively predicts changes in volatility, which rises as prices adjust after a period of over- or under-valuation; and (3) sentiment positively predicts changes in credit spreads, which widen as sentiment-driven equity misvaluation reverts. Overall, these results indicate that sentiment indexes with higher sentiment statistics not only perform well in predicting returns, but also tend to forecast other key financial indicators, highlighting the broader applicability of the joint conditions framework.

The findings in this paper have two main implications for the empirical finance literature. First, employing measures that accurately capture sentiment-induced misvaluation is essential for examining the effects of sentiment across financial assets and markets. Second, the proposed conditions provide a unified framework for evaluating such measures. These implications are especially relevant today, as social media allows investors to express their sentiment instantly and at scale. In addition, rapid advances in AI and LLMs are making it easier to develop new empirical measures of sentiment, which makes assessing their effectiveness increasingly crucial.

1.1 Related Literature

This paper relates to a broad range of behavioral finance literature, encompassing both theoretical and empirical studies that examine how misvaluation arises in asset prices. Theoretical models in this literature primarily investigate which psychological biases lead to distorted beliefs and, consequently, misvaluation. Most of these models adopt some variation of the extrapolation framework (Barberis et al., 1998; Bordalo et al., 2018), in which investors overweight recent performance or fundamentals and form biased beliefs about future outcomes. Barberis (2018) provides a comprehensive review of this class of models. Beyond extrapolation, Daniel et al. (1998, 2001) propose an overconfidence framework in which investors underweight public fundamental signals in favor of their private signals, leading to biased beliefs about a firm's intrinsic value. The simple model used in this paper is designed to represent this entire class of belief-based models, regardless of the underlying psychological biases that generate misvaluation.[3] To this end, I focus on their common component—the resulting misvaluation—and incorporate several assumptions to ensure that the misvaluation generated by my model is consistent with those documented in the literature.

This paper also relates to a class of empirical behavioral finance literature focused on identifying investor sentiment (Baker and Wurgler, 2006; Tetlock, 2007; Karabulut, 2013; Huang et al., 2015; Bybee, 2023). However, there is little consensus on how to evaluate the efficacy of these measures. Existing studies typically assess sentiment based on its ability to predict return reversals, consistent with the idea that misvaluation eventually corrects. I emphasize, however, that sentiment-induced misvaluation implies additional testable relationships beyond return predictability. In particular, changes in investor sentiment should contemporaneously relate to returns, and the volatility of those changes should contemporaneously relate to market volatility. The new measures proposed in this paper are constructed on this basis, ensuring that they reflect sentiment contemporaneously.

The main finding of Baker and Wurgler (2006) is that investor sentiment only predicts stock returns in the cross-section, particularly for small, hard-to-value, and non-dividend-paying firms. In contrast, Huang et al. (2015) propose a modified version of the index, aggregated using partial least squares (PLS) instead of principal component analysis (PCA), and show that it predicts S&P 500 returns in the time series across 1-month and longer horizons. Both the original and PLS versions are included in my analysis, which extends the sample periods beyond those examined in the original studies. Both indexes have been periodically updated by the authors since their initial publication. Using their updated measures, and consistent with Huang et al. (2015), I find that investor sentiment can be a consistent and significant predictor of market returns *through time*; however, the PLS version performs poorly in comparison to other measures, particularly at longer horizons. Notably, the BW index exhibits time-series performance predictability after 1998, supporting the time-series predictability of investor sentiment.[4]

Due to the persistence of sentiment-induced misvaluation, the second part of the Predictabil-

[3]In addition to belief-based explanations, another stream of research employs preference-based frameworks to model misvaluation in financial assets; examples include Kahneman and Tversky (2013); Grinblatt and Han (2005); Barberis and Huang (2008); Barberis et al. (2016). However, this paper focuses on belief-based frameworks because they are more directly aligned with the definition of investor sentiment adopted here: biased investor beliefs.

[4]The predictability of investor sentiment is distinct from the short-term reversal phenomenon, in which aggregate market returns exhibit negative autocorrelation at weekly or monthly frequencies (Lehmann, 1990; Jegadeesh, 1990). Short-term reversal is primarily attributed to temporary liquidity imbalances, market frictions, or trading dynamics, rather than biased investor beliefs. Moreover, the analysis in this paper examines predictability of investor sentiment over horizons often extending beyond one month.

ity condition requires that investor sentiment's forecasting power increases over longer horizons; therefore, I examine multiple horizons in this paper. The new measures proposed in this paper exhibit this pattern, and interestingly, the original BW index does as well in recent years, unlike its PLS version. Greenwood and Shleifer (2014) report a similar long-horizon pattern, showing that consumer confidence surveys are negatively related to future returns over 12- and 36-month horizons. Furthermore, they document that periods of high (low) market returns raise (lower) the expectations of survey respondents regarding future returns. Consistent with their results, I find that survey-based measures often relate to contemporaneous returns.

Aside from the behavioral finance literature, this paper directly relates to the return predictability literature, particularly the seminal study by Welch and Goyal (2008). They demonstrate that established return predictors often fail to forecast returns out-of-sample (OOS), highlighting their instability. The new sentiment measures exhibit comparable OOS R^2 values at short horizons to other measures and superior performance compared to them over longer horizons. Thus, I conduct a bootstrapping test to further assess the consistency of predictive performance, ensuring that OOS results are not driven by a few influential outliers. This analysis resamples multiple 120-observation subsets to test whether forecasting power persists across different samples. I find that MPsy, which exhibits the highest sentiment statistics, is the only measure that consistently predicts returns. This suggests that this approach provides a sharper test of predictor stability.

Lastly, beyond returns, the empirical literature examines how investor sentiment affects a broad set of corporate decisions and asset pricing dynamics.[5] This literature, which frequently relies on surveys or the BW index, primarily explores the contemporaneous effects of sentiment. For example, Stambaugh et al. (2012) use the BW index to identify periods of high and low sentiment and find that anomaly profits are stronger when sentiment is high. Mian and Sankaraguruswamy (2012) document that positive earnings surprises generate stronger immediate price reactions when contemporaneous sentiment (also measured using the BW index) is elevated, and weaker reactions when sentiment is low. These and similar studies highlight the implications of the Contemporaneity condition for empirical research. If a measure does not capture investor sentiment in a timely manner, the results may be biased, particularly over shorter windows. Accordingly, I show that the newly proposed measures contemporaneously relate to, and also forecast, three other financial outcomes known to be impacted by sentiment—active equity fund flows, implied volatility (VIX), and aggregate credit spreads—across both short and long horizons.

[5] See, e.g. Tang and Yan (2010); Stambaugh et al. (2012); Da et al. (2015); López-Salido et al. (2017); Mian and Sankaraguruswamy (2012); Stambaugh and Yuan (2017); Du and Hu (2020); Li et al. (2023).

2 Joint Conditions of Sentiment-Induced Misvaluation

Various theoretical models in the behavioral finance literature demonstrate how asset prices can reflect misvaluation driven by investor beliefs. Notable examples include the overconfidence framework (Daniel et al., 1998; Daniel and Hirshleifer, 2015), as well as models of extrapolative beliefs (Barberis, 2018; Barberis et al., 2018) and diagnostic expectations (Bordalo et al., 2018, 2019). To derive the conditions implied by the potential existence of misvaluation, I employ a simple signaling framework that draws on these prominent models. The model is agnostic about the underlying drivers of investor sentiment and treats sentiment as an exogenous signal. Several assumptions are incorporated to ensure that the misvaluation it generates is consistent with patterns documented in the literature. In this mechanism, initially positive (negative) investor sentiment—representing biased beliefs of a representative investor—pushes asset prices above (below) their intrinsic value, creating misvaluation. Sentiment then persists, allowing misvaluation to deepen until a final fundamental signal arrives, correcting the prior misvaluation as sentiment dissipates.

Consider a single-stock economy, in which the stock pays a terminal liquidating dividend. The stock price is determined by a representative agent's expectation of the terminal dividend.[6] At time t, the agent receives two signals: $s_{s,t}$, a sentiment (irrational) signal, and $s_{f,t}$, a fundamental (rational) signal. The two signals are assumed to be jointly bivariate normal and positively correlated:

$$
\begin{aligned}
& s_{s,t} = \bar{S} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_s^2) \Rightarrow s_{s,t} \sim \mathcal{N}(\bar{S}, \sigma_s^2) \\
& s_{f,t} = \bar{\theta} + u, \quad u \sim \mathcal{N}(0, \sigma_f^2) \Rightarrow s_{f,t} \sim \mathcal{N}(\bar{\theta}, \sigma_f^2) \\
& \Rightarrow \begin{bmatrix} s_{s,t} \\ s_{f,t} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{S} \\ \bar{\theta} \end{bmatrix}, \begin{bmatrix} \sigma_s^2 & \rho\sigma_s\sigma_f \\ \rho\sigma_s\sigma_f & \sigma_f^2 \end{bmatrix} \right), \quad \text{where} \quad \rho > 0
\end{aligned}
$$

The assumption of a positive correlation between sentiment and fundamental signals is motivated by leading behavioral theoretical models, where misvaluation is often driven by such a correlation.[7] This positive correlation implies that, for instance, when s_f is above its average, s_s

[6] Alternatively, one could assume the presence of both rational and sentiment-driven investors, with their relative proportions determining the aggregate sentiment. For simplicity, I follow Barberis et al. (1998) and assume a representative agent.

[7] In most extrapolation-based frameworks (Barberis et al., 2015; Barberis, 2018; Barberis et al., 2018), including experience effects (Malmendier and Nagel, 2011; Malmendier et al., 2020) and diagnostic expectations (Bordalo et al., 2018, 2019), the repeated occurrence of good (bad) returns or fundamental news generates high (low) sentiment and leads to overvaluation (undervaluation), indicating a positive relationship between fundamentals/performance and sentiment. In the overconfidence framework (Daniel et al., 1998; Daniel and Hirshleifer, 2015), overconfident investors overweight their private signals, thereby pushing prices higher (lower) than intrinsic value with good (bad) signals. This positive alignment between sentiment and fundamentals is also evident empir-

is positive (given that sentiment is mean-zero, i.e. $\bar{S} = 0$).[8]

The equilibrium price at time t equals the expectation of the representative agent of the terminal dividend:

$$P_t = E_t(s_{f,t}|s_{s,t} = \bar{S} + \varepsilon) = \bar{\theta} + \rho\frac{\sigma_f\sigma_s}{\sigma_s^2}(s_{s,t} - \bar{S}) = \bar{\theta} + \rho\frac{\sigma_f}{\sigma_s}\hat{S}_t$$

The expression above shows that the initial price equals the expected value of the fundamental signal plus a misvaluation term, given by $\rho\frac{\sigma_f}{\sigma_s}\hat{S}_t$, which represents the misvaluation embedded in the initial price. Since $\rho > 0$, the misvaluation term is positive when $\hat{S}_t > 0$, and negative when $\hat{S}_t < 0$. That is, a positive sentiment signal increases the initial price above its intrinsic value, while a negative sentiment signal pushes it below its intrinsic value.

For simplicity, my model omits the intermediate periods between t and $t + h$, as they are not essential for generating the conditions. However, it is assumed that the misvaluation continues to deepen with successive sentiment signals at $t + 1, t + 2, t + 3, \ldots$. This assumption is motivated by other theoretical settings, which show that misvaluation may persist or even deepen before eventually correcting.[9]

At time $t + h$, the arrival of conclusive information reveals the terminal dividend. Without loss of generality, I assume this dividend equals $\bar{\theta}$, so the terminal price is $P_{t+h} = \bar{\theta}$. Thus, the difference between the initial and terminal prices over horizon h is $\rho\frac{\sigma_f}{\sigma_s}\hat{S}_t$, which represents the misvaluation embedded in the initial price. Figure 1 illustrates the resulting prices over time, as implied by the model.

First, I consider the contemporaneous relationship between changes in prices and investor sentiment over horizon h. The price change is given by $\Delta P = P_{t+h} - P_t = -\rho\frac{\sigma_f}{\sigma_s}\hat{S}_t$. The change in sentiment is $\Delta\hat{S} = \hat{S}_{t+h} - \hat{S}_t = -\hat{S}_t$, because investor sentiment reverts to zero at time $t+h$. Since $\rho > 0$, both ΔP and $\Delta\hat{S}$ share the same sign, implying a positive contemporaneous relationship between changes in price—and thus returns—and changes in investor sentiment.

ically, as periods of strong (weak) economic conditions are typically accompanied by high (low) aggregate investor sentiment.

[8] Modeling sentiment as mean-zero emphasizes its transitory nature and aligns with the intuition that sentiment-induced misvaluations are temporary, thereby predictive of future returns. In contrast, non-mean-zero investor sentiment would imply persistent misvaluations, contradicting the assumption that markets eventually correct such deviations.

[9] For example, in the overconfident investor framework of Daniel et al. (1998), self-attribution bias leads overconfident investors to overweight confirming public signals. As a result, good (bad) news that reinforces their prior beliefs pushes prices higher (lower), further deepening misvaluation before it is eventually corrected. Shleifer and Vishny (1997) argue that because arbitrage is costly and risky, deviations from rational prices driven by sentiment can persist. Barberis et al. (1998) propose a framework in which investors underreact to fundamentals upon their arrival, allowing sentiment to push prices away from intrinsic values and generate momentum until correction occurs. This mechanism is prevalent in almost all extrapolation frameworks, as discussed in detail by Barberis (2018).

Similarly, $\rho > 0$ also implies a positive relationship between the volatility of sentiment changes and the volatility of price changes. As additional sentiment signals arrive between t and $t+h$, shifts in investor sentiment can push the agent's expectation of terminal dividend further away from intrinsic value, leading to additional price changes. Therefore, higher volatility in sentiment changes ($\sigma^2_{\Delta\hat{S}}$) over horizon h results in greater volatility in price changes.[10] Considering the aforementioned contemporaneous relationships, I propose the following:

CONTEMPORANEITY CONDITION: If sentiment-induced misvaluation is present in the market, then:

1. Contemporaneous changes in investor sentiment levels are positively related to changes in prices, and thus returns.

2. The contemporaneous volatility of changes in investor sentiment is positively related to the volatility of price changes, and thus return volatility.

Next, I consider the relationship between investor sentiment and future price changes. Since $\Delta P = -\rho\frac{\sigma_f}{\sigma_s}\hat{S}_t$, a positive (negative) investor sentiment at time t ($\hat{S}_t$) leads to a negative (positive) change in price (ΔP). More specifically, the holding period return over horizon h is given by:

$$r_{t\rightarrow t+h} = \frac{P_{t+h}}{P_t} - 1 = \frac{\bar{\theta}}{\bar{\theta} + \rho\frac{\sigma_f}{\sigma_s}\hat{S}_t} - 1 = \frac{-\rho\frac{\sigma_f}{\sigma_s}\hat{S}_t}{\bar{\theta} + \rho\frac{\sigma_f}{\sigma_s}\hat{S}_t}$$

This expression illustrates the negative return predictability associated with investor sentiment levels, a relationship frequently emphasized in the empirical behavioral finance literature: when $\hat{S}_t$ is positive, the return over the future horizon h is negative, and vice versa.

Importantly, the return predictability of sentiment relies on the correction of misvaluation, prompted in this model by the arrival of conclusive information about the terminal dividend at time $t+h$. However, due to the persistence of sentiment-induced misvaluation—this correction may not occur over shorter horizons. In other words, the likelihood of misvaluation being corrected increases with time (the length of the return horizon). Building on these predictive patterns, I propose the following:

[10] More formally, under sentiment-induced misvaluation, prices—and by extension, returns—can be decomposed into rational and irrational components. The volatility of returns can then be expressed as:

$$\sigma^2 = \sigma_r^2 + \sigma_i^2 + 2\sigma_{r,i}$$

where the subscripts r and i refer to the rational and irrational components, respectively. Since the covariance term is positive, it follows from the expression above that σ^2 is increasing in σ_i, which is the volatility of the irrational component driven by the volatility of changes in investor sentiment $\sigma_i^2 = \sigma^2_{\Delta\hat{S}}$. Thus, the sentiment volatility ($\sigma^2_{\Delta\hat{S}}$) is positively related to return volatility.

PREDICTABILITY CONDITION: If sentiment-induced misvaluation is present in the market, then:

1. Investor sentiment levels are negatively related to (predict) future price changes, and thus returns.
2. The predictive power of investor sentiment increases with the length of the return horizon.[11]

Finally, I examine the consistency of the sentiment-induced misvaluation for positive and negative investor sentiments. As discussed earlier, the investor sentiment can either increase or decrease the price relative to its intrinsic value, inducing positive or negative misvaluations, respectively, as illustrated in Figure 1. Accordingly, the contemporaneous and predictive relationships derived from the model should hold *regardless* of the direction of sentiment level or its changes. Based on this premise, I propose the following:

CONSISTENCY CONDITION: If sentiment-induced misvaluation is present in the market, then:

1. The Contemporaneity condition holds for both positive and negative changes in investor sentiment levels.
2. The Predictability condition holds for both positive and negative investor sentiment levels.

Notably, the model yields symmetrical effects for positive and negative sentiments, which contradicts empirical evidence and the findings in the literature.[12] Thus, rather than imposing symmetry, the Consistency condition simply states that both positive and negative investor sentiments generate misvaluation and, consequently, lead to changes in prices. The magnitude and persistence of these effects, however, may differ across the two sentiment types, as suggested by the literature.

The schematic in Figure 1 visually conveys the core intuitions behind the three joint conditions implied by the sentiment-misvaluation:

1. Contemporaneity: The investor sentiment and prices move together, and larger swings in sentiment generate greater price volatility.

[11] While the first part of the Predictability condition is extensively studied in the empirical literature, the second part is often overlooked. This omission may stem from the fact that asset pricing studies typically use monthly returns, and thus the forecasting power of investor sentiment is typically tested using a 1-month-ahead basis. To address this gap, this paper evaluates the return predictability of investor sentiment over multiple horizons. Given the persistence of investor sentiment, longer-horizon returns may even provide a more appropriate setting for assessing its predictive power.

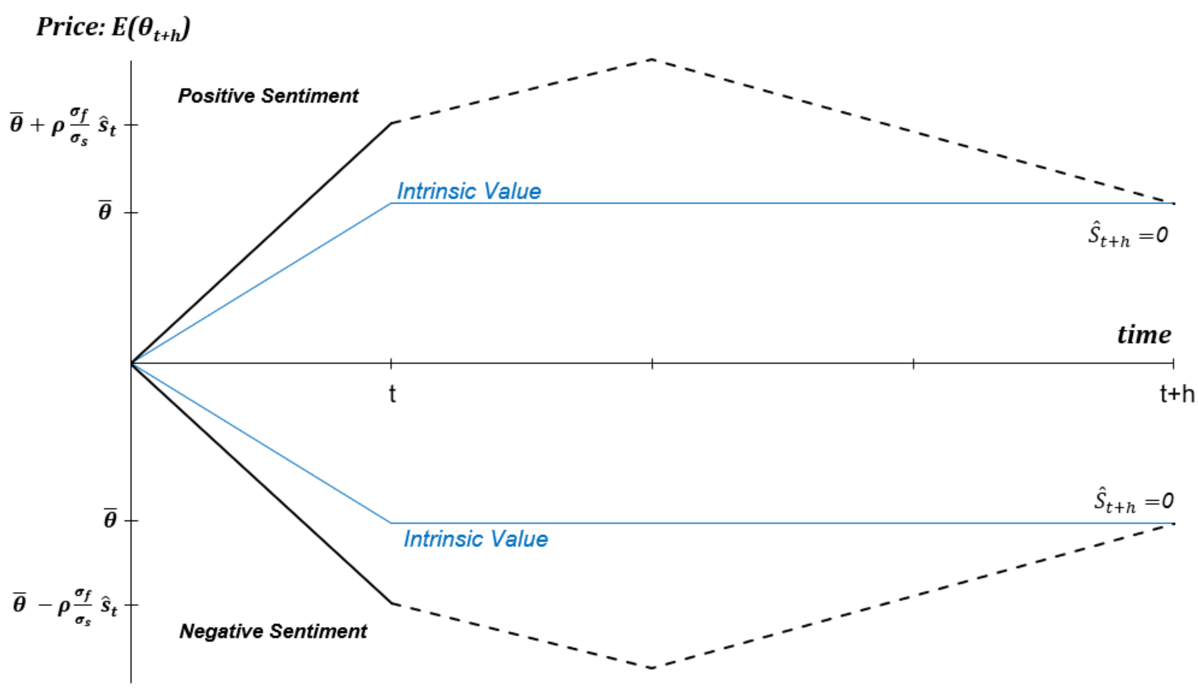
[12] See, e.g. Shleifer and Vishny (1997); Abreu and Brunnermeier (2002); Hong and Stein (2003). The asymmetry is often attributed to different constraints during the boom and busts of economic cycles, which tend to coincide with periods of high and low investor sentiment, respectively. When sentiment is high, equity short-selling constraints limit arbitrage, preventing the correction of overvaluation. Conversely, during periods of low sentiment, borrowing constraints hinder the correction of undervaluation.

2. Predictability: A positive (negative) sentiment level at time t, indicating overvaluation (undervaluation), predicts a negative (positive) return over horizon h.

3. Consistency: Both positive and negative sentiments induce misvaluation, indicating that the Contemporaneity and Predictability conditions hold under overvaluation and undervaluation regimes alike.

Lastly, it is important to note that while the three joint conditions are necessary implications of sentiment-induced misvaluation in my model, I make no claims regarding their sufficiency. The necessity implies that if an empirical measure fails to satisfy these conditions, it cannot be regarded as a reliable proxy for the aggregate investor sentiment and, by extension, for sentiment-induced misvaluation in financial markets.

Figure 1: The expected price over time under sentiment-induced misvaluation



Notes: The figure illustrates the mechanism of sentiment-induced misvaluation implied by the model under both negative and positive sentiment regimes. The misvaluation term, $\rho \frac{\sigma_f}{\sigma_s} \hat{S}_t$, initially shifts the stock price above or below its intrinsic value depending on the sign of investor sentiment at time t ($\hat{S}_t$), and dissipates once sentiment reverts to zero at time $t+h$. For simplicity, the liquidating dividend at time $t+h$ is assumed to be $\bar{\theta}$ in both sentiment regimes. The figure also assumes additional sentiment signals between t and $t+h$ that deepen the misvaluation (dashed lines), however, these signals are omitted from the formal model for tractability.

3 Data

I employ multiple data sources in this paper. The test assets used to evaluate the joint conditions and return predictability include the CRSP value-weighted and S&P 500 index returns, sourced from CRSP and Bloomberg. Other return predictors, as outlined in Welch and Goyal (2008), are obtained from Amit Goyal's website. Monthly excess returns are computed using risk-free rates from Ken French's website, and the cyclically adjusted price-to-earnings (CAPE) ratio is taken from Robert Shiller's website. The Mutual fund flow data, used for additional predictability tests, is drawn from the CRSP mutual fund database, with fund classifications provided by Morningstar Direct. The data on the VIX and credit spreads, also used in additional tests, are obtained from the Federal Reserve Economic Data (FRED) database.

I examine a variety of sentiment measures, including both surveys and empirical indexes commonly cited in the literature. Survey-based measures include the University of Michigan's Consumer Sentiment, AAII's investor sentiment, Yale's stock market confidence, and The Conference Board's consumer confidence and expectations— all sourced from Bloomberg or their respective websites. The original Baker and Wurgler (BW) sentiment index and its components are obtained from Jeffery Wurgler's website. I also include an alternative version of the BW index, which is reconstructed via partial least squares ($\mathrm{BW}_{\mathrm{PLS}}$) by Huang et al. (2015), and the manager sentiment index ($\mathrm{SENT}_{\mathrm{Mng}}$) by Jiang et al. (2019), both from Guofu Zhou's website. In addition to these, I introduce three new investor sentiment measures. Details on their construction follow.

3.1 Constructing New Measures of Investor Sentiment

In this paper, I construct three new empirical measures of investor sentiment. The first uses London Stock Exchange Group (LSEG) MarketPsych analytics, which apply natural language processing (NLP) to news and social media content to generate various sentiment scores. The second is an adjusted version of the BW index ($\mathrm{BW}_{\mathrm{adj}}$), constructed using the raw components of the original index. The third is a principal component analysis (PCA) combining $\mathrm{BW}_{\mathrm{adj}}$ with the Conference Board's Consumer Confidence survey. During the construction of the latter two measures, I mainly focus on the Contemporaneity conditions. My goal is to demonstrate that readily available data can be leveraged to develop sentiment indexes that better align with the joint conditions of sentiment-induced misvaluation (forecast returns *and* explain them contemporaneously) than commonly used alternatives in the literature.

3.1.1 MarketPsych's Investor Sentiment Index (MPsy)

LSEG's MarketPsych Analytics[13] provides granular data with a broad range of tone indicators (sentiment, positive, negative, optimism, pessimism, joy, fear, greed, etc.) measured at both the firm and macro levels from news and social media sources. The data are normalized on a scale from −1 to 1 and reported at a daily frequency. In this paper, I use the positive and negative measures from social sources to aggregate data across both firm and macro levels, and from daily to monthly frequency. I then construct the overall sentiment index as the difference between the two. More specifically, I proceed in the following steps:

1. Constructing monthly measures from firm-level data

   I begin with the daily firm-level data. Stock-day observations with fewer than 20 mentions in a given month are dropped due to sparsity. I then standardize observations across each stock (mean zero, standard deviation one) and compute the firm-level index as the cumulative sum of monthly averages of these standardized stock-day observations:

$$M_{j,k,\tau}^{firm} = \sum_{t=1}^{t=\tau} \left(\frac{\sum_{i=1}^{N} m_{j,k,i}^{firm}}{N} \right)_t$$

   where $M^{\text{firm}} j, k, \tau$ denotes the monthly aggregate index based on firm-level observations from source j (social media), sentiment type k (positive or negative), and month τ; $m^{\text{firm}} j, k, i$ represents the i-th standardized stock-day observation of sentiment type k in month t from source j; and N is the total number of stock-day observations in month t.

2. Constructing monthly measures from macro-level data

   Following a similar procedure to the previous step, I average the standardized daily macro-level observations over each month, and the cumulative sum of these averages forms the macro-level index.

$$M_{j,k,\tau}^{macro} = \sum_{t=1}^{t=\tau} \left(\frac{\sum_{1}^{N} m_{j,k}^{macro}}{N} \right)_t$$

   where $M_{j,k,\tau}^{macro}$ denotes the monthly aggregate index based on macro-level observations from source j (social media), sentiment type k (positive or negative), and month τ; $m_{j,k,i}^{macro}$ represents the i-th standardized daily macro-level observation of type k in month t from source j; and N is the total number of days in month t.

[13] MarketPsych was previously owned by Refinitiv and Thomson Reuters

3. Constructing the final sentiment index

   Two firm- and macro-level sentiment indexes are defined as the difference between their respective positive and negative indexes. The final sentiment index is then calculated as the equal-weighted average of the firm- and macro-level sentiment indexes.

$$S^{firm}_{j,\tau} = M^{firm}_{j,k=positive,\tau} - M^{firm}_{j,k=negative,\tau} \quad \text{and} \quad S^{macro}_{j,\tau} = M^{macro}_{j,k=positive,\tau} - M^{macro}_{j,k=negative,\tau}$$

$$S_\tau = \frac{\sum_{j=1}^{2} S^{macro}_{j,\tau} + \sum_{j=1}^{2} S^{firm}_{j,\tau}}{2} \quad \text{and} \quad \Delta S_\tau = S_\tau - S_{\tau-1}$$

   where S_τ and ΔS_τ represent the level and monthly change of the final investor sentiment index at month τ, respectively.

3.1.2 Adjusted Baker and Wurgler's Sentiment Index ($\mathbf{BW_{adj}}$)

Baker and Wurgler (2006, 2007)'s sentiment index (BW index) is one of the most commonly used measures of investor sentiment in the literature. Because investor sentiment is latent and believed to manifest indirectly through multiple correlated proxies, they employ Principal Component Analysis (PCA) to summarize the common underlying variation. Specifically, the BW index is constructed as the first principal component (PC), capturing the shared component of six individual proxies believed to reflect investor sentiment. These proxies are as follows.

- Trading Volume (*turn*): log of market turnover, calculated as the ratio of trading volume to the number of shares listed on the NYSE, minus a five-year moving average. However, *turn* is later excluded from the index[14], and the index is currently constructed using the remaining five indicators.

- Dividend Premium (*pdnd*): 12-month lagged of the difference between the average market to book ratio of dividend payers and nonpayers firms, following Baker and Wurgler (2004b,a).

- Closed-End Fund Discount (*cefd*): the average discount rate of closed-end funds.

- IPO-Volume (*nipo*): total of number of IPOs over the prior twelve months.

[14]The exclusion is due to the fact that "turnover does not mean what it once did, given the explosion of institutional high-frequency trading and the migration of trading to a variety of venues" (source: Jefferey Wurgler's website).

- IPO First-Day Returns (*ripo*): 12-month lagged of *nipo*-weighted average of monthly measures of IPO's first-day returns over the prior 12 months.
- Equity Issues Over Total New Issues (*s*): total volume of equity issues over the prior 12 months divided by the total volume of equity and debt issues over the prior 12 months, following Baker and Wurgler (2000).

Upon closer examination of the BW index, it becomes apparent that the extremes of the BW index do not align with major market events, commonly believed to be exacerbated by investor sentiment. For instance, the peak of the dot-com bubble—as reflected in the S&P 500 index (SPX)—was in August 2000, whereas the BW index consistently increases through 2000 and peaks in March 2001, as shown in Figure 3. Subsequently, as shown in section 4, I find that changes in the level of the BW index are negatively correlated with contemporaneous market returns. Baker and Wurgler (2007) acknowledge this and argue that, because the the components of the index exhibit different levels of noisiness when moving from levels to changes, the appropriate way to test the contemporaneous relationship with returns is to use the first principal component (PC) of the changes in the components, rather than simply differencing the index itself. [15] However, my analysis (not reported here) indicates that the first PC of changes in the original five components also fails to satisfy the contemporaneous relationship with returns. Moreover, utilizing two different indexes is not ideal. The objective is to have an investor sentiment index that accurately reflects the sentiment-induced misvaluation in the market, and its changes, indicative of sentiment shocks, should similarly reflect changes of misvaluation.

The delayed response of the BW index to major investor sentiment shocks suggests that either its components react sluggishly to sentiment, or that modifications in its construction introduce additional lags. To address this, I propose three adjustments aimed at producing a sentiment index that satisfies the contemporaneous relationship with returns. First, I reduce the lags of *ripo* and *pdnd* from 12 to 6 months.[16] Second, I exclude *nipo* and *s*, as these indicators inherently lag investor sentiment and typically follow favorable market conditions. Third, I extract the first PC from the changes, rather than the levels, of the remaining components: *ripo*, *pdnd*, and *cefd*.[17]

[15] The Baker and Wurgler (2007) do not provide a formal test of their suggestion. The authors present a figure of the "investor sentiment changes index" and argue, based on visual inspection, that sentiment volatility in the new index coincides with speculative episodes in the market.

[16] One reason Baker and Wurgler (2006) use a 12-month lag is to orthogonalize the index with macroeconomic variables, often reported annually. However, they show the orthogonalized version closely resembles the original, making the effort largely redundant.

[17] Some may raise concerns about using only three components to identify investor sentiment. Baker and Wurgler (2006) mention that combining multiple proxies helps isolate sentiment—assumed to influence all compo-

The final index, $\mathrm{BW_{adj}}$, is defined as the cumulative value of the identified first PC and captures the cumulative influence of the common factor—changes in investor sentiment—driving variation across the three underlying components.

Figure 2 compares the original BW index with the adjusted version, $\mathrm{BW_{adj}}$. Both measures are standardized to have a mean of zero and a standard deviation of one. Although the two indexes exhibit similar patterns—with a correlation of 66%—they differ in performance. As demonstrated in Section 4, $\mathrm{BW_{adj}}$ satisfies the Contemporaneity condition better. This can also be visually confirmed in Figure 3, which highlights the run-up and subsequent crash of the dot-com bubble. During this period, the $\mathrm{BW_{adj}}$ index aligns more closely with movements in the S&P 500 index.

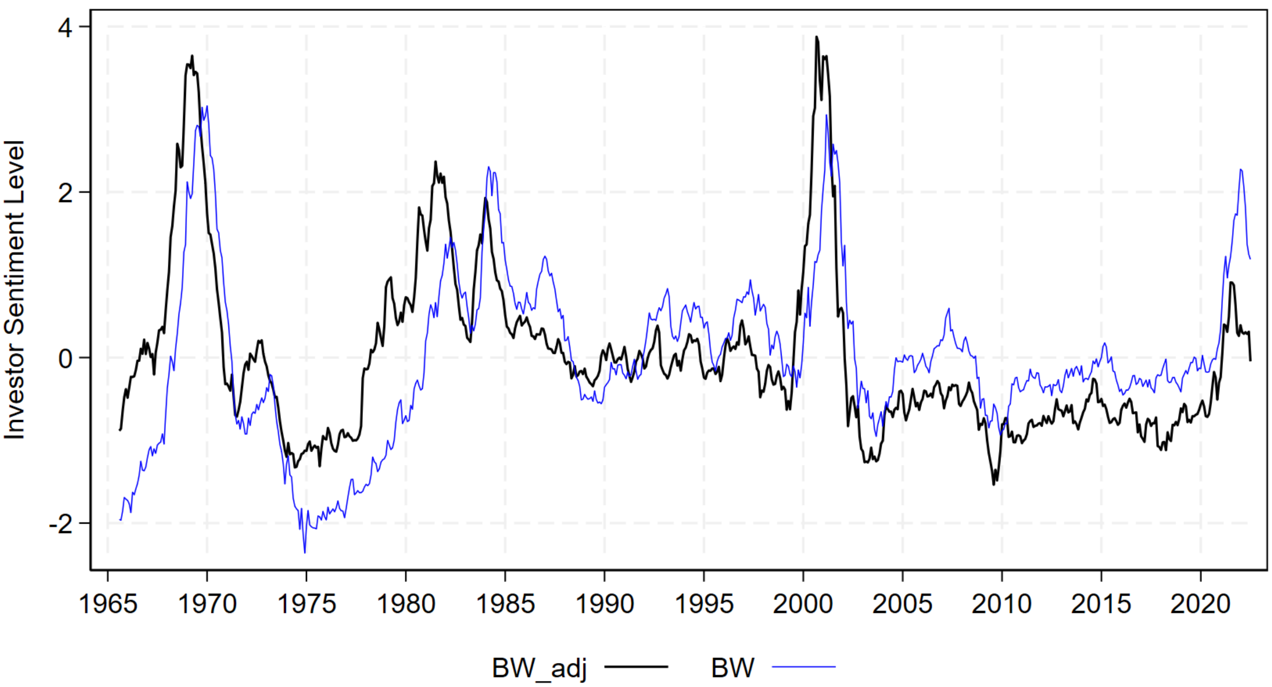
3.1.3 Consumer Confidence and $\mathrm{BW_{adj}}$ Sentiment Index ($\mathrm{CBW_{adj}}$)

Since the $\mathrm{BW_{adj}}$ index is based on only three components, I seek to augment it with an additional, readily available measure. To do so, I combine the $\mathrm{BW_{adj}}$ with the Conference Board's Consumer Confidence survey using principal component analysis (PCA), as before. This survey measures whether consumers feel optimistic or pessimistic about their expected financial situation. I use this index because, unlike the $\mathrm{BW_{adj}}$ index, it performs well in satisfying the Contemporaneity condition over horizons longer than 12 months. The resulting index ($\mathrm{CBW_{adj}}$) has a correlation of 84% with $\mathrm{BW_{adj}}$; however, the inclusion of this survey improves the performance in satisfying the joint conditions further, especially over longer horizons.

Figure 4 compares the three investor sentiment measures proposed in this paper: MPsy, $\mathrm{BW_{adj}}$, and $\mathrm{CBW_{adj}}$.

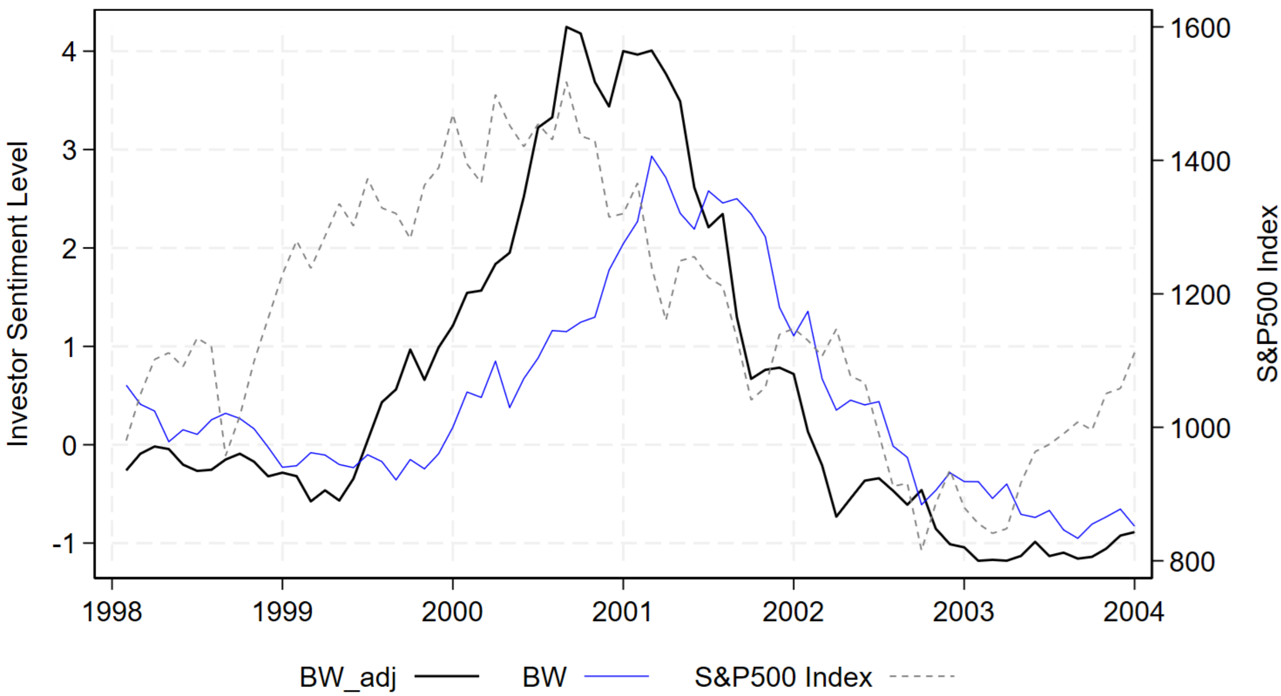
nents—while filtering out unrelated factors. However, one can similarly argue that there could be a trade-off between number of components and accuracy. My goal is to show that modest adjustments to existing data can yield a sentiment measure that meets the joint conditions proposed in this paper without sacrificing return predictability. Future research may explore incorporating additional components to enhance the index further.

Figure 2: The BW and BW_{adj} sentiment indexes



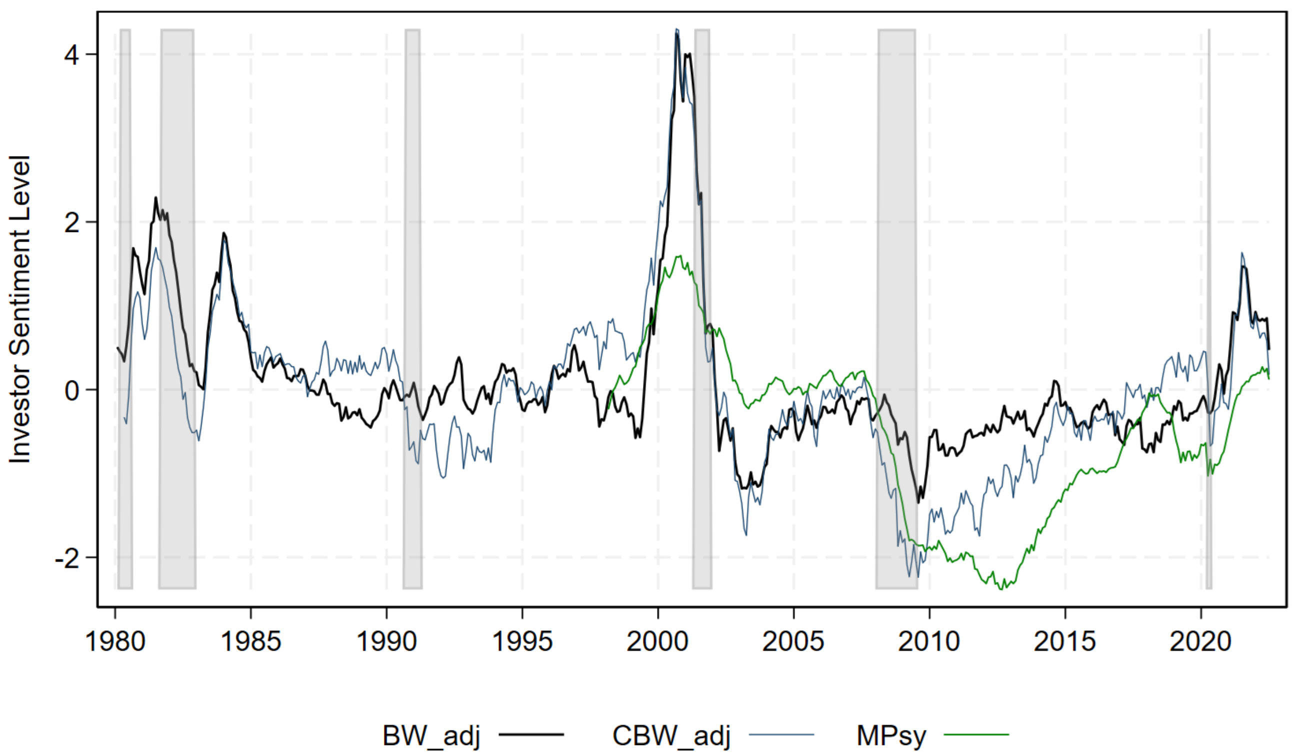
Notes: This figure compares the original Baker and Wurgler (2006) investor sentiment index (BW) with its adjusted version (BW_{adj}), which I propose in this paper, over the period from July 1965 through June 2022.

Figure 3: Investor Sentiment During the Dot-Com Bubble



Notes: This figure compares the original Baker and Wurgler (2006) investor sentiment index with its adjusted version (BW_{adj}) during the run-up and subsequent collapse of the Dot-Com bubble, and shows that BW_{adj} aligns more closely with the S&P 500 index—consistent with the Contemporaneity condition.

Figure 4: Proposed Investor Sentiment Indexes



Notes: This figure compares the three investor sentiment indexes developed in this paper. The BW_{adj} index, shown in black, is an adjusted version of the original Baker and Wurgler (2006) sentiment index. The CBW_{adj} index, shown in navy, is derived as the first PC of changes of the BW_{adj} index and the Conference Board's Consumer Confidence Index. The MPsy index, shown in green, is constructed from LSEG's MarketPsych Analytics data and begins in February 1998.

3.2 Empirical Measures of Investor Sentiment

In this paper, I evaluate the performance of 14 indexes that are commonly used or assumed as proxies for investor sentiment. Of these, eight are survey-based indexes, and three have been introduced in the literature as empirical sentiment measures. I construct the remaining three as described in earlier sections. Table 1 lists the names, labels, and coverage periods of these indexes. As shown in the table, the indexes span different time periods, which pose challenges for direct comparison. To address this issue, all analyses in the paper are conducted over two distinct sample periods. The first spans from either 1965 or 1980 through June 2022, depending on data availability, while the second runs from 1998 through June 2022. An index is evaluated in both periods if data are available before 1998; otherwise, it is only analyzed over the latter period.

The summary statistics of the sentiment indexes and their changes are presented in Table 2. Following standard practice in the literature, I standardize all measures (if not already standardized), ensuring that each has a mean of zero and a standard deviation of one. The correlations between each pair of these measures are shown in Table 3. As indicated in this table, several of these indexes exhibit high correlations with one another. However, a key finding of this paper is that high correlation between sentiment indexes does not necessarily translate into similar performance. For instance, although UMCSENT and CBCONF have a correlation of 92%, the results in Section 4 show that UMCSENT performs substantially worse in satisfying the joint conditions.

Table 1: Empirical Measures of Investor Sentiment

Notes: This table presents the names, descriptions, labels, and coverage periods of the 14 investor sentiment indexes evaluated in this paper. All measures are reported at a monthly frequency.

Index Name and Description	**Acronym**	**Coverage Period**	**Construction Notes**
Panel A: Surveys			
Univ. of Michigan's Consumer Sentiment	UMCSENT	01/1978 – 03/2023	
AAII Sentiment Index	AAII	01/1987 – 03/2023	Constructed as the difference between AAII bullish and bearish indexes
Yale's Market Confidence (institutionals)	$\text{MCONF}_{\text{Inst}}$	06/2001 – 03/2023	
Yale's Market Confidence (individuals)	$\text{MCONF}_{\text{Ind}}$	06/2001 – 03/2023	
Yale's Stock Market Confidence (institutionals)	$\text{SMCONF}_{\text{Inst}}$	06/2001 – 03/2023	
Yale's Stock Market Confidence (individuals)	$\text{SMCONF}_{\text{Ind}}$	06/2001 – 03/2023	
Conference Board's Consumer Expectations	CBEXP	03/1980 – 03/2023	
Conference Board's Consumer Confidence	CBCONF	03/1980 – 03/2023	
Panel B: Proposed in the Literature			
Baker and Wurgler (BW) sentiment index (Baker and Wurgler, 2006)	BW	07/1965 – 06/2022	
Reconstructed BW sentiment index using PLS (Huang et al., 2015)	BW_{PLS}	07/1965 – 12/2023	
Managers' sentiment index (Jiang et al., 2019)	SENT_{Mng}	01/2003 – 12/2017	
Panel C: Presented in this paper			
MarketPsych's aggregate investor sentiment index	MPsy	02/1998 – 12/2022	Constructed from LSEG's MarketPsych Analytics, see section 3.1.1
Adjusted BW sentiment index	BW_{adj}	07/1965 – 06/2022	Constructed from three of the BW's original components, see section 3.1.2
Conference Board's Consumer Confidence and BW_{adj}	CBW_{adj}	04/1980 – 06/2022	Constructed from BW_{adj} and Conf. Board's Cons. Confidence index, see section 3.1.3

Table 2: Summary Statistics

Notes: This table presents summary statistics of the 14 indexes, and their changes (first differences), evaluated in this paper. For better comparison, all measures are standardized (if not already) to have a mean of zero and a standard deviation of one. The first row for each index reports the statistics of the level series, while the second row reports the statistics of its changes.

Index	N	Min	Max	Mean	sd	p(10)	p(50)	p(90)
UMCSENT	543	-2.69	2.04	0.00	1.00	-1.47	0.29	1.10
	542	-4.29	4.31	0.00	1.00	-1.13	-0.01	1.20
AAII	429	-2.71	3.09	0.00	1.00	-1.22	-0.01	1.31
	428	-3.51	3.08	0.00	1.00	-1.33	0.05	1.21
CFO	72	-3.00	1.72	0.00	1.00	-1.23	0.08	1.15
	71	-3.43	2.63	0.00	1.00	-0.98	0.19	1.05
$\mathbf{MCONF_{Inst}}$	262	-2.54	2.03	0.00	1.00	-1.51	0.06	1.23
	261	-2.08	3.24	0.00	1.00	-1.31	-0.04	1.31
$\mathbf{MCONF_{Ind}}$	262	-2.34	2.00	0.00	1.00	-1.42	0.20	1.16
	261	-2.33	2.11	0.00	1.00	-1.33	0.09	1.39
$\mathbf{SMCONF_{Inst}}$	262	-2.22	2.80	0.00	1.00	-1.31	-0.06	1.55
	261	-3.62	4.57	0.00	1.00	-1.13	0.06	1.04
$\mathbf{SMCONF_{Ind}}$	262	-2.53	2.36	0.00	1.00	-1.29	-0.04	1.25
	261	-2.46	2.64	0.00	1.00	-1.26	-0.01	1.34
CBEXP	517	-3.92	2.07	0.00	1.00	-1.26	0.11	1.15
	516	-3.49	5.02	0.00	1.00	-1.10	0.00	1.18
CBCONF	517	-2.78	2.00	0.00	1.00	-1.45	0.12	1.34
	516	-5.42	3.54	0.00	1.00	-1.15	0.01	1.15
BW	684	-2.36	3.04	0.00	1.00	-1.32	-0.10	1.25
	683	-0.72	0.76	0.00	1.00	-0.17	0.01	0.17
$\mathbf{BW_{PLS}}$	702	-1.45	3.76	0.00	1.00	-0.90	-0.28	1.54
	701	-0.61	0.64	0.00	1.00	-0.15	0.00	0.15
$\mathbf{SENT_{Mng}}$	180	-4.15	1.97	0.00	1.00	-1.13	0.14	1.00
	179	-1.64	3.52	0.00	1.00	-0.57	0.02	0.44
$\mathbf{BW_{adj}}$	684	-1.54	3.88	0.00	1.00	-0.98	-0.21	1.38
	683	-6.10	5.32	0.00	1.00	-1.03	0.00	1.04
$\mathbf{CBW_{adj}}$	506	-2.24	4.30	0.00	1.00	-1.22	0.00	1.07
	506	-4.58	3.53	0.00	1.00	-1.08	0.00	1.18
MPsy	299	-1.88	2.11	0.00	1.00	-1.49	0.27	1.23
	298	-5.16	2.56	0.00	1.00	-1.27	0.13	1.19

Table 3: Pairwise Correlations

Notes: This table reports the pairwise correlations among the investor sentiment measures evaluated in this paper.

	UMCSENT	**AAII**	**$\mathbf{MCONF_{Inst}}$**	**$\mathbf{MCONF_{Indiv}}$**	**$\mathbf{SMCONF_{Inst}}$**	**$\mathbf{SMCONF_{Indiv}}$**	**CBEXP**	**CBCONF**	**BW**	**$\mathbf{BW_{PLS}}$**	**$\mathbf{SENT_{Mng}}$**	**$\mathbf{BW_{adj}}$**	**$\mathbf{CBW_{adj}}$**	**MPsy**
UMCSENT	1.00													
AAII	0.36	1.00												
$\mathbf{MCONF_{Inst}}$	0.02	0.04	1.00											
$\mathbf{MCONF_{Indiv}}$	0.00	-0.01	0.82	1.00										
$\mathbf{SMCONF_{Inst}}$	0.43	0.15	-0.07	-0.05	1.00									
$\mathbf{SMCONF_{Indiv}}$	0.59	0.22	-0.01	0.02	0.83	1.00								
CBEXP	0.89	0.41	0.03	0.01	0.42	0.56	1.00							
CBCONF	0.92	0.27	-0.04	-0.08	0.53	0.58	0.87	1.00						
BW	0.35	0.21	0.04	0.00	0.56	0.43	0.30	0.53	1.00					
$\mathbf{BW_{PLS}}$	-0.40	-0.11	0.13	0.11	-0.51	-0.36	-0.42	-0.57	-0.40	1.00				
$\mathbf{SENT_{Mng}}$	-0.12	0.10	0.11	-0.01	0.11	-0.06	-0.09	0.07	0.35	-0.29	1.00			
$\mathbf{BW_{adj}}$	0.26	0.07	0.01	-0.02	0.49	0.43	0.14	0.39	0.68	-0.43	0.25	1.00		
$\mathbf{CBW_{adj}}$	0.85	0.25	-0.03	-0.07	0.59	0.61	0.77	0.96	0.65	-0.61	0.13	0.64	1.00	
MPsy	0.54	0.01	-0.04	-0.08	0.63	0.53	0.40	0.74	0.63	-0.50	0.33	0.50	0.76	1.00

4 Empirical Findings

In this section, I present empirical findings from my paper. I begin by evaluating a broad set of empirical measures of investor sentiment, testing whether each satisfies the joint conditions derived in Section 2. I then proceed to a more detailed analysis of the return predictability performance of these measures across different horizons, and by considering how well they meet the joint conditions. Finally, I extend the analysis beyond returns to explore the predictive power of investor sentiment for other financial outcomes, such as market volatility, mutual fund flows, and aggregate credit spread.

4.1 Evaluating Empirical Measures of Investor Sentiment

In Section 2, I derived a set of necessary conditions implied by the existence of sentiment-induced misvaluation in the market, and subsequently argued that any empirical measure intended to reflect investor sentiment should satisfy these conditions jointly. The remaining task is how to empirically test whether different measures of investor sentiment meet these joint conditions. To that end, I first express these conditions formally through Regressions (1) to (3), and then estimate these regressions jointly using a first-stage Generalized Method of Moments (GMM) approach.[18]

$$r_{t\to t+h} = \alpha + \beta_1 \Delta S^{+}_{t\to t+h} + \beta_2 \Delta S^{-}_{t\to t+h} + \varepsilon_{t\to t+h} \tag{1}$$

$$\sigma_{t\to t+h} = \alpha + \beta_3 \sigma_{\Delta S, t\to t+h} + \varepsilon_{t\to t+h} \tag{2}$$

$$r_{t\to t+h} = \alpha + \beta_4 S^{+}_{t} + \beta_5 S^{-}_{t} + \varepsilon_{t\to t+h} \tag{3}$$

$$\text{subject to necessary conditions:} \quad \beta_1, \beta_2, \beta_3 > 0 \quad \text{and} \quad \beta_4, \beta_5 < 0$$

where $r_{t\to t+h}$ denotes the compounded monthly market (CRSP value-weighted) excess return (total return minus the risk-free rate) over horizon h, and $\sigma_{t\to t+h}$ is the annualized standard deviation of monthly market excess returns over the same horizon. S_t and $\Delta S_{t\to t+h}$ represent the investor sentiment level and its change over horizon h, respectively. $\sigma_{\Delta S, t\to t+h}$ denotes the annualized standard deviation of monthly changes in investor sentiment over horizon h (when $h \geq 12$). Superscripts $+$ and $-$ refer to positive and negative investor sentiment values, respectively. The first regression tests whether positive and negative changes in sentiment levels are positively related to contemporaneous returns over the same horizon, addressing the first parts of the Contemporaneity and Consistency conditions. The second regression examines whether the volatility of sentiment

[18] I use GMM in place of OLS due to the potential correlation of error terms across the regressions. Moreover, since the conditions are intended to hold jointly, estimating them simultaneously using GMM deems more appropriate.

changes is positively related to contemporaneous return volatility, capturing the second part of the Contemporaneity condition. Lastly, the third regression tests whether positive and negative sentiment levels negatively predict returns over the next horizon, corresponding to the second parts of the Predictability and Consistency conditions.

To assess statistical significance, standard errors are estimated using the Newey–West adjustment with 12-month lags. Because the sentiment indexes differ in their coverage periods, I conduct the evaluation over two time frames: July 1965 to June 2022, and February 1998 to June 2022. Indexes with coverage beginning before 1998 are evaluated in both periods; otherwise, they are included only in the latter period.

The GMM is performed over multiple return horizons: 1- and 3-month horizons for Regressions (1) and (3), and 12- and 24-month horizons for Regressions (1), (2), and (3). Following the estimation, I compute two test statistics—hereafter referred to as *sentiment statistics*—for each sentiment measure:

$$\text{joint-score} = \frac{\sum_{h=1}^{H}(\frac{n}{df})_h - \sum_{h=1}^{H}(\frac{m}{df})_h}{H} \tag{4}$$

$$\text{Mean Wald statistic} = \frac{\sum_{h=1}^{H}(\text{Wald statistic})_h}{H} \tag{5}$$

where n and m are the numbers of statistically significant coefficients with the correct and incorrect signs, respectively; df represents the number of GMM restrictions (i.e., estimated coefficients, either 4 or 5 depending on the horizon), and H denotes the number of horizons tested (equal to 4). Thus, the sentiment score (*joint-score*) for any given empirical measure can range between -1 and 1, with a value of 1 indicating full compliance with the joint conditions.

While the mean Wald statistic is conventionally associated with statistical significance under a chi-squared distribution, it should not be interpreted as such in this application, for two main reasons. First, the number of degrees of freedom varies across horizons, making it challenging to assess the significance of the average. Second, the estimated Wald statistics do not account for the direction (sign) of the coefficients. That is, the betas in Equations (1) through (3) may be statistically significant, but their signs may not align with the necessary conditions. Thus, I use the mean Wald statistic for relative comparison across sentiment measures rather than as a formal test of joint significance.[19]

[19]Furthermore, the Wald- (and by extension the F-) statistics can be severely oversized when one or two coefficients have extremely small standard errors (parameters are estimated with high precision). Thus, a statistically

The evaluation results are presented in Table 4 for both sample periods. For conciseness, the detailed regression results are provided in Appendix A.1. Overall, the findings indicate that, based on both joint-scores and mean Wald statistics, the newly proposed measures rank higher than those commonly used in the literature. Across both periods, MPsy and CBW_{adj} emerge as the best-performing sentiment measures, followed by BW_{adj}, CBCONF, and BW. Notably, the BW index exhibits improved performance in recent years, with a higher joint-score in the latter period compared to the earlier one (0.65 vs. 0.25); however, its mean Wald statistic remains lower than that of the three newly proposed indexes. While BW_{adj} scores lower than the other two new measures, it still outperforms the original BW index (when dating back to 1965), BW_{PLS}, and Sent_{Mng}. Both BW_{PLS} and Sent_{Mng} display return predictability but often fail the Contemporaneity and Consistency conditions (see Tables A.1 and A.2). Among the surveys, AAII and the two Conference Board surveys perform relatively better than the others, whereas the four Yale surveys record the lowest sentiment statistics.

The key question arising from these results is how the effects of measures with higher sentiment statistics differ from those with lower statistics. To address this, the following sections delve deeper into the predictability performance of various empirical measures of investor sentiment.

4.2 Return Predictability

The literature focused on identifying investor sentiment often explores how the identified indexes predict future returns, either in the cross-section or the time-series. As discussed previously, the return predictability of investor sentiment is implied by the existence of misvaluation in the markets: based on the intuition that any misvaluation will eventually correct, market returns tend to be in the opposite direction of prior investor sentiment. Since the previous findings reveal that investor sentiment measures commonly used in the literature do not meet the joint conditions suggested by the existence of market misvaluation, I proceed to examine their return predictability by conducting several horse race tests between them and those that satisfy the joint conditions.

Since the prior analysis testing for the joint conditions included in-sample return predictability, it is expected that measures with higher sentiment scores exhibit stronger in-sample forecasting power. To confirm a direct relationship with these scores, I extend the analysis by conducting a series of additional horse race exercises focused on forecasting S&P 500 excess returns—commonly used in the return predictability literature. Specifically, I extend the prediction horizons to include 36-, 48-, and 60-month horizons; report out-of-sample R^2 following Welch and Goyal (2008); apply

significant Wald statistic for multiple coefficients would only indicate that at least one of those coefficients is non-zero, and does not necessarily indicate statistical significance for all tested coefficients.

Table 4: Evaluating Empirical Measures of Investor Sentiment

Notes: The table reports the performance of various investor sentiment indexes in satisfying the joint conditions implied by the existence of misvaluation in the market. joint-scores and mean Wald statistics are calculated using Equations (4) and (5), based on the estimation of Regressions (1) through (3) via first-stage GMM. The measures are listed in descending order of joint-scores. The new sentiment measures proposed in this paper are shown in black.

Index Name and Description	Index Label	joint-score	Mean Wald-stat
Panel A: 07/1965 to 06/2022			
Conf. Board's Cons. Confidence and adjusted BW	$\mathbf{CBW_{adj}}$	**0.76**	**59.01**
Adjusted BW	$\mathbf{BW_{adj}}$	**0.55**	**55.56**
Conference Board's Consumer Confidence	CBCONF	0.49	27.46
AAII Sentiment Index	AAII	0.45	40.08
Conference Board's Consumer Expectations	CBEXP	0.44	35.58
Univ. of Michigan 's Consumer Sentiment	UMCSENT	0.34	27.86
BW_{PLS} sentiment index	BW_{PLS}	0.26	30.72
Baker and Wurgler's sentiment index	BW	0.25	24.17
Panel B: 02/1998 to 06/2022			
MarketPysch's aggregate sentiment index	**MPsy**	**0.84**	**159.54**
Adjusted BW	$\mathbf{BW_{adj}}$	**0.65**	**84.21**
Baker and Wurgler's sentiment index	BW	0.65	61.55
Conf. Board's Cons. Confidence and adjusted BW	$\mathbf{CBW_{adj}}$	**0.59**	**110.25**
Conference Board's Consumer Confidence	CBCONF	0.59	42.64
Conference Board's Consumer Expectations	CBCEXP	0.50	54.38
AAII Sentiment Index	AAII	0.45	33.56
BW_{PLS} sentiment inde	BW_{PLS}	0.43	51.77
Univ. of Michigan 's Consumer Sentiment	UMCSENT	0.40	22.04
Yale's Stock Market Confidence (individuals)	$SMCONF_{Indiv}$	0.33	15.97
Managers' sentiment index	$SENT_{Mng}$	0.29	19.43
Yale's Stock Market Confidence (institutionals)	$SMCONF_{Inst}$	0.16	7.73
Yale's Market Confidence (individuals)	$MCONF_{Indiv}$	0.10	9.21
Yale's Market Confidence (institutionals)	$MCONF_{Inst}$	0.05	8.02

bootstrap resampling to assess the stability of predictions over time; and control for other known predictors of market returns. This analysis covers two consistent sample periods, 1980–2022 and 1998–2022, with the starting year of the first period adjusted to 1980 (instead of 1965) to ensure a more balanced number of observations across different sentiment indexes. I include eight indexes in this exercise: AAII, UMCSENT [20], BW, $\mathrm{BW_{PLS}}$, and $\mathrm{BW_{Mng}}$, along with three new indexes introduced in this paper: $\mathrm{BW_{adj}}$, $\mathrm{CBW_{adj}}$, and MPsy.

4.2.1 In-Sample Tests

I begin with in-sample (IS) regressions, where standard errors are estimated using the Newey–West kernel with 12 lags. The results are presented in Tables 5 and 6. Over the 1980–2022 period, in which six indexes are tested against each other, $\mathrm{CBW_{adj}}$ exhibits the highest return predictability, followed by $\mathrm{BW_{adj}}$, $\mathrm{BW_{PLS}}$, and BW. Over the 1998–2022 window, where eight indexes are evaluated, MPsy ranks highest in return predictability, followed by $\mathrm{CBW_{adj}}$, BW, $\mathrm{BW_{adj}}$, $\mathrm{BW_{PLS}}$, and $\mathrm{SENT_{Mng}}$. Notably, the return predictability of all indexes improves in the latter period, with the BW index in particular exhibiting stronger performance—consistent with its higher joint-score in this period observed earlier. While BW, $\mathrm{BW_{PLS}}$, and $\mathrm{SENT_{Mng}}$ demonstrate meaningful return predictability over short horizons (under 12 months) in the 1998–2022 period, their performance weakens at longer horizons. In contrast, the widely used survey-based measures, AAII and UMCSENT, which perform the worst across both periods, display no meaningful return predictability over sort horizons shorter than 12 months. Overall, MPsy, $\mathrm{CBW_{adj}}$, and $\mathrm{BW_{adj}}$ consistently exhibit strong in-sample return predictability across all horizons, as reflected in their higher coefficient magnitudes, t-statistics, and R^2 values.[21] These results suggest that higher sentiment statistics are associated with stronger and more robust return predictability.

According to the second part of the Predictability condition, the predictive power of investor sentiment is expected to strengthen as the return horizon increases. While this condition is not

[20] The prior analysis indicated weak performance of survey-based measures; however, because AAII and UMCSENT are commonly used in the literature as proxies for investor sentiment, I include both in this analysis to provide better comparison benchmarks for the other measures.

[21] The new measures exhibit relatively high R^2 values, particularly over longer horizons. For example, over the 1998–2022 period and across the 36-month horizon, the R^2 values for MPsy, $\mathrm{CBW_{adj}}$, and $\mathrm{BW_{adj}}$ are 64%, 39%, and 26%, respectively, higher than BW at 21% and the next best-performing measure at just 12%. High in-sample R^2 values for return predictors at long horizons are commonly reported. Fama and French (1988) explore this using the dividend-price ratio and propose a two-part explanation: "(1) High autocorrelation causes the variance of expected returns to grow faster than the return horizon. (2) The growth of the variance of unexpected returns with the return horizon is attenuated by a discount-rate effect." Nevertheless, in this paper, R^2 is used solely as a comparative metric across sentiment measures. Furthermore, a high R^2 does not imply that investor sentiment alone drives return predictability. Instead, it suggests that a properly timed sentiment measure should capture misvaluation *in addition* to changes in valuation arising from shifts in objective risk premia, as sentiment cycles often coincide with economic cycles.

formally tested in Regressions (1)–(3), the results offer supporting evidence. CBW_{adj}, which had the highest sentiment statistics in the first period, and MPsy, which ranked highest in the second period, exhibit this pattern more strongly than the other measures: as the return horizon extends, their coefficient magnitudes and corresponding R^2 values increase consistently. This observation supports the conclusion that higher sentiment statistics from the GMM framework are indicative of measures that more accurately reflect investor sentiment and, as a result, yield stronger market return predictability.

Table 5: In-sample Return Predictability (1980–2022)

Notes: This table presents the results in-sample return predictability test for six investor sentiment indexes. The dependent variable is S&P 500 excess return over the specified horizon. The sample period spans from January 1980 to June 2022. The t-statistics, in parentheses, are estimated based on Newey-West standard error with 12 lags. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	$\mathbf{BW_{PLS}}$	$\mathbf{SENT_{Mng}}$	$\mathbf{BW_{adj}}$	$\mathbf{CBW_{adj}}$	**MPsy**
1	β	-0.002	-0.000	-0.005**	-0.006***		-0.007***	-0.005***	
		(-0.91)	(-0.11)	(-2.24)	(-3.51)		(-4.05)	(-2.59)	
	R^2	0.00	0.00	0.00	0.02		0.01	0.01	
	N	429	519	510	522		510	507	
3	β	-0.005	-0.001	-0.017***	-0.017***		-0.020***	-0.016***	
		(-0.88)	(-0.18)	(-2.63)	(-3.22)		(-4.90)	(-3.05)	
	R^2	0.00	0.00	0.01	0.04		0.04	0.04	
	N	429	519	510	520		510	507	
12	β	-0.022	-0.004	-0.063***	-0.050**		-0.073***	-0.059***	
		(-1.31)	(-0.20)	(-2.82)	(-2.22)		(-6.38)	(-4.63)	
	R^2	0.01	0.00	0.02	0.08		0.11	0.12	
	N	421	511	510	511		510	507	
24	β	-0.038*	-0.013	-0.072	-0.057		-0.112***	-0.093***	
		(-1.70)	(-0.37)	(-1.60)	(-1.60)		(-5.09)	(-4.12)	
	R^2	0.01	0.00	0.00	0.05		0.11	0.13	
	N	409	499	499	499		499	496	
36	β	-0.041	-0.044	-0.059	-0.044		-0.117***	-0.121***	
		(-1.53)	(-0.99)	(-1.15)	(-1.07)		(-4.33)	(-4.74)	
	R^2	0.01	0.00	0.00	0.02		0.06	0.13	
	N	397	487	487	487		487	484	
48	β	-0.053	-0.099*	-0.095*	-0.049		-0.133***	-0.166***	
		(-1.63)	(-1.91)	(-1.90)	(-1.11)		(-4.45)	(-7.15)	
	R^2	0.00	0.02	0.00	0.02		0.05	0.16	
	N	385	475	475	475		475	472	
60	β	-0.120***	-0.179***	-0.096	-0.035		-0.133***	-0.207***	
		(-3.20)	(-3.23)	(-1.32)	(-0.64)		(-3.18)	(-5.94)	
	R^2	0.02	0.06	0.00	0.01		0.03	0.17	
	N	373	463	463	463		463	460	

Table 6: In-sample Return Predictability (1998–2022)

Notes: This table presents the results in-sample return predictability test for eight investor sentiment indexes. The dependent variable is S&P 500 excess return over the specified horizon. The sample period spans from February 1998 to June 2022. The t-statistics, in parentheses, are estimated based on Newey-West standard error with 12 lags. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	$\mathbf{BW_{PLS}}$	$\mathbf{SENT_{Mng}}$	$\mathbf{BW_{adj}}$	$\mathbf{CBW_{adj}}$	**MPsy**
1	β	-0.000	-0.000	-0.008***	-0.007***	-0.008**	-0.006***	-0.004**	-0.007***
		(-0.02)	(-0.13)	(-3.09)	(-4.35)	(-2.10)	(-3.29)	(-2.06)	(-3.62)
	R^2	0.00	0.00	0.02	0.02	0.04	0.02	0.02	0.03
	N	303	303	294	306	180	294	294	299
3	β	-0.003	-0.002	-0.028***	-0.022***	-0.021**	-0.019***	-0.014***	-0.021***
		(-0.50)	(-0.19)	(-4.12)	(-4.96)	(-2.36)	(-4.56)	(-2.63)	(-4.05)
	R^2	0.00	0.00	0.06	0.07	0.08	0.06	0.06	0.09
	N	303	303	294	304	180	294	294	299
12	β	-0.032	-0.022	-0.125***	-0.086***	-0.048**	-0.081***	-0.065***	-0.088***
		(-1.54)	(-0.84)	(-8.10)	(-5.89)	(-2.21)	(-9.00)	(-6.53)	(-5.56)
	R^2	0.03	0.01	0.24	0.23	0.08	0.22	0.22	0.34
	N	295	295	294	295	180	294	294	294
24	β	-0.069**	-0.084**	-0.212***	-0.110***	-0.058*	-0.143***	-0.124***	-0.169***
		(-2.54)	(-2.09)	(-5.02)	(-2.99)	(-1.70)	(-10.47)	(-8.94)	(-6.88)
	R^2	0.05	0.06	0.26	0.19	0.05	0.31	0.36	0.55
	N	283	283	283	283	180	283	283	282
36	β	-0.076**	-0.140***	-0.233***	-0.092**	-0.064*	-0.158***	-0.158***	-0.221***
		(-2.53)	(-2.89)	(-3.36)	(-2.06)	(-1.80)	(-4.50)	(-6.04)	(-7.69)
	R^2	0.04	0.10	0.21	0.12	0.03	0.26	0.39	0.64
	N	271	271	271	271	180	271	271	270
48	β	-0.085**	-0.201***	-0.272***	-0.081	-0.096**	-0.170***	-0.192***	-0.271***
		(-2.37)	(-3.71)	(-3.01)	(-1.63)	(-1.97)	(-3.98)	(-4.92)	(-8.38)
	R^2	0.02	0.14	0.19	0.08	0.04	0.21	0.39	0.67
	N	259	259	259	259	180	259	259	258
60	β	-0.157***	-0.279***	-0.291***	-0.073	-0.041	-0.173***	-0.222***	-0.332***
		(-3.67)	(-5.26)	(-2.65)	(-1.15)	(-0.56)	(-3.69)	(-4.22)	(-8.43)
	R^2	0.05	0.20	0.14	0.06	0.00	0.16	0.38	0.71
	N	247	247	247	247	180	247	247	246

4.2.2 Out-of-Sample Tests

Following the IS analysis, I use out-of-sample (OOS) tests to re-evaluate the return predictability of various investor sentiment indexes—an approach widely adopted in the return predictability literature, most notably by Welch and Goyal (2008), among others. The key advantage of OOS analysis lies in its ability to assess the stability and robustness of return predictors over time. In this procedure, an IS regression is first estimated recursively up to time t, and the resulting coefficient estimates (IS regression slope) are then used to generate one-step-ahead forecasts of market returns for time $t+1$ through the end of the sample. Predictive performance is evaluated using the R^2_{OS} statistic proposed by Campbell and Thompson (2008), which compares the mean squared forecast error (MSE) of the predictive regression to a benchmark of the historical average market return:

$$R^2_{OOS} = 1 - \frac{\sum_{t=k}^{\tau-1}(R_{t+1} - \hat{R}_{t+1})^2}{\sum_{t=k}^{\tau-1}(R_{t+1} - \bar{R}_{t+1})^2} \tag{6}$$

where k denotes the number of periods that are initially used to estimate the first IS regression. The out-of-sample R^2_{OS} lies in the range of $(-\infty, 1]$, with values greater than zero indicating that the predictor outperforms the historical average. For the first sample period (1980 to 2022), I set $k = 240$ months. For the shorter second period (1998 to 2022), I use $k = 60$ months. The results of this analysis are reported in Tables 7 and 8.

Starting with Table 7 and examining the R^2_{OOS} statistics across all horizons, $\mathrm{CBW_{adj}}$ outperforms all other measures, followed by $\mathrm{BW_{adj}}$, $\mathrm{BW_{PLS}}$, and the BW index. While $\mathrm{BW_{PLS}}$ and BW exhibit the highest R^2_{OS} values at the 1-and 3-month horizons, their predictive power declines substantially at longer horizons. For example, at the 12-month horizon, $\mathrm{CBW_{adj}}$ and $\mathrm{BW_{adj}}$ achieve R^2_{OS} values of 25.63% and 24.90%, respectively, compared to only 7.65% for $\mathrm{BW_{PLS}}$ and a mere 0.11% for the BW index.

During the latter period (see Table 8), all sentiment measures exhibit improved performance relative to the earlier period, consistent with the IS results. At shorter horizons (less than 12 months), predictive performance is relatively comparable across measures, with $\mathrm{BW_{adj}}$, BW, and $\mathrm{SENT_{Mng}}$ showing the highest R^2_{OS} values, followed closely by MPsy and $\mathrm{BW_{PLS}}$. However, as the return horizon is extended, the three newly proposed measures, along with the original BW index, demonstrate the strongest predictive power. In contrast, the forecasting ability of $\mathrm{SENT_{Mng}}$ disappears entirely beyond the 12-month horizon.

Consistent with the IS analysis, AAII and UMCSENT are the worst performers across both periods, exhibiting return predictability only at horizons greater than 12 months. Overall, the OOS tests broadly confirm the robustness of the IS results.

Table 7: Out-of-sample Return Predictability (1980–2022)

Notes: This table presents the results out-of-sample return predictability test for six investor sentiment indexes. The dependent variable is S&P 500 excess return over the specified horizon. The sample period spans from January 1980 to June 2022. The average β and N report the average estimated coefficients from the out-of-sample predictions and the total number of predictions, respectively.

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	$\mathbf{BW_{PLS}}$	$\mathbf{SENT_{Mng}}$	$\mathbf{BW_{adj}}$	$\mathbf{CBW_{adj}}$	**MPsy**
	R^2_{OOS}	-0.73%	-0.45%	0.88%	1.60%		-0.02%	0.13%	
1	Avg. β	-0.0016	-0.0003	-0.0052	-0.0060		-0.0067	-0.0059	
	N	177	267	267	267		267	264	
	R^2_{OOS}	-0.27%	-0.35%	3.46%	5.10%		2.91%	3.14%	
3	Avg. β	-0.0029	-0.0028	-0.0146	-0.0169		-0.0207	-0.0186	
	N	175	265	265	265		265	262	
	R^2_{OOS}	2.90%	-1.54%	8.26%	10.92%		15.13%	16.36%	
12	Avg. β	-0.0142	-0.0019	-0.0419	-0.0442		-0.0691	-0.0599	
	N	166	256	256	256		256	253	
	R^2_{OOS}	5.27%	3.71%	0.11%	7.65%		24.90%	25.63%	
24	Avg. β	-0.0224	-0.0036	-0.031	-0.048		-0.101	-0.095	
	N	154	244	244	244		244	241	
	R^2_{OOS}	3.87%	3.35%	-5.78%	3.23%		21.30%	27.08%	
36	Avg. β	-0.0268%	-0.0371%	0.000	-0.031		-0.096	-0.120	
	N	142	232	232	232		232	229	
	R^2_{OOS}	4.36%	14.24%	2.63%	2.88%		20.61%	27.81%	
48	Avg. β	-0.0439	-0.0959	-0.022	-0.034		-0.099	-0.160	
	N	130	220	220	220		220	217	
	R^2_{OOS}	12.38%	26.16%	0.44%	0.57%		14.98%	22.02%	
60	Avg. β	-0.1082	-0.1724	0.004	-0.013		-0.086	-0.191	
	N	118	208	208	208		208	205	

Table 8: Out-of-sample Return Predictability (1998–2022)

Notes: This table presents the results out-of-sample return predictability test for eight investor sentiment indexes. The dependent variable is S&P 500 excess return over the specified horizon. The sample period spans from February 1998 to June 2022. The average β and N report the average estimated coefficients from the out-of-sample predictions and the total number of predictions, respectively.

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	$\mathbf{BW_{PLS}}$	$\mathbf{SENT_{Mng}}$	$\mathbf{BW_{adj}}$	$\mathbf{CBW_{adj}}$	**MPsy**
1	R^2_{OOS}	0.20%	-2.15%	1.25%	1.15%	1.77%	1.62%	0.91%	0.17%
	Avg. β	0.0020	-0.0002	-0.0102	-0.0078	-0.0097	-0.0059	-0.0045	-0.0100
	N	231	231	231	231	119	231	231	230
3	R^2_{OOS}	-0.73%	-2.4%	5.14%	3.26%	4.84%	5.55%	4.14%	3.32%
	Avg. β	0.0004	-0.0004	-0.0298	-0.0229	-0.0259	-0.0193	-0.0149	-0.0306
	N	229	229	229	229	117	229	229	228
12	R^2_{OOS}	2.5%	2.78%	22.54%	3.69%	1.43%	15.43%	13.94%	11.44%
	Avg. β	-0.0244	-0.0292	-0.1317	-0.0840	-0.0584	-0.0796	-0.0645	-0.1309
	N	220	220	220	220	108	220	220	219
24	R^2_{OOS}	5.86%	10.21%	33.77%	8.57%	-5.25%	27.76%	26.41%	36.63%
	Avg. β	-0.434	-0.1125	-0.1819	-0.0954	-0.0736	-0.1382	-0.1270	-0.2329
	N	208	208	208	208	96	208	208	207
36	R^2_{OOS}	3.59%	15.41%	27.68%	7.25%	-5.33%	31.02%	38.88%	57.97%
	Avg. β	-0.0437	-0.1891	-0.1556	-0.0560	-0.0813	-0.1325	-0.1440	-0.2503
	N	196	196	196	196	84	196	196	195
48	R^2_{OOS}	3.52%	27.56%	25.03%	3.54%	-2.72%	27.22%	43.12%	65.24%
	Avg. β	-0.0467	-0.2481	-0.1586	-0.0313	-0.1201	-0.1248	-0.1517	-0.2446
	N	184	184	184	184	72	184	184	183
60	R^2_{OOS}	11.83%	39.41%	19.62%	1.89%	-1.34%	20.24%	38.97%	69.35%
	Avg. β	-0.1022	-0.2869	-0.1323	-0.0078	-0.0499	-0.1136	-0.1473	-0.2606
	N	172	172	172	172	60	172	172	171

4.2.3 Bootstrapping

The OOS tests aimed to assess the stability of sentiment indexes in predicting market returns relative to a benchmark of historical average returns. To evaluate the consistency of predictors, Welch and Goyal (2008) adopt graphical plots that track the OOS performance over time. However, given that the horse race tests compare the return predictability of eight sentiment indexes across multiple horizons and time periods, a graphical approach would be impractical. Instead, to assess the robustness and consistency of these indexes, I employ a bootstrapping technique. This analysis addresses potential concerns that the earlier IS and OOS results may have been driven by a small number of influential or outlier observations.

In this analysis, I draw a sample of 120 monthly observations (with replacement) to re-estimate the return predictability of each sentiment measure. This resampling procedure is repeated 10,000 times across different horizons and for both periods, as in the earlier tests. Since the returns are assessed over different horizons, each sampled monthly observation includes its associated return over the relevant horizon.[22] The results of the bootstrapping analysis are presented in Tables 9 and 10.

Over the 1980–2022 period (Table 9), no sentiment measure demonstrates return predictability at the 1-month horizon. Only BW_{adj} and CBW_{adj} exhibit meaningful predictability at the other horizons, clearly outperforming the other measures. Consistent with earlier findings, CBW_{adj} shows stronger forecasting power than BW_{adj}. Both BW and BW_{PLS} display weak return predictability beyond the 12-month horizon, as reflected in their relatively low R^2 values.

Over the 1998-2022 period (Table 10), MPsy is the only measure exhibiting return predictability across all horizons. SENT_{Mng} yields the highest R^2 values at the 1- and 3-month horizons among all measures; however, these values are lower than those achieved by other indexes at longer horizons. BWCP, BW_{adj}, and BW follow MPsy in terms of overall performance across all horizons. Consistent with earlier results, AAII and UMCSENT show no return predictability over short horizons.

Once again, the findings indicate that MPsy, CBW_{ajd}, and BW_{adj} exhibit more consistent and stable performance than other indexes.

4.2.4 Comparison with Known Predictors of Return

The empirical finance literature identifies several predictors of market returns that are not directly linked to investor sentiment. In this section, I examine whether controlling for these well-

[22] For instance, the observation for March 2005 is linked to its 3-month return (through May 2005), 12-month return (through February 2006), and so forth.

Table 9: Bootstrapping Return Predictability (1980–2022)

Notes: The table presents the beta coefficient and R^2 values from the bootstrapping analysis of six investor sentiment indexes. In each iteration, 120 monthly observations are sampled with replacement to re-estimate the return predictability of each index. This resampling procedure is repeated 10,000 times. The dependent variable is S&P 500 excess return over the specified horizon. The sample period spans from February 1980 to June 2022. The bootstrapping z-statistics are reported in parentheses. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	**BW$_{\mathbf{PLS}}$**	**SENT$_{\mathbf{Mng}}$**	**BW$_{\mathbf{adj}}$**	**CBW$_{\mathbf{adj}}$**	**MPsy**
	β	-0.0017	-0.0001	-0.0057	-0.0061		-0.0061	-0.0052	
1		(-0.39)	(-0.02)	(-0.93)	(-1.38)		(-1.30)	(-1.14)	
	R^2	0.0015	0.0000	0.0086	0.0163		0.0160	0.0145	
	β	-0.0046	-0.0017	-0.0168*	-0.0169**		-0.0185***	-0.0163**	
3		(-0.63)	(-0.18)	(-1.67)	(-2.26)		(-2.64)	(-2.22)	
	R^2	0.0036	0.0004	0.0247	0.0423		0.0495	0.0462	
	β	-0.0215	-0.0038	-0.0633***	-0.0506***		-0.0674***	-0.0591***	
12		(-1.25)	(-0.22)	(-3.35)	(-2.85)		(-5.20)	(-4.83)	
	R^2	0.0183	0.0005	0.0798	0.0853		0.1482	0.1369	
	β	-0.0384	-0.0135	-0.0721**	-0.0567**		-0.1046***	-0.0933***	
24		(-1.61)	(-0.52)	(-2.27)	(-2.35)		(-5.68)	(-5.34)	
	R^2	0.0253	0.0031	0.0459	0.0509		0.1758	0.1686	
	β	-0.0405	-0.0436	-0.0594	-0.0445		-0.1107***	-0.1205***	
36		(-1.40)	(-1.49)	(-1.57)	(-1.56)		(-5.22)	(-6.34)	
	R^2	0.0167	0.0206	0.0196	0.0201		0.1262	0.1774	
	β	-0.0525	-0.0988***	-0.0947***	-0.0487*		-0.1276***	-0.1659***	
48		(-1.48)	(-2.86)	(-2.69)	(-1.67)		(-5.65)	(-9.33)	
	R^2	0.0181	0.0698	0.0333	0.0161		0.1116	0.2252	
	β	-0.1202***	-0.1792***	-0.0957*	-0.0348		-0.1297***	-0.2069***	
60		(-3.03)	(-4.67)	(-1.96)	(-0.99)		(-4.43)	(-8.08)	
	R^2	0.0659	0.1612	0.0241	0.0058		0.0815	0.2503	

Table 10: Bootstrapping Return Predictability (1998–2022)

Notes: The table presents the beta coefficient and R^2 values from the bootstrapping analysis of eight investor sentiment indexes. In each iteration, 120 monthly observations are sampled with replacement to re-estimate the return predictability of each index. This resampling procedure is repeated 10,000 times. The dependent variable is S&P 500 excess return over the specified horizon. The sample period spans from February 1998 to June 2022. The bootstrapping z-statistics are reported in parentheses. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	**BW_{PLS}**	**$SENT_{Mng}$**	**BW_{adj}**	**BWCB**	**MPsy**
	β	0.0008	-0.0001	-0.0090	-0.0076	-0.0076**	-0.0063	-0.0044	-0.0071*
1		(0.18)	(-0.02)	(-1.43)	(-1.53)	(-2.19)	(-1.28)	(-1.11)	(-1.75)
	R^2	0.0003	0.0000	0.0216	0.0229	0.0389	0.0179	0.0138	0.0256
	β	-0.0029	-0.0023	-0.0272***	-0.0222***	-0.0210***	-0.0200***	-0.0143**	-0.0209***
3		(-0.38)	(-0.25)	(-2.75)	(-3.00)	(-3.40)	(-3.07)	(-2.32)	(-3.41)
	R^2	0.0014	0.0009	0.0652	0.0645	0.0871	0.0600	0.0479	0.0738
	β	-0.0320*	-0.0221	-0.1271***	-0.0870***	-0.0481***	-0.0865***	-0.0653***	-0.0882***
12		(-1.77)	(-1.22)	(-8.56)	(-8.26)	(-3.84)	(-7.60)	(-7.40)	(-8.87)
	R^2	0.0371	0.0174	0.3019	0.2114	0.0987	0.2402	0.2118	0.2783
	β	-0.0685***	-0.0840***	-0.2121***	-0.1098***	-0.0581***	-0.1550***	-0.1240***	-0.1687***
24		(-2.87)	(-3.45)	(-6.83)	(-5.31)	(-3.26)	(-11.45)	(-12.12)	(-12.55)
	R^2	0.0746	0.1086	0.3141	0.1448	0.0663	0.3574	0.3539	0.4735
	β	-0.0759***	-0.1401***	-0.2328***	-0.0925***	-0.0642***	-0.1759***	-0.1584***	-0.2212***
36		(-2.78)	(-5.51)	(-4.81)	(-3.70)	(-3.31)	(-5.75)	(-9.82)	(-14.84)
	R^2	0.0613	0.2048	0.2409	0.0690	0.0545	0.3094	0.3863	0.5558
	β	-0.0852**	-0.2013***	-0.2716***	-0.0809***	-0.0957***	-0.1930***	-0.1921***	-0.2714***
48		(-2.55)	(-7.01)	(-4.60)	(-3.08)	(-3.49)	(-5.44)	(-9.02)	(-17.78)
	R^2	0.0511	0.2876	0.2302	0.0370	0.0733	0.2609	0.3961	0.5864
	β	-0.1569***	-0.2789***	-0.2906***	-0.0732**	-0.0412	-0.2013***	-0.2225***	-0.3318***
60		(-4.21)	(-9.36)	(-4.56)	(-2.24)	(-0.96)	(-5.64)	(-8.08)	(-17.88)
	R^2	0.1212	0.3857	0.1886	0.0216	0.0088	0.2030	0.3790	0.6326

established predictors reduces the statistical and economic significance of investor sentiment in forecasting market returns. To that end, I estimate the following regression:

$$r_{t \to t+h} = \alpha + \beta S_t + \delta X_t + \varepsilon_{t \to t+h} \tag{7}$$

where $r_{t \to t+h}$ and S_t represent the h-month-ahead market excess return and the investor sentiment index at month t, respectively, and X_t denotes an additional predictor used as a control variable. The control variables are drawn from Welch and Goyal (2008) and include the dividend-price ratio (dp), book-to-market ratio (bm), stock variance ($svar$), long-term return (ltr), inflation ($infl$), investment-to-capital ratio (ik), and the consumption–wealth–income ratio (cay). In addition to these variables, I also include Shiller's cyclically adjusted price-to-earnings (CAPE) ratio in the analysis. Once more, the regressions are estimated over two periods: January 1980 to June 2020, and January 1998 to June 2022. Statistical significance of the sentiment indexes is evaluated using Newey–West standard errors with 12 lags (4 lags when controlling for ik and cay, which are available quarterly).

To conserve space, allow for the inclusion of quarterly controls, and maintain consistency with Welch and Goyal (2008), I restrict the analysis to horizons longer than 12 months. Results for the 24- and 48-month horizons are omitted, as they are similar to other horizons. The results are reported in Tables 11 and 12.

Over the 1980–2022 period (Table 11), $\mathrm{BW}_{\mathrm{adj}}$ and $\mathrm{CBW}_{\mathrm{adj}}$ exhibit the strongest performance. Among the control variables, only the ik removes the forecasting ability of these two sentiment measures at horizons longer than 12 months. In contrast, the predictive performance of BW and $\mathrm{BW}_{\mathrm{PLS}}$ is frequently diminished at longer horizons when controlling for bm, $svar$, ltr, $infl$, ik, and cay.

Over the 1998–2022 period (Table 12), MPsy once again emerges as the strongest performer among all investor sentiment indexes, maintaining statistical significance across all predictors and horizons. Its superior performance is followed by $\mathrm{CBW}_{\mathrm{adj}}$, $\mathrm{BW}_{\mathrm{adj}}$, and BW. However, the predictive power of these latter measures often diminishes in the presence of ik and bm, especially at longer return horizons; and in the presence of dp and $CAPE$ over the 60-month horizon. Considering both periods, the newly proposed indexes again demonstrate the strongest performance, while AAII and UMCSENT continue to show the weakest.

Next, instead of treating other predictors as controls, I include $CAPE$, ik, and cay in direct IS horse race tests against the investor sentiment indexes. For consistency, this analysis is conducted using quarterly observations for all variables, including the sentiment measures. Table 15 presents the findings.

Over the 1980–2022 period (Panel A), ik exhibits the strongest forecasting power among the three non-sentiment predictors and outperforms both CBW_{adj} and BW_{adj} at horizons beyond 3 months; however, at the 3-month horizon, its coefficient is not statistically significant. In the 1998–2022 period (Panel B), ik again ranks as the top-performing non-sentiment predictor, though it does not surpass MPsy over any horizon, and only outperforms CBW_{adj}, BW_{adj}, and BW at the 48- to 60-month horizons. Overall, the return predictability of investor sentiment measures during the 1998–2022 period is clearly superior to that of non-sentiment predictors. For example, at the 12-month horizon, the R^2 values are as follows: MPsy = 35%, BW = 25%, CBW_{adj} and BW_{adj} = 22%, BW_{PLS} = 20%, $CAPE$ = 4%, ik = 7%, and cay = 3%.

Table 11: Return Predictability While Controlling for Other Predictors (1980-2022)

Notes: This table reports the estimated β coefficients from Regression (7), where the return predictability of investor sentiment is assessed while controlling for other well-known return predictors. The dependent variable is the S&P 500 excess return over the specified horizon. The sample period spans January 1980 to June 2022. The t-statistics, shown in parentheses, are computed using Newey–West standard errors with 12 lags, except for specifications that include *ik* and *cay*, which are available quarterly and use 4 lags. Significance levels are denoted as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Control Predictor	AAII	UMCSENT	BW	$\mathrm{BW_{PLS}}$	$\mathrm{SENT_{Mng}}$	$\mathrm{BW_{adj}}$	$\mathrm{CBW_{adj}}$	MPsy
Panel A: 12-month horizon								
dp	-0.013	0.001	-0.069***	-0.062***		-0.081***	-0.060***	
	(-0.78)	(0.04)	(-3.48)	(-3.80)		(-6.89)	(-4.87)	
bm	-0.012	-0.001	-0.066***	-0.063***		-0.081***	-0.059***	
	(-0.70)	(-0.05)	(-3.22)	(-3.92)		(-7.43)	(-4.77)	
svar	-0.019	-0.001	-0.062***	-0.051**		-0.073***	-0.058***	
	(-1.18)	(-0.05)	(-2.73)	(-2.23)		(-6.30)	(-4.39)	
ltr	-0.021	-0.004	-0.064***	-0.051**		-0.073***	-0.059***	
	(-1.26)	(-0.20)	(-2.88)	(-2.29)		(-6.48)	(-4.59)	
infl	-0.028*	-0.009	-0.059**	-0.046**		-0.068***	-0.054***	
	(-1.71)	(-0.48)	(-2.55)	(-2.00)		(-5.97)	(-3.99)	
CAPE	-0.013	0.007	-0.063***	-0.054***		-0.073***	-0.058***	
	(-0.83)	(0.30)	(-3.27)	(-2.85)		(-6.04)	(-3.95)	
ik	-0.008	0.029	-0.043*	-0.031		-0.058***	-0.050*	
	(-0.65)	(1.20)	(-1.69)	(-1.30)		(-3.06)	(-1.91)	
cay	-0.015	0.005	-0.058**	-0.043*		-0.067***	-0.053***	
	(-1.20)	(0.23)	(-2.27)	(-1.84)		(-4.99)	(-4.33)	
Panel B: 36-month horizon								
dp	-0.012	-0.039	-0.074	-0.064*		-0.135***	-0.123***	
	(-0.45)	(-0.90)	(-1.47)	(-1.67)		(-5.24)	(-5.19)	
bm	-0.020	-0.056	-0.062	-0.053		-0.127***	-0.121***	
	(-0.88)	(-1.27)	(-1.14)	(-1.17)		(-4.85)	(-4.81)	
svar	-0.040	-0.043	-0.059	-0.045		-0.117***	-0.121***	
	(-1.50)	(-0.96)	(-1.14)	(-1.08)		(-4.30)	(-4.73)	
ltr	-0.039	-0.043	-0.061	-0.045		-0.117***	-0.121***	
	(-1.42)	(-0.98)	(-1.18)	(-1.10)		(-4.33)	(-4.74)	
infl	-0.047*	-0.049	-0.057	-0.040		-0.112***	-0.117***	
	(-1.72)	(-1.18)	(-1.10)	(-0.97)		(-4.05)	(-4.51)	
CAPE	-0.004	-0.010	-0.071*	-0.060*		-0.119***	-0.111***	
	(-0.16)	(-0.24)	(-1.65)	(-1.77)		(-4.95)	(-4.87)	
ik	0.003	0.031	0.056	0.034		-0.008	-0.010	
	(0.14)	(1.05)	(1.21)	(0.99)		(-0.23)	(-0.28)	
cay	-0.019	-0.044	-0.053	-0.033		-0.102***	-0.114***	
	(-0.85)	(-0.92)	(-1.04)	(-0.80)		(-3.48)	(-4.36)	
Panel C: 60-month horizon								
dp	-0.051	-0.171***	-0.133*	-0.079**		-0.169***	-0.215***	
	(-1.48)	(-3.26)	(-1.94)	(-2.00)		(-3.85)	(-5.12)	
bm	-0.056*	-0.199***	-0.113	-0.064		-0.159***	-0.211***	
	(-1.76)	(-3.90)	(-1.59)	(-1.41)		(-4.15)	(-5.60)	
svar	-0.119***	-0.181***	-0.094	-0.036		-0.132***	-0.207***	
	(-3.25)	(-3.22)	(-1.31)	(-0.67)		(-3.17)	(-5.89)	
ltr	-0.121***	-0.179***	-0.098	-0.036		-0.133***	-0.207***	
	(-3.06)	(-3.20)	(-1.36)	(-0.67)		(-3.19)	(-5.93)	
infl	-0.135***	-0.189***	-0.093	-0.029		-0.124***	-0.203***	
	(-3.64)	(-3.90)	(-1.29)	(-0.53)		(-2.96)	(-5.82)	
CAPE	-0.041	-0.143***	-0.114*	-0.062*		-0.134***	-0.185***	
	(-1.21)	(-2.63)	(-1.84)	(-1.65)		(-3.07)	(-4.68)	
ik	-0.066*	-0.073	0.110	0.097**		0.087*	-0.004	
	(-1.75)	(-1.38)	(1.57)	(2.26)		(1.83)	(-0.07)	
cay	-0.108***	-0.184***	-0.079	-0.032		-0.114**	-0.191***	
	(-2.71)	(-3.04)	(-1.00)	(-0.61)		(-2.57)	(-4.68)	

Table 12: Return Predictability While Controlling for Other Predictors (1998-2022)

Notes: This table reports the estimated β coefficients from Regression (7), where the return predictability of investor sentiment is assessed while controlling for other well-known return predictors. The dependent variable is the S&P 500 excess return over the specified horizon. The sample period spans January to June 2022. The t-statistics, shown in parentheses, are computed using Newey–West standard errors with 12 lags, except for specifications that include *ik* and *cay*, which are available quarterly and use 4 lags. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Control Predictor	AAII	UMCSENT	BW	BW_{PLS}	$SENT_{Mng}$	BW_{adj}	CBW_{adj}	MPsy
Panel A: 12-month horizon								
dp	-0.002	0.032	-0.097***	-0.061***	-0.054**	-0.052***	-0.035***	-0.064**
	(-0.10)	(1.18)	(-5.10)	(-4.77)	(-2.41)	(-4.45)	(-2.69)	(-2.54)
bm	-0.010	0.037	-0.091***	-0.051**	-0.050***	-0.045**	-0.024	-0.057**
	(-0.46)	(1.26)	(-4.89)	(-2.31)	(-2.87)	(-2.51)	(-0.74)	(-1.99)
svar	-0.027	-0.016	-0.126***	-0.092***	-0.051**	-0.081***	-0.063***	-0.087***
	(-1.30)	(-0.55)	(-8.38)	(-6.66)	(-2.26)	(-9.40)	(-5.91)	(-5.50)
ltr	-0.032	-0.023	-0.127***	-0.087***	-0.048**	-0.081***	-0.065***	-0.088***
	(-1.46)	(-0.83)	(-8.03)	(-6.02)	(-2.20)	(-8.96)	(-6.49)	(-5.48)
infl	-0.039*	-0.023	-0.123***	-0.085***	-0.047**	-0.077***	-0.062***	-0.085***
	(-1.92)	(-0.88)	(-7.81)	(-5.96)	(-2.23)	(-8.76)	(-5.65)	(-5.35)
CAPE	-0.019	0.035	-0.116***	-0.071***	-0.044**	-0.069***	-0.078***	-0.093***
	(-0.97)	(1.14)	(-5.78)	(-4.58)	(-2.05)	(-5.31)	(-5.31)	(-2.76)
ik	-0.019	0.042	-0.096***	-0.055**	-0.005	-0.049**	-0.034	-0.069***
	(-1.32)	(1.29)	(-5.49)	(-2.37)	(-0.30)	(-1.97)	(-1.09)	(-2.76)
cay	-0.027*	-0.023	-0.137***	-0.081***	-0.028	-0.082***	-0.067***	-0.088***
	(-1.66)	(-0.81)	(-6.38)	(-5.35)	(-1.45)	(-5.45)	(-7.28)	(-5.57)
Panel B: 36-month horizon								
dp	0.018	-0.003	-0.097	-0.015	-0.077**	-0.070**	-0.082**	-0.169***
	(0.97)	(-0.05)	(-1.52)	(-0.35)	(-2.39)	(-2.16)	(-2.46)	(-4.24)
bm	-0.013	-0.011	-0.075	0.022	-0.067***	-0.063*	-0.086**	-0.214***
	(-0.53)	(-0.24)	(-0.93)	(0.46)	(-2.88)	(-1.95)	(-2.29)	(-2.91)
svar	-0.070**	-0.138***	-0.232***	-0.100**	-0.071*	-0.158***	-0.156***	-0.220***
	(-2.24)	(-2.70)	(-3.41)	(-2.31)	(-1.88)	(-4.59)	(-6.04)	(-7.87)
ltr	-0.074**	-0.139***	-0.236***	-0.094**	-0.064*	-0.159***	-0.159***	-0.221***
	(-2.31)	(-2.88)	(-3.34)	(-2.10)	(-1.78)	(-4.48)	(-6.11)	(-7.80)
infl	-0.083***	-0.138***	-0.230***	-0.095**	-0.063*	-0.156***	-0.157***	-0.219***
	(-2.87)	(-2.85)	(-3.40)	(-2.20)	(-1.80)	(-4.47)	(-5.97)	(-7.66)
CAPE	-0.019	0.075	-0.120**	-0.023	-0.054*	-0.064**	-0.084**	-0.186***
	(-0.85)	(1.22)	(-2.09)	(-0.59)	(-1.76)	(-2.13)	(-2.07)	(-3.36)
ik	-0.023	0.011	0.025	0.050	-0.001	0.039	0.034	-0.092**
	(-1.56)	(0.26)	(0.94)	(1.37)	(-0.02)	(1.36)	(0.94)	(-2.14)
cay	-0.033	-0.184***	-0.185**	-0.048	-0.051	-0.127***	-0.150***	-0.191***
	(-1.22)	(-4.93)	(-2.52)	(-1.13)	(-1.60)	(-3.29)	(-4.00)	(-6.00)
Panel C: 60-month horizon								
dp	0.004	-0.050	-0.019	0.069	-0.073	0.013	-0.033	-0.199***
	(0.20)	(-0.65)	(-0.25)	(1.49)	(-1.47)	(0.33)	(-0.59)	(-4.92)
bm	-0.039*	-0.095	0.049	0.158***	-0.047	0.037	-0.031	-0.260***
	(-1.92)	(-1.28)	(0.58)	(3.01)	(-1.32)	(1.19)	(-0.57)	(-2.98)
svar	-0.145***	-0.279***	-0.290***	-0.092	-0.055	-0.174***	-0.217***	-0.330***
	(-3.51)	(-5.12)	(-2.78)	(-1.64)	(-0.71)	(-3.94)	(-4.36)	(-9.31)
ltr	-0.162***	-0.279***	-0.293***	-0.074	-0.041	-0.174***	-0.223***	-0.332***
	(-3.60)	(-5.25)	(-2.65)	(-1.17)	(-0.55)	(-3.69)	(-4.24)	(-8.43)
infl	-0.174***	-0.275***	-0.286***	-0.078	-0.039	-0.170***	-0.219***	-0.328***
	(-4.38)	(-5.20)	(-2.68)	(-1.30)	(-0.53)	(-3.65)	(-4.17)	(-8.41)
CAPE	-0.060*	-0.067	-0.092	0.051	-0.024	0.014	-0.044	-0.273***
	(-1.76)	(-0.55)	(-1.24)	(0.86)	(-0.38)	(0.34)	(-0.85)	(-4.19)
ik	-0.107***	-0.123	0.077	0.118**	0.023	0.121***	0.043	-0.231***
	(-2.64)	(-1.48)	(1.07)	(2.18)	(0.29)	(2.61)	(0.60)	(-3.10)
cay	-0.134***	-0.321***	-0.218*	-0.027	-0.044	-0.131**	-0.196***	-0.295***
	(-2.85)	(-7.18)	(-1.86)	(-0.46)	(-0.61)	(-2.54)	(-3.01)	(-5.81)

Table 13: Return Predictability Against Other Predictors

Notes: This table reports the IS return predictability of investor sentiment indexes along with well-known predictors of return: $CAPE$, ik, and cay. All conversations are quarterly. The dependent variable is the S&P 500 excess return over the specified horizon. The t-statistics, shown in parentheses, are computed using Newey–West standard errors with 4 lags. Significance levels are denoted as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	**BW_{PLS}**	**$SENT_{Mng}$**	**BW_{adj}**	**CBW_{adj}**	**MPsy**	**CAPE**	**ik**	**cay**
Panel A: 01/1980 – 06/2022												
3	β	-0.002	0.000	-0.016**	-0.014**		-0.020***	-0.016***		-0.001	-3.450	-0.092
		(-0.40)	(0.03)	(-2.45)	(-2.41)		(-5.43)	(-3.41)		(-0.69)	(-1.63)	(-0.91)
	R^2	0.00	0.00	0.01	0.03		0.04	0.04		0.03	0.03	0.00
	N	138	168	168	168		168	167		168	168	168
12	β	-0.016	0.004	-0.058**	-0.042*		-0.065***	-0.052***		-0.002	-12.794*	-0.181
		(-1.21)	(0.18)	(-2.29)	(-1.75)		(-4.88)	(-4.27)		(-0.91)	(-1.85)	(-0.40)
	R^2	0.01	0.00	0.01	0.06		0.09	0.10		0.08	0.11	0.00
	N	138	168	168	168		168	167		168	168	168
24	β	-0.030*	-0.009	-0.063	-0.044		-0.098***	-0.086***		-0.005	-27.527**	0.032
		(-1.92)	(-0.24)	(-1.32)	(-1.19)		(-3.98)	(-3.64)		(-1.04)	(-2.57)	(0.06)
	R^2	0.01	0.00	0.00	0.04		0.08	0.11		0.16	0.21	0.02
	N	135	165	165	165		165	164		165	165	165
36	β	-0.022	-0.045	-0.056	-0.030		-0.102***	-0.114***		-0.008	-49.258***	-0.922
		(-0.96)	(-0.96)	(-1.09)	(-0.72)		(-3.44)	(-4.38)		(-1.36)	(-4.88)	(-0.46)
	R^2	0.00	0.00	0.00	0.01		0.05	0.12		0.22	0.29	0.09
	N	131	161	161	161		161	160		161	161	161
48	β	-0.037	-0.113**	-0.089*	-0.033		-0.117***	-0.163***		-0.011*	-71.786***	-1.260
		(-1.26)	(-2.05)	(-1.68)	(-0.72)		(-3.60)	(-6.16)		(-1.89)	(-8.78)	(-0.46)
	R^2	0.00	0.02	0.00	0.01		0.03	0.15		0.24	0.32	0.13
	N	127	157	157	157		157	156		157	157	157
60	β	-0.108***	-0.187***	-0.085	-0.018		-0.109**	-0.192***		-0.016***	-85.257***	-3.037
		(-2.67)	(-3.14)	(-1.05)	(-0.32)		(-2.25)	(-4.69)		(-2.62)	(-9.22)	(-0.91)
	R^2	0.01	0.06	0.01	0.00		0.01	0.14		0.25	0.35	0.14
	N	123	153	153	153		153	152		153	153	153
Panel B: 02/1998 – 06/2022												
3	β	-0.004	-0.002	-0.028***	-0.020***	-0.015*	-0.020***	-0.015***	-0.021***	-0.002*	-5.883**	-0.198**
		(-0.50)	(-0.17)	(-5.20)	(-4.57)	(-1.65)	(-6.45)	(-3.59)	(-3.84)	(-1.71)	(-2.25)	(-2.30)
	R^2	0.00	0.00	0.06	0.07	0.03	0.07	0.06	0.10	0.01	0.02	0.02
	N	96	96	96	96	60	96	96	96	96	96	96
12	β	-0.030*	-0.024	-0.127***	-0.078***	-0.034	-0.079***	-0.063***	-0.090***	-0.010***	-23.180***	-0.549
		(-1.78)	(-0.87)	(-8.08)	(-4.87)	(-1.41)	(-7.56)	(-7.25)	(-5.41)	(-3.30)	(-3.01)	(-1.11)
	R^2	0.03	0.01	0.25	0.20	0.03	0.22	0.22	0.35	0.04	0.07	0.03
	N	96	96	96	96	60	96	96	96	96	96	96
24	β	-0.054***	-0.086*	-0.203***	-0.094**	-0.050	-0.134***	-0.120***	-0.166***	-0.021***	-45.990***	-0.738
		(-2.80)	(-1.89)	(-4.04)	(-2.31)	(-1.24)	(-7.81)	(-8.06)	(-6.34)	(-5.17)	(-4.91)	(-0.92)
	R^2	0.03	0.05	0.24	0.15	0.02	0.27	0.34	0.53	0.08	0.12	0.02
	N	93	93	93	93	60	93	93	93	93	93	93
36	β	-0.053**	-0.142***	-0.217***	-0.073	-0.073	-0.143***	-0.149***	-0.216***	-0.029***	-70.369***	-7.597***
		(-2.13)	(-2.63)	(-3.06)	(-1.54)	(-1.55)	(-4.08)	(-5.07)	(-6.77)	(-6.83)	(-9.33)	(-3.70)
	R^2	0.02	0.09	0.18	0.09	0.02	0.22	0.36	0.62	0.11	0.16	0.04
	N	89	89	89	89	60	89	89	89	89	89	89
48	β	-0.068**	-0.219***	-0.253***	-0.060	-0.128**	-0.153***	-0.183***	-0.271***	-0.038***	-88.177***	-10.597***
		(-2.20)	(-3.90)	(-2.72)	(-1.10)	(-1.97)	(-3.63)	(-4.15)	(-7.40)	(-7.91)	(-10.71)	(-3.82)
	R^2	0.01	0.15	0.16	0.06	0.04	0.17	0.36	0.66	0.14	0.20	0.01
	N	85	85	85	85	60	85	85	85	85	85	85
60	β	-0.142***	-0.292***	-0.262**	-0.053	-0.083	-0.154***	-0.207***	-0.330***	-0.046***	-93.834***	-13.570***
		(-3.10)	(-5.31)	(-2.32)	(-0.73)	(-0.87)	(-3.21)	(-3.67)	(-7.43)	(-6.96)	(-9.25)	(-3.64)
	R^2	0.04	0.20	0.11	0.04	0.01	0.12	0.33	0.69	0.17	0.25	0.00
	N	81	81	81	81	60	81	81	81	81	81	81

4.2.5 Further Analysis and Discussion

Considering the full set of results, the best overall return predictability among the new investor sentiment indexes is achieved by $\mathrm{CBW}_{\mathrm{adj}}$ over the 1980–2022 period and by MPsy over the 1998–2022 period, consistent with their high sentiment statistics reported in Table 4.

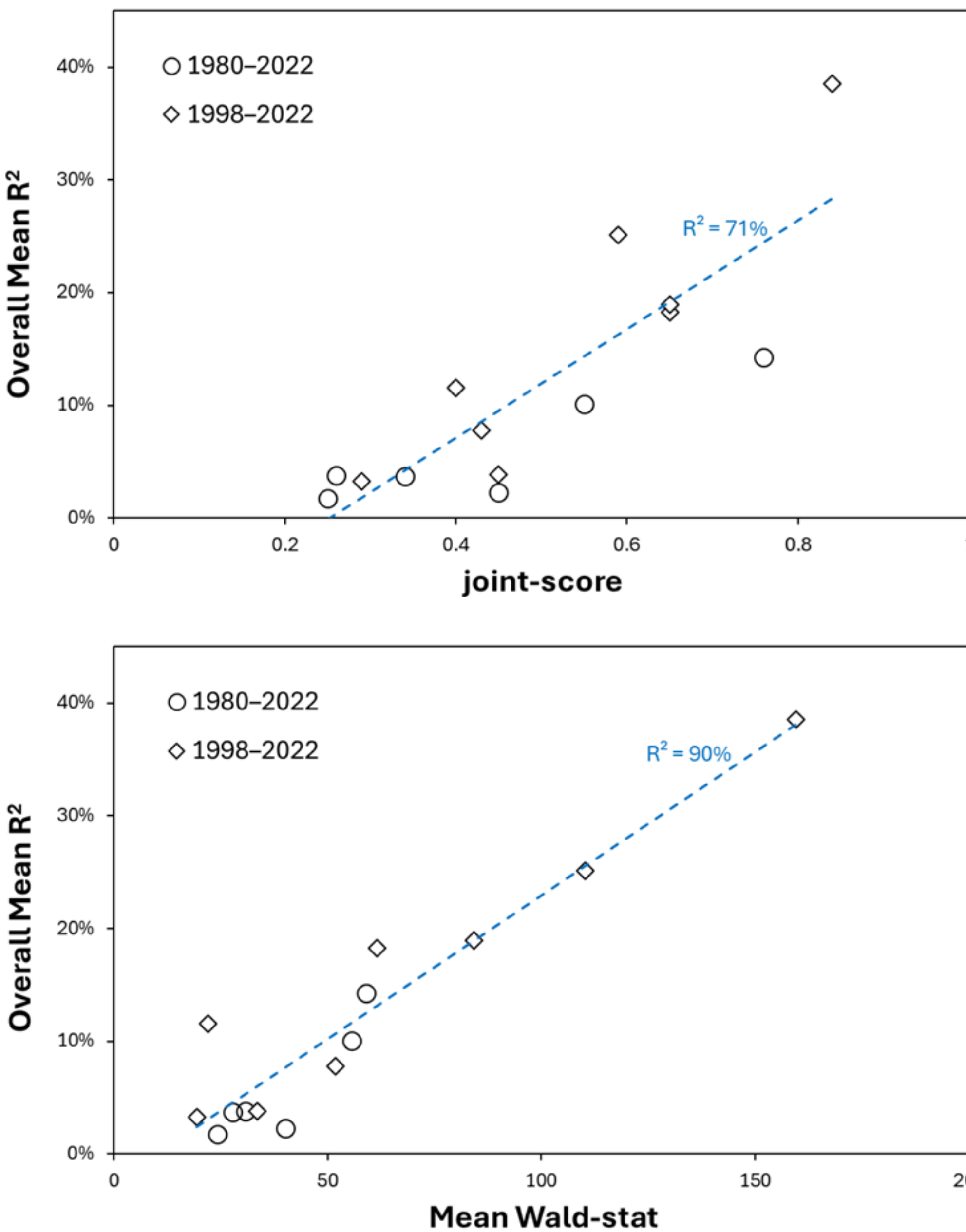
To synthesize prior evidence and formally link return predictability to sentiment statistics, Table 14 reports the mean R^2 values (averaged across different horizons) from earlier analyses, alongside the joint-scores and mean Wald statistics from Table 4. The results show that R^2 values from different analyses generally increase with both sentiment statistics. This positive relationship is even more apparent Figure 5, which plots the overall mean R^2 values against the joint-scores and mean Wald statistics. As both sentiment statistics increase across measures, the overall mean R^2 values rise accordingly. Consistent with the sentiment-induced misvaluation mechanism discussed earlier, this pattern suggests that the forecasting power of sentiment measures strengthens when predictability is accompanied by the Contemporaneity and Consistency conditions, as reflected in higher sentiment statistics.

Among the two sentiment statistics, the mean Wald statistic exhibits a stronger correlation with the overall mean R^2. However, two important caveats must be noted. First, although the Wald statistic appears to be a better overall indicator of alignment with the joint conditions, as illustrated in Figure 5, the relationship flattens when its value falls below 50, suggesting limited explanatory power in that range. Second, as discussed earlier (section 4.1) while the Wald statistic reflects joint statistical significance under a chi-squared distribution, its mean value in this context cannot be interpreted as a proper test statistic because the degrees of freedom vary across horizons. Moreover, the statistical significance alone does not ensure full alignment with the joint conditions as some sentiment measures exhibit strong contemporaneous relationships that inflate the joint Wald statistic, even in the absence of predictive power (see Tables A.1 and A.2). Thus, both sentiment statistics should be considered when evaluating the extent to which a measure satisfies the proposed conditions.

A notable observation is the relatively strong short-horizon (1–3 month) return predictability of $\mathrm{BW}_{\mathrm{PLS}}$ and $\mathrm{SENT}_{\mathrm{Mng}}$. If these indexes do not fully satisfy the joint conditions—implying they are not proper proxies for investor sentiment—why do they exhibit such predictive power in the short term? Relatedly, if the newly proposed indexes better satisfy the joint conditions while others do not, why don't they exhibit the strongest short-term return predictability?

One possible explanation is that some of the observed predictability reflects a Type I error driven by influential outliers. A more compelling explanation, however, lies in the construction

Figure 5: Overall Mean R^2 vs. Sentiment Statistics



Notes: This figure plots the overall mean R^2—calculated as the average of R^2 values from the IS, OOS, and bootstrapping analyses across multiple horizons—against each sentiment measure's joint-score (top panel) and mean Wald statistic (bottom panel). A linear trendline is overlaid (blue dashed line) in each panel. The positive slope in both plots suggests that higher sentiment statistics, which reflect greater alignment with the joint conditions, are associated with stronger overall return predictability.

Table 14: Overall Return Predictability vs. Sentiment Statistics

Notes: This table presents the mean R^2 values from the IS, OOS, and bootstrapping analyses across multiple horizons for each sentiment measure. These values are shown alongside the corresponding joint-score and mean Wald statistic from Table 4, for two sample periods: 1980–2022 (Panel A) and 1998–2022 (Panel B). The overall mean R^2 for each measures is calculated as the average of its IS, OOS, and bootstrapping mean R^2 values.

	AAII	**UMCSENT**	**BW**	$\mathbf{BW_{PLS}}$	$\mathbf{SENT_{Mng}}$	$\mathbf{BW_{adj}}$	$\mathbf{CBW_{adj}}$	**MPsy**
PANEL A: 01/1980 - 06/2022								
Mean IS R^2	0.71%	1.14%	0.43%	3.43%		5.86%	10.86%	
Mean OOS R^2	3.97%	6.45%	1.43%	4.56%		14.26%	17.45%	
Mean Bootstrapping R^2	2.13%	3.65%	3.37%	3.38%		10.13%	14.56%	
Overall Mean R^2	2.27%	3.75%	1.74%	3.79%		10.08%	14.29%	
joint-score	0.45	0.34	0.25	0.36		0.55	0.76	
mean Wald-stat	40.08	27.86	24.17	30.72		55.56	59.01	
PANEL B: 02/1998 - 06/2022								
Mean IS R^2	2.71%	7.29%	16.00%	11.00%	4.57%	17.71%	26.00%	43.29%
Mean OOS R^2	3.82%	12.97%	19.29%	4.19%	-0.94%	18.41%	23.77%	34.87%
Mean Bootstrapping R^2	4.96%	14.36%	19.46%	8.16%	6.11%	20.70%	25.55%	37.51%
Overall Mean R^2	3.83%	11.54%	18.25%	7.78%	3.25%	18.94%	25.11%	38.56%
joint-score	0.45	0.4	0.65	0.43	0.29	0.65	0.59	0.84
mean Wald-stat	33.56	22.04	61.55	51.77	19.43	84.21	110.25	159.54

of the indexes, particularly in their scaling, timing, and noisiness. Specifically, if the raw components of an index reflect a delayed response to investor sentiment, the measure may be mistimed, effectively capturing lagged rather than contemporaneous sentiment. This hypothesis can be examined through an additional analysis comparing the contemporaneous and predictive relationships of 3-month lagged BW_{adj} and CBW_{adj} indexes with their original counterparts. The results are presented in 15.

I begin with the contemporaneous regressions. While changes in the original measures are positively and significantly related to contemporaneous returns across all horizons, the lagged measures perform worse, particularly at shorter horizons. $L3.BW_{adj}$ and $L3.CBW_{adj}$ only become positively related to contemporaneous returns when the horizon extends beyond 36 and 12 months, respectively. Even then, their relationships are weaker than those of their original counterparts, as reflected in both the β coefficients and R^2 values. For instance, over the 12-month horizon, changes in BW_{adj} are positively related to contemporaneous returns with an R^2 of 7.25%, whereas its lagged version shows a non-significant relationship with an R^2 of 0.8%. Similarly, changes in CBW_{adj} have an R^2 of 18.70% over the same horizon, compared to a R^2 of 4.83% for its lagged version.

Conversely, the predictive regressions show that lagged measures can forecast returns across all horizons. Over the 1–3 month horizons, their predictive power is comparable to that of the original measures. However, once the return horizon extends beyond 12 months, the lagged measures again exhibit weaker performance. Overall, these results suggest that mistimed sentiment measures retain predictive power, particularly over short horizons, while their contemporaneous relationship with returns is substantially weakened. Similar patterns are observed in this paper when comparing the newly proposed measures to those from the existing literature, indicating that existing measures may capture investor sentiment with a delay.

Furthermore, the second part of the Predictability condition states that return predictability should strengthen as the forecast horizon increases. In this view, a proper sentiment proxy should exhibit stable and growing predictive power over longer horizons, rather than peak performance in the short term. The three new indexes better conform to this pattern compared to the others. Figures 6 and 7 illustrate this by comparing the in-sample performance of BW and $\mathrm{BW}_{\mathrm{PLS}}$ with the new indexes, plotting actual market returns against fitted values for the 1980–2022 and 1998–2022 periods. As the forecast horizon extends, the new indexes show a closer alignment with the 45-degree line, indicating improved fit. The enhanced performance of BW in the post-1998 sample, consistent with earlier findings, is also visible in Figure 6.

Lastly, the second part of the Consistency condition posits that both positive and negative investor sentiments should forecast returns. While the literature often reports asymmetrical effects of positive and negative sentiment shocks,[23], this does not necessarily contradict the Consistency condition, as discussed in Section 2. To examine this issue more closely, I conduct an additional analysis that distinguishes the return predictability of extreme positive and negative investor sentiment observations. Specifically, I classify observations in the top and bottom quintiles of each sentiment index as high and low sentiments, respectively. I then estimate the following regression:

$$r_{t \to t+h} = \alpha + \beta_1 S_t \times low + \beta_2 S_t \times high + \varepsilon_{t \to t+h} \tag{8}$$

where *low* and *high* are dummy variables equal to one when sentiment falls in the lowest and highest quintiles, respectively.

The results are presented in Table 16. Over the 1980–2022 period, BW exhibits poor return predictability under both low and high sentiment conditions. Notably, $\mathrm{BW}_{\mathrm{PLS}}$'s predictive power arises almost entirely from high sentiment observations—its low-sentiment coefficients consistently

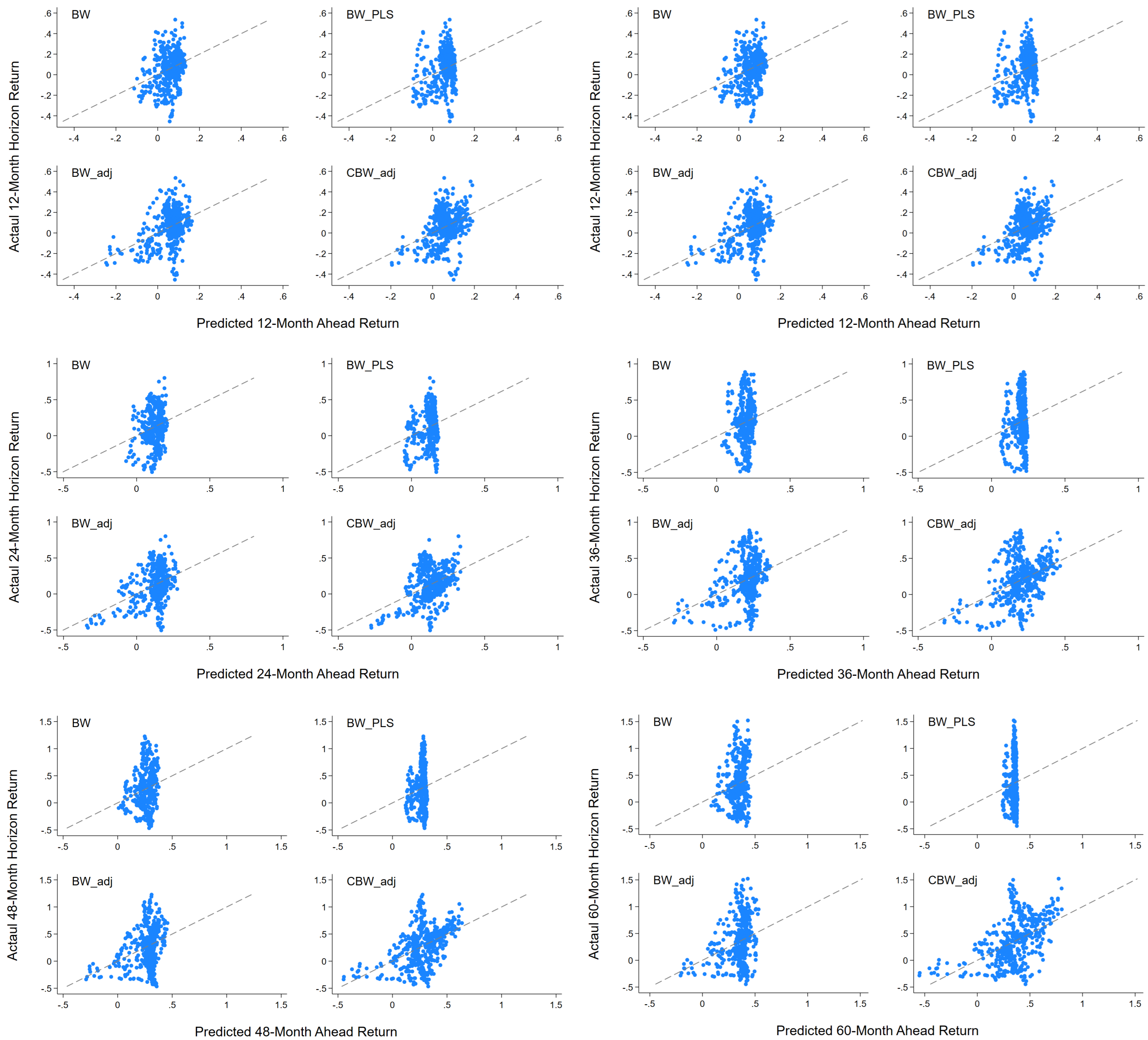
[23]The asymmetry in effects is largely due to different arbitrage constraints during good (boom) and bad (bust) times of the economic cycles—which typically coincide with periods of positive and negative investor sentiment, respectively.(Shleifer and Vishny, 1997; Abreu and Brunnermeier, 2002; Hong and Stein, 2003)

Table 15: Return Predictability of Mistimed Indexes

Notes: This table reports the IS contemporaneous and predictive relationships of 3-month lagged $\mathrm{BW_{adj}}$ and $\mathrm{CBW_{adj}}$, denoted by $L3$, with their original counterparts. The dependent variable is the S&P 500 excess return over the specified horizon. The contemporaneous regressions test the relationship between changes in investor sentiment levels and returns over the same period, while the predictive regressions assess the relationship between investor sentiment levels and future returns. The sample period spans January 1980 to June 2022. The t-statistics, shown in parentheses, are computed using Newey–West standard errors with 12 lags. Significance levels are denoted as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

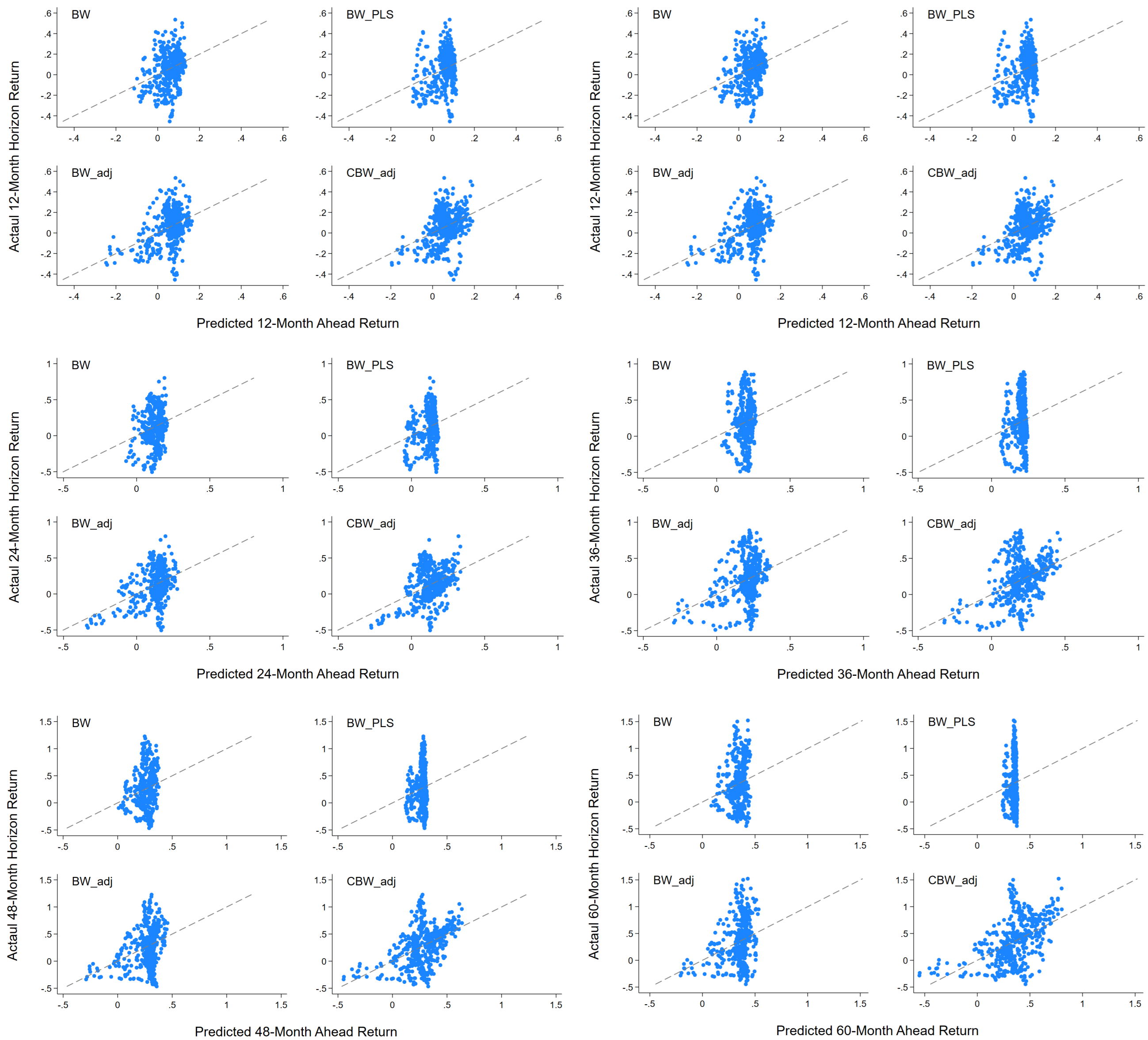
		Contemporaneous Regressions				***Predictive Regressions***			
Horizon	**Statistic**	**L3.BW$_{\mathrm{adj}}$**	**BW$_{\mathrm{adj}}$**	**L3.CBW$_{\mathrm{adj}}$**	**CBW$_{\mathrm{adj}}$**	**L3.BW$_{\mathrm{adj}}$**	**BW$_{\mathrm{adj}}$**	**L3.CBW$_{\mathrm{adj}}$**	**CBW$_{\mathrm{adj}}$**
1	β	0.006	0.004*	-0.004	0.005**	-0.007***	-0.007***	-0.005***	-0.005***
		(0.48)	(1.82)	(-0.44)	(2.37)	(-4.66)	(-4.05)	(-2.89)	(-2.59)
	R^2 (%)	0.05	0.65	0.03	1.57	1.43	1.39	1.25	1.40
	N	509	510	506	507	510	510	507	507
3	β	0.005	0.005**	-0.002	0.012***	-0.021***	-0.020***	-0.016***	-0.016***
		(0.31)	(2.19)	(-0.14)	(4.83)	(-5.84)	(-4.90)	(-3.68)	(-3.05)
	R^2 (%)	0.05	1.58	0.00	9.41	4.79	4.43	4.32	4.46
	N	507	508	504	505	510	510	507	507
12	β	0.017	0.009**	0.041*	0.016***	-0.067***	-0.073***	-0.054***	-0.059***
		(0.66)	(2.42)	(1.74)	(4.00)	(-4.37)	(-6.38)	(-4.26)	(-4.63)
	R^2 (%)	0.80	7.25	4.83	18.70	9.28	11.43	9.98	12.05
	N	498	499	495	496	508	510	505	507
24	β	0.052	0.014**	0.087***	0.021***	-0.094***	-0.112***	-0.083***	-0.093***
		(1.51)	(2.38)	(3.12)	(4.57)	(-3.68)	(-5.09)	(-3.42)	(-4.12)
	R^2 (%)	4.74	9.85	14.95	24.11	7.56	10.92	10.96	13.43
	N	486	487	483	484	496	499	493	496
36	β	0.066*	0.017***	0.118***	0.027***	-0.096***	-0.117***	-0.110***	-0.121***
		(1.94)	(2.92)	(4.21)	(5.56)	(-3.29)	(-4.33)	(-4.36)	(-4.74)
	R^2 (%)	3.90	7.97	18.28	25.88	3.98	6.39	10.97	13.10
	N	474	475	471	472	484	487	481	484
48	β	0.111***	0.026***	0.185***	0.039***	-0.113***	-0.133***	-0.160***	-0.166***
		(3.51)	(5.01)	(6.04)	(7.51)	(-3.49)	(-4.45)	(-6.32)	(-7.15)
	R^2 (%)	5.82	10.12	28.29	34.56	3.16	4.81	14.74	15.67
	N	462	463	459	460	472	475	469	472
60	β	0.136***	0.032***	0.220***	0.047***	-0.101**	-0.133***	-0.189***	-0.207***
		(4.08)	(5.46)	(6.19)	(7.53)	(-2.14)	(-3.18)	(-4.81)	(-5.94)
	R^2 (%)	4.76	8.79	27.92	35.02	1.18	2.65	14.38	16.92
	N	450	451	447	448	460	463	457	460

Figure 6: Predicted vs. Actual Market Returns from IS regressions (1980-2022)



Notes: This figure compares the return predictability of BW and BW_{PLS} with that of BW_{adj} and CBW_{adj} by plotting actual market returns against fitted values from IS regressions. The dashed line represents the 45-degree reference line. The sample period spans from January 1980 to June 2022.

Figure 7: Predicted vs. Actual Market Returns from IS regressions (1998-2022)



Notes: This figure compares the return predictability of BW and BW_{PLS} with that of CBW_{adj} and Mpsy by plotting actual market returns against fitted values from IS regressions. The dashed line represents the 45-degree reference line. The sample period spans February 1998 to June 2022.

exhibit the incorrect (positive) sign. For SENT_{Mng}, low sentiment predicts returns only over short horizons (1- and 3-month), whereas beyond the 12-month horizon, only high sentiment retains predictive power. CBW_{adj} and BW_{adj} deliver the strongest performance overall; however, their low-sentiment coefficients lack predictive strength at shorter horizons.

In the 1998–2022 period, MPsy is the only measure that shows statistically significant return predictability across all horizons for both high and low sentiments. A consistent pattern emerges across BW_{adj}, CBW_{adj}, and MPsy in both sample periods: at shorter horizons (under 12 months), return predictability is primarily driven by high sentiment, as reflected in the magnitude and significance of the coefficients; as the return horizon increases, the predictive contribution of low sentiment becomes more pronounced. In terms of overall performance, BW ranks just below MPsy and CBW_{adj}; however, unlike the new measures, its return predictability under low sentiment is concentrated at shorter horizons. Once again, BW_{PLS} produces high-sentiment coefficients with the incorrect (positive) sign.

Consistent with prior literature, the results from the new indexes reveal a clear asymmetry between the effects of positive and negative investor sentiment. High sentiment primarily drives short-horizon return predictability, whereas low sentiment becomes more influential over longer horizons. Overall, these findings suggest that misvaluation associated with low sentiment (undervaluation) is more substantial and persistent than that linked to high sentiment (overvaluation).

Table 16: Return Predictability of High and Low Investor Sentiment

Notes: This table presents the β_1 and β_2 coefficients from IS regression (8) for different investor sentiment indexes over two sample periods: 1980–2022 (Panel A) and 1998–2022 (Panel B). The low and high columns correspond to extreme negative and positive sentiment levels, defined using the first and fifth quintiles, respectively. The dependent variable is S&P500 excess return over the specified return horizon. The t-statistics, shown in parenthesis, are estimated using Newey-West standard errors with 12 lags. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon	**AAII**		**UMCSENT**		**BW**		**BW_{PLS}**		**$SENT_{Mng}$**		**BW_{adj}**		**CBW_{adj}**		**MPsy**	
(month)	*low*	*high*	*low*	*high*	*low*	*high*	*low*	*high*	*low*	*high*	*low*	*high*	*low*	*high*	*low*	*high*
Panel A: 01/1980 to 06/2022																
1	0.002	-0.004	0.002	-0.001	-0.021	-0.004	0.004	-0.009***			-0.005	-0.008***	-0.002	-0.008***		
	(0.37)	(-0.94)	(0.55)	(-0.25)	(-1.32)	(-1.16)	(0.97)	(-4.26)			(-0.91)	(-4.58)	(-0.44)	(-3.89)		
3	-0.006	-0.000	0.004	-0.004	-0.047	-0.013	0.008	-0.026***			-0.023**	-0.025***	-0.006	-0.024***		
	(-0.60)	(-0.04)	(0.29)	(-0.44)	(-1.31)	(-1.50)	(0.83)	(-3.92)			(-1.98)	(-5.50)	(-0.50)	(-4.90)		
12	-0.008	-0.025	-0.005	-0.001	-0.113*	-0.056*	0.055*	-0.081***			-0.048	-0.090***	-0.039*	-0.084***		
	(-0.29)	(-1.14)	(-0.19)	(-0.02)	(-1.96)	(-1.93)	(1.70)	(-2.76)			(-1.41)	(-6.04)	(-1.86)	(-6.50)		
24	-0.036	-0.036	-0.003	-0.043	-0.120	-0.069	0.190***	-0.107***			-0.068	-0.132***	-0.081**	-0.121***		
	(-1.08)	(-1.05)	(-0.07)	(-0.61)	(-1.09)	(-1.20)	(2.75)	(-2.63)			(-1.40)	(-5.41)	(-2.46)	(-4.69)		
36	-0.042	-0.034	-0.023	-0.103	-0.147	-0.040	0.265***	-0.110**			-0.143**	-0.132***	-0.122***	-0.136***		
	(-1.09)	(-0.82)	(-0.45)	(-1.16)	(-1.26)	(-0.60)	(2.69)	(-2.49)			(-2.27)	(-4.17)	(-2.89)	(-3.67)		
48	-0.026	-0.055	-0.036	-0.216***	-0.290*	-0.091	0.297***	-0.120**			-0.260***	-0.138***	-0.208***	-0.155***		
	(-0.54)	(-1.31)	(-0.49)	(-2.88)	(-1.70)	(-1.59)	(2.98)	(-2.43)			(-2.73)	(-4.13)	(-4.05)	(-4.29)		
60	-0.090	-0.130***	-0.116	-0.291***	-0.265	-0.088	0.379***	-0.113*			-0.183	-0.142***	-0.266***	-0.166***		
	(-1.57)	(-2.96)	(-1.30)	(-4.16)	(-0.82)	(-1.22)	(3.24)	(-1.74)			(-1.03)	(-3.43)	(-3.76)	(-4.01)		
Panel B: 02/1998 to 06/2022																
1	0.003	-0.002	0.006	-0.002	-0.033**	-0.008**	0.003	-0.010***	-0.006***	-0.014	-0.007	-0.006***	-0.003	-0.006***	-0.006*	-0.016**
	(0.44)	(-0.46)	(1.11)	(-0.55)	(-2.57)	(-2.14)	(0.63)	(-4.33)	(-3.31)	(-1.02)	(-1.18)	(-3.26)	(-0.62)	(-3.06)	(-1.80)	(-2.53)
3	-0.001	-0.001	0.015	-0.011	-0.066*	-0.023***	0.007	-0.031***	-0.019***	-0.033	-0.030**	-0.019***	-0.010	-0.019***	-0.016**	-0.046***
	(-0.06)	(-0.17)	(0.83)	(-1.20)	(-1.83)	(-2.78)	(0.55)	(-4.85)	(-4.39)	(-1.00)	(-2.43)	(-4.07)	(-0.84)	(-3.87)	(-1.96)	(-2.90)
12	-0.003	-0.047*	0.001	-0.033	-0.129*	-0.118***	0.055	-0.130***	-0.033*	-0.052**	-0.064	-0.087***	-0.059**	-0.075***	-0.047**	-0.195***
	(-0.09)	(-1.69)	(0.01)	(-0.78)	(-1.95)	(-7.14)	(1.51)	(-6.19)	(-1.75)	(-1.99)	(-1.48)	(-5.66)	(-2.44)	(-5.81)	(-2.37)	(-5.52)
24	-0.057	-0.066	-0.089*	-0.094	-0.187	-0.182***	0.228**	-0.187***	-0.027	-0.114*	-0.095	-0.152***	-0.118***	-0.136***	-0.091***	-0.341***
	(-1.27)	(-1.57)	(-1.66)	(-1.34)	(-1.64)	(-4.55)	(2.21)	(-5.66)	(-1.09)	(-1.88)	(-1.50)	(-7.07)	(-2.94)	(-6.66)	(-3.19)	(-8.44)
36	-0.073	-0.073	-0.145**	-0.159*	-0.240*	-0.158***	0.305**	-0.176***	-0.042	-0.102	-0.213***	-0.142***	-0.192***	-0.142***	-0.140***	-0.363***
	(-1.55)	(-1.51)	(-2.30)	(-1.85)	(-1.92)	(-3.36)	(2.19)	(-4.41)	(-1.18)	(-1.59)	(-2.65)	(-3.98)	(-3.62)	(-3.76)	(-3.82)	(-4.56)
48	-0.055	-0.086*	-0.205**	-0.218***	-0.443**	-0.157***	0.314**	-0.180***	-0.058	-0.152**	-0.366***	-0.129***	-0.303***	-0.123***	-0.211***	-0.324***
	(-0.89)	(-1.65)	(-2.50)	(-2.74)	(-2.52)	(-2.98)	(2.35)	(-3.83)	(-1.21)	(-2.23)	(-3.41)	(-3.71)	(-4.66)	(-3.18)	(-4.35)	(-3.70)
60	-0.095	-0.169***	-0.302***	-0.275***	-0.420	-0.176***	0.473***	-0.210***	0.030	-0.181*	-0.279	-0.148***	-0.361***	-0.128***	-0.271***	-0.315***
	(-1.18)	(-3.50)	(-3.13)	(-3.89)	(-1.24)	(-2.94)	(3.42)	(-3.89)	(0.47)	(-1.93)	(-1.51)	(-3.56)	(-4.13)	(-2.82)	(-4.62)	(-3.03)

4.3 Broader Financial Predictability

Earlier findings suggest that return predictability improves when investor sentiment measures align with the joint conditions implied by the presence of misvaluation. Specifically, sentiment indexes become more powerful predictors when they not only forecast future returns but also contemporaneously explain returns and volatility. In this section, I extend the analysis beyond return predictability to examine whether sentiment measures can also forecast the changes in three other financial outcomes that have been shown to be affected by investor sentiment: fund flows (Brown et al., 2003; Brown and Cliff, 2004; Ben-Rephael et al., 2012; Greenwood and Shleifer, 2014), market volatility (Chen et al., 2021; Ding et al., 2021; Da et al., 2015), and credit spreads (Tang and Yan, 2010; López-Salido et al., 2017). In addition to these three financial outcomes, I also examine the predictability of investor sentiment for returns of call option indexes from Constantinides et al. (2013). However, as the IS results closely mirror those observed for equity returns, they are not reported in detail.[24]

While the return predictability of investor sentiment measures is linked to sentiment-induced misvaluation and its eventual correction, as discussed in Section 2, the same mechanism may not necessarily explain their relationship with the additional financial outcomes examined here. A detailed exploration of these mechanisms is left for future research. Nonetheless, without asserting causality, I hypothesize that investor sentiment may still possess predictive power for these outcomes and outline the following general hypotheses for each:

1. During periods of high (low) investor sentiment, fund flows increase (decrease) as investors become more optimistic (pessimistic) about future equity returns. Once sentiment reverses and prior misvaluation begins to correct, flows are expected to decrease (increase). Therefore, investor sentiment levels should negatively predict future aggregate fund flows.[25]

2. During periods of high (low) investor sentiment, market volatility tends to decrease (increase) as prices rise (fall).[26] Once sentiment reverses and prior misvaluation begins to correct, volatility is expected to increase (decrease) as prices adjust. Thus, investor sentiment levels should positively predict future changes in market volatility.

[24] Overall, the newly proposed measures display a positive contemporaneous relationship with call option returns, unlike most measures from the existing literature. Furthermore, their short-horizon forecasting power is comparable to that of the other measures, however, they outperform over longer horizons.

[25] Equity mutual fund flows are often accompanied by inverse bond fund flows (Da et al., 2015). For simplicity, this analysis focuses on equity fund flows; however, the inverse relationship is also evident in my sample.

[26] The inverse relationship between returns and volatility is well-documented in the empirical literature. Two leading economic explanations are the Leverage Effect (Black, 1976; Christie, 1982) and the Volatility-Feedback Effect (French et al., 1987; Campbell and Hentschel, 1992).

3. During periods of high (low) investor sentiment, often coinciding with strong (weak) economic conditions, credit spreads tend to narrow (widen) as equity prices become overvalued (undervalued) and investors underestimate (overestimate) default risk. Once sentiment reverses and prior misvaluation begins to correct, typically alongside a shift in the economic cycle, spreads are expected to widen (narrow) as equity prices and investors' estimation of default risk adjust. Therefore, investor sentiment levels should positively predict future changes in aggregate credit spread.[27]

[27] More precisely, investor sentiment affects the default risk perception of investors. Optimistic investors underestimate default risk, compressing credit spreads below levels justified by fundamentals, while pessimistic investors overestimate default risk, widening spreads (Gennaioli et al., 2015; Bordalo et al., 2018).

4.3.1 Fund Flows

To test whether investor sentiment measures can predict future changes in equity fund flows, I begin by classifying mutual fund data into four categories: index mutual funds, actively managed mutual funds, index ETFs, and actively managed ETFs. This classification is based on a merged dataset constructed from CRSP's mutual fund database and Morningstar's fund summary data. Also, to improve classification accuracy, I apply keyword-based filters to fund names. Specifically, I identify index funds using keywords such as Vanguard, S&P 500, and Index, and flag leveraged or inverse products using terms like Bull, Bear, Inverse, Short, Leveraged, Ultra, 1.25x, 2x, 3x, 4x, and 5x. ETFs and ETNs are identified using the corresponding terms in fund names. ETNs and leveraged products are excluded from the analysis. Then, I estimate the following regression:

$$\Delta Flow_{j,t\to t+h} = \alpha + \beta S_t + \delta RET_{t-12\to t} + \varepsilon_{t\to t+h} \tag{9}$$

where $\Delta Flow_{j,t\to t+h}$ represents the change in aggregate flow, expressed as a percentage of aggregate assets under management (AUM), into equity fund type j over horizon h. Subscript j refers to one of four fund categories: index mutual funds, actively managed mutual funds, index ETFs, and actively managed ETFs. The control variable $RET_{t-12\to t}$ is the S&P 500 excess return over the prior 12 months. As before, statistical significance is evaluated using Newey–West standard errors with 12 lags.

The results show no significant predictability for flows into index mutual funds, index ETFs, or actively managed ETFs from any of the sentiment indexes.[28] However, some sentiment measures are found to predict changes in aggregate flow into actively managed mutual funds. Table 17 presents the forecasting results for this category.

MPsy significantly predicts future changes in active mutual fund flows across all horizons (1–24 months). BW_{adj} and CBW_{adj} also show predictive power at horizons beyond 1 month. BW, AAII, and UMCSENT display limited predictability over longer horizons, while BW_{PLS} and SENT_{Mng} exhibit no predictive power. In all cases where beta coefficients are statistically significant, their signs are negative, consistent with the first hypothesis: high (low) investor sentiment forecasts a decrease (increase) in fund flows. Overall, these results suggest that the predictive performance of the new indexes surpasses that of the other sentiment measures.[29]

[28] Closer examination reveals a strong positive trend in flows into these fund types over the sample period, consistent with the growing popularity of ETFs and passive investing. Future research may explore whether investor sentiment is related to innovations in flows into these funds.

[29] Appendix A.2 presents additional analyses examining the contemporaneous relationships between changes in investor sentiment and the three financial outcomes discussed in this section. As shown in Table A.3, MPsy, AAII, and UMCSENT exhibit positive and statistically significant correlations with contemporaneous changes in fund flows. BW, BW_{adj}, and CBW_{adj} also show positive relationships, though they are not statistically significant.

Table 17: Fund Flow Predictability

Notes: This table presents the results of in-sample regression (9), which tests the forecasting power of eight investor sentiment indexes for changes in aggregate fund flow. The dependent variable is the change in aggregate flow (as percentage of aggregate AUM) into actively-managed equity mutual funds over the specified horizon. The sample period spans from January 1999 to June 2022. The t-statistics, in parentheses, are estimated based on Newey-West standard error with 12 lags. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	$\mathbf{BW_{PLS}}$	$\mathbf{SENT_{Mng}}$	$\mathbf{BW_{adj}}$	$\mathbf{CBW_{adj}}$	**MPsy**
1	β	-0.025	0.005	-0.024	-0.015	-0.012	-0.009	-0.016	-0.030**
		(-1.19)	(0.20)	(-1.28)	(-0.92)	(-0.64)	(-0.79)	(-1.54)	(-2.40)
	R^2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	N	279	279	279	279	180	279	279	279
3	β	-0.039	-0.016	-0.055	-0.035	-0.042	-0.042**	-0.045***	-0.080***
		(-1.04)	(-0.53)	(-1.46)	(-1.08)	(-1.07)	(-2.09)	(-2.63)	(-2.96)
	R^2	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	N	279	279	279	279	180	279	279	279
12	β	-0.114**	-0.129***	-0.185**	-0.071	-0.088	-0.073**	-0.092***	-0.161**
		(-2.40)	(-3.33)	(-2.48)	(-0.92)	(-0.92)	(-2.06)	(-2.88)	(-2.44)
	R^2	0.07	0.09	0.06	0.07	0.06	0.06	0.06	0.06
	N	279	279	279	279	180	279	279	279
24	β	-0.136*	-0.136***	-0.099	0.049	-0.056	-0.084*	-0.107**	-0.200**
		(-1.81)	(-3.06)	(-0.98)	(0.38)	(-0.61)	(-1.70)	(-2.49)	(-2.17)
	R^2	0.03	0.04	0.00	0.03	0.01	0.01	0.02	0.03
	N	275	275	275	275	180	275	275	275

4.3.2 Market Volatility

Next, I evaluate the forecasting power of investor sentiment measures for market volatility. I consider two proxies: realized volatility, measured as the standard deviation of monthly S&P 500 excess returns, and implied volatility, measured by changes in the CBOE Volatility Index (VIX). I estimate the following regression:

$$\Delta Vol_{t \to t+h} = \alpha + \beta S_t + \delta VIX_t + \varepsilon_{t \to t+h} \tag{10}$$

where $\Delta Vol_{t \to t+h}$ denotes either the standard deviation of monthly S&P 500 excess returns over horizon or the change in the VIX, over horizon h. The control variable VIX_t represents the VIX level at time t. As before, statistical significance is evaluated using Newey–West standard errors with 12 lags.

The results do not show meaningful differences across sentiment measures in forecasting realized volatility, as BW, BW_{PLS}, BW_{adj}, CBW_{adj}, and MPsy all positively predict realized volatility across all horizons (6–24 months). Thus, I do not present or discuss these results further. However, the sentiment measures exhibit more variation in their ability to forecast changes in implied volatility, as shown in Table 18. MPsy, CBW_{adj}, and BW_{adj} consistently predict changes in the VIX across all horizons (6–24 months). BW and BW_{PLS} demonstrate limited predictive power at shorter horizons (6–12 months), while SENT_{Mng} shows weak predictability over longer horizons (12–24 months). AAII and UMCSENT exhibit little to no predictive power. In all cases where the coefficients are statistically significant, their signs are positive, consistent with the second hypothesis: high (low) investor sentiment forecasts an increase (decrease) in the VIX. Once more, the findings suggest that the three new indexes outperform other sentiment measures in predicting market-implied volatility.[30]

4.3.3 Credit Spreads

Lastly, I test the forecasting power of sentiment measures in predicting future changes in aggregate corporate credit spread (using Moody's BAA 10-year corporate spread). I estimate the following

In contrast, BW_{PLS} and SENT_{Mng} display negative, but statistically insignificant, relationships. Overall, these findings support the general intuition behind the hypothesis: sentiment measures that exhibit some degree of predictability also tend to be positively correlated with changes in fund flows.

[30] Additionally, as shown in Table A.4, MPsy, CBW_{adj}, BW_{adj}, AAII, and UMCSENT exhibit negative and statistically significant correlations with contemporaneous changes in the VIX. BW also shows negative relationships, though they are not statistically significant. In contrast, BW_{PLS} and SENT_{Mng} display positive, but statistically insignificant, relationships. Overall, only the new investor sentiment indexes support the general intuition behind the hypothesis: investor sentiment levels positively forecast changes in the VIX and exhibit a significant negative contemporaneous relationship with them.

Table 18: Market Implied Volatility Predictability

Notes: This table presents the results of in-sample regression (10), which tests the forecasting power of eight investor sentiment indexes for changes in market's implied volatility. The dependent variable is the change in CBOE Volatility Index (VIX) over the specified horizon. The sample period spans from February 1998 to June 2022. The t-statistics, in parentheses, are estimated based on Newey-West standard error with 12 lags. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	**BW_{PLS}**	**$SENT_{Mng}$**	**BW_{adj}**	**CBW_{adj}**	**MPsy**
	β	-0.830	-0.135	2.139***	1.898***	1.206	1.325***	0.903**	1.053**
6		(-0.69)	(-0.26)	(3.36)	(3.01)	(1.29)	(3.62)	(2.01)	(2.08)
	R^2	0.31	0.30	0.33	0.34	0.31	0.33	0.32	0.32
	N	290	290	290	290	180	290	290	290
	β	0.126	0.522	3.139***	2.338***	1.644*	1.622***	1.421***	1.598**
12		(0.14)	(0.57)	(3.60)	(2.94)	(1.73)	(2.88)	(2.94)	(2.53)
	R^2	0.39	0.39	0.44	0.43	0.37	0.41	0.42	0.41
	N	290	290	290	290	180	290	290	290
	β	1.499*	-0.131	1.958	0.434	1.799*	1.849***	1.996***	2.351***
24		(1.81)	(-0.17)	(1.45)	(0.34)	(1.75)	(2.99)	(3.97)	(2.86)
	R^2	0.49	0.48	0.49	0.48	0.53	0.50	0.52	0.52
	N	286	286	286	286	180	286	286	286

regression:

$$\Delta Spread_{t\to t+h} = \alpha + \beta S_t + \delta infl_t + \varepsilon_{t\to t+h} \tag{11}$$

where $\Delta Spread_{t\to t+h}$ denotes the change in the aggregate credit spread (Moody's BAA 10-year corporate spread) over horizon h. The control variable $infl_t$ represents the inflation measure from Welch and Goyal (2008), observed at time t.[31] As before, statistical significance is assessed using Newey–West standard errors with 12 lags.

The results are presented in Table 19. Overall, the predictability performance of investor sentiment measures for changes in credit spreads is modest. MPsy is the only measure that significantly forecasts changes in credit spreads across all horizons (1–24 months). CBW_{adj} and BW follow in performance, showing predictability at three of the four horizons. BW_{adj} displays significance only over the 12–24-month horizon, whereas $SENT_{Mng}$ predicts credit spreads only at the 1-3-month horizons. AAII, UMCSENT, and BW_{PLS} exhibit little to no predictive power. In all cases where the β coefficients are statistically significant, their signs are positive, consistent with the third

[31] The inflation variable (*infl*) is the monthly inflation rate based on the change in the Consumer Price Index (CPI). Inflation is closely related to monetary policy and expected economic conditions. Generally, credit spreads widen following monetary tightening, which is often triggered by rising inflation (Gilchrist and Zakrajšek, 2012; Bekaert et al., 2013; Kang and Pflueger, 2015). In my sample, lagged levels of *infl* are positively related to future aggregate spreads over all horizons, and the relationship is statistically significant at 5% level for 12-, 24-, and 36-month horizons.

hypothesis: high (low) investor sentiment forecasts an increase (decrease) in credit spreads. These findings suggest that MPsy, CBW_{adj}, and BW outperform other sentiment measures in forecasting changes in credit spreads.[32]

Table 19: Credit Spread Predictability

Notes: This table presents the results of in-sample regression (11), which tests the forecasting power of eight investor sentiment indexes for changes in aggregate credit spread. The dependent variable is the change in Moody's BAA 10-year corporate spread over the specified horizon. The sample period spans from February 1998 to June 2022. The t-statistics, in parentheses, are estimated based on Newey-West standard error with 12 lags. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	$\mathbf{BW_{PLS}}$	$\mathbf{SENT_{Mng}}$	$\mathbf{BW_{adj}}$	$\mathbf{CBW_{adj}}$	**MPsy**
1	β	0.013	-0.007	0.038*	0.014*	0.038*	0.019	0.022*	0.024*
		(0.63)	(-0.46)	(1.73)	(1.78)	(1.86)	(1.59)	(1.68)	(1.88)
	R^2	0.00	0.00	0.02	0.00	0.04	0.01	0.02	0.01
	N	292	292	292	292	180	292	292	292
3	β	0.039	0.013	0.099*	0.031	0.106**	0.051	0.060	0.066*
		(0.62)	(0.31)	(1.73)	(1.65)	(2.02)	(1.57)	(1.55)	(1.86)
	R^2	0.01	0.00	0.03	0.01	0.05	0.02	0.03	0.03
	N	292	292	292	292	180	292	292	292
12	β	0.175	0.154	0.287**	0.032	0.181	0.154**	0.195**	0.193**
		(1.48)	(1.39)	(2.54)	(0.35)	(1.50)	(2.16)	(2.33)	(2.45)
	R^2	0.10	0.09	0.12	0.06	0.10	0.09	0.13	0.11
	N	292	292	292	292	180	292	292	292
24	β	0.357***	0.220**	0.290	-0.100	0.230	0.207**	0.303***	0.311**
		(3.46)	(2.03)	(1.48)	(-0.59)	(1.36)	(2.44)	(3.09)	(2.27)
	R^2	0.14	0.07	0.07	0.04	0.06	0.07	0.14	0.12
	N	287	287	287	287	180	287	287	287

[32] As shown in Table A.5, MPsy, CBW_{adj}, UMCSENT, and AAII exhibit negative and statistically significant contemporaneous correlations with changes in credit spreads. BW and BW_{adj} also display negative relationships, though not statistically significant. In contrast, BW_{PLS} and $SENT_{Mng}$ tend to show positive but insignificant relationships. Overall, only MPsy and CBW_{adj} support the third hypothesis: investor sentiment positively forecasts changes in aggregate credit spreads, and rising sentiment is associated with narrowing spreads contemporaneously.

4.3.4 Discussion of Broader Financial Predictability Results

To synthesize the findings across prior sections, MPsy and $\mathrm{CBW}_{\mathrm{adj}}$ consistently outperform other sentiment measures in predicting all three financial outcomes. $\mathrm{BW}_{\mathrm{adj}}$ ranks just behind them in performance, failing only to predict fund flows at the 1-month horizon and credit spread at the 1- and 3-month horizons. In contrast, BW and $\mathrm{SENT}_{\mathrm{Mng}}$ fail to predict fund flows (except for BW at the 12-month horizon) but display some predictive power for changes in the VIX and aggregate credit spread. $\mathrm{BW}_{\mathrm{PLS}}$, AAII, and UMCSENT perform the weakest among the eight sentiment indexes evaluated.

In addition to the predictability tests, Appendix A.2 reports results on the contemporaneous relationships between changes in investor sentiment and the three financial outcomes. In summary, MPsy, AAII, and UMCSENT consistently exhibit statistically significant relationships, while $\mathrm{CBW}_{\mathrm{adj}}$ fails only to relate to fund flows over the 1–12-month horizons. $\mathrm{BW}_{\mathrm{adj}}$ and BW generally display the expected signs according to the three hypotheses, although their coefficients are rarely statistically significant. The remaining sentiment measures exhibit weak contemporaneous relationships overall.

To synthesize the findings across prior sections, MPsy and $\mathrm{CBW}_{\mathrm{adj}}$ consistently outperform other sentiment measures in predicting all three financial outcomes. $\mathrm{BW}_{\mathrm{adj}}$ ranks just behind them in performance, failing only to predict fund flows at the 1-month horizon and credit spread at the 1- and 3-month horizons. In contrast, BW and $\mathrm{SENT}_{\mathrm{Mng}}$ fail to predict fund flows (except for BW at the 12-month horizon) but display some predictive power for changes in the VIX and aggregate credit spread. $\mathrm{BW}_{\mathrm{PLS}}$, AAII, and UMCSENT perform the weakest among the eight sentiment indexes evaluated.

In addition to the predictability tests, Appendix A.2 reports results on the contemporaneous relationships between changes in investor sentiment and the three financial outcomes. In summary, MPsy, AAII, and UMCSENT consistently exhibit statistically significant relationships, while $\mathrm{CBW}_{\mathrm{adj}}$ fails only to relate to fund flows over the 1–12-month horizons. $\mathrm{BW}_{\mathrm{adj}}$ and BW generally display the expected signs according to the three hypotheses, although their coefficients are rarely statistically significant. The remaining sentiment measures exhibit weak contemporaneous relationships overall.

Considering both the predictability results and the contemporaneous relationship findings, MPsy and $\mathrm{CBW}_{\mathrm{adj}}$ rank highest in overall performance, aligning most closely with all three hypotheses. $\mathrm{BW}_{\mathrm{adj}}$ and BW follow, demonstrating partial alignment. As previously noted, these four sentiment measures also exhibit the highest sentiment statistics (Table 4), suggesting they best satisfy the theoretical implications of sentiment-induced misvaluation in the market. Once again, this highlights a positive relationship between the sentiment statistics of different measures and their forecasting performance.

5 Conclusion

Aligning with belief-based theoretical models in the behavioral finance literature, this paper proposes a set of joint conditions implied by the potential existence of sentiment-induced misvaluation in financial markets. These conditions offer a general framework for constructing and evaluating empirical measures of investor sentiment.

The findings reveal that indexes often considered to represent investor sentiment in the literature do not fully satisfy the joint conditions, particularly that predictive power is not accompanied by a contemporaneous relationship. In contrast, I introduce new sentiment measures that better align with the joint conditions and demonstrate stronger and more consistent predictive performance. Overall, the results highlight that sentiment measures satisfying the joint conditions not only serve as more reliable predictors of returns but also as robust indicators of broader market behavior.

Several worthy scopes for future research remain, including determining the true economic effects of investor sentiment using properly identified measures, especially its contemporaneous role within asset pricing models. Extending this research to corporate finance applications may also enhance our understanding of the equity cost of capital, corporate decision-making, and managerial behavior.

References

Abreu, D. and Brunnermeier, M. K. (2002). Synchronization risk and delayed arbitrage. *Journal of Financial Economics*, 66(2-3):341–360.

Baker, M. and Wurgler, J. (2000). The equity share in new issues and aggregate stock returns. *the Journal of Finance*, 55(5):2219–2257.

Baker, M. and Wurgler, J. (2004a). Appearing and disappearing dividends: The link to catering incentives. *Journal of financial economics*, 73(2):271–288.

Baker, M. and Wurgler, J. (2004b). A catering theory of dividends. *The Journal of finance*, 59(3):1125–1165.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4):1645–1680.

Baker, M. and Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of economic perspectives*, 21(2):129–151.

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of behavioral economics: applications and foundations 1*, volume 1, pages 79–175. Elsevier.

Barberis, N., Greenwood, R., Jin, L., and Shleifer, A. (2015). X-capm: An extrapolative capital asset pricing model. *Journal of financial economics*, 115(1):1–24.

Barberis, N., Greenwood, R., Jin, L., and Shleifer, A. (2018). Extrapolation and bubbles. *Journal of Financial Economics*, 129(2):203–227.

Barberis, N. and Huang, M. (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review*, 98(5):2066–2100.

Barberis, N., Mukherjee, A., and Wang, B. (2016). Prospect theory and stock returns: An empirical test. *The review of financial studies*, 29(11):3068–3107.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of financial economics*, 49(3):307–343.

Bekaert, G., Hoerova, M., and Duca, M. L. (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788.

Ben-Rephael, A., Kandel, S., and Wohl, A. (2012). Measuring investor sentiment with mutual fund flows. *Journal of financial Economics*, 104(2):363–382.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings from the American statistical association, business and economic statistics section*, page 177.

Bordalo, P., Gennaioli, N., Porta, R. L., and Shleifer, A. (2019). Diagnostic expectations and stock returns. *The Journal of Finance*, 74(6):2839–2874.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2018). Diagnostic expectations and credit cycles. *The Journal of Finance*, 73(1):199–227.

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

Brown, S. J., Goetzmann, W. N., Hiraki, T., Shiraishi, N., and Watanabe, M. (2003). Investor sentiment in japanese and us daily mutual fund flows.

Bybee, J. L. (2023). The ghost in the machine: Generating beliefs with large language models. *Working Paper*.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of financial Economics*, 31(3):281–318.

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

Chen, Z., Liang, C., and Umar, M. (2021). Is investor sentiment stronger than vix and uncertainty indices in predicting energy volatility? *Resources Policy*, 74:102391.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, 10(4):407–432.

Constantinides, G. M., Jackwerth, J. C., and Savov, A. (2013). The puzzle of index option returns. *Review of Asset Pricing Studies*, 3(2):229–257.

Cookson, J. A., Engelberg, J. E., and Mullins, W. (2023a). Echo chambers. *The Review of Financial Studies*, 36(2):450–500.

Cookson, J. A., Fox, C., Gil-Bazo, J., Imbet, J. F., and Schiller, C. (2023b). Social media as a bank run catalyst. *Available at SSRN*, 4422754.

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

Daniel, K. and Hirshleifer, D. (2015). Overconfident investors, predictable returns, and excessive trading. *Journal of Economic Perspectives*, 29(4):61–88.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance*, 53(6):1839–1885.

Daniel, K. D., Hirshleifer, D., and Subrahmanyam, A. (2001). Overconfidence, arbitrage, and equilibrium asset pricing. *The Journal of Finance*, 56(3):921–965.

Ding, W., Mazouz, K., and Wang, Q. (2021). Volatility timing, sentiment, and the short-term profitability of vix-based cross-sectional trading strategies. *Journal of Empirical Finance*, 63:42–56.

Du, D. and Hu, O. (2020). Why does stock-market investor sentiment influence corporate investment? *Review of Quantitative Finance and Accounting*, 54(4):1221–1246.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of financial economics*, 22(1):3–25.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of financial Economics*, 19(1):3–29.

Gennaioli, N., Shleifer, A., and Vishny, R. (2015). Neglected risks: The psychology of financial crises. *American Economic Review*, 105(5):310–314.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American economic review*, 102(4):1692–1720.

Greenwood, R. and Shleifer, A. (2014). Expectations of returns and expected returns. *The Review of Financial Studies*, 27(3):714–746.

Grinblatt, M. and Han, B. (2005). Prospect theory, mental accounting, and momentum. *Journal of financial economics*, 78(2):311–339.

Hong, H. and Stein, J. C. (2003). Differences of opinion, short-sales constraints, and market crashes. *The Review of Financial Studies*, 16(2):487–525.

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

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of finance*, 45(3):881–898.

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

Kahneman, D. and Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific.

Kang, J. and Pflueger, C. E. (2015). Inflation risk in corporate bonds. *Journal of Finance*, 70(1):115–162.

Karabulut, Y. (2013). Can facebook predict stock market activity? In *AFA 2013 San Diego Meetings Paper*.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, 105(1):1–28.

Li, S., Hoque, H., and Liu, J. (2023). Investor sentiment and firm capital structure. *Journal of Corporate Finance*, 80:102426.

López-Salido, D., Stein, J. C., and Zakrajšek, E. (2017). Credit-market sentiment and the business cycle. *The Quarterly Journal of Economics*, 132(3):1373–1426.

Malmendier, U. and Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The quarterly journal of economics*, 126(1):373–416.

Malmendier, U., Pouzo, D., and Vanasco, V. (2020). Investor experiences and financial market dynamics. *Journal of Financial Economics*, 136(3):597–622.

Mian, G. M. and Sankaraguruswamy, S. (2012). Investor sentiment and stock market response to earnings news. *The Accounting Review*, 87(4):1357–1384.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *The Journal of finance*, 52(1):35–55.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of financial economics*, 104(2):288–302.

Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors. *The review of financial studies*, 30(4):1270–1315.

Tang, D. Y. and Yan, H. (2010). Market conditions, default risk and credit spreads. *Journal of Banking & Finance*, 34(4):743–753.

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

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

A Appendix

A.1 GMM Results

The GMM results from Section 4.1 are reported in Tables A.1 and A.2 for the 1965–2022 and 1998–2022 periods, respectively. Over the full 1965–2022 period, CBW_{adj} and BW_{adj} emerge as the best-performing indexes, achieving the highest joint-scores (0.76 and 0.55, respectively) and mean Wald statistics (59.01 and 55.56, respectively). In contrast, UMCSENT and BW record the lowest joint-scores (0.34 and 0.38) and the lowest mean Wald statistics (27.81 and 24.17, respectively). BW_{PLS} performs slightly better than BW, but both measures fail to satisfy the consistency condition: their β_1 coefficients have the incorrect sign (negative), implying that positive changes in sentiment are associated with lower market returns, and their β_5 coefficients are always insignificant and often incorrectly signed (positive), suggesting that negative sentiment predicts lower returns. Among the four survey-based indexes (UMCSENT, AAII, CBEXP, and CBCONF) all satisfy the first part of the contemporaneity condition by relating to returns contemporaneously. However, they consistently fail the second part (β_3 is always insignificant) and exhibit very limited predictability, as reflected in the β_4 and β_5 coefficients. Overall, the Conference Board surveys, CBEXP and CBCONF, are the best performing survey-based indexes.

Over the full 1998–2022 period, which includes six additional sentiment indexes, MPsy emerges as the best-performing measure, with a joint-score of 0.84 and a mean Wald statistic of 159.54. Based on the mean Wald statistic, CBW_{adj} and BW_{adj} follow, with values of 110.25 and 84.21, respectively. However, when ranked by joint-score, BW and BW_{adj} are next in line, with scores of 0.71 and 0.65, respectively. The improvement of BW in satisfying the joint conditions over this period is notable, as it surpasses both BW_{PLS} and SENT_{Mng}. Once again, the coefficients on β_3 are never statistically significant for the survey-based measures, indicating their failure to satisfy the second component of the contemporaneity condition. Among these, CBEXP and CBCONF remain the best-performing survey-based indexes out of the eight evaluated.

Table A.1: GMM Regressions Results (07/1965 to 06/2022)

Notes: This table reports regression results evaluating the ability of eight investor sentiment indexes to satisfy the joint conditions implied by the existence of market misvaluation. The conditions are tested jointly using first-stage GMM: regressions (1) and (3) are used for the 1- and 3-month horizons, while regressions (1), (2), and (3) are used for the 12- and 24-month horizons. Statistical inference is based on Newey–West standard errors with 12 lags: ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The sample period is from July 1965 to June 2022.

	UMCSENT	AAII	CBEXP	CBCONF	BW	BW_PLS	BW_{adj}	CBW_{adj}
Panel A: 1-Month Estimation Window/Horizon								
β_1	0.006*	0.014***	0.007**	0.008**	-0.069***	-0.027	0.004	0.006**
β_2	0.007	0.018***	0.008	0.002	0.035	-0.001	0.006**	0.007
β_4	-0.001	-0.002	-0.005	-0.001	-0.008***	-0.008***	-0.009***	-0.007***
β_5	-0.002	-0.001	-0.001	-0.005	0.003	-0.001	0.006	-0.005
W-stat	10.09	51.99	11.43	8.68	21.26	34.7	34.49	21.64
joint-score$_{(h=1)}$	0.25	0.50	0.25	0.25	0.00	0.25	0.50	0.50
Panel B: 3-Month horizon								
β_1	0.023***	0.025***	0.020***	0.018***	-0.045**	-0.037**	0.002	0.009**
β_2	0.016***	0.018***	0.019***	0.016***	0.043	0.029	0.008***	0.017***
β_4	-0.002	0.002	-0.012	-0.001	-0.022***	-0.024***	-0.026***	-0.021***
β_5	-0.006	-0.013	-0.006	-0.018	0.009	0.002	0.017	-0.015
W-stat	51.78	37.33	44.15	33.83	13.71	30.64	42.31	64.91
joint-score$_{(h=3)}$	0.50	0.50	0.50	0.50	0.00	0.00	0.50	0.75
Panel C: 12-Month horizon								
β_1	0.036***	0.071***	0.036***	0.031***	-0.043	-0.045**	0.001	0.012*
β_2	0.020**	0.032*	0.034***	0.026***	0.081**	0.061**	0.017***	0.020***
β_3	0.003	-0.002	0.005	0.006	0.090***	0.080***	0.006**	0.006*
β_4	0.026	-0.002	-0.014	0.003	-0.065**	-0.067***	-0.091***	-0.066***
β_5	-0.047	-0.054	-0.032	-0.066***	0.021	-0.002	0.048	-0.064**
W-stat	35.38	53.51	63.47	37.47	27.70	20.45	73.50	84.48
joint-score$_{(h=12)}$	0.40	0.40	0.40	0.60	0.60	0.40	0.60	1.00
Panel D: 24-Month horizon								
β_1	0.036***	0.103***	0.047***	0.036***	-0.032	-0.066***	-0.002	0.017
β_2	0.016	0.006	0.026*	0.025***	0.112***	0.100***	0.022***	0.023***
β_3	0	-0.003	0.003	0.004	0.087***	0.083***	0.008***	0.006**
β_4	0.037	0.02	0.016	-0.004	-0.051	-0.067**	-0.118***	-0.077*
β_5	-0.077	-0.122***	-0.080*	-0.102***	0.021	0.014	0.043	-0.133***
W-stat	14.18	17.47	23.25	29.87	34.02	37.11	71.95	65.02
joint-score$_{(h=24)}$	0.20	0.40	0.60	0.60	0.40	0.40	0.60	0.80
Panel E: Average across all horizons								
Average W-stat	27.86	40.08	35.58	27.46	24.17	30.72	55.56	59.01
joint-score	0.34	0.45	0.44	0.49	0.38	0.26	0.55	0.76

Table A.2: GMM Regressions Results (02/1998 to 06/2022)

Notes: This table reports regression results evaluating the ability of 14 investor sentiment indexes to satisfy the joint conditions implied by the existence of market misvaluation. The conditions are tested jointly using first-stage GMM: regressions (1) and (3) are used for the 1- and 3-month horizons, while regressions (1), (2), and (3) are used for the 12- and 24-month horizons. Statistical inference is based on Newey–West standard errors with 12 lags: ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The sample period is from February 1998 to June 2022.

	UMCSENT	**AAII**	**MCONF$_{\text{inst}}$**	**MCONF$_{\text{indiv}}$**	**SMCONF$_{\text{inst}}$**	**SMCONF$_{\text{indiv}}$**	**CBEXP**	**CBCONF**	**BW**	**BW$_{\text{PLS}}$**	**SENT$_{\text{Mng}}$**	**BW$_{\text{adj}}$**	**CBW$_{\text{adj}}$**	**MPsy**
Panel A: 1-Month horizon														
β_1	0.002	0.010*	0.004	-0.003	-0.003	0.011**	0.006	0.009*	-0.071**	-0.057	0.009**	0.006*	0.005	0.017***
β_2	0.005	0.021***	0.008	0.001	0.012	0.005	0.011	0.003	0.052	0.01	-0.009	0.006	0.009	0.028***
β_4	-0.012**	-0.002	0.007	0.013*	0.002	0.004	-0.013**	-0.006	-0.007*	-0.010**	-0.008	-0.006***	-0.007***	-0.016**
β_5	0.009	0.005	-0.002	-0.003	-0.002	-0.006	0.005	-0.001	-0.025**	-0.005	-0.008***	-0.008	-0.003	-0.004
W-stat	6.37	58.21	9.87	5.47	2.52	7.85	10.33	7.78	16.94	24.13	16.49	17.12	16.92	146.05
joint-score$_{(h=1)}$	0.25	0.50	0.00	0.00	0.00	0.25	0.25	0.25	0.25	0.25	0.50	0.50	0.25	0.75
Panel B: 3-Month horizon														
β_1	0.022***	0.021**	-0.01	-0.009	0.007	0.019***	0.022***	0.022***	-0.032	-0.032	0.027***	0.003	0.008	0.009*
β_2	0.015***	0.017*	0.014	0.007	0.012	0.007	0.023***	0.018***	0.078*	0.052	-0.025**	0.009***	0.020***	0.027***
β_4	-0.033**	0	0.021*	0.017	0.014	0.017	-0.031**	-0.016	-0.020**	-0.032***	-0.015	-0.018***	-0.020***	-0.047***
β_5	0.022	-0.007	-0.003	0.004	-0.026**	-0.027*	0.009	-0.006	-0.075**	-0.003	-0.025***	-0.032	-0.012	-0.012*
W-stat	35.64	29.96	8.69	4.13	6.92	14.9	41.14	28.76	29.6	24.85	22.68	29.92	61.19	74.07
joint-score$_{(h=3)}$	0.75	0.50	0.00	0.00	0.25	0.25	0.75	0.50	0.75	0.25	0.25	0.50	0.50	1.00
Panel C: 12-Month horizon														
β_1	0.036***	0.062**	0.013	0.01	0.025*	0.036***	0.023	0.053***	-0.009	-0.047	0.049**	0.004	0.011	0.010*
β_2	0.029***	0.050**	-0.017	-0.02	0.009	0.007	0.052***	0.029***	0.157***	0.111***	-0.012	0.019***	0.024***	0.023***
β_3	0.018	0.002	-0.006	-0.007	0.001	0	0.032***	0.027***	0.100***	0.152***	-0.022**	0.011***	0.024***	0.030***
β_4	-0.063	-0.045	0.035	0.034	0.065**	0.05	-0.053	-0.032	-0.106***	-0.137***	-0.034	-0.076***	-0.076***	-0.186***
β_5	0.008	-0.026	-0.03	-0.021	-0.108**	-0.113**	-0.009	-0.063*	-0.266**	0.033	-0.059***	-0.138**	-0.067**	-0.053**
W-stat	27.75	34.52	7.06	9.97	6.58	30.69	100.17	59.81	117.31	79.43	24.71	110.06	107.27	166.96
joint-score$_{(h=12)}$	0.40	0.40	0.00	0.00	0.20	0.40	0.40	0.80	0.80	0.60	0.20	0.80	0.80	0.80
Panel D: 24-Month horizon														
β_1	0.051***	0.085**	0.068	0.072*	0.027	0.048***	0.046***	0.039***	0.005	-0.069	0.049*	0.002	0.007	0.002
β_2	0.026	0.043	-0.057	-0.048	0.018	0.002	0.057***	0.043***	0.191***	0.118***	0.109	0.025***	0.030***	0.024***
β_3	0.02	-0.001	-0.006	-0.006	0.001	0.005	0.044***	0.033***	0.102***	0.177***	-0.015*	0.012***	0.026***	0.030***
β_4	-0.093	-0.05	0.095	0.115	0.027	0.06	-0.083	-0.063	-0.159***	-0.198***	-0.036	-0.127***	-0.125***	-0.318***
β_5	-0.089	-0.099*	-0.099*	-0.113**	-0.194***	-0.199***	-0.072	-0.132***	-0.430***	0.144	-0.075**	-0.250***	-0.144***	-0.107***
W-stat	18.39	11.54	6.46	17.27	14.89	10.43	65.89	74.19	82.33	78.66	13.83	179.74	255.63	251.07
joint-score$_{(h=24)}$	0.20	0.40	0.20	0.40	0.20	0.40	0.60	0.80	0.80	0.60	0.20	0.80	0.80	0.80
Panel E: Average across all horizons														
Average W-stat	22.04	33.56	8.02	9.21	7.73	15.97	54.38	42.64	61.55	51.77	19.43	84.21	110.25	159.54
joint-score	0.40	0.45	0.0.05	0.10	0.16	0.33	0.50	0.59	0.65	0.43	0.29	0.65	0.59	0.84

A.2 Contemporaneous Relationships with Other Financial Outcomes

In addition to the predictability tests, I examine the contemporaneous relationships between changes in investor sentiment and changes in the financial outcomes using the following regression:

$$\Delta Y_{t\to t+h} = \alpha + \beta \Delta S_{t\to t+h} + \varepsilon_{t\to t+h} \tag{12}$$

where $\Delta Y_{t\to t+h}$ denotes the change over horizon h in one of the three financial outcomes: flows into actively managed mutual funds, the VIX, or the aggregate credit spread (Moody's BAA 10-year corporate spread). $\Delta S_{t\to t+h}$ represents the change in investor sentiment levels over the same horizon. As in the prior analysis, statistical significance is evaluated using Newey–West standard errors with 12 lags.

The results are presented in Tables A.3 to A.5.

Table A.3: Contemporaneous Relationship with Fund Flows

Notes: This table presents the results of in-sample regression (12), which examines the contemporaneous relationship between changes of eight investor sentiment indexes and changes in aggregate fund flow. The dependent variable is the change in aggregate flow (as percentage of aggregate AUM) into actively managed equity mutual funds over the specified horizon. The sample period spans from January 1999 to June 2022. The t-statistics, in parentheses, are estimated based on Newey-West standard error with 12 lags. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	**BW$_{\text{PLS}}$**	**SENT$_{\text{Mng}}$**	**BW$_{\text{adj}}$**	**CBW$_{\text{adj}}$**	**MPsy**
1	β	0.240**	0.076**	0.103	-0.241	-0.027	-0.117	0.052	1.223***
		(1.97)	(2.51)	(0.52)	(-0.87)	(-0.72)	(-1.00)	(0.27)	(2.77)
	R^2	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.02
	N	280	280	280	280	179	280	280	280
3	β	0.194*	0.079*	0.015	-0.171	-0.038	0.055	0.120	0.515**
		(1.95)	(1.76)	(0.12)	(-1.27)	(-1.11)	(0.62)	(1.36)	(2.14)
	R^2	0.02	0.02	0.00	0.01	0.00	0.00	0.01	0.02
	N	280	280	279	280	177	279	279	280
12	β	0.358***	0.220***	0.026	-0.030	-0.086	0.029	0.122	0.318*
		(2.90)	(5.49)	(0.31)	(-0.41)	(-0.87)	(0.81)	(1.64)	(1.67)
	R^2	0.15	0.14	0.00	0.00	0.01	0.00	0.03	0.04
	N	279	279	270	280	168	270	270	276
24	β	0.337**	0.286***	0.062	-0.082	-0.030	0.068	0.166*	0.290*
		(2.26)	(5.19)	(0.58)	(-1.04)	(-0.31)	(1.32)	(1.67)	(1.76)
	R^2	0.16	0.17	0.01	0.01	0.00	0.01	0.07	0.07
	N	267	267	258	276	156	258	258	264

Table A.4: Contemporaneous Relationship with Market Implied Volatility

Notes: This table presents the results of in-sample regression (12), which examines the contemporaneous relationship between changes of eight investor sentiment and changes in market's implied volatility. The dependent variable is the change in CBOE Volatility Index (VIX) over the specified horizon. The sample period spans from February 1998 to June 2022. The t-statistics, in parentheses, are estimated based on Newey-West standard error with 12 lags. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	$\mathbf{BW_{PLS}}$	$\mathbf{SENT_{Mng}}$	$\mathbf{BW_{adj}}$	$\mathbf{CBW_{adj}}$	**MPsy**
6	β	-1.664***	-2.216***	-0.976	0.899	0.371	-0.218	-0.702**	-0.498**
		(-4.26)	(-4.20)	(-0.61)	(0.58)	(0.36)	(-1.21)	(-2.42)	(-2.33)
	R^2	0.14	0.10	0.00	0.00	0.00	0.01	0.07	0.04
	N	291	291	287	291	174	287	287	291
12	β	-1.767***	-3.283***	-1.566	0.392	0.263	-0.223*	-0.560**	-0.384**
		(-5.32)	(-5.53)	(-1.22)	(0.36)	(0.16)	(-1.71)	(-2.12)	(-2.13)
	R^2	0.23	0.18	0.01	0.00	0.00	0.02	0.09	0.07
	N	290	290	281	291	168	281	281	287
24	β	-1.712***	-3.390***	-0.880	1.294	-1.093	-0.158*	-0.448**	-0.288*
		(-3.09)	(-3.41)	(-0.74)	(1.31)	(-1.03)	(-1.75)	(-2.15)	(-1.88)
	R^2	0.26	0.14	0.01	0.02	0.01	0.01	0.09	0.08
	N	278	278	269	287	156	269	269	275

Table A.5: Contemporaneous Relationship with Aggregate Credit Spread

Notes: This table presents the results of in-sample regression (12), which examines the contemporaneous relationship between changes of eight investor sentiment and changes in aggregate credit spread. The dependent variable is the change in Moody's BAA 10-year corporate spread over the specified horizon. The sample period spans from February 1998 to June 2022. The t-statistics, in parentheses, are estimated based on Newey-West standard error with 12 lags. Significance levels are denoted as follows: *** p<0.01, **p<0.05, * p<0.1 .

Horizon (month)	**Statistic**	**AAII**	**UMCSENT**	**BW**	$\mathbf{BW_{PLS}}$	$\mathbf{SENT_{Mng}}$	$\mathbf{BW_{adj}}$	$\mathbf{CBW_{adj}}$	**MPsy**
1	β	-0.012	-0.034***	-0.076	0.087	-0.005	-0.011	-0.023*	-0.037**
		(-1.22)	(-3.74)	(-0.67)	(0.81)	(-0.28)	(-1.12)	(-1.96)	(-2.14)
	R^2	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.03
	N	291	291	291	291	179	291	291	291
3	β	-0.066***	-0.051***	-0.143	0.147	0.010	-0.007	-0.042**	-0.056**
		(-3.32)	(-2.68)	(-1.06)	(1.03)	(0.25)	(-0.51)	(-2.30)	(-2.29)
	R^2	0.06	0.02	0.01	0.01	0.00	0.00	0.04	0.07
	N	291	291	290	291	177	290	290	291
12	β	-0.145***	-0.218***	-0.182	0.118	0.035	-0.013	-0.049*	-0.045*
		(-3.19)	(-3.75)	(-1.20)	(0.90)	(0.19)	(-0.86)	(-1.71)	(-1.83)
	R^2	0.21	0.11	0.03	0.02	0.00	0.01	0.09	0.12
	N	290	290	281	291	168	281	281	287
24	β	-0.173***	-0.274**	-0.119	0.197*	-0.170	-0.009	-0.046*	-0.036*
		(-2.62)	(-2.45)	(-0.75)	(1.78)	(-1.33)	(-0.61)	(-1.66)	(-1.73)
	R^2	0.29	0.10	0.01	0.06	0.02	0.00	0.11	0.13
	N	278	278	269	287	156	269	269	275