Journal of Behavioral and Experimental Finance 37 (2023) 100787

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Finance

journal homepage: www.elsevier.com/locate/jbef

Full length article

# Is sentiment the solution to the risk–return puzzle? A (cautionary) note[☆]

Sze Nie Ung[a], Bartosz Gebka[b,*], Robert D.J. Anderson[b]

[a] Alliance Manchester Business School, Booth Street West, M15 6PB, Manchester, UK
[b] Newcastle University Business School, 5 Barrack Road, NE1 4SE, Newcastle upon Tyne, UK

ARTICLE INFO

*Article history:*
Received 17 January 2022
Received in revised form 20 December 2022
Accepted 6 January 2023
Available online 10 January 2023

*JEL classification:*
G10
G40

*Keywords:*
Risk–return trade-off
Investor sentiment
Market returns
Volatility risk

ABSTRACT

The risk–return relationship in stock markets is often found to be negative or non-existent, in contrast with fundamental finance theories. In this note we investigate if one proposed solution to this puzzle, which states that high irrational investor sentiment disrupts the otherwise positive risk–return nexus, is robust across popular sentiment proxies and therefore empirically comprehensively validated. We find that it is not robust, as most individual sentiment proxies fail to support the hypothesised negative impact of sentiment on the risk–return relationship. Only when a common component of individual proxies is extracted to form a single sentiment measure do we find robust support for the notion that high sentiment impedes rational asset pricing.


## 1. Introduction

One of the cornerstones of finance theory is a positive relationship between volatility risk and expected returns (Merton, 1980); however, corresponding empirical evidence has long been mixed and inconclusive (see discussions in, e.g., Guo and Whitelaw, 2006; Pástor et al., 2008). Yu and Yuan (2011, YY thereafter) proposed a now widely accepted resolution to this puzzle by demonstrating that the risk–return relationship is positive, but only in the absence of strong irrational optimism driving investor behaviour; specifically, they pose that when irrational investor sentiment is low, stocks are priced rationally and the positive risk–return relationship prevails, whereas if sentiment is high, irrationally-motivated trading causes prices to deviate from their fundamentals and the risk–return relationship to break down.[1] YY provide supportive empirical evidence for this conjecture using the Baker and Wurgler (2006) investor sentiment measure.[2,3]

However, the literature on sentiment highlights that various sentiment proxies are only loosely related to one another and, hence, might measure different aspects of sentiment, or worse, some might not even be capturing the underlying sentiment at all (e.g., Huang et al., 2015; Ferrer et al., 2016; Chan et al., 2017).[4] Therefore, a question naturally arises: were the results

☆ We thank the Editor (Stefan Palan), an anonymous reviewer, and Vasileios Kallinterakis for their helpful comments and suggestions. All remaining errors are authors' responsibility.

* Corresponding author.
*E-mail addresses:* sze-nie.ung@manchester.ac.uk (S.N. Ung), b.t.gebka@newcastle.ac.uk (B. Gebka), robert.anderson@newcastle.ac.uk (R.D.J. Anderson).

1 This asymmetric impact of high versus low sentiment on stock prices is derived from a number of premises (Yu and Yuan, 2011): firstly, acting on low sentiment involves shorting of stocks which is costly and can be limited due to restricted availability of stocks available for borrowing, whereas acting on high sentiment simply involves going into less expensive and more easily available long positions; secondly, sentiment traders are inexperienced and mis-estimate the volatility risk, demanding insufficient risk compensation, hence when they trade more (during high sentiment periods) the risk–return relationship is more likely to break down; thirdly, sentiment traders, being irrational by definition, sacrifice rational risk compensation to pursue (mis)anticipated gains driven by their cognitive biases, e.g., overinvest in lottery-type stocks. In sum, stock prices are expected to be affected by high but not low sentiment.

2 Shen et al. (2017), Wang (2018) and Wang and Duxbury (2021) also find high investor sentiment to disrupt the risk–return relationship, in line with YY's hypothesis.

3 Other attempts at explaining the mixed results on the sign of the risk–return relationship include, e.g., conditioning on days with important economic news announcements (Savor and Wilson, 2014) or on whether the firm's investors face prior gains or losses (Wang et al., 2017), attributing negative risk coefficients to incorrectly specified volatility (Ghysels et al., 2005) or to the desire to hedge changes in investment opportunities (Guo and Whitelaw, 2006), focusing on long-run risk and return measures (Bandi and Perron, 2008), or conditioning on the stage of the business cycle (Harvey, 2001), to name just a few. However, in this study we explore exclusively the robustness and viability of the sentiment-related explanation by YY.

4 Gregory (2021) shows that investor and managerial sentiment indices are linked to rational factors as well, e.g., macroeconomic variables and common risk

https://doi.org/10.1016/j.jbef.2023.100787
2214-6350/

in YY and related studies exclusively due to the choice of a particular, if widely used, sentiment measure which might not be capturing the true sentiment, or is investor sentiment genuinely a robust explanation of the risk–return puzzle? In this study, we investigate if the sentiment-based explanation by YY holds empirically for a wide range of sentiment proxies, as well as of volatility measures. If YY's hypothesis is correct, and if each of the sentiment proxies widely employed in the literature does capture the underlying unobservable sentiment to some extent, it should be validated empirically across a wide range of sentiment (and volatility) measures and corresponding sample periods. However, should we find mixed support for the YY hypothesis using different sentiment measures, it could constitute a rejection of the YY hypothesis, or be due to the deficient ability of some of those proxies to effectively capture sentiment; in this case, we shall attempt to extract the common sentiment factor underlying all individual proxies and will construct and employ those resulting, more robust and reliable, sentiment measures to test the empirical validity of the YY's explanation for the risk–return puzzle.

To give a glimpse into our findings, our first contribution to this branch of the literature is to uncover that most of the popular sentiment measures do not support the risk–return puzzle resolution proposed by YY, with some even generating significantly opposite results. However, when we extract the common sentiment component underlying all these individual measures, which constitutes our second contribution, the results strongly support the theorising behind YY's original results. Hence, sentiment can help explain the risk–return puzzle, but one needs to be extremely prudent and use robust aggregate sentiment measures to arrive at reliable conclusions, in this and any other research area involving latent investor sentiment.

## 2. Data and variable construction

As an equally-weighted stock market index does not underweight small stocks, where sentiment's impact would be most pronounced, as a measure of market returns we employ returns data on an equally-weighted S&P500 index, including all distributions, from CRSP.[5] Monthly data is employed for regression analysis, though daily data is also obtained for calculations of volatility measures, as discussed below. To calculate the excess return, we proxy the risk-free return by the 3-month treasury bill rate from the FRED database.

To measure stock market volatility ($Vol_t$), we employ six proxies: realised volatility ($RV_t$), three proxies based on GARCH family models ($GARCHV_t$, $TGARCHV_t$, and $EGARCHV_t$), sample variance ($VAR_t$), and inter-quantile volatility ($IQV_t$). Table 1, Panel A presents a detailed description of these variables.

We employ the following eleven empirical proxies for the unobserved investor sentiment $S_t$: Baker and Wurgler (2007) aggregate sentiment proxy ($BW_t$), Huang et al. (2015) orthogonalised index ($PLS_t$), Conference Board's Consumer Confidence Index ($CCI_t$), University of Michigan Index of Consumer Sentiment ($ICS_t$), American Association of Individual Investors' index ($AAII_t$), NYSE strength index ($STRENGHT_t$), New York Times Sentiment index ($NYTS_t$) based on García (2013), CBOE Volatility Index ($VIX_t$), Manager Sentiment Index ($MSI_t$) from Jiang et al. (2019), FEARS index ($FEARS_t$) of Da et al. (2015), and Gao et al. (2020) sentiment index ($GIS_t$). In line with the related literature, monthly values are obtained as averages of higher frequency data within each month unless stated otherwise. Table 1, Panel B presents a detailed description of our sentiment proxies. A broader discussion of the concept of investor sentiment and its measurement is presented in Appendix.

Summary statistics for the above are tabulated in Table 2 while plots of the volatility measures, excess market return and sentiment measures are in Figs. 1–3, respectively.

## 3. Methodology

Following Pástor et al. (2008), we regress the realised excess market return at time $t$+1, $r_{t+1}$, a proxy for the expected future excess market return ($E_t(r_{t+1})$), on a market volatility proxy, $Vol_t$:

$$r_{t+1} = a + bVol_t + \varepsilon_{t+1} \tag{1}$$

Under the established theories of rational asset pricing, we should observe a positive and significant relationship between (volatility) risk and expected return, i.e., $b > 0$. To investigate if this relationship is affected by sentiment, we follow YY and estimate the following model:

$$r_{t+1} = a_1 + b_1Vol_t + a_2D_t + b_2Vol_tD_t + \varepsilon_{t+1}. \tag{2}$$

$D_t$ is a dummy variable to indicate high-sentiment regimes constructed following YY:[6] within each year, $D_t$ is set to one if that year is identified as a high-sentiment regime (i.e., if the average sentiment proxy value within the *previous* year was higher (lower for VIX and FEARS) than its full sample mean).[7] As YY argue, $b_1$ should be expected to be positive, as a positive risk–return relationship in low-sentiment regimes should not be impaired by the impact of irrational sentiment; however, during high-sentiment periods that irrational trading is expected to bias the risk–return relationship, giving rise to negative $b_2$.

## 4. Results

We estimate models (1) and (2) using monthly returns data for the period dictated by the availability of $BW_t$, i.e., 07/1965 – 12/2018; some variables are only available for shorter intervals, as listed in Table 1. We first look at correlations between proxies for both volatility and sentiment (tabulated in Tables 3 and 4). Our volatility measures seem to be capturing the same phenomenon, as correlations are very high, all being above 89%. For sentiment proxies, correlations are more dispersed, ranging from almost 0 to 75% (in absolute terms); this indicates support for the notion that sentiment proxies available in the literature capture different aspects of a common sentiment component, or

factors. Relatedly, Zhang et al. (2018) show that their sentiment proxy affects returns in some but not all countries considered, casting further doubts onto the universal reliability of any empirical proxy in consistently capturing the latent investor sentiment.

5 The focus on small stocks follows the reasoning in YY and is motivated by the well-known phenomenon that sentiment affects small stocks more than their larger counterparts (e.g., Baker and Wurgler, 2006, 2007), therefore, any sentiment-related effects will manifest themselves more clearly when using an equally-weighted rather than value-weighted index. Indeed, when we repeat our analysis using a value-weighted index which relatively under-weighs small stocks, the results (available on request) are qualitatively virtually identical, albeit less pronounced, in line with YY.

6 Unless stated otherwise, sentiment dummies are calculated based on each proxy-specific sample dictated by data availability; for volatilities we utilise the entire sample period throughout, to maximise the number of observations and therefore obtain the most precise volatility estimates.

7 The BW index is, by construction, centred around zero. Therefore, YY were able to differentiate between high and low sentiment regimes based on the sign of that index. However, the construction methods of other sentiment indexes differ and they are not necessarily naturally centred around zero. Hence, it is necessary to proxy their (latent) steady-state value (from which deviations can be classified as high or low sentiment); this is done by using the mean value of each series as an unbiased proxy of the long-run steady state around which sentiment is anchored. It is worth noting that employing this method on BW makes no difference empirically since the average of this measure is zero (see Table 2).

**Table 1**
Definitions of variables employed.

| Variable name | Symbol | Description |
|---|---|---|
| *Panel A: Proxies of stock market volatility* | | |
| Realised volatility | $\boldsymbol{RV_t}$ | Monthly sum of daily squared market returns (Merton, 1980). |
| GARCH-based volatility | $\boldsymbol{GARCHV_t}$ | GARCH (1,1) predicted daily conditional volatility using daily market returns. Summed over the calendar month. |
| Threshold-GARCH-based volatility | $\boldsymbol{TGARCHV_t}$ | A GARCH specification (Glosten et al., 1993) accounting for asymmetric impacts of past shocks on conditional volatility. |
| Exponential-GARCH-based volatility | $\boldsymbol{EGARCHV_t}$ | An alternative GARCH specification (Nelson, 1991) to allow for an asymmetric impact of past shocks on conditional volatility. |
| Sample variance | $\boldsymbol{VAR_t}$ | Computed monthly based on daily market returns. |
| Inter-quantile volatility | $\boldsymbol{IQV_t}$ | Estimated following Pearson and Tukey (1965) as 95th minus 5th quantiles of return distribution divided by 3.25. Monthly values of $\boldsymbol{IQV_t}$ obtained based on the distribution of daily market returns in each calendar month. |
| *Panel B: Proxies of investor sentiment* | | |
| Baker and Wurgler (2007) aggregate sentiment proxy | $\boldsymbol{BW_t}$ | First principal component of sentiment proxies: value-weighted dividend premium, first-day IPO returns, IPO volume, closed-end fund discount, and equity shares in new issues. Each orthogonalised using macroeconomic variables. Monthly: 07/1965-12/2018. |
| Huang et al. (2015) orthogonalised index | $\boldsymbol{PLS_t}$ | Employs variables underlying the $\boldsymbol{BW_t}$ index, over the same time period, combined using partial least squares (PLS). Monthly: 07/1965-12/2018. |
| Conference Board's Consumer Confidence Index | $\boldsymbol{CCI_t}$ | Survey of consumers' opinions on present conditions and expectations about the future of the economy. Data from DataStream (mnemonic: USCNFCONQ). Monthly: 02/1967-. |
| University of Michigan Index of Consumer Sentiment | $\boldsymbol{ICS_t}$ | Household level survey data. Annual series employed: 1961- (sentiment regimes calculated using annual series to maximise sample period). |
| American Association of Individual Investors' index | $\boldsymbol{AAII_t}$ | Individual investor responses on stock market expectations in the next six months as bullish, bearish or neutral. Calculated as *%Bullish - %Bearish*. Weekly: 24/07/1987-. |
| NYSE strength index | $\boldsymbol{STRENGHT_t}$ | Proportion of shares where prices have risen, minus the proportion which have fallen. Equivalent to TICK index. DataStream (NYSTRGT(SI)). Daily: 02/01/1969-27/05/2016. |
| New York Times Sentiment index | $\boldsymbol{NYTS_t}$ | García (2013): the number of positive minus negative words identified in the relevant NYT columns, scaled by the total number of words. Daily: 03/01/1905-03/01/2006. |
| CBOE Volatility Index | $\boldsymbol{VIX_t}$ | Future expected stock market volatility derived from observed option prices; DataStream (CBOEVIX). Daily: 02/01/1990-. |
| Manager Sentiment Index | $\boldsymbol{MSI_t}$ | Jiang et al. (2019), based on the textual tone of corporate financial disclosures. Monthly: 01/2003-12/2017. |
| FEARS index of Da et al. (2015) | $\boldsymbol{FEARS_t}$ | Based on internet searches for terms with negative economic connotations for households' finances. Daily: 01/07/2004-30/12/2011. |
| Gao et al. (2020) sentiment index | $\boldsymbol{GIS_t}$ | Extends Da et al. (2015) by additionally including searches for non-economic conditions, such as sport outcomes, weather, terrorism, etc. Weekly: 04/07/2004-21/12/2014. |

maybe even other, unrelated phenomena. Likewise, the direction of the correlation is not always of the expected sign: for example, we would expect VIX and PLS to be negatively correlated, given that the former is a "fear index" and hence captures the inverse of positive sentiment, yet they are strongly positively correlated. If the YY hypothesis possesses strong empirical validity and if sentiment measures capture the same underlying latent phenomenon in terms of the prevailing sentiment regime, one would expect the YY hypothesis to manifest itself across a range of those different sentiment measures.

Table 5 reports results from estimating model (1). In line with a substantial number of papers, including YY, there is no support for the positive risk–return relationship: rather than being positive and significant, $b$ is negative in all but one case and insignificant for all six volatility measures.

Table 6 presents the results from estimating model (2) across the six volatility proxies and eleven sentiment measures. Looking across all cases, there is very little support for the YY's hypothesis that risk–return relationship remains positive in low-sentiment regimes (as measured by $b_1$) but is negatively affected by high sentiment (as measured by $b_2$): out of 66 cases, there are only 13 where both $b_1 > 0$ and $b_2 < 0$ and both are significant (at the 10% level). Admittedly, this is slightly higher than the 6–7 one would expect to observe by chance at this significance level, however, even when looking at coefficients' signs and ignoring significance there are only 29 cases, i.e., fewer than 50%, where the signs are as hypothesised ($b_1 > 0$ and $b_2 < 0$).

Interestingly, for the Baker and Wurgler (2007) sentiment measure, an annual version of which YY employed, the results strongly support their sentiment-driven explanation: we observe

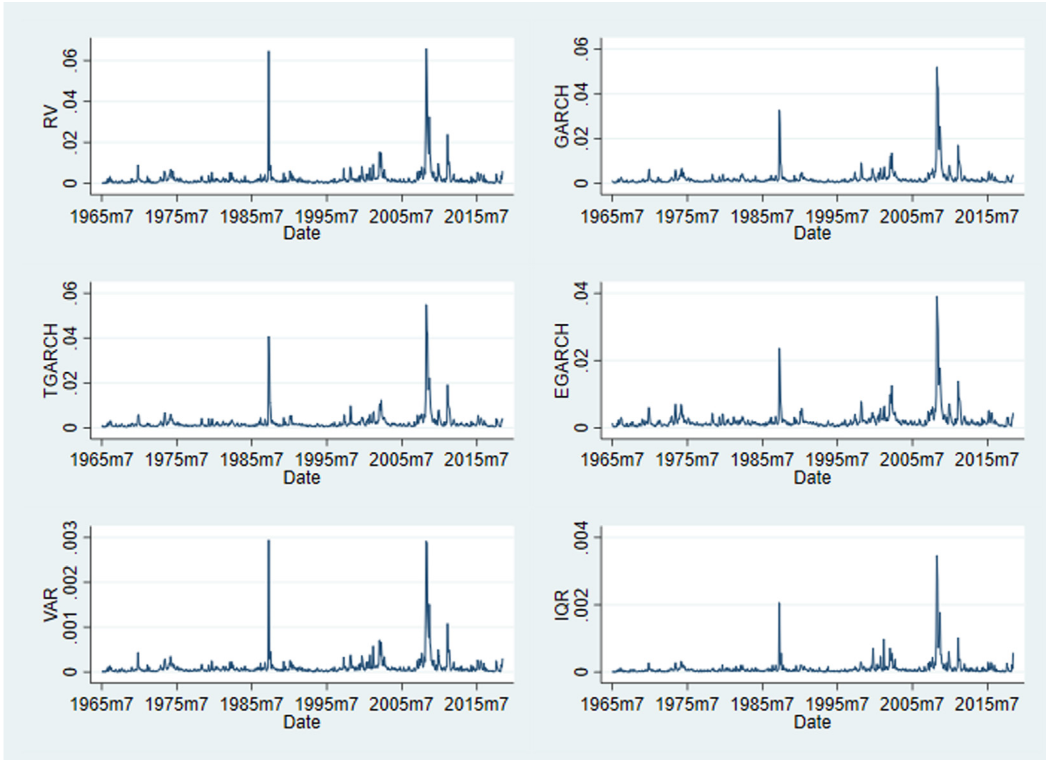


**Fig. 1.** Volatility measures.

**Table 2**
Summary statistics of variables used.

| Variable | Obs | Mean | Std. dev. | Min | Max |
|---|---|---|---|---|---|
| *Excess market return* | | | | | |
| $r_{t+1}$ | 642 | 0.0069 | 0.0496 | −0.2612 | 0.2261 |
| *Volatility measures* | | | | | |
| $RV_t$ | 642 | 0.0023 | 0.0050 | 0.0001 | 0.0656 |
| $GARCH_t$ | 642 | 0.0023 | 0.0041 | 0.0004 | 0.0518 |
| $TGARCH_t$ | 642 | 0.0023 | 0.0042 | 0.0004 | 0.0548 |
| $EGARCH_t$ | 642 | 0.0022 | 0.0031 | 0.0003 | 0.0390 |
| $VAR_t$ | 642 | 0.0001 | 0.0002 | 0.0000 | 0.0029 |
| $IQR_t$ | 642 | 0.0001 | 0.0002 | 0.0000 | 0.0035 |
| *Sentiment measures* | | | | | |
| $BW_t$ | 642 | 0.0000 | 1.0000 | −2.4220 | 3.1974 |
| $PLS_t$ | 642 | 0.0000 | 1.0008 | −1.7654 | 3.7628 |
| $CCI_t$ | 623 | 94.8039 | 24.8470 | 25.3000 | 144.7000 |
| $AAII_t$ | 411 | 0.0753 | 0.1474 | −0.4100 | 0.5047 |
| $STR_t$ | 569 | 1.4715 | 9.4608 | −31.4909 | 51.6000 |
| $NYTS_t$ | 486 | −0.0102 | 0.0040 | −0.0228 | −0.0006 |
| $VIX_t$ | 348 | 19.2683 | 7.4992 | 10.1255 | 62.2535 |
| $MSI_t$ | 180 | 0.0000 | 1.0028 | −4.1460 | 1.9660 |
| $FEARS_t$ | 90 | 0.0023 | 0.0369 | −0.1389 | 0.1162 |
| $GIS_t$ | 126 | 0.0033 | 0.0565 | −0.1358 | 0.2274 |
| $ICS_t$ | 642 | 85.8187 | 11.5390 | 63.7000 | 107.6000 |

**Table 3**
Correlations between volatility measures.

| | $RV_t$ | $GARCH_t$ | $TGARCH_t$ | $EGARCH_t$ | $VAR_t$ | $IQR_t$ |
|---|---|---|---|---|---|---|
| $RV_t$ | 1 | | | | | |
| $GARCH_t$ | 0.9227 | 1 | | | | |
| $TGARCH_t$ | 0.9288 | 0.9909 | 1 | | | |
| $EGARCH_t$ | 0.9252 | 0.9866 | 0.9837 | 1 | | |
| $VAR_t$ | 0.9955 | 0.9245 | 0.9295 | 0.9292 | 1 | |
| $IQR_t$ | 0.9532 | 0.8924 | 0.8925 | 0.9009 | 0.9611 | 1 |

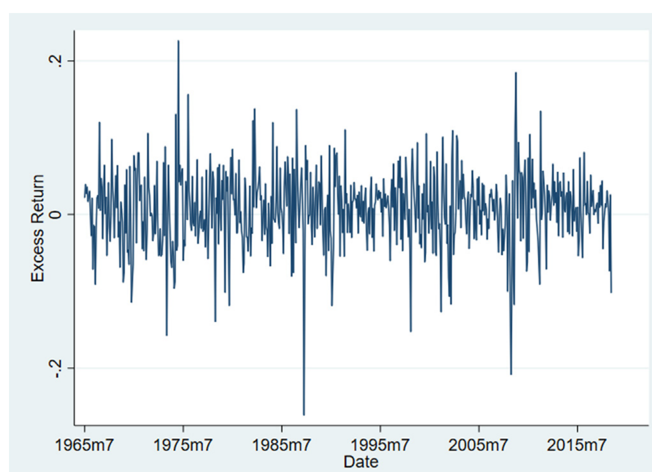


**Fig. 2.** Excess market returns.

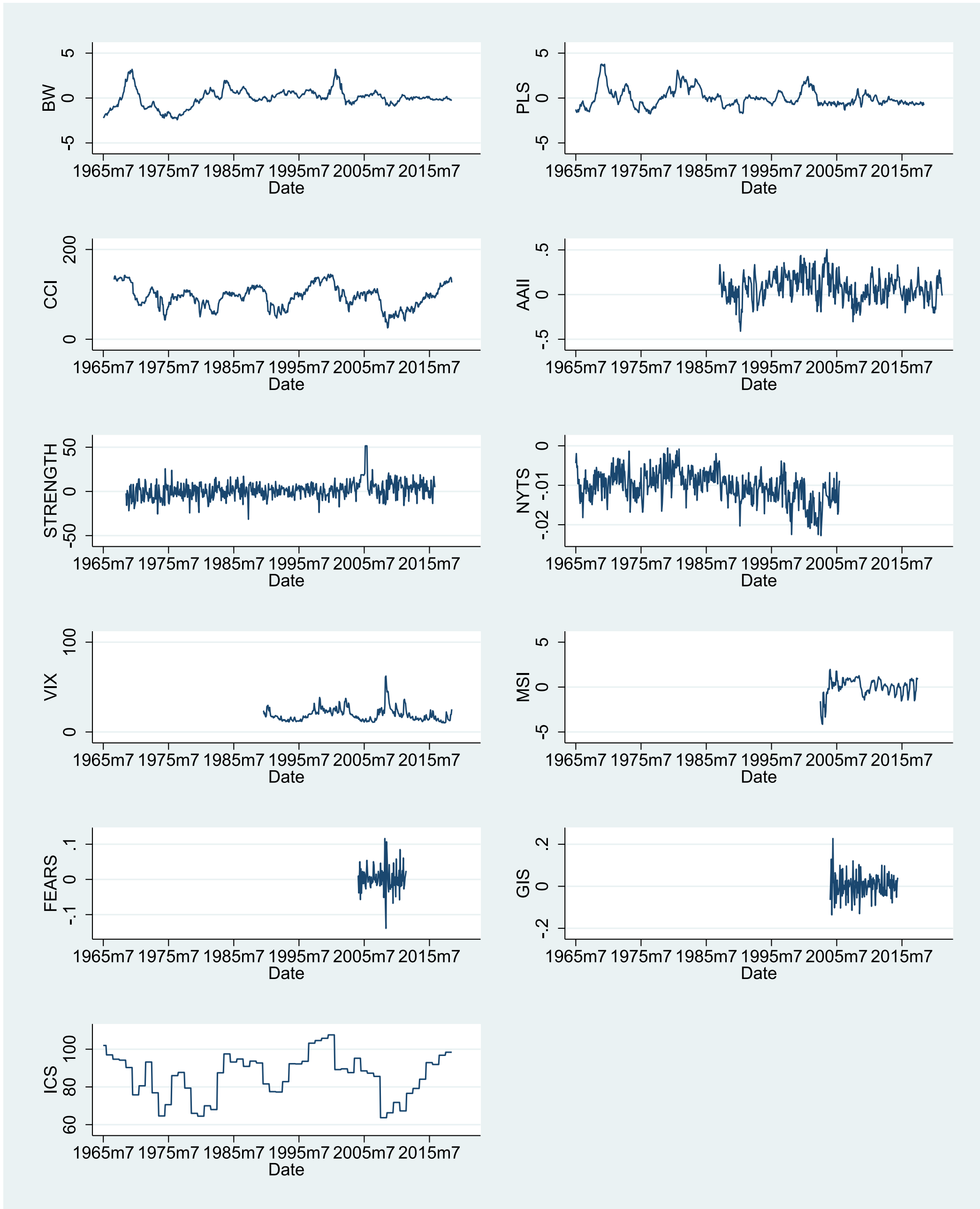


**Fig. 3.** Sentiment measures.

**Table 4**
Correlations between sentiment measures.

| | $BW_t$ | $PLS_t$ | $CCI_t$ | $AAII_t$ | $STR_t$ | $NYTS_t$ | $VIX_t$ | $MSI_t$ | $FEARS_t$ | $GIS_t$ | $ICS_t$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $BW_t$ | 1 | | | | | | | | | | |
| $PLS_t$ | 0.3611 | 1 | | | | | | | | | |
| $CCI_t$ | −0.5782 | −0.5125 | 1 | | | | | | | | |
| $AAII_t$ | −0.2495 | −0.1777 | 0.1119 | 1 | | | | | | | |
| $STR_t$ | 0.7465 | 0.1230 | −0.3777 | −0.0876 | 1 | | | | | | |
| $NYTS_t$ | −0.2425 | −0.2765 | 0.4318 | 0.4971 | −0.0044 | 1 | | | | | |
| $VIX_t$ | 0.0599 | 0.2917 | −0.3397 | −0.4054 | −0.3901 | −0.6691 | 1 | | | | |
| $MSI_t$ | −0.6119 | −0.3623 | 0.4254 | 0.1553 | −0.4567 | −0.1549 | 0.0371 | 1 | | | |
| $FEARS_t$ | 0.0911 | −0.1322 | 0.1184 | −0.0024 | 0.1276 | 0.1846 | −0.0407 | −0.1208 | 1 | | |
| $GIS_t$ | 0.0529 | −0.0877 | 0.1207 | 0.0944 | −0.1658 | 0.2703 | −0.0798 | −0.1183 | −0.3457 | 1 | |
| $ISC_t$ | −0.2563 | 0.1883 | −0.1549 | 0.4052 | −0.6251 | −0.1496 | 0.5470 | 0.1770 | −0.1519 | 0.2435 | 1 |

**Table 5**
Selected estimation results for model (1).

| Volatility measure | $RV_t$ | $GARCHV_t$ | $TGARCHV_t$ | $EGARCHV_t$ | $VAR_t$ | $IQV_t$ |
|---|---|---|---|---|---|---|
| Parameter: $b$ | −0.5228 | −0.0004 | −0.1267 | 0.0343 | −9.9880 | −4.6926 |
| | (0.6230) | (0.8560) | (0.7655) | (1.1317) | (12.8094) | (14.6053) |

Note: Table represents estimating (1) in the period of 07/1965 to 12/2018. Standard errors are shown below each parameter estimate in parenthesis. $^{***}$, $^{**}$, $^{*}$ denotes significance at 1%, 5%, 10% level, respectively.

**Table 6**
Selected estimation results for model (2).

| Volatility measure/ Parameter estimate (SE) | $RV_t$ | | $GARCHV_t$ | | $TGARCHV_t$ | | $EGARCHV_t$ | | $VAR_t$ | | $IQV_t$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ |
| $BW_t$ | 1.9709 | −3.0541$^{**}$ | **3.1701**$^{**}$ | **−3.9758**$^{***}$ | **2.9271**$^{*}$ | **−3.6986**$^{**}$ | **3.9911**$^{**}$ | **−5.0954**$^{***}$ | 41.5096 | −62.9655$^{**}$ | **44.4389**$^{*}$ | **−63.9439**$^{**}$ |
| | (1.4537) | (1.4926) | (1.2827) | (1.3764) | (1.5101) | (1.5918) | (1.6924) | (1.8367) | (29.0795) | (29.5507) | (24.6630) | (25.3644) |
| $PLS_t$ | −1.3524$^{***}$ | 3.9798$^{***}$ | −1.0944$^{**}$ | 4.7695$^{***}$ | −1.0457$^{**}$ | 4.6829$^{***}$ | −1.5369$^{**}$ | 6.0829$^{***}$ | −27.1789$^{***}$ | 82.2862$^{***}$ | −25.5712$^{***}$ | 81.7382$^{***}$ |
| | (0.3132) | (1.2791) | (0.4918) | (1.2992) | (0.5134) | (1.4621) | (0.6351) | (1.7160) | (6.6357) | (25.5591) | (6.0475) | (21.2214) |
| $CCI_t$ | 0.1713 | −1.1593 | 1.7321 | −2.6046$^{*}$ | 1.1209 | −1.9710 | 2.2020 | −3.1197 | 3.5051 | −21.5670 | 16.9598 | −31.2756 |
| | (1.1386) | (1.3342) | (1.2762) | (1.4726) | (1.1359) | (1.3746) | (1.7400) | (2.0811) | (24.0596) | (27.9292) | (29.8771) | (33.0671) |
| $ICS_t$ | −0.7096 | 0.4981 | −0.4252 | 1.6601 | −0.6186 | 1.5786 | −0.5650 | 2.1810 | −13.5190 | 9.9369 | −7.3981 | 10.5198 |
| | (0.8313) | (1.0776) | (0.9026) | (1.3954) | (0.8153) | (1.2125) | (1.2039) | (1.8120) | (16.6229) | (22.5420) | (17.2465) | (25.9454) |
| $AAII_t$ | −0.7690 | 1.3344 | −0.5857 | 1.8717 | −0.7766 | 2.1682 | −0.8706 | 2.6676 | −14.7962 | 28.6386 | −7.0798 | 14.3344 |
| | (0.9178) | (1.7410) | (0.9399) | (2.0130) | (0.8393) | (2.0083) | (1.2291) | (2.2655) | (18.1716) | (35.5610) | (18.8879) | (33.3478) |
| $STRENGHT_t$ | **3.3991**$^{***}$ | **−4.6042**$^{***}$ | **4.5727**$^{***}$ | **−5.4978**$^{***}$ | **4.6779**$^{***}$ | **−5.5350**$^{***}$ | **5.7099**$^{***}$ | **−6.8992**$^{***}$ | **67.7774**$^{***}$ | **−91.9588**$^{***}$ | **63.4891**$^{***}$ | **−88.0389**$^{***}$ |
| | (0.9861) | (0.9636) | (1.1206) | (1.0157) | (1.3336) | (1.1924) | (1.4201) | (1.3320) | (21.4391) | (21.3530) | (15.2658) | (16.3956) |
| $NYTS_t$ | 1.4594 | −1.7187 | 1.3783 | 0.4428 | 1.4813 | −0.2090 | 1.4089 | 1.2491 | 31.2546 | −35.9688 | 27.5384 | −29.6156 |
| | (1.4457) | (1.5970) | (2.1840) | (2.6652) | (2.1018) | (2.4067) | (2.1079) | (3.0301) | (29.7409) | (33.5787) | (21.9213) | (31.3404) |
| $VIX_t$ | 1.9170 | −3.4717$^{**}$ | **2.7889**$^{*}$ | **−4.3366**$^{***}$ | 2.6709 | −4.2294$^{***}$ | **3.5072**$^{*}$ | **−5.5928**$^{***}$ | 38.2826 | −67.9112$^{**}$ | **40.7899**$^{*}$ | **−66.6415**$^{***}$ |
| | (1.4679) | (1.4618) | (1.4575) | (1.4598) | (1.6319) | (1.6305) | (1.9248) | (1.9284) | (29.4364) | (29.5792) | (24.0763) | (23.9482) |
| $MSI_t$ | −1.7179$^{*}$ | 1.0057 | −0.3048 | −0.2477 | −0.5424 | −0.1887 | 0.0404 | −0.8854 | −34.7457 | 21.1822 | −40.4407 | 34.4877 |
| | (1.0380) | (1.4381) | (1.7099) | (1.9769) | (1.5912) | (1.8226) | (1.9483) | (2.3478) | (23.0851) | (30.3435) | (25.8892) | (33.0739) |
| $FEARS_t$ | −0.7298 | −1.8376 | −0.5284 | −0.7445 | −0.7225 | −0.7168 | −0.8272 | −0.2327 | −13.8528 | −40.9095 | −5.6889 | −56.9716$^{*}$ |
| | (0.9939) | (1.3807) | (1.0003) | (1.9918) | (0.8836) | (1.8043) | (1.3127) | (2.4505) | (19.5962) | (28.9907) | (21.5256) | (31.6859) |
| $GIS_t$ | −1.5691$^{***}$ | 4.6765$^{***}$ | −1.3512$^{***}$ | 4.9041$^{***}$ | −1.3986$^{***}$ | 4.9598$^{**}$ | −1.7806$^{***}$ | 6.4825$^{***}$ | −29.7927$^{**}$ | 89.4988$^{**}$ | −25.6021$^{***}$ | 85.3498$^{***}$ |
| | (0.4898) | (1.6723) | (0.3975) | (1.5983) | (0.3840) | (1.9992) | (0.5704) | (2.4292) | (11.5559) | (37.7736) | (9.6563) | (28.0076) |

Note: Table represents estimating (2) across a maximised sample period for each sentiment proxy. Standard errors are shown below each parameter estimate in parenthesis. $^{***}$, $^{**}$, $^{*}$ denotes significance at 1%, 5%, 10% level, respectively. Cases which support the YY hypothesis ($b_1 > 0$ and $b_2 < 0$ and significant) are in bold.

$b_1 > 0$ and $b_2 < 0$ for all six volatility proxies, with most coefficients being significant. Some other sentiment proxies generate similarly supportive results for the hypothesis of YY, especially $STRENGTH_t$ and $VIX_t$. However, several other sentiment measures generate insignificant results throughout, while $PLS_t$ and $GIS_t$ consistently and statistically significantly indicate that the opposite of what YY implied holds.[8,9]

This apparent failure to uniformly support the YY hypothesis across a large set of alternative sentiment proxies could be indicative of one of the following two possibilities: (*i*) either the YY hypothesis that high sentiment leads to a break-down of the otherwise positive risk–return relationship is incorrect (which manifests itself as lack of robustness to the use of different sentiment proxies, and indeed appears significantly rejected when employing $PLS_t$ and $GIS_t$), or (*ii*) the YY hypothesis is indeed correct but the available empirical sentiment measures are heterogeneous in their respective ability to accurately capture the true latent sentiment, resulting in mixed empirical evidence as observed here. If the latter is the correct explanation, using the "right" sentiment proxy would help address the hypothesis testing problem here; however, there is no consensus in the literature as to which sentiment measure is the correct one or which is more driven by noise or other, unrelated phenomena. To address this issue, we draw on the rationale underlying the construction of Baker and Wurgler (2006, 2007) and Huang et al. (2015) composite sentiment indices, and extract the common component underlying our employed individual sentiment measures; as those authors have argued, such a common component constitutes an improved proxy for the sentiment itself.

Sentiment extraction is performed using two alternative methods, the principal component analysis (PCA) as in Baker and Wurgler (2006, 2007) and the partial least squares (PLS) method as in Huang et al. (2015). PCA allows for the creation of variables, principal components, each of which captures a (different) commonality in the underlying variable set and which are uncorrelated with one another. PLS is designed to avoid a potential pitfall of the PCA in misidentifying the common measurement error as a 'genuine' common economic factor. Each approach is used to generate an aggregate index for all underlying variables based around their common factor. If the YY hypothesis is correct, we would expect the sentiment measures resulting from PCA and PLS, which should capture the latent sentiment more accurately than the individual variables used so far, to generate results in line with that hypothesis, i.e., $b_1 > 0$ and $b_2 < 0$. Both methods are applied to a set of five sentiment proxies: $BW_t$, $PLS_t$, $STRENGTH_t$, $CCI_t$, and $NYTS_t$, which are available at the monthly frequency for the longest possible common period in our sample, i.e., 01/1969 – 12/2005. Measures of both volatility and sentiment

[8] Our finding that only few sentiment measures support the YY hypothesis is robust to the sample period chosen, as shown by the results in Table A.1 in the appendix: for the common period 01/1969 – 12/2005 for which there exist monthly observations for five out of 11 sentiment proxies, we find that only two proxies (the same as in Table 6), $BW_t$ and $STRENGTH_t$, generate results in line with the YY hypothesis. For the sake of this common sample analysis, we re-estimated all volatilities and sentiment proxies utilising only data from within this reduced sample period.

[9] Windsorising volatilities proxies at the top 5%, as a robustness check to ensure that more extreme volatilities are not unduly influencing our results, yields results qualitatively identical to those in Table 6.

**Table 7**
Selected estimation results for model (2): Sentiment aggregates.

| Volatility measure/ Parameter estimate (SE) | $RV_t$ | | $GARCHV_t$ | | $TGARCHV_t$ | | $EGARCHV_t$ | | $VAR_t$ | | $IQV_t$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ |
| $PC1_t$ | **5.8644**** | **−6.2294**** | **11.9366***** | **−10.8114***** | **9.1291***** | **−8.2099**** | **9.6857***** | **−8.0112**** | **142.4016**** | **−149.3308***** | **157.4700**** | **−158.8182**** |
| | (2.4912) | (2.5645) | (3.1888) | (3.3695) | (3.3085) | (3.4278) | (3.0238) | (3.3807) | (55.1451) | (56.7167) | (63.2473) | (65.9762) |
| $PC2_t$ | **2.8328*** | **−3.5569**** | 3.0986 | −1.8357 | 3.5323 | −2.6986 | 3.5037 | −1.4692 | **61.7332** * | **−77.2899**** | **49.2085** ** | **−65.7067**** |
| | (1.5808) | (1.6148) | (2.4441) | (2.7284) | (2.3526) | (2.5189) | (2.4876) | (3.1928) | (33.4277) | (34.2669) | (23.9349) | (27.3657) |
| $CPLS_t$ | **2.6540*** | **−3.3383**** | 3.9591 | −2.9910 | **4.6341*** | **−4.1275*** | 4.5416 * | −3.1026 | **58.1811** * | **−73.5066**** | **48.4581** ** | **−65.6733**** |
| | (1.5751) | (1.6544) | (2.5457) | (2.8316) | (2.3631) | (2.5007) | (2.5550) | (3.2524) | (32.7128) | (34.4073) | (23.8189) | (29.7991) |

Note: Table represents estimating (2) in a common sample period of 01/1969 to 12/2005 for the various measures of volatility, and including 5 sentiment measures, *BW*, *PLS*, *CCI*, *STRENGTH* and *NYTS* in the calculation of the first and second principal components (PC1 and PC2) and CPLS. Standard errors are shown below each parameter estimate in parenthesis. ***, **, * denotes significance at 1%, 5%, 10% level, respectively. Cases which support the YY hypothesis ($b_1 > 0$ and $b_2 < 0$ and significant) are in bold.

entering model (2) are recalculated to only utilise data from within this common sample.

Results from the PCA analysis (untabulated) indicate that there are two principal components, $PC1_t$ and $PC2_t$, which are orthogonal to one another and which make significant contributions to explaining the total variance of the underlying variables (with eigenvalues significantly higher than one; Kaiser, 1960), jointly explaining around 60% of the total variance (at 36.46% and 23.62%, respectively).[10] $PC1_t$ is mostly driven by $BW_t$ and $PLS_t$ whereas $PC2_t$ shows highest loadings from $STRENGTH_t$ and $NYTS_t$, with $CCI_t$ contributing equally to both. Hence, there appear to be potentially two sentiment aspects captured by our data and manifesting themselves in those two distinct principal components. As there is no *a priori* theoretical indication as to which one, $PC1_t$ or $PC2_t$ (or both), is the "correct" approximation for the latent sentiment, we employ each of them in turn to measure sentiment in model (2). However, following the rationale underlying the BW index, one would expect only the first principal component to represent common sentiment; we discuss later on whether $PC2_t$ is also proxying for sentiment.

Estimation results reported in Table 7 for both principal components clearly indicate empirical support for the YY hypothesis: for all six volatility proxies used, we find $b_1 > 0$ and $b_2 < 0$, with this result being significant for all applications with $PC1_t$ and half of cases where $PC2_t$ was used as a sentiment measure. To the extent that each of those principal components captures a common factor underlying all our individual sentiment proxies, hence resulting in more precise and reliable measures of the latent sentiment, this finding yields strong support for the YY hypothesis. We also note that $PC1_t$ outperforms $PC2_t$ in this context, suggesting that the former is a superior measure of sentiment.

The result obtained when employing the PLS composite index ($CPLS_t$, constructed by employing our five longest sentiment variables), as shown in the bottom row of Table 7, also supports the YY hypothesis, as we find $b_1 > 0$ and $b_2 < 0$ for all volatility measures, with most of those coefficients being significant.[11]

The importance of sentiment revealed here could be argued to be spurious, as sentiment could simply coincide with the true factor affecting the risk–return relationship, such as financial stress and market uncertainty. If this was the case, conditioning on high/low value of such a factor would render the sentiment coefficient insignificant in model (2). To investigate such an alternative explanation, we divide our common sample into high/low financial stress and uncertainty regimes, using mean values of the Financial Stress Index by Püttmann (2018) and realised volatility, respectively. The results (unreported to conserve space but available on request) firstly show that the risk–return nexus is insignificant only in high stress and uncertainty regimes; this is where the puzzle is concentrated and where sentiment could show its explanatory power. When we estimate model (2) in those high stress/uncertainty subsamples, we obtain a heterogeneous picture in terms of how each sentiment proxy affects the risk–return relationship, with some generating results in line with the YY hypothesis and results (e.g., $BW_t$, $STRENGTH_t$), while other proxies continuing to indicate the opposite effect ($PLS_t$, $GIS_t$), in line with our baseline results in Table 6. More importantly, for our proposed methods of aggregation of sentiment proxies, the results for PCA-based and CPLS sentiment indices are also in line with our baseline findings (Table 7) in that all coefficients are consistent ($b_1 > 0$, $b_2 < 0$), which supports the YY hypothesis that high sentiment disrupts the otherwise positive risk–return relationship.

Overall, results from both sentiment aggregation approaches, PCA and PLS, which should yield more robust and reliable empirical approximations of the unobservable sentiment (Baker and Wurgler, 2006, 2007; Huang et al., 2015), strongly support the hypothesis of YY that the positive risk–return relationship prevails when investor sentiment is low, but is negatively affected by high values of this sentiment.[12,13] Methodologically, the fact

[10] Their less-than-perfect coverage of the total variance further highlights the notion that these individual sentiment measures are rather noisy proxies of the true latent sentiment, as indicated by their low correlations already noted, and speaks in favour of employing a wide range, or an aggregate, of individual measures, rather than relying on a single one.

[11] Again, we conduct additional robustness checks of these results as shown in Table A.2 in the appendix. Firstly, to strengthen the robustness of PCA- and PLS-derived sentiment proxies we add an additional variable for the common sample analysis, i.e. $ICS_t$, excluded previously due to the mixed survey sampling frequency (quarterly in the 1969–1978 period, monthly thereafter) used to construct the annual index value (following YY). Next, we conduct the PCA and PLS analyses on such a wider set of six sentiment measures over a common (but shorter) sample period of 01/1969 – 12/2005. These results strongly support our baseline finding in Table 7 that aggregate sentiment proxies ($PC1_t$ and $CPLS_t$) widely support the YY hypothesis. In addition (results untabulated), estimations employing windsorised volatilities (at top 5%) strongly support the YY hypothesis when using $PC1_t$, but less so for $CPLS_t$ and especially for $PC2_t$.

[12] For further robustness, we repeat our analysis on a shorter sample of 2004–2011, for which data on most sentiment proxies fully exists (all except for $NYTS_t$). We find that (*i*) the risk–return relationship is consistently insignificant, (*ii*) different proxies yield heterogeneous, oftentimes contradictory to one another, results from model (2), in line with our baseline finding that one cannot rely on a randomly chosen proxy to test for the YY proposition, and (*iii*) the aggregate sentiment proxy $PC1_t$ consistently yields coefficient values in support of YY but the other two aggregate measures do not. The last finding could be due to the fact that this sub-sample is relatively short, therefore estimates are less reliable and precise, and it was rather unrepresentative of a longer US stock market history with the Global Financial Crisis of 2007–9 and its aftermath exerting unique effects on investors' behaviour and asset prices.

[13] The absolute values of $b_1$ and $b_2$ are almost identical across Table 7 for each combination of volatility and sentiment proxies, implying that high sentiment

that both aggregate measures yield qualitatively identical results, whereas individual measures showed a high degree of disagreement, further supports the proposition that the latter are plagued by noise and/or capture phenomena other than sentiment; therefore, aggregate sentiment measures should be used in empirical research instead.

## 5. Conclusions

When investor sentiment is measured using individual proxies, there is only very weak support for the notion proposed by YY that irrational sentiment explains the risk–return puzzle. However, when we control for the imperfect ability of individual proxies to capture sentiment by extracting their common factor, i.e., the unobservable sentiment, we obtain supportive results for the YY hypothesis, in line with empirical findings in YY and related literature (Shen et al., 2017; Wang, 2018; Wang and Duxbury, 2021): high sentiment causes the otherwise positive risk–return relationship to break down. Hence, this note provides robust evidence in support of sentiment as an explanation for the risk–return puzzle.

More generally, our results indicate that any conclusions in the literature regarding the importance of investor sentiment can be highly sensitive to the choice of the sentiment measure: while some individual proxies generate results in line with YY's sentiment-based explanation, most do not, even if they have been shown elsewhere to be superior sentiment measures, such as the $PLS_t$ index. In fact, it would have been easy to reject the YY hypothesis, had just one, "wrong" sentiment measure, been employed. The implication from this analysis is therefore that researchers should be wary of any results involving investor sentiment obtained using just one individual variable to capture sentiment; rather, one should be employing aggregate sentiment proxies which are more likely to reveal the true underlying sentiment component.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix. Investor sentiment (measures) and asset prices

Market efficiency had been the reigning paradigm in finance (Fama, 1970; Jensen, 1978) until significant evidence of irrational behaviour started to emerge (Banz, 1981; Shiller, 1984; De Bondt and Thaler, 1985; Black, 1986). Investor sentiment was proposed as a potential causal force behind these irrational movements in asset prices, being broadly defined as investors' believes about asset values not fully justified by hard facts.[14] This notion gained ground following the theoretical work by De Long et al. (1990): they demonstrated analytically that irrational investors can drive the equilibrium price away from its fundamental value, and the mere existence of these "noise traders" can limit and delay the corrective actions by rational arbitrageurs, hence resulting in potentially long-lasting mispricing. Empirical studies proceeded to test if high (low) sentiment causes contemporaneous excessive price increases (declines) and longer-term price reversals towards their fundamental values; hence, a negative relationship between current sentiment and future stock returns is accepted as an indicator of the former having systematic impact on stock prices. However, as investor sentiment is not directly observable, a plethora of empirical approximations for that latent factor have been proposed in the literature. These can be divided into survey-, market-based, and textual (media- and search-based) sentiment measures.

Survey-based measures of sentiment attempt to capture the respondents' views by gathering information on their expectations of the future stock market developments and general economic conditions (e.g., De Bondt, 1993; Fisher and Statman, 2000; Brown and Cliff, 2004, 2005; Lemmon and Portniaguina, 2006; Greenwood and Shleifer, 2014). The surveyed agents can be stock investor themselves (American Association of Individual Investors' index) or, more broadly, consumers (Conference Board's Consumer Confidence Index, University of Michigan Index of Consumer Sentiment). Most empirical studies tend to find a negative relationship between such measured sentiment and future returns. However, there has been some criticism of, e.g., consumer confidence indexes as proxies for investor sentiment (Ferrer et al., 2016; Fisher and Statman, 2003; Jansen and Nahuis, 2003; Otoo, 1999), as general consumers' views do not necessarily translate directly into investors' trading decisions (Binswanger and Salm, 2017).

Market-based measures, on the other hand, utilise observable market data which is co-determined by latent investor sentiment. These proxies include the options-implied volatility (VIX) index (e.g., Cheon and Lee, 2018; Da et al., 2015; Kaplanski and Levy, 2010; Lutz, 2016) and other derivatives-related variables (Bathia and Bredin, 2013; Dennis and Mayhew, 2002; Wang et al., 2006), the closed-end fund discount (Doukas and Milonas, 2004; Gemmill and Thomas, 2002; Lee et al., 1991; Neal and Wheatley, 1998), equity issuance including IPOs (Baker and Wurgler, 2000; Brown and Cliff, 2004), trading intensity (Baker and Stein, 2004), and the dividend premium (Baker and Wurgler, 2004, 2007). Given a wide range of these potential proxies of one single underlying factor, i.e., investor sentiment, Baker and Wurgler (2006) proposed to extract it by means of the principal component analysis utilising the closed-end fund discount, two IPO-related measures, share of equity issues, dividend premium, and trading volume (the latter was subsequently dropped as a less reliable proxy), resulting in a widely popular aggregate US sentiment index. Other studies have followed their approach, constructing investor sentiment indexes for other stock markets (e.g., Chen et al., 2010, for Hong Kong; Finter et al., 2012, for Germany; Hu and Wang, 2012, Li, 2015, and Yang and Zhou, 2016, for China, Ryu et al., 2016, for Korea). The Baker and Wurgler (2006) index has also been widely employed in various financial contexts, for instance to investigate stock market anomalies (Stambaugh et al., 2012; Antoniou et al., 2013), mean–variance relation (Yu and Yuan, 2011), pricing of macro-risk (Shen et al., 2017), and the slope of the security market line (Antoniou et al., 2016). More recently, Huang et al. (2015) argued that the method by Baker and Wurgler captures a common error component rather than sentiment, and constructed an alternative proxy using a different methodological approach (partial least squares). They found that such an investor sentiment index significantly predicts short-term future aggregate stock market returns.

is not just marginally but rather fully able to eradicate the otherwise positive risk–return relationship, highlighting the significant role of investor sentiment in asset price behaviour.

[14] The definitions of sentiment vary but appear closely aligned to one another; for instance, Morck et al. (1990) observe that "There seem to be good theoretical as well as empirical reasons to believe that investor sentiment, also referred to as fads and fashions, affects stock prices. By investor sentiment we mean beliefs held by some investors that cannot be rationally justified. Such investors are sometimes referred to as noise traders". Similarly, Shleifer and Summers (1990) note that "[…] not all demand changes appear to be so rational; some seem to be a response to changes in expectations or sentiment that are not fully justified by information. […] Although these changes in demand are unwarranted by fundamentals, they can be related to fundamentals, as in the case of overreaction to news." Baker and Wurgler (2006) are more generic: "One might also define investor sentiment as optimism or pessimism about stocks in general", but Baker and Wurgler (2007) also highlight the detachment from hard facts as a distinguishing feature of sentiment: "Investor sentiment, defined broadly, is a belief about future cash flows and investment risks that is not justified by the facts at hand."

A more recent branch of the literature focuses on text-based investor sentiment measures, comprising of media- and search-based approaches. Media-based measures are computed by analysing the content published by traditional or social media, such as positive and/or negative words in newspaper columns (e.g., Tetlock, 2007; García, 2013), opinions on internet stock message boards (e.g., Antweiler and Frank, 2004; Das and Chen, 2007; Kim and Kim, 2014), Twitter posts (e.g., Bollen and Mao, 2011), and opinions from Seeking Alpha (Chen et al., 2014). Meanwhile, the Thomson Reuters MarketPhysc indices employed by Sun et al. (2016) and Eierle et al. (2022)[15] as a sentiment measure contains textual information from both traditional and social media. These measures are believed to reflect or shape investor sentiment as investors express or follow the news and opinions, and react accordingly. Nevertheless, evidence on the ability of these media-based sentiment measures to predict stock (market) returns is inconclusive.

The possible reason why media-based sentiment measures produce mixed results is also their main drawback: they are not elicited from actual investor actions, and hence may potentially capture stated (hence largely hypothetical) but not revealed (via actual trades) preferences/sentiment, and therefore do not capture that important 'skin-in-the-game' element of investor behaviour. Besides that, Eierle et al. (2022) suggest that social media generated indices are capturing firm prospects, rather than sentiment *per se*. The classification of media news is also complex, given that terms can be interpreted in many ways, with terms having opposing implications depending on the overall context within which they are used. Hence, these methods, although relying on complex dictionaries to quantify the occurrence of negative or positive words or phrases, by their nature can fail to also appreciate the tone, the optimistic or pessimistic state, of the article in which they are used.

Search-based sentiment measures are constructed mainly based on the Google Search Volume Index (SVI). For instance, the FEARS index of Da et al. (2015), formed by aggregating the number of searches for the words that express household concerns, e.g. "unemployment", "recession" and "bankruptcy", reveals that investor pessimism is associated with low contemporaneous returns but higher future returns. Joseph et al. (2011) utilise the number of searches for stock tickers and also find that investor sentiment predicts the return reversal over longer horizons (i.e., beyond two weeks) for stocks that are hard to arbitrage and of high volatility. While Dimpfl and Jank (2016) mention that the number of searches for the stock index is mainly driven by noise traders, Da et al. (2011) claim that an increase in investor attention could also be caused by investors paying attention to genuine news, and find that the SVI is weakly correlated with other media-based sentiment measures. Hence one can question if conceptually the intensity of internet searches should even be considered as a sentiment indicator.

### A.1. Robustness checks

See Tables A.1 and A.2.

[15] Eierle et al. (2022) only use the social media measures.

**Table A.1**
Selected estimation results for model (2), common sample Period (01/1969 – 12/2005).

| Volatility measure/ Parameter estimate (SE) | $RV_t$ | | $GARCHV_t$ | | $TGARCHV_t$ | | $EGARCHV_t$ | | $VAR_t$ | | $IQV_t$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ |
| $BW_t$ | **3.8087*** | −**4.1580*** | **7.2174**** | −**6.0660*** | **7.1009**** | −**6.2209**** | **7.7085**** | −**6.1400*** | **97.9852*** | −**105.3592**** | 72.5284 | −74.6492 |
| | (2.1467) | (2.2470) | (2.9704) | (3.1809) | (2.9598) | (3.1009) | (3.0967) | (3.4720) | (50.0103) | (52.0637) | (46.4604) | (50.9344) |
| $PLS_t$ | −0.3185 | 2.1461 | 1.6884 | 0.8420 | 1.1425 | 1.8254 | 2.7016 | 0.4229 | −6.1829 | 45.3739 | −1.3313 | 40.2967 |
| | (0.7124) | (1.6600) | (1.5039) | (2.8698) | (1.1325) | (2.6801) | (2.2829) | (3.3669) | (16.0426) | (35.4852) | (23.0548) | (32.8248) |
| $CCI_t$ | −0.6844 | 3.3383** | 0.9682 | 2.9909 | 0.5066 | 4.1275* | 1.4389 | 3.1026 | −15.3255* | 73.5066** | −17.2153 | 65.6733** |
| | (0.4158) | (1.6544) | (1.1704) | (2.8317) | (0.8015) | (2.5007) | (1.9702) | (3.2523) | (8.8233) | (34.4073) | (15.6129) | (29.7991) |
| $STRENGHT_t$ | **3.0842*** | −**3.4656*** | **6.8916***** | −**5.9792**** | **6.9112***** | −**6.1950***** | **6.9477***** | −**5.5242*** | **65.8289*** | −**74.0882*** | **53.8441***** | −**63.4274**** |
| | (1.7438) | (1.9338) | (2.1138) | (2.3790) | (1.9564) | (2.1436) | (2.3737) | (2.8819) | (34.1759) | (38.6420) | (20.6225) | (31.0588) |
| $NYTS_t$ | 1.5516 | −1.8797 | 1.4123 | 0.5082 | 1.4932 | −0.0850 | 1.4976 | 1.7351 | 32.9710 | −39.3634 | 29.7761 | −33.6719 |
| | (1.4710) | (1.6068) | (2.3661) | (2.8233) | (2.2613) | (2.5680) | (2.3573) | (3.3559) | (30.2398) | (33.6507) | (21.9534) | (30.8556) |

Note: Table represents estimating (2) in a common sample period of 01/1969 to 12/2005 for the five sentiment proxies which have consistent data during this period. Standard errors are shown below each parameter estimate in parenthesis. ***, **, * denotes significance at 1%, 5%, 10% level, respectively. Cases which support the YY hypothesis (b1>0 and b2<0 and significant) are in bold.

**Table A.2**
Selected estimation results for model (2): Sentiment aggregates, common sample Period (01/1969 - 12/2005).

| Volatility measure/ Parameter estimate (SE) | $RV_t$ | | $GARCHV_t$ | | $TGARCHV_t$ | | $EGARCHV_t$ | | $VAR_t$ | | $IQV_t$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ | $b_1$ | $b_2$ |
| $PC1_t$ | **5.8406**** | −**6.2428***** | **9.9152***** | −**8.8472**** | **9.5553***** | −**8.6896**** | **8.9559**** | −**7.3778*** | **143.7280***** | −**151.7801***** | **153.0770***** | −**155.4136***** |
| | (2.2700) | (2.3513) | (3.6430) | (3.7947) | (3.6425) | (3.7507) | (3.7342) | (4.0354) | (54.3361) | (56.0032) | (52.1514) | (55.5410) |
| $PC2_t$ | −0.2448 | 1.4004 | 1.7698 | 0.0394 | 1.2125 | 0.9814 | 2.8541 | −0.5937 | −4.4405 | 30.7826 | 2.4189 | 29.5738 |
| | (0.8102) | (1.8217) | (1.5948) | (2.9179) | (1.2134) | (2.7119) | (2.3606) | (3.4733) | (18.2811) | (40.9621) | (25.6326) | (46.4489) |
| $CPLS_t$ | 2.4263 | −3.0002* | 3.5741 | −2.3562 | 4.4043* | −3.7747 | 4.2975 * | −2.4612 | **53.3078*** | −**65.9574*** | 43.2759* | −54.2638 |
| | (1.5096) | (1.6376) | (2.4193) | (2.8246) | (2.2883) | (2.4733) | (2.4771) | (3.3729) | (31.5306) | (34.4446) | (22.3014) | (33.1045) |

Note: Table represents estimating (2) in a common sample period of 01/1969 to 12/2005 for the various measures of volatility, and including 6 sentiment measures: *BW*, *PLS*, *CCI*, *STRENGTH*, *NYTS* and *ICS* (instead of 5), in the calculation of the first and second principal components (PC1 and PC2) and CPLS. Standard errors are shown below each parameter estimate in parenthesis. ***, **, * denotes significance at 1%, 5%, 10% level, respectively. Cases which support the YY hypothesis ($b_1 > 0$ and $b_2 < 0$ and significant) are in bold.

## References


Antoniou, C., Doukas, J.A., Subrahmanyam, A., 2013. Cognitive dissonance, sentiment, and momentum. J. Finan. Quant. Anal. 48 (1), 245–275.
Antoniou, C., Doukas, J.A., Subrahmanyam, A., 2016. Investor sentiment, beta, and the cost of equity capital. Manage. Sci. 62 (2), 347–367.
Antweiler, W., Frank, M.Z., 2004. Is all that talk just noise? The information content of internet stock message boards. J. Finance 59 (3), 1259–1294.
Baker, M., Stein, J.C., 2004. Market liquidity as a sentiment indicator. J. Financial Mark. 7 (3), 271–299.
Baker, M., Wurgler, J., 2000. The equity share in new issues and aggregate stock returns. J. Finance 55 (5), 2219–2257.
Baker, M., Wurgler, J., 2004. A catering theory of dividends. J. Finance 59 (3), 1125–1165.
Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. J. Finance 61 (4), 1645–1680.
Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. J. Econ. Perspect. 21 (2), 129–151.
Bandi, F.M., Perron, B., 2008. Long-run risk-return trade-offs. J. Econometrics 143 (2), 349–374.
Banz, R.W., 1981. The relationship between return and market value of common-stocks. J. Finan. Econ. 9 (1), 3–18.
Bathia, D., Bredin, D., 2013. An examination of investor sentiment effect on G7 stock market returns. Eur. J. Finance 19 (9), 909–937.
Binswanger, J., Salm, M., 2017. Does everyone use probabilities? The role of cognitive skills. Eur. Econ. Rev. 98, 73–85.
Black, F., 1986. Noise. J. Finance 41 (3), 529–543.
Bollen, J., Mao, H.N., 2011. Twitter mood as a stock market predictor. Computer 44 (10), 90–93.
Brown, G.W., Cliff, M.T., 2004. Investor sentiment and the near-term stock market. J. Empir. Finance 11 (1), 1–27.
Brown, G.W., Cliff, M.T., 2005. Investor sentiment and asset valuation. J. Bus. 78 (2), 405–440.
Chan, F., Durand, R.B., Khuu, J., Smales, L.A., 2017. The validity of investor sentiment proxies. Int. Rev. Financ. 17 (3), 473–477.
Chen, H., Chong, T.T.-L., Duan, X., 2010. A principal-component approach to measuring investor sentiment. Quant. Finance 10 (4), 339–347.
Chen, H., De, P., Hu, Y., Hwang, B.-H., 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. Rev. Financ. Stud. 27 (5), 1367–1403.
Cheon, Y.-H., Lee, K.-H., 2018. Maxing out globally: Individualism, investor attention, and the cross section of expected stock returns. Manage. Sci. 64 (12), 5807–5831.
Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. J. Finance 66, 1461–1499.
Da, Z., Engelberg, J., Gao, P., 2015. The sum of all FEARS investor sentiment and asset prices. Rev. Financ. Stud. 28 (1), 1–32.
Das, S.R., Chen, M.Y., 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. Manage Sci. 53 (9), 1375–1388.
De Bondt, W.F.M., 1993. Betting on trends: Intuitive forecasts of financial risk and return. Int. J. Forecast. 9 (3), 355–371.
De Bondt, W.F.M., Thaler, R., 1985. Does the stock market overreact? J. Finance 40 (3), 793–805.
De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Noise trader risk in financial markets. J. Polit. Econ. 98 (4), 703–738.
Dennis, P., Mayhew, S., 2002. Risk-neutral skewness: Evidence from stock options. J. Finan. Quant. Anal. 37 (3), 471–493.
Dimpfl, T., Jank, S., 2016. Can internet search queries help to predict stock market volatility? Eur. Finan. Manage. 22 (2), 171–192.
Doukas, J.A., Milonas, N.T., 2004. Investor sentiment and the closed-end fund puzzle: Out-of-sample evidence. Eur. Finan. Manage. 10 (2), 235–266.
Eierle, B., Klamer, S., Muck, M., 2022. Does it really pay off for investors to consider information from social media? Int. Rev. Finan. Anal. 81.
Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. J. Finance 25 (2), 383–417.
Ferrer, E., Salaber, J., Zalewska, A., 2016. Consumer confidence indices and stock markets meltdowns. Eur. J. Finance 22 (3), 195–220.
Finter, P., Niessen-Ruenzi, A., Ruenzi, S., 2012. The impact of investor sentiment on the German stock market. Z. Betriebswirtschaft 82 (2), 133–163.
Fisher, K.L., Statman, M., 2000. Investor sentiment and stock returns. Financ. Anal. J. 56 (2), 16–23.
Fisher, K.L., Statman, M., 2003. Consumer confidence and stock returns. J. Portfol. Manage. 30 (1), 115–127.
Gao, Z.Y., Ren, H.H., Zhang, B.H., 2020. Googling investor sentiment around the world. J. Finan. Quant. Anal. 55 (2), 549–580.
García, D., 2013. Sentiment during recessions. J. Finance 68 (3), 1267–1300.
Gemmill, G., Thomas, D.C., 2002. Noise trading, costly arbitrage, and asset prices: Evidence from closed-end funds. J. Finance 57 (6), 2571–2594.
Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return trade-off after all. J. Finan. Econ. 76 (3), 509–548.
Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. J. Finance 48 (5), 1779–1801.
Greenwood, R.M., Shleifer, A., 2014. Expectations of returns and expected returns. Rev. Financ. Stud. 27 (3), 714–746.
Gregory, R.P., 2021. What determines manager and investor sentiment? J. Behav. Exp. Financ. 30.
Guo, H., Whitelaw, R.F., 2006. Uncovering the risk-return relation in the stock market. J. Finance 61 (3), 1433–1463.
Harvey, C.R., 2001. The specification of conditional expectations. J. Empir. Finance 8 (5), 573–637.
Hu, C., Wang, Y., 2012. Investor sentiment and assets valuation. Syst. Eng. Procedia 3, 166–171.
Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. Rev. Financ. Stud. 28 (3), 791–837.
Jansen, W.J., Nahuis, N.J., 2003. The stock market and consumer confidence: European evidence. Econom. Lett. 79 (1), 89–98.
Jensen, M.C., 1978. Some anomalous evidence regarding market efficiency. J. Finan. Econ. 6 (2–3), 95–101.
Jiang, F.W., Lee, J., Martin, X.M., Zhou, G.F., 2019. Manager sentiment and stock returns. J. Finan. Econ. 132 (1), 126–149.
Joseph, K., Wintoki, M.B., Zhang, Z.L., 2011. Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. Int. J. Forecast. 27 (4), 1116–1127.
Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educ. Psychol. Meas. 20 (1), 141–151.
Kaplanski, G., Levy, H., 2010. Sentiment and stock prices: The case of aviation disasters. J. Finan. Econ. 95 (2), 174–201.
Kim, S.-H., Kim, D., 2014. Investor sentiment from internet message postings and the predictability of stock returns. J. Econ. Behav. Organ. 107, 708–729.
Lee, C.M.C., Shleifer, A., Thaler, R.H., 1991. Investor sentiment and the closed-end fund puzzle. J. Finance 46 (1), 75–109.
Lemmon, M., Portniaguina, E., 2006. Consumer confidence and asset prices: Some empirical evidence. Rev. Financ. Stud. 19 (4), 1499–1529.
Li, J., 2015. The asymmetric effects of investor sentiment and monetary policy on stock prices. Appl. Econ. 47 (24), 2514–2522.
Lutz, C., 2016. The asymmetric effects of investor sentiment. Macroecon. Dynam. 20 (6), 1477–1503.
Merton, R.C., 1980. On estimating the expected return on the market - An exploratory investigation. J. Finan. Econ. 8 (4), 323–361.
Morck, R., Shleifer, A., Vishny, R.W., Shapiro, M., Poterba, J.M., 1990. The stock market and investment: Is the market a sideshow? Brookings Pap. Econ. Act. 1990 (2).
Neal, R., Wheatley, S.M., 1998. Do measures of investor sentiment predict returns?. J. Finan. Quant. Anal. 33 (4), 523–547.
Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns - A new approach. Econometrica 59 (2), 347–370.
Otoo, M.W., 1999. Consumer Sentiment and the Stock Market (60). Federal Reserve Board of Governors, [Online]. Available at: https://www.federalreserve.gov/pubs/feds/1999/199960/199960pap.pdf.
Pástor, L̆., Sinha, M., Swaminathan, B., 2008. Estimating the intertemporal risk-return tradeoff using the implied cost of capital. J. Finance 63 (6), 2859–2897.
Pearson, E.S., Tukey, J.W., 1965. Approximate means and standard deviations based on distances between percentage points of frequency curves. Biometrika 52 (3/4), 523–546.
Püttmann, L., 2018. Patterns of panic: Financial crisis language in historical newspapers. SSRN Electron. J..
Ryu, D., Kim, H., Yang, H., 2016. Investor sentiment, trading behavior and stock returns. Appl. Econ. Lett. 24 (12), 826–830.
Savor, P., Wilson, M., 2014. Asset pricing: A tale of two days. J. Finan. Econ. 113 (2), 171–201.
Shen, J., Yu, J., Zhao, S., 2017. Investor sentiment and economic forces. J. Monet. Econ. 86, 1–21.
Shiller, R.J., 1984. Stock prices and social dynamics. Brookings Pap. Econ. Act. 1984 (2).
Shleifer, A., Summers, L.H., 1990. The noise trader approach to finance. J. Econ. Perspect. 4 (2), 19–33.
Stambaugh, R.F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. J. Finan. Econ. 104 (2), 288–302.
Sun, L., Najand, M., Shen, J., 2016. Stock return predictability and investor sentiment: A high-frequency perspective. J. Bank. Finance 73 (C), 147–164.
Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. J. Finance 62 (3), 1139–1168.
Wang, W.Z., 2018. The mean–variance relation and the role of institutional investor sentiment. Econom. Lett. 168, 61–64.
Wang, W.Z., Duxbury, D., 2021. Institutional investor sentiment and the mean–variance relationship: Global evidence. J. Econ. Behav. Organ. 191, 415–441.

Wang, Y.-H., Keswani, A., Taylor, S.J., 2006. The relationships between sentiment, returns and volatility. Int. J. Forecast. 22 (1), 109–123.
Wang, H.J., Yan, J.H., Yu, J.F., 2017. Reference-dependent preferences and the risk-return trade-off. J. Finan. Econ. 123 (2), 395–414.
Yang, C., Zhou, L., 2016. Individual stock crowded trades, individual stock investor sentiment and excess returns. N. Amer. J. Econ. Finance 38, 39–53.
Yu, J., Yuan, Y., 2011. Investor sentiment and the mean–variance relation. J. Finan. Econ. 100 (2), 367–381.
Zhang, W., Wang, P.F., Li, X., Shen, D.H., 2018. Twitter's daily happiness sentiment and international stock returns: Evidence from linear and nonlinear causality tests. J. Behav. Exp. Financ. 18, 50–53.