Volume 19, issue 2, March 2012
ISSN 0927-5398

ELSEVIER

# Journal of EMPIRICAL FINANCE



CONTENTS



Available online at www.sciencedirect.com
SciVerse ScienceDirect

www.elsevier.com/locate/jempfin
Access point to full text and abstracts of 80 journals in economics, econometrics and finance

Journal of Empirical Finance 19 (2012) 217–240

Contents lists available at SciVerse ScienceDirect

Journal of Empirical Finance

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

# When does investor sentiment predict stock returns?

San-Lin Chung [a,1], Chi-Hsiou Hung [b,2], Chung-Ying Yeh [c,*]

[a] *Department of Finance, National Taiwan University, 1, Section 4, Roosevelt Road, Taipei 106, Taiwan*
[b] *Durham Business School, Durham University, Mill Hill Lane, Durham, DH1 3LB, UK*
[c] *Department of Finance, National Chung Hsing University, 250, Kuo-Kuang Road, Taichung 402, Taiwan*

ARTICLE INFO

*Article history:*
Received 13 November 2010
Received in revised form 28 December 2011
Accepted 19 January 2012
Available online 27 January 2012

*JEL classification:*
E32
G11
G12
G14

*Keywords:*
Investor sentiment
Return predictability
Multivariate Markov-switching model
Bootstrap

ABSTRACT

We examine the asymmetry in the predictive power of investor sentiment in the cross-section of stock returns across economic expansion and recession states. We test the implication of behavioral theories and evidence that the return predictability of sentiment should be most pronounced in an expansion state when investors' optimism increases. We segregate economic states according to the NBER business cycles and further implement a multivariate Markov-switching model to capture the unobservable dynamics of the changes in the economic regime. The evidence suggests that only in the expansion state does sentiment perform both in-sample and out-of-sample predictive power for the returns of portfolio formed on size, book-to-market equity ratio, dividend yield, earnings-to-price ratio, age, return volatility, asset tangibility, growth opportunities, and 11 widely documented anomalies. In a recession state, however, the predictive power of sentiment is generally insignificant.



## 1. Introduction

Behavioral theories posit that investors may form erroneous stochastic beliefs, either with excessive optimism or pessimism, and therefore incorrectly evaluate asset values, causing asset prices to deviate from their intrinsic values (see, e.g., De Long et al. (1990), Lee et al. (1991), and Kumar and Lee (2006)). The mispricing gets corrected as the economic fundamentals are revealed and sentiment wanes. The pricing correction results in a negative relation between investor sentiment and future stock returns. As a consequence, investor sentiment exhibits predictive power for stock returns. Lemmon and Portniaguina (2006) show that investor sentiment can predict the returns on small size stocks. Baker and Wurgler (2006) present evidence that the pattern of the return predictive effect of sentiment varies with stock characteristics such as firm size, volatility, and age.

Recently, Stambaugh et al. (in press) document that the long-short and the short-legs of the anomaly strategies are more profitable in months following high sentiment, while the long-legs of the strategies have similar returns following high and low sentiment. Yu and Yuan (2011) find that the relation between the expected return and volatility of the U.S. stock market hinges on investor sentiment. Ho and Hung (2009) show that incorporating investor sentiment in modeling the dynamics of risk exposures enhances the explanatory power of asset pricing models for stock returns. Brown and Cliff (2004) document a contemporaneous

* Corresponding author. Tel.: +886 4 2285 5308.
*E-mail addresses:* chungs@management.ntu.edu.tw (S.-L. Chung), d.c.hung@durham.ac.uk (C.-H. Hung), cyyeh1@dragon.nchu.edu.tw (C.-Y. Yeh).
[1] Tel.: +886 2 3366 1084.
[2] Tel.: +44 191 334 5498.

0927-5398/$ – see front matter 
doi:10.1016/j.jempfin.2012.01.002

relation between changes in investor sentiment and U.S. stock market returns. Schmeling (2009) reports that when consumer confidence is high, future stock returns tend to be lower in most of the 18 industrialized countries.

This paper contributes to the literature by studying, across different states of the economy, cross-sectional predictability patterns of investor sentiment in stock returns. Our goal is to capture the asymmetry in the predictive power of investor sentiment in stock returns in times of flourishing economic environments when investors are becoming more optimistic and in times of economic downturns when investors become more pessimistic.

We separate the state of the economy into expansion and contraction regimes according to the business cycles designated by the NBER. We further segregate the economic regimes using a two-state Markov-switching model of stock returns with a time-varying state transition matrix (see, also, Perez-Quiros and Timmermann (2000), and Ozoguz (2009)).[3] We remove the potential impact of shifts in sentiment on regime shifts in the process of identifying regimes. [4] Specifically, we orthogonalize the monthly returns of Fama and French (1993) market, SMB, and HML portfolios from the variation in sentiment, and then use the residuals of the portfolio returns to estimate the parameters of the Markov-switching model.

Next, we control for the state of the regime to examine the return predictive power of sentiment. We treat the state of the regime as an exogenous input and estimate predictive regressions using the regime-sorted data (regimes identified by the NBER and the Markov-switching model). To completely control for the effect of regime shifts, we eliminate the observations at the turning points where the regime switches from one state to another. We test whether the state of the regime affects the significance of the regression coefficient on sentiment in the model.

We employ the monthly orthogonalized sentiment index of Baker and Wurgler (2006) as a proxy for investor sentiment[5] to predict monthly returns of the equal-weighted portfolios that are long and short, respectively, in stocks with high and low values of firm characteristics. We consider a wide range of firm characteristics including size, book-to-market ratio, dividend yield, earnings-to-price ratio, age, return volatility, R&D expense-to-assets ratio, fixed assets, sales growth, and external finance-to-assets. In the predictive regressions we follow Stambaugh et al. (in press) to include control variables that may affect stock returns including firm size, and book-to-market ratio.[6]

Our results shed light on the relation between the state of the economy and the predictive ability of sentiment in stock returns. The most striking finding of the paper is that only in an economic expansion state does investor sentiment show a significant and robust predictive power in stock returns. In contrast, when the economy is in contraction, the return predictive ability of sentiment is generally insignificant. Furthermore, the return predictive power of sentiment is not only regime dependent but also exhibits a cross-sectional pattern. When the economy is in expansion, higher sentiment is associated with lower subsequent stocks returns of firms with small size, young age, low book-to-market ratio, high return volatility, non-earnings, non-dividend-paying status, high intangible assets, and high growth opportunities.[7]

Note that in contrast to Baker and Wurgler (2006) who use the annual orthogonalized sentiment index to predict monthly stock returns, in this study we utilize their monthly sentiment index. Using the annual sentiment measure to predict monthly stock returns may not timely reflect the variation in sentiment, and may also inappropriately fit into the regime-switching framework which models random shifts in the economic regime.

Our study is therefore able to reveal the return predictive ability of sentiment in stock returns over short-term horizons. For example, in the expansion state as classified by the NBER, one unit increase in the orthogonalized sentiment index (which equals one standard deviation increase, because the indexes are standardized; see, Baker and Wurgler (2006)) in one month is associated with $+1.96\%$ of the return in the next month on the portfolio that is long old age stocks and short young stocks, after controlling for other determinants of stock returns.

In order to detect the source of the predictive power of sentiment, we apply a conditional beta model to examine whether the predictive effect of sentiment is attributable to time-varying systematic risk or mispricing. We find that the predictability patterns reflect mispricing and the subsequent pricing correction, rather than time-variation in the market beta, even after conditioning on the state of the economy.

Finally, we implement several analyses for robustness checks. First of all, we further elaborate on the asymmetry in the predictive pattern using the Michigan Consumer Confidence Index which has been used in prior research to study the relation between investor sentiment and stock returns (see, Lemmon and Portniaguina (2006), among others). Secondly, we also conduct a predictive regression with regime dummy variables. This predictive regression includes regime dummies independently in

[3] Many studies have demonstrated that the Markov-switching model provides a rather flexible filter to extract the latent regime from observed time-series data and is rather useful to characterize the evolution of regime shifts related to the business cycle (see, e.g., Whitelaw (2000), Ang and Bekaert (2002), Ang and Chen (2002), and Guidolin and Timmermann (2008)). Moreover, the results from using the Markov-switching model are helpful to confirm the main results from using the NBER index, and thus enhance the robustness of our findings.

[4] The reason for doing so is to mitigate the concern that, in addition to changing fundamentals, regime shifts may be also due to changes in investor sentiment.

[5] Baker and Wurgler (2006) construct the orthogonalized sentiment index by regressing each of their six raw sentiment proxies on macroeconomic variables and then obtaining the first principal component of the regression residuals. The descriptions of the Baker and Wurgler's (2006) orthogonalized sentiment index data are given in Section *3.1*. The sentiment index data are available for both annual and monthly frequencies from Jeffrey Wurgler's website: http://pages.stern.nyu.edu/~jwurgler/.

[6] Our conclusions remain unchanged when we add extra liquidity and momentum factors. To confirm the robustness of our findings, we also compute the bootstrapped *p*-value based on the procedure of Kosowski et al. (2006) for testing. Our conclusions continue to hold.

[7] As suggested by the associate editor, we further conduct tests using value-weighted portfolios. The results (available upon request) are qualitatively similar and do not change our conclusions. It is worth to note that the return predictability of investor sentiment is generally weaker using value-weighted portfolios. This is consistent with the argument of Baker and Wurgler (2006) that large firms are less affected by sentiment, and thus value-weighted returns may obscure the predictive effect of sentiment.

addition to making sentiment loadings conditional on dummies. Thirdly, we further control for macroeconomic variables, including the yield spread, default premium, dividend-to-price ratio, the growth rate of industrial production, and the growth rate of personal consumption expenditures in durables, nondurables and services, in the predictive regressions. Fourthly, for comprehensive coverage on the cross-section of stock returns, we also examine 11 asset pricing anomalies that firmly survive even after adjusting for exposures to the three factors of Fama and French (1993). Overall, the results confirm our main findings that the return predictive power of investor sentiment on these anomalies is significant only in an economic expansion state, regardless of using either the NBER classification or the regime-switching framework to identify the state of the economy.[8]

The evidence in this paper suggests the two-regime pattern of the return predictive ability of sentiment. It is important to note that our analysis does not completely rule out other sources that could drive the two-regime pattern. Stambaugh et al. (in press), for example, give an insightful discussion on the possible ways of time-variation in the cross-sectional dispersion of investors' views. Investigating the statistical properties over time of the cross-sectional distribution of investors' views could be a fruitful path for future research.

The rest of the paper is organized as follows. We develop our hypothesis in Section 2. Section 3 presents the multivariate Markov-switching model and characterizes the economy regimes. Section 4 describes the stock data and sentiment measures, tests our hypothesis and reports the findings. Section 5 conducts robustness checks, and further analyzes 11 well-documented asset pricing anomalies. Section 6 concludes. The Appendix contains the details of the bootstrap procedure.

## 2. Hypothesis development

The hypothesis in this paper is based on the following arguments. First, prior studies have shown that overpricing occurs in good times and underpricing appears in bad times. For example, Daniel et al. (1998) show that a string of good (bad) news related to the economy or firms leads to overpricing (underpricing). Gervais and Odean (2001) argue that agents attribute success heavily to superior ability rather than luck. Since most investors are long in stocks, in times when the market gains, the aggregate overconfidence becomes higher, leading to more aggressive trading. Further, these effects may rise late in a bull market and attract more investment capital, pushing prices even higher.

Indeed, our evidence (detailed in Section 3.3 and reported in Table 2) is consistent with the findings in the literature, and suggests that when the economy is in expansion, investors' optimism grows as reflected by the increase in sentiment. In contrast, investor sentiment tends to decrease when the economy is in contraction. As argued by Brown and Cliff (2005), when investor sentiment increases with the market price, the build-up of optimism leads to an extended period of market overvaluation. In contrast, investors' growing pessimistic beliefs in bad times may result in assets being underpriced.[9]

Second, empirical evidence and behavioral models demonstrate that the sentiment-driven overpricing is more prevalent than underpricing due to the limits of arbitrage and short sales constraints. De Long et al. (1990) analyze the limits of arbitrage, and demonstrate that arbitrageurs not only bear fundamental risk but also face the noise trader risk. Arbitrageurs' positions are deterred by the additional risk that investors' optimism could become more extreme in the near future, and that stock prices could increase even more significantly. Shleifer and Vishny (1997) show that investors may withdraw capital from institutional arbitrageurs when it is most needed. Further, arbitrageurs also face the financing risk of meeting margin calls (Mitchell et al. (2002)).

Importantly, short sales constraints keep negative opinions off the market, and thus allow substantial overpricing (Ofek et al. (2004) and Chang et al. (2007)).[10] Jones and Lamont (2002), for example, show that the level of shares sold short in the U.S. market is rather low.[11] Nagel (2005) finds that most sophisticated professional investors never sell short and therefore cannot trade against overpricing. Moreover, some studies report that market efficiency is weakened during the booming or bubble periods. For example, Lamont and Thaler (2003) identify violations of the law of one price during the rapid rise of technology stocks prices. Ofek and Richardson (2003) also provide evidence against market efficiency during the booming period in the late 1990s.

In contrast, executing buying trades is straightforward when pessimistic investors depress stock prices below fundamental values. The long-only institutional investors such as most mutual funds can increase holdings of underpriced stocks. Furthermore, in response to drastic declines in stock prices or in times of crisis, regulators often place severe restrictions on short sales and discourage securities lending (Lamont (2005)), thereby preventing large negative changes in the market price (Bris et al. (2007)).

The above evidence, taken together, suggests that the sentiment-driven overpricing, which most likely occurs in an economic expansion state, is relatively difficult to be arbitraged away than underpricing which probably appears in an economic contraction state. The extent of pricing deviations from fundamentals, and hence the correction for mispricing, is greater during economic expansions than during contractions. The immediate implication, and our hypothesis, is that the return predictive ability of investor

[8] In unreported results (available upon request) we further use two Markov-switching models with one adding the growth rate of industrial production to the returns on market, SMB, and HML portfolios, and the other using the industrial production growth rate, and the growth rate of personal consumption expenditures in durables, nondurables and services. The overall results are consistent with our main findings.

[9] Baker and Wurgler (2006 and 2007) also demonstrate that as investors value assets subjectively, stocks of firms that are hard to value are most likely to be affected by shifts in sentiment, and hence mispriced.

[10] The constraints for taking short positions include the risks, costs, legal and institutional restrictions, and the need of sufficient stock supply from investors who are willing to lend. See Lamont (2005) for detailed discussions.

[11] For the U.S. stock market, Figlewski and Webb (1993) also report that, on average, only 0.2% of shares outstanding was sold short for the 1973–1983 period. Dechow et al. (2001) show that those stocks having short interest greater than 5% of shares outstanding account for less than 2% of all stocks for the 1976–1993 period.

sentiment may exhibit an asymmetric pattern across different states of the economy, and should be most pronounced in a flourishing economic state.

The closest research related to ours is Stambaugh et al. (in press) who hypothesize that anomalies in the cross-section of stock returns may reflect mispricing in the presence of sentiment effects and short-sale impediments, where overpricing is more prevalent when market-wide sentiment is high than underpricing when sentiment is low. They classify a high (low)-sentiment month as the one in which the sentiment index value in the previous month is above (below) the sample median. They conclude that overpricing happens in high-sentiment periods.

Our study differs from theirs in our hypothesis and test designs. We consider overpricing to be more likely to occur when the economy is in expansion (either classified by the NBER or by the Markov-switching model) because of investors' growing optimism during these periods and the presence of the limits of arbitrage and short sales constraints. In contrast, although underpricing might happen when the economy is in contraction, it is less likely. We then separately examine the return predictive effect of sentiment during economic expansionanry periods and contractionary periods. In our sample, during economic expansions the sentiment index has a lower average value but is generally increasing; whereas during economic contractions the sentiment index has a higher average value but is generally decreasing. This pattern is consistent with the evidence that investor sentiment is mean-reverting (see, Baker and Wurgler (2006), and Yu, and Yuan (2011)).

Intuitively, during good times investors are likely to face the arrival of a series of good news pointing toward more prosperous economy outlook and better prospects of future cash flows of companies, while during bad times news of economic downturns or disappointing information related to firms often comes into light. Thus, it is natural to observe the increase in investor sentiment over the period of economic expansions, and vice versa. The wave of increasing sentiment during economic expansions results in overpricing, and such an effect, as argued by Baker and Wurgler (2006), is particularly significant on stocks that are difficult to arbitrage and hard to value.

## 3. Economic regimes and investor sentiment

### 3.1. NBER business cycles and a multivariate Markov-switching model

We first use the NBER business cycles to classify the state of the economy. We also characterize economic regimes by estimating a Markov-switching model with time-varying regime transition probabilities. The state of the regime switches at random times but is driven by a latent regime variable following a Markov chain that is assumed to change over time. Since recent research documents that size and value premiums vary with the state of economic regime (Perez-Quiros and Timmermann (2000) and Gulen et al. (2008), among others), we characterize economic regimes in the joint process of portfolio returns on the market, size, and value factors of Fama and French (1993) as in Guidolin and Timmermann (2008). These factors are the monthly returns on the CRSP value-weighted market index in excess of the one-month T-bill rate and the monthly returns on the SMB and HML factors.[12] Including the size and value factors has the advantage of utilizing more information to characterize regimes than that contained in the excess market return and thereby reduces the noise of the smoothed probabilities.

We remove the sentiment variation from the factor portfolio returns prior to the estimation of the Markov-switching model. Specifically, we regress the factor portfolio returns on the sentiment proxy—the Baker and Wurgler's (2006) orthogonalized sentiment index. The regressions residuals, labeled with a superscript $\perp$, are the factor portfolio returns orthogonalized to the sentiment variation. The sample of orthogonalized factor portfolios covers the 504-month period from January 1966 to December 2007, which is dictated by the availability of the sentiment index. The mean returns of the unorthogonalized market, SMB, and HML portfolios are, respectively, around 0.4%, 0.2%, and 0.5% per month with volatility of 4.5%, 3.3%, and 2.9% per month. The portfolio returns are all skewed and leptokurtic. The orthogonalized factors retain the appealing properties (such as skew and leptokurtosis) of the unorthogonalized ones.

We model the joint distribution of the vector of orthogonalized returns of the 3 factor portfolios, $\mathbf{r}_t^{\perp}$, as a multivariate Markov-switching process driven by a common discrete regime variable $s_t$ which takes two integer values {1,2} as:

$$\mathbf{r}_t^{\perp} = \mu_{s_t} + \Phi_{s_t}\mathbf{X}_{t-1} + \varepsilon_t, \tag{1}$$

where $\mathbf{X}_{t-1}$ is the vector of publicly available information for predicting stock returns, $\mu_{s_t}$ is the $3\times1$ vector of the regime-dependent intercepts, $\Phi_{s_t}$ is the $3\times3$ matrix of the regime-dependent coefficients. The vector of return innovations $\varepsilon_t \sim N(0,\Omega_{s_t})$ is assumed to follow a multivariate normal distribution with zero means and a regime-dependent variance–covariance matrix $\Omega_{s_t}$. The discrete regime variable $s_t$ is assumed to follow a two-state first-order Markov chain governed by a $2\times2$ transition probability matrix with time-varying elements in which

$$\begin{aligned} p_{11,t} &= \Pr(s_t = 1 | s_{t-1} = 1, \Delta\mathrm{CLI}_{t-1}) = N(a_1 + b_1\Delta\mathrm{CLI}_{t-1}), \\ p_{22,t} &= \Pr(s_t = 2 | s_{t-1} = 2, \Delta\mathrm{CLI}_{t-1}) = N(a_2 + b_2\Delta\mathrm{CLI}_{t-1}), \end{aligned} \tag{2}$$

where $\Delta\mathrm{CLI}_{t-1}$ is the one-month lagged value of the change in log composite leading indicator and $N(\cdot)$ is the cumulative density function of a standard normal variable.[13] The regime variable $s_t$, as a latent variable, can be statistically inferred by the realized

[12] These factors can be downloaded from Kenneth French's data library at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french.

[13] The composite leading indicator data are taken from the OECD's web-page at: http://stats.oecd.org/Index.aspx?DatasetCode=MEI_CLI.

**Table 1**
Parameter estimates of the Markov-switching model for the orthogonalized market, SMB, and HML returns.

| Parameters | Market | SMB | HML |
|---|---|---|---|
| *Panel A: Mean parameters* | | | |
| Constant, Regime 1 | −0.031 | 0.014 | 0.017 |
| Constant, Regime 2 | −0.002 | −0.015 | −0.001 |
| Default premium($Def_{t-1}$), Regime 1 | 38.181** | 12.556 | −2.391 |
| Default premium($Def_{t-1}$), Regime 2 | 20.607** | 9.826 | −3.195 |
| Interest rate($I_{t-1}$), Regime 1 | −7.202** | −2.576 | 2.242 |
| Interest rate($I_{t-1}$), Regime 2 | −2.787** | −2.658** | 0.020 |
| Dividend yield($Div_{t-1}$), Regime 1 | 0.797 | −0.466 | −0.551 |
| Dividend yield($Div_{t-1}$), Regime 2 | 0.016 | 0.620** | 0.031 |
| *Panel B: Correlations/Volatilities* | | | |
| Regime 1 | | | |
| Market | 0.068** | | |
| SMB | 0.282** | 0.055** | |
| HML | −0.358** | −0.295** | 0.047** |
| Regime 2 | | | |
| Market | 0.035** | | |
| SMB | 0.312** | 0.023** | |
| HML | −0.387** | −0.213* | 0.022** |

Panel C: Transition matrix parameters

| | Regime 1 | Regime 2 |
|---|---|---|
| Constant | 0.505 | 1.502** |
| Leading indicator($\Delta CLI_{t-1}$) | −54.464** | 19.955** |

Note: This table reports the results of the parameter estimates for the Markov-switching model:

$$\mathbf{r}_t^{\perp} = \mu_{s_t} + \Phi_{s_t}\mathbf{X}_{t-1} + \varepsilon_t,$$

where $\mu_{s_t}$ is the 3×1 intercept vector in regime $s_t$, $\Phi_{s_t}$ is the 3×3 regime-dependent coefficients, $\mathbf{X}_{t-1}$ is the vector of dividend yield ($Div_{t-1}$), default premium ($Def_{t-1}$) and interest rate ($I_{t-1}$), and $\varepsilon_t \sim \mathcal{N}(0, \Omega_{s_t})$ is the 3×1 innovation vector of returns. $s_t$ is an unobserved state variable driven by a two-state first-order Markov chain governed by a 2×2 transition probability matrix with time-varying elements

$$\Pr(s_t = i | s_{t-1} = i, \Delta CLI_{t-1}) = N(a_i + b_i \Delta CLI_{t-1}), i = 1, 2,$$

where $\Delta CLI_{t-1}$ is the one-month lagged value of the change in log composite leading indicator and $N(\cdot)$ is the cumulative density function of a standard normal variable. The three series are excess returns on the value-weighted market portfolio and returns on Fama and French's (1993) SMB and HML portfolios that have been orthogonalized to sentiment. The sample period is from January 1966 to December 2007. Values reported on the diagonals of the correlation matrices are volatilities. All estimates are monthly. * and ** denote significance at 5% and 1% levels, respectively.

observations. The possibilities of the regimes at each time point can be characterized by filtered probabilities $\Pr(s_t|\mathbf{Y}^t)$ and smoothed probabilities $\Pr(s_t|\mathbf{Y}^T)$, where $\mathbf{Y}^t$ is the information set at time $t$ and $\mathbf{Y}^T$ is the complete information set.

The vector of $\mathbf{X}_{t-1}$ comprises dividend yield, default premium, and the short-term interest rate, following Perez-Quiros and Timmermann (2000) and Ozoguz (2009). The dividend yield is defined as the dividends on the valued-weighted CRSP index over the past 12 months. We construct the dividend payout series using the value-weighted return including dividends, and the price index series associated with the value-weighted return excluding dividends. The dividend series is the sum of dividend payout over the past 12 months. The default premium is the yield spread between Baa and Aaa corporate bonds. The short-term interest rate is defined as the 90-day T-bill rate. We obtain the Baa and Aaa corporate bond yields and the short-term interest rate data from the web-page of the Federal Reserve at St. Louis.[14]

### 3.2. Empirical results of the multivariate Markov-switching model

Denote the set of parameters by $\Theta = (\mu_1, \mu_2, \Phi_1, \Phi_2, \Omega_1, \Omega_2, a_1, a_2, b_1, b_2)$. Because the number of parameters is large and will cause severe problems in numerical optimization, we are not able to estimate the Markov-switching model of Eqs. (1) and (2) by maximum likelihood method directly. We adopt a pragmatic approach instead. We divide the set of parameters into two subsets: $\Theta_1 = (\mu_1, \mu_2, \Phi_1, \Phi_2, \Omega_1, \Omega_2)$ and $\Theta_2 = (a_1, a_2, b_1, b_2)$. Given an initial value of $\Theta_1^{(0)}$, we use the maximum likelihood estimation to obtain the estimates of $\Theta_2^{(1)}$. We then take these estimates as inputs and use the EM algorithm of Hamilton (1990) to estimate $\Theta_1^{(1)}$. We repeat these procedures iteratively until we achieve convergence. We compute the standard deviations of the parameter estimates following the standard convention of the maximum likelihood approach. Table 1 reports the results from the estimation of the two-state Markov-switching model. Fig. 1 plots the historical patterns of the smoothed probabilities for two regimes.

[14] The web-page is at: http://research.stlouisfed.org/fred2/categories/22/.

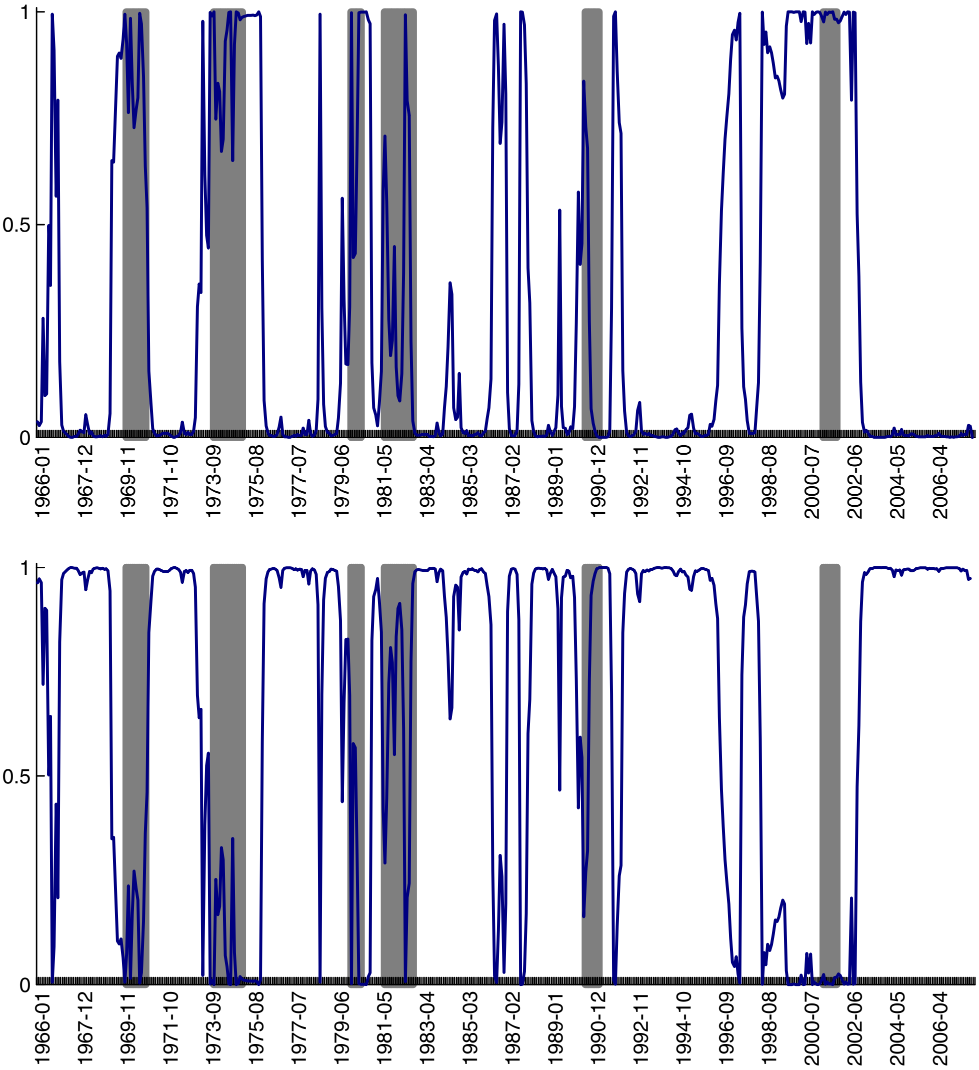


**Fig. 1.** The smoothed probabilities of the two-state Markov-switching model for the orthogonalized market, SMB, and HML returnsThis figure plots the smoothed probabilities for the two-state Markov-Switching model comprising monthly excess returns on the value-weighted market portfolio and return series on Fama and French's (1993) SMB and HML portfolios that have been orthogonalized to sentiment variation. The upper (lower) panel displays the smoothed probabilities of Regime 1 (Regime 2). The sample period is from January 1966 to December 2007. Parameters estimates underlying these plots are reported in Table 1. Regime 1 is a high-volatility recession state that captures episodes of sharp declines in stock prices since 1960, such as the two oil shocks in the 1970s, the Gulf War in the beginning of the 1990s, the default of the Russian sovereign bonds and the collapse of Long Term Capital Management surrounding the crash in 1998, the internet bubble burst and corporate malfeasance in the beginning of the 2000s. Most of the periods classified as regime 1 occur in the NBER recessions (the shaded areas). Regime 2 is a low-volatility expansion state that covers most of the bull markets with growing stock prices since the mid-1960s, including the run-ups in the 1980s and 1990s.

The decision criterion for inferring the state of the regime at each time point is that the regime has smoothed probability above 0.5.[15] This decision rule is reasonable because very few of the smoothed probabilities as shown in Fig. 1 lie between 0.3 and 0.7. We identify regime 1 as a high-volatility economic contraction state and regime 2 as a low-volatility economic expansion state. The upper panel of Fig. 1, which plots the historical patterns of the smoothed probabilities of regime 1, shows that this regime captures most of the NBER recessions[16] (the shaded areas) and episodes of sharp declines in stock prices since the 1960s. These include the 1969–1970 recession, the oil crisis in 1973, the stock market crashes in 1973–1974 and 1987, the Gulf War in the early 1990s, the default of the Russian sovereign bonds and the near collapse of Long Term Capital Management in 1998, the Internet bubble burst and corporate malfeasance in the beginning of the 2000s. In contrast, as displayed on the lower panel of Fig. 1, regime 2 covers most of the bull markets with growing stock prices since the mid-1960s and the run-ups in the 1980s and 1990s.[17]

[15] According to this criterion, there are 136 points classified as regime 1 and 326 points classified as regime 2. There are 42 points classified as regime turning points.

[16] The data is taken from NBER's web-page at: http://www.nber.org/cycles/cyclesmain.html.

[17] In unreported results and figures (available upon request), we also analyze whether the expansion (recession) periods characterized by Markov-switching model cover bull (bear) markets using the definitions of bull and bear markets of Hardouvelis and Theodossiou (2002). Indeed, the identified expansion (recession) regime covers most of the bull (bear) markets.

Panel C of Table 1 displays the parameter estimates of $a_1$, $a_2$, $b_1$, and $b_2$ associated with the transition probabilities in Eq. (2). The coefficient $b_1$ ($b_2$) on the change in the composite leading indicator is significantly negative (positive), indicating that an increase in the leading indicator decreases (increases) the probability of staying in regime 1 (regime 2). We also sort the growth rate of industrial production (obtained from the web-page of the Federal Reserve at St. Louis) based on the identified regimes. The results (available upon request) show that the average monthly growth rate of industrial production is 0.04% in regime 1 and 0.32% in regime 2. The volatility of the industrial production growth rate in regime 1 is higher than that in regime 2. These pieces of evidence, taken together, confirm the economic interpretation that the regimes characterized by the Markov-switching model are indeed associated with underlying economic fundamentals.

We proceed to the mean parameter estimates in Eq. (1). Panel A of Table 1 suggests that the coefficient estimates of the excess market return on the default premium are highly significant and positive in both regimes. The coefficient estimate of SMB on the default premium is significantly positive in regime 1. The magnitude of the coefficient estimates on the default premium in the contraction state is larger than that in the expansion state, suggesting that the default premium is more important during economic recessions or bearish markets and is particular relevant to the size premium. These findings are consistent with those of Perez-Quiros and Timmermann (2000). The coefficient estimates of the excess market return on the lagged interest rate are statistically significant and negative in both regimes. For the SMB factor portfolio, the coefficient on the lagged interest rate is significantly negative in regime 2.

Panel B of Table 1 presents the estimates of volatilities and correlations in the diagonals and the off-diagonals of the correlation matrices, respectively, showing that in regime 1 (regime 2) the monthly volatility of the excess market return is 6.8% (3.5%). This result is consistent with the finding reported by Schwert (1990) that stock market volatility tends to be high during economic recessions. The correlation between the market portfolio and the SMB portfolio is positive, while the HML portfolio is negatively correlated with both the market portfolio and the SMB portfolio, suggesting that HML may serve as a hedge against the market portfolio.

### 3.3. The link between economic regimes and investor sentiment

We find that the smoothed probability of the recession state (regime 1) moves in the same direction with the NBER recession indicator. The Spearman's correlation coefficient between the NBER recession indicator and the smoothed probability of regime 1 (as shown in Table 2) is about 0.35. The regression coefficient estimate of the smoothed probability of regime 1 on the NBER recession indicator gives a positive value of 0.33 and is significant at the 1% level.

We use the orthogonalized, monthly sentiment index of Baker and Wurgler (2006) as the measure of investor sentiment. The top panel of Fig. 2 plots the historical pattern of the sentiment index between January 1966 and December 2007. The sentiment index shows a spike before the 1970s and then turns into negative for a long period during the 1970s which might be attributable, in part, to a series of oil crises. The sentiment index became positive in the 80s until the Gulf War in the early 90s. Investor sentiment reached a spike before the Dot-Com bubble burst and then became negative afterwards until 2003 when the market recovered.

Table 2 shows that the correlation coefficient between the level of investor sentiment and NBER recessions is positive 0.11. Similarly, the correlation coefficient between the level of investor sentiment and the recession regime 1 identified by the two-state Markov-switching model is also positive 0.27. These indicate that the Baker and Wurgler (2006) index has a higher average value during recessions than that during expansions.

Notice that investor sentiment tends to drop when the NBER recession index (and the smoothed probabilities of recession regime 1) indicates a recession state. Investor sentiment exhibits an increasing pattern when the NBER index (and the smoothed probabilities of expansion regime 2) indicates an expansion state.

The NBER recession index (or the smoothed probabilities of recession regime 1) is negatively correlated with the change in sentiment, −0.20 (or −0.11). These negative correlations indicate that when the economy is in recession (expansion) investor

**Table 2**
Correlations between the NBER recession index, smoothed probability of Regime 1, sentiment, and change in sentiment.

| | Spearman's rank correlations | | | |
|---|---|---|---|---|
| | $NBER_t$ | $\Pr(s_t=1\|\mathbf{Y}^T)$ | $SENTIMENT_t^{\perp}$ | $\Delta SENTIMENT_t^{\perp}$ |
| $NBER_t$ | 1.00 | | | |
| | (−) | | | |
| $\Pr(s_t=1\|\mathbf{Y}^T)$ | 0.35** | 1.00 | | |
| | (0.00) | (−) | | |
| $SENTIMENT_t^{\perp}$ | 0.11** | 0.27** | 1.00 | |
| | (0.01) | (0.00) | (−) | |
| $\Delta SENTIMENT_t^{\perp}$ | −0.20** | −0.11** | 0.10** | 1.00 |
| | (0.00) | (0.02) | (0.02) | (−) |

Note: This table reports the Spearman's rank correlations matrix for the NBER recession index $NBER_t$, the smoothed probability of Regime 1 (a contraction state) $\Pr(s_t=1|\mathbf{Y}^T)$, the sentiment index $SENTIMENT^{\perp}$, and the change in the sentiment index $\Delta SENTIMENT^{\perp}$. The $p$-values are in parentheses and * and ** denote significance at 5% and 1% levels, respectively.

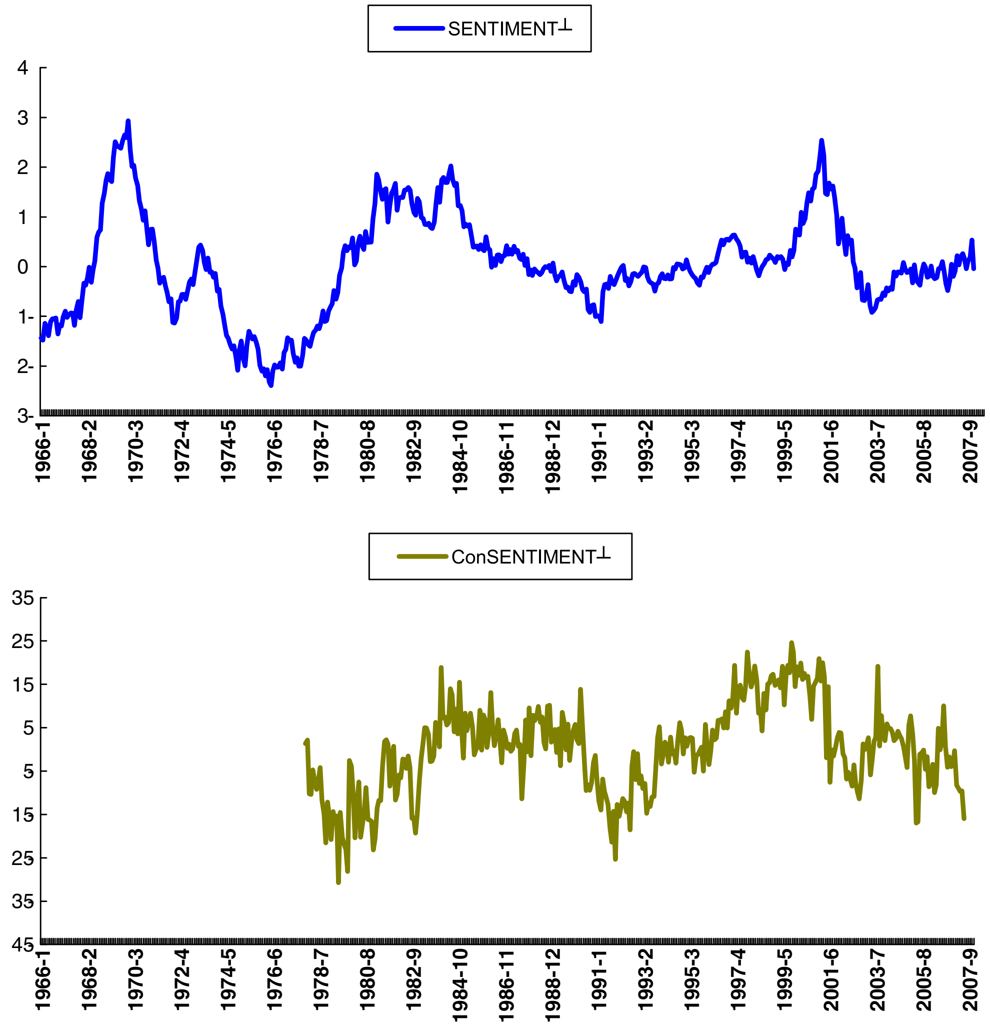


**Fig. 2.** The historical patterns of the sentiment proxiesThe upper panel plots the Baker and Wurgler's (2006) monthly orthogonalized sentiment index for the period from January 1966 to December 2007 and the lower panel displays an orthogonalized consumer sentiment measure of the University of Michigan consumer confidence index for the period from January 1978 to December 2007.

sentiment tends to decrease (increase). Thus, while during recessions investor sentiment is relatively high, on average, it exhibits a tendency to drop. On the other hand, during expansions investor sentiment is relatively low, on average, it exhibits a tendency to increase.[18]

## 4. Predictive regressions for long-short portfolios

### 4.1. Data and sample

In our predictive regressions, we use the equally weighted portfolios formed on firm characteristics: (i) size (*ME*), (ii) book-to-market ratio (*BE*/*ME*), (iii) dividend yield (*D*/*P*), (iv) earnings-to-price ratio (*E*/*P*), (v) firm age (*AGE*), (vi) return volatility (*SIGMA*), (vii) R&D expense-to-assets (*RD*/*A*), (viii) fixed assets (measured by property, plant and equipment over assets, *PPE*/*A*), (ix) sales growth (*GS*), and (x) external finance-to-assets (*EF*/*A*).[19] *ME* is the market equity at the end of each June. *BE*/*ME* is book equity at the last fiscal year end of the prior calendar year $t-1$ divided by market equity at the end of December in year $t-1$. *D*/*P* is the total dividends paid from July of year $t-1$ to June of year $t$ divided by market equity in June of year $t$. *E*/*P* is earnings before extraordinary items at the last fiscal year end of the prior calendar year $t-1$ divided by market equity at the end of December in year $t-1$.*AGE* is measured by the period of time since firm founding and is computed based on Jay Ritter's historical founding dates data for 9089 IPOs in the U.S. during 1975–2009.[20] *SIGMA* is

[18] One of the referees suggests us to check for any causality relation between economic state and sentiment, such as lead-lag relation. We thus utilize a VAR model and conduct the Granger causality test. The results do not show a clear lead-lag or Granger causality relation between sentiment and economic regimes.

[19] The data of the portfolios formed on size, book-to-market ratio, dividend yield, and earnings-to-price ratio are from Kenneth French's data library.

[20] The age measure based on a firm's founding date captures the actual survival period of a firm and is widely used in the corporate finance literature. In contrast, the number of years since the firm's first appearance on CRSP mainly reflects the period of time since listing, but not precisely the age of a firm. Jay Ritter provides more detailed information on the data of founding dates and their applications at: http://bear.warrington.ufl.edu/ritter/ipodata.htm.

**Table 3**
Summary statistics for the portfolio returns.

| Decile | ≤0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Portfolios formed on size | | | | | | | | | | | |
| Mean | | 1.43 | 1.09 | 1.13 | 1.09 | 1.13 | 1.06 | 1.11 | 1.04 | 1.01 | 0.89 |
| Standard Deviation | | 6.72 | 6.46 | 6.23 | 6.01 | 5.80 | 5.48 | 5.31 | 5.18 | 4.76 | 4.58 |
| Skewness | | 0.33 | −0.06 | −0.19 | −0.32 | −0.35 | −0.38 | −0.29 | −0.30 | −0.22 | −0.22 |
| Excess kurtosis | | 2.83 | 2.83 | 2.40 | 2.26 | 2.51 | 1.94 | 2.23 | 1.68 | 1.56 | 1.87 |
| Portfolios formed on book-to-market | | | | | | | | | | | |
| Mean | | 0.65 | 1.00 | 1.11 | 1.24 | 1.28 | 1.40 | 1.48 | 1.51 | 1.65 | 1.86 |
| Standard Deviation | | 7.59 | 6.47 | 6.09 | 5.76 | 5.44 | 5.29 | 5.12 | 5.17 | 5.49 | 6.29 |
| Skewness | | 0.02 | −0.22 | −0.30 | −0.32 | −0.26 | −0.16 | −0.08 | 0.14 | 0.07 | 0.61 |
| Excess Kurtosis | | 2.78 | 1.86 | 2.51 | 3.02 | 3.53 | 3.62 | 3.91 | 4.27 | 4.42 | 4.96 |
| Portfolios formed on dividend yield | | | | | | | | | | | |
| Mean | 1.30 | 1.18 | 1.28 | 1.24 | 1.31 | 1.28 | 1.36 | 1.34 | 1.38 | 1.29 | 1.18 |
| Standard Deviation | 7.59 | 5.79 | 5.35 | 5.13 | 4.95 | 4.72 | 4.57 | 4.32 | 4.19 | 3.84 | 4.07 |
| Skewness | 0.20 | −0.53 | −0.52 | −0.36 | −0.49 | −0.39 | −0.48 | −0.42 | −0.22 | 0.10 | 1.07 |
| Excess kurtosis | 2.31 | 2.23 | 3.21 | 3.70 | 3.74 | 3.80 | 3.82 | 4.48 | 4.59 | 4.42 | 8.47 |
| Portfolios formed on earnings/price | | | | | | | | | | | |
| Mean | 1.34 | 0.95 | 1.16 | 1.15 | 1.21 | 1.27 | 1.31 | 1.36 | 1.42 | 1.56 | 1.67 |
| Standard Deviation | 8.62 | 6.89 | 5.95 | 5.56 | 5.24 | 5.09 | 4.95 | 4.80 | 4.76 | 4.97 | 5.74 |
| Skewness | 0.60 | −0.15 | −0.35 | −0.36 | −0.38 | −0.42 | −0.31 | −0.28 | 0.02 | 0.02 | 0.14 |
| Excess kurtosis | 3.51 | 1.91 | 2.33 | 3.00 | 4.26 | 4.20 | 4.48 | 4.34 | 4.92 | 4.77 | 4.45 |
| Portfolios formed on age | | | | | | | | | | | |
| Mean | | 0.63 | 1.22 | 1.06 | 1.24 | 1.43 | 1.59 | 1.59 | 1.27 | 1.24 | 1.27 |
| Standard Deviation | | 8.83 | 8.53 | 8.08 | 8.08 | 7.32 | 7.54 | 6.99 | 6.31 | 5.78 | 5.28 |
| Skewness | | 0.21 | 0.23 | 0.13 | 0.28 | −0.12 | 0.13 | 0.04 | −0.19 | −0.47 | −0.79 |
| Excess kurtosis | | 3.81 | 2.75 | 1.97 | 2.38 | 1.91 | 1.77 | 2.07 | 2.64 | 2.69 | 4.70 |
| Portfolios formed on sigma | | | | | | | | | | | |
| Mean | | 1.48 | 1.53 | 1.56 | 1.62 | 1.64 | 1.62 | 1.67 | 1.64 | 1.75 | 1.79 |
| Standard Deviation | | 3.14 | 3.79 | 4.30 | 4.88 | 5.33 | 6.05 | 6.81 | 7.43 | 8.53 | 9.76 |
| Skewness | | −0.56 | −0.59 | −0.76 | −0.50 | −0.49 | −0.26 | −0.03 | 0.09 | 0.37 | 0.60 |
| Excess kurtosis | | 9.07 | 6.56 | 6.75 | 5.93 | 5.05 | 4.49 | 3.65 | 3.53 | 3.59 | 4.56 |
| Portfolios formed on R&D expense/assets | | | | | | | | | | | |
| Mean | | 1.50 | 1.42 | 1.50 | 1.59 | 1.71 | 1.70 | 1.87 | 1.91 | 1.91 | 2.05 |
| Standard Deviation | | 5.82 | 5.97 | 5.54 | 5.41 | 6.29 | 6.86 | 7.66 | 8.30 | 8.95 | 10.40 |
| Skewness | | −0.09 | −0.12 | −0.18 | −0.47 | 0.00 | 0.03 | 0.28 | 0.22 | 0.50 | 0.88 |
| Excess Kurtosis | | 5.45 | 3.37 | 5.64 | 5.13 | 4.01 | 3.47 | 3.42 | 2.64 | 3.54 | 5.44 |
| Portfolios formed on fixed assets | | | | | | | | | | | |
| Mean | | 1.11 | 1.14 | 1.17 | 1.20 | 1.37 | 1.23 | 1.38 | 1.36 | 1.38 | 1.39 |
| Standard Deviation | | 6.99 | 7.30 | 7.29 | 7.23 | 5.69 | 7.13 | 5.55 | 5.36 | 5.11 | 4.87 |
| Skewness | | 0.01 | 0.04 | 0.04 | 0.03 | −0.29 | 0.01 | −0.30 | −0.37 | −0.41 | −0.44 |
| Excess kurtosis | | 2.46 | 2.36 | 2.47 | 2.49 | 3.30 | 2.51 | 3.51 | 3.57 | 3.85 | 3.60 |
| Portfolios formed on sales growth | | | | | | | | | | | |
| Mean | | 1.63 | 1.58 | 1.55 | 1.52 | 1.49 | 1.10 | 1.04 | 0.93 | 0.81 | 0.60 |
| Standard Deviation | | 7.99 | 7.08 | 6.57 | 6.21 | 6.01 | 6.70 | 7.04 | 7.40 | 7.94 | 8.60 |
| Skewness | | 0.45 | 0.26 | 0.17 | 0.05 | −0.03 | −0.19 | −0.15 | −0.08 | 0.03 | 0.09 |
| Excess kurtosis | | 2.78 | 2.68 | 3.05 | 3.18 | 3.25 | 2.56 | 2.51 | 2.51 | 2.58 | 2.62 |
| Portfolios formed on external finance/assets | | | | | | | | | | | |
| Mean | | 1.78 | 1.66 | 1.56 | 1.52 | 1.49 | 1.02 | 0.93 | 0.83 | 0.69 | 0.47 |
| Standard Deviation | | 6.83 | 6.22 | 5.72 | 5.55 | 5.47 | 6.42 | 6.73 | 7.17 | 7.81 | 8.86 |
| Skewness | | 0.20 | −0.01 | −0.15 | −0.19 | −0.22 | −0.15 | −0.08 | −0.02 | 0.07 | 0.22 |
| Excess kurtosis | | 3.19 | 3.20 | 3.34 | 3.55 | 3.61 | 2.79 | 2.77 | 2.79 | 2.92 | 3.33 |

Note: This table reports the summary statistics of the monthly equal-weighted portfolio returns formed on size, book-to-market, dividend yield, earnings/price, age, volatility, R&D expense/assets, fixed assets, sales growth, and external finance/assets. All portfolios are constructed at the end of each June. At the end of June in year $t$ we match all NYSE, AMEX, and NASDAQ stocks based on *ME*, *BE*/*ME*, *D*/*P*, *E*/*P*, *AGE*, *SIGMA*, *RD*/*A*, *PPE*/*A*, *GS*, and *EF*/*A*. *ME* is the June market equity of year $t$. *BE*/*ME* is book equity at the last fiscal year end of the prior calendar year $t-1$ divided by market equity at the end of December of the prior year $t-1$. *D*/*P* is the total dividends paid from July of the prior year $t-1$ to June of the present year $t$ divided by market equity at June of the present year $t$. *E*/*P* is earnings before extraordinary at the last fiscal year end of the prior calendar year $t-1$ divided by market equity at the end of December of the prior year $t-1$. *AGE* is measured by the period of time since firm founding and is computed based on Jay Ritter's historical founding dates data for 9089 firms going public in the U.S. during 1975–2009. *SIGMA* is the standard deviation of monthly returns over the 12 months ending in June of year $t$ and we only consider firms that have at least 10 returns to estimate sigma. *RD*/*A* is the R&D expense at the last fiscal year end of the prior calendar year $t-1$ divided by total assets at the end of December of the prior year $t-1$. *PPE*/*A* (fixed assets) is measured by property, plant expense divided by total assets at the end of December of the prior year $t-1$. *GS* (sales growth) is the change in net sales divided by prior-year net sales. *EF*/*A* is the change in assets minus the change in retained earnings divided by divided by total assets at the end of December of the prior year $t-1$. For size, book-to-market, age, sigma, R&D expense/assets, there are 10 portfolios corresponding to each decile. There are 11 portfolios for dividend yield and earnings/price in which "≤0" represents the portfolios for non-dividend-paying stocks and non-earning stocks. The values of portfolio returns are in terms of percentage. The sample period for the size, book-to-market, earnings, dividend, volatility, and fixed assets portfolios is from January 1966 to December 2007. The R&D, age, sales growth, and external finance portfolios begin by January 1975, January 1980, July 1967, and July 1967, respectively, and end by December 2007.

**Table 4**
Predictive regressions for long-short portfolio returns (with Newey–West p values).

| | | NBER recession index | | Markov-switching model | |
|---|---|---|---|---|---|
| Long-short | All | Rec. | Exp. | Regime 1 (Rec.) | Regime 2 (Exp.) |
| *Portfolios formed on size* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | 0.71$^{**}$ (0.01) | −0.06 (0.93) | 1.03$^{**}$ (0.00) | 0.13 (0.79) | 1.04$^{**}$ (0.00) |
| 5 − 1 | 0.39$^{*}$ (0.02) | −0.14 (0.71) | 0.55$^{**}$ (0.01) | 0.08 (0.76) | 0.51$^{*}$ (0.02) |
| Cont1rolling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | 0.34$^{**}$ (0.01) | 0.10 (0.58) | 0.51$^{**}$ (0.00) | 0.17 (0.46) | 0.36$^{**}$ (0.01) |
| 5 − 1 | 0.30$^{*}$ (0.02) | 0.00 (0.98) | 0.36$^{*}$ (0.05) | 0.19 (0.34) | 0.27 (0.12) |
| *Portfolios formed on book-to-market* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | 0.26 (0.25) | 0.46 (0.48) | 0.19 (0.50) | 0.16 (0.73) | 0.44$^{*}$ (0.02) |
| 5 − 1 | 0.51$^{**}$ (0.01) | 0.58 (0.23) | 0.48 (0.07) | 0.30 (0.21) | 0.54$^{**}$ (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | 0.01 (0.97) | −0.12 (0.63) | 0.14 (0.25) | −0.22 (0.30) | 0.24$^{*}$ (0.04) |
| 5 − 1 | 0.23$^{*}$ (0.02) | 0.16 (0.43) | 0.33$^{**}$ (0.01) | 0.17 (0.37) | 0.31$^{**}$ (0.00) |
| *Portfolios formed on dividend yield* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − = 0 | 1.03$^{**}$ (0.00) | 1.07$^{**}$ (0.01) | 0.97$^{**}$ (0.01) | 0.82 (0.07) | 1.13$^{**}$ (0.00) |
| 5 − = 0 | 0.83$^{**}$ (0.00) | 0.62 (0.13) | 0.92$^{**}$ (0.00) | 0.72 (0.06) | 0.86$^{**}$ (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − = 0 | 0.51$^{**}$ (0.00) | 0.74$^{**}$ (0.01) | 0.51$^{**}$ (0.01) | 0.34 (0.14) | 0.47$^{*}$ (0.03) |
| 5 − = 0 | 0.44$^{**}$ (0.00) | 0.44 (0.06) | 0.56$^{**}$ (0.00) | 0.43 (0.06) | 0.40$^{**}$ (0.01) |
| *Portfolios formed on earnings/price* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − ≤0 | 0.86$^{**}$ (0.00) | 1.12$^{**}$ (0.01) | 0.82$^{**}$ (0.01) | 0.97$^{**}$ (0.01) | 0.77$^{**}$ (0.00) |
| 5 − ≤0 | 0.95$^{**}$ (0.00) | 0.87 (0.12) | 1.08$^{**}$ (0.00) | 0.88 (0.07) | 0.90$^{**}$ (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − ≤0 | 0.54$^{**}$ (0.00) | 0.95$^{**}$ (0.00) | 0.54$^{*}$ (0.02) | 0.68$^{**}$ (0.01) | 0.42 (0.06) |
| 5 − ≤0 | 0.57$^{**}$ (0.00) | 0.75 (0.08) | 0.70$^{*}$ (0.02) | 0.60 (0.10) | 0.47$^{*}$ (0.05) |
| *Portfolios formed on age* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | 2.03$^{**}$ (0.00) | 1.52 (0.15) | 2.57$^{**}$ (0.00) | 3.87$^{**}$ (0.01) | 1.42$^{**}$ (0.00) |
| 5 − 1 | 1.58$^{**}$ (0.00) | 1.36$^{*}$ (0.04) | 1.74$^{*}$ (0.02) | 3.18$^{**}$ (0.00) | 0.76$^{*}$ (0.05) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | 1.21$^{**}$ (0.00) | −0.44 (0.53) | 1.96$^{**}$ (0.00) | 1.00 (0.40) | 1.09$^{**}$ (0.00) |
| 5 − 1 | 1.26$^{**}$ (0.00) | 0.41 (0.58) | 1.56$^{*}$ (0.02) | 1.80 (0.13) | 0.78$^{*}$ (0.05) |
| *Portfolios formed on sigma* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | −1.77$^{**}$ (0.00) | −1.67 (0.25) | −1.91$^{**}$ (0.00) | −2.51$^{*}$ (0.03) | −1.54$^{**}$ (0.00) |
| 10 − 5 | −1.34$^{**}$ (0.00) | −1.52 (0.14) | −1.45$^{**}$ (0.00) | −2.01$^{*}$ (0.02) | −1.05$^{**}$ (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | −1.08$^{**}$ (0.00) | −0.54(0.60) | −1.34$^{**}$ (0.00) | −1.20 (0.13) | −0.92$^{**}$ (0.00) |
| 10 − 5 | −0.94$^{**}$ (0.00) | −0.38 (0.70) | −1.12$^{**}$ (0.00) | −1.14 (0.09) | −0.75$^{**}$ (0.01) |
| *Portfolios formed on R&D expense/assets* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | −0.61 (0.11) | −1.86 (0.12) | −0.34 (0.46) | −1.12 (0.20) | −0.58 (0.10) |
| 10 − 5 | −0.60$^{*}$ (0.04) | −0.63 (0.37) | −0.52 (0.16) | −0.82 (0.22) | −0.72$^{*}$ (0.03) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | −0.00 (1.00) | −0.63 (0.37) | 0.16 (0.69) | −0.11 (0.87) | −0.21 (0.40) |
| 10 − 5 | −0.13 (0.59) | −0.15 (0.80) | −0.08 (0.79) | −0.13 (0.79) | −0.33 (0.12) |
| *Portfolios formed on fixed assets* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | 0.42 (0.06) | 0.52 (0.33) | 0.33 (0.23) | 0.61 (0.16) | 0.23 (0.38) |
| 5 − 1 | 0.23 (0.11) | 0.29 (0.37) | 0.16 (0.34) | 0.33 (0.22) | 0.08 (0.64) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | 0.11 (0.45) | 0.31 (0.23) | 0.09 (0.65) | 0.33 (0.17) | −0.17 (0.26) |
| 5 − 1 | 0.08 (0.47) | 0.17 (0.40) | 0.05 (0.72) | 0.19 (0.31) | −0.12 (0.32) |
| *Portfolios formed on sales growth* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | −0.28 (0.16) | −0.22 (0.58) | −0.34 (0.21) | −0.26 (0.48) | −0.36$^{*}$ (0.02) |
| 10 − 5 | −0.72$^{**}$ (0.00) | −0.69 (0.13) | −0.77$^{**}$ (0.01) | −0.78 (0.08) | −0.71$^{**}$ (0.00) |

**Table 4** (*continued*)

| Long-short | All | NBER recession index | | Markov-switching model | |
|---|---|---|---|---|---|
| | | Rec. | Exp. | Regime 1 (Rec.) | Regime 2 (Exp.) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | −0.17 (0.25) | 0.11 (0.67) | −0.37 (0.06) | −0.02 (0.92) | −0.33* (0.02) |
| 10 − 5 | −0.44** (0.00) | −0.33 (0.23) | −0.63** (0.00) | −0.41 (0.09) | −0.47** (0.00) |
| *Portfolios formed on external finance/assets* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | −0.52* (0.02) | −0.59 (0.25) | −0.55 (0.08) | −0.70 (0.11) | −0.43** (0.00) |
| 10 − 5 | −0.85** (0.00) | −0.89 (0.12) | −0.93** (0.01) | −1.07* (0.04) | −0.70** (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | −0.28* (0.05) | −0.20 (0.52) | −0.45* (0.02) | −0.34 (0.18) | −0.31** (0.01) |
| 10 − 5 | −0.48** (0.00) | −0.49 (0.18) | −0.69** (0.00) | −0.62* (0.04) | −0.39** (0.00) |

Note: This table contains the results of (i) regressions of value-weighted long-short portfolio returns on the lagged SENTIMENT$^{\perp}$,

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1} + \varepsilon_{i,t},$$

and (ii) regressions of valued-weighted long-short portfolio returns on the lagged SENTIMENT$^{\perp}$, the market factor (RMRF), and the Fama–French factors (HML and SMB),

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1} + \beta_{i,\circ}\mathrm{RMRF}_t + \gamma_{i,2}\mathrm{SMB}_t$$

where $r_{(i,k_2),t} - r_{(i,k_1),t}$ is the long-short portfolio return that longs portfolio $k_2$ and shorts portfolio $k_1$ with firm characteristic $i$ (including size, book-to-market, dividend yield, earnings/price, age, sigma, R&D expense/assets, fixed assets, sales growth, and external finance/assets) at time $t$, and $k_1,k_2 \in \{\leq 0, = 0, 1, 2, \ldots, 10\}$. "1", "2", …, and "10" indicate the portfolios in the 1st (the lowest characteristic value), the 2nd, …, and the 10th (the highest characteristic value) deciles, respectively. "≤0" and "= 0" represent the portfolios for non-earnings stocks and non-dividend-paying stocks. This table only reports the parameter estimates of $\gamma_{i,1}$. The sample period for the portfolios formed on size, book-to-market, dividend yield, earnings/price, sigma, and fixed assets is from January 1966 to December 2007 and those of the portfolios formed on R&D expense/assets, age, sales growth, and external finance/assets begin by January 1975, January 1980, July 1967, and July 1967, respectively, and end by December 2007. The portfolio returns are in percentage. SENTIMENT$^{\perp}$ is the Baker and Wurgler's orthogonalized sentiment proxy. The column "All" reports the results without regime sorting. The other columns "Rec." and "Exp." ("Regime $j$", $j = 1,2$) show the results based on NBER dummy-sorted (regime-sorted) observations as recessions and expansions (regime $= j$). The Newey–West $p$-values (lagged terms are determined by $0.75T^{1/3}$, where $T$ is the sample length) are in parentheses. * and ** denote significance at 5% and 1% levels, respectively.

measured by the standard deviation of monthly stock returns over the 12 months ending in June of year $t$, for firms with at least 10 return observations. Following Baker and Wurgler (2006), we measure asset tangibility by $RD/A$ and $PPE/A$ and growth opportunities by $GS$ and $EF/A$. $RD/A$ is the R&D expense at the last fiscal year end of the prior calendar year $t-1$ divided by total assets at the last fiscal year end of the prior calendar year $t-1$. $PPE/A$ is measured by property, plant, and equipment at the last fiscal year end of the prior calendar year $t-1$ divided by total assets at the last fiscal year end of the prior calendar year $t-1$. $GS$ is the change in net sales divided by prior-year net sales. $EF/A$ is the change in assets minus the change in retained earnings divided by divided by total assets at the end of December of the prior year $t-1$.

At the end of June in year $t$ we sort all NYSE, AMEX, and NASDAQ stocks based on $ME$, $BE/ME$, $D/P$, $E/P$, $AGE$, $SIGMA$, $RD/A$, $PPE/A$, $GS$, and $EF/A$, and then, for each of the characteristics, allocate them into 10 groups. The decile portfolios are equally weighted. Monthly portfolio returns are then calculated from July of year $t$ through June of year $t+1$. Note that the decile portfolios formed on $D/P$ and $E/P$ only include stocks with positive $D/P$ and $E/P$, respectively. We include an additional portfolio of non-dividend-paying stocks ("$D/P = 0$") and an additional portfolio of non-earnings stocks ("$E/P \leq 0$").

The sample of portfolio returns begins from January 1966 or later, depending on firm characteristics, but all end in December 2007. The returns of portfolios formed on $AGE$ begin from January 1980, which is restricted by the availability of a comprehensive firm coverage in Jay Ritter's historical founding dates data. The returns of portfolios formed on $RD/A$ begin from January 1975 because only until 1974 the Financial Accounting Standards Board started to require R&D costs to be expensed. The returns of portfolios formed on $GS$ and $EF/A$ begin from July 1967 due to the limit of our data sample.

Table 3 reports summary statistics for the portfolio returns in percentage. Average returns are higher on small size stocks than large size stocks, high book-to-market (value) stocks than low book-to-market (growth) stocks, and high earnings-to-price stocks than low earnings-to-price stocks. Non-earnings stocks earn high average returns of 1.34% per month over the next year. There is no obviously cross-sectional effect of dividend yield. The mean returns of the middle $AGE$ deciles are higher than those of the top and bottom $AGE$ deciles. The mean returns of $SIGMA$, $RD/A$, $GS$ and $EF/A$ portfolios generally exhibit a monotonic pattern. The Jarque–Bera statistic for testing return normality rejects the null at the 5% level for all portfolios. The property of non-normal distribution is empirically related to the regime-switching feature of stock returns and could invalidate the conventional $t$-test in the predictive regression.

### *4.2. The regression models*

Following Baker and Wurgler (2006), we use the lagged sentiment measure to predict the equal-weighted returns of the long-short portfolios that are long in stocks with high characteristic values and short in stocks with low characteristic values. We run

**Table 5**
Predictive regressions for long-short portfolio returns (with bootstrapped p values).

| Long-short | All | NBER recession index | | Markov-switching model | |
|---|---|---|---|---|---|
| | | Rec. | Exp. | Regime 1 (Rec.) | Regime 2 (Exp.) |
| *Portfolios formed on size* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | 0.71** (0.00) | − 0.06 (0.55) | 1.03** (0.00) | 0.13 (0.33) | 1.04** (0.00) |
| 5 − 1 | 0.39**(0.00) | − 0.14(0.74) | 0.55** (0.00) | 0.08 (0.30) | 0.51** (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | 0.34** (0.00) | 0.10(0.17) | 0.51** (0.00) | 0.17(0.11) | 0.36** (0.00) |
| 5 − 1 | 0.30** (0.00) | 0.00(0.49) | 0.36** (0.00) | 0.19*(0.02) | 0.27** (0.00) |
| *Portfolios formed on book-to-market* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | 0.26*(0.03) | 0.46 (0.16) | 0.19(0.06) | 0.16 (0.30) | 0.44** (0.00) |
| 5 − 1 | 0.51** (0.00) | 0.58* (0.03) | 0.48** (0.00) | 0.50 (0.23) | 0.54** (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | 0.01(0.49) | − 0.12(0.83) | 0.14* (0.02) | − 0.22(0.99) | 0.24** (0.00) |
| 5 − 1 | 0.23** (0.00) | 0.16 (0.07) | 0.33** (0.00) | 0.17(0.07) | 0.31** (0.00) |
| *Portfolios formed on dividend yield* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − = 0 | 1.03** (0.00) | 1.07** (0.00) | 0.97** (0.00) | 0.82** (0.01) | 1.13** (0.00) |
| 5 − = 0 | 0.83** (0.00) | 0.62** (0.00) | 0.92** (0.00) | 0.72*(0.03) | 0.86** (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − = 0 | 0.51** (0.00) | 0.74** (0.00) | 0.51** (0.00) | 0.34** (0.01) | 0.47** (0.00) |
| 5 − = 0 | 0.44** (0.00) | 0.44** (0.00) | 0.56** (0.00) | 0.43** (0.00) | 0.40** (0.01) |
| *Portfolios formed on earnings/price* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − ≤0 | 0.86** (0.00) | 1.12** (0.00) | 0.82** (0.00) | 0.97** (0.00) | 0.77** (0.00) |
| 5 − ≤0 | 0.95** (0.00) | 0.87** (0.00) | 1.08** (0.00) | 0.88*(0.05) | 0.90** (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − ≤0 | 0.54** (0.00) | 0.95** (0.00) | 0.54** (0.00) | 0.68** (0.00) | 0.42** (0.00) |
| 5 − ≤0 | 0.57** (0.00) | 0.75** (0.00) | 0.70** (0.00) | 0.60** (0.00) | 0.47** (0.00) |
| *Portfolios formed on age* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | 2.03** (0.00) | 1.52** (0.01) | 2.57** (0.00) | 3.87**(0.01) | 1.42** (0.00) |
| 5 − 1 | 1.58** (0.00) | 1.36** (0.00) | 1.74** (0.00) | 3.18**(0.00) | 0.76** (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | 1.21** (0.00) | − 0.44 (0.87) | 1.96** (0.00) | 1.00(0.14) | 1.09** (0.00) |
| 5 − 1 | 1.26** (0.00) | 0.41 (0.23) | 1.56** (0.00) | 1.80*(0.02) | 0.78** (0.00) |
| *Portfolios formed on sigma* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | − 1.77** (0.00) | − 1.67* (0.02) | − 1.91** (0.00) | − 2.51*(0.03) | − 1.54** (0.00) |
| 10 − 5 | − 1.34** (0.00) | − 1.52** (0.01) | − 1.45** (0.00) | − 2.01*(0.03) | − 1.05** (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | − 1.08** (0.00) | − 0.54(0.18) | − 1.34** (0.00) | − 1.20** (0.01) | − 0.92** (0.00) |
| 10 − 5 | − 0.94** (0.00) | − 0.38(0.20) | − 1.12** (0.00) | − 1.14** (0.01) | − 0.75** (0.01) |
| *Portfolios formed on R&D expense/assets* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | − 0.61*(0.02) | − 1.86*(0.02) | − 0.34(0.06) | − 1.12** (0.00) | − 0.58** (0.00) |
| 10 − 5 | − 0.60** (0.00) | − 0.63*(0.04) | − 0.52** (0.00) | − 0.82(0.07) | − 0.72** (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | − 0.00(0.55) | − 0.63(0.12) | 0.16(0.82) | − 0.11(0.35) | − 0.21* (0.05) |
| 10 − 5 | − 0.13(0.19) | − 0.15(0.33) | − 0.08(0.34) | − 0.13(0.31) | − 0.33** (0.00) |
| *Portfolios formed on fixed assets* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | 0.42** (0.00) | 0.52* (0.02) | 0.33** (0.00) | 0.61** (0.00) | 0.23*(0.05) |
| 5 − 1 | 0.23** (0.00) | 0.29* (0.03) | 0.16** (0.00) | 0.33** (0.00) | 0.08 (0.18) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | 0.11(0.06) | 0.31* (0.02) | 0.09(0.15) | 0.33 (0.17) | − 0.17(0.26) |
| 5 − 1 | 0.08(0.06) | 0.17* (0.05) | 0.05(0.21) | 0.19 (0.31) | − 0.12(0.32) |
| *Portfolios formed on sales growth* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | − 0.28** (0.00) | − 0.22 (0.16) | − 0.34** (0.00) | − 0.26 (0.09) | − 0.36** (0.00) |
| 10 − 5 | − 0.72** (0.00) | − 0.69** (0.01) | − 0.77** (0.00) | − 0.78** (0.00) | − 0.71** (0.00) |

**Table 5** (*continued*)

| Long-short | All | NBER recession index: Rec. | NBER recession index: Exp. | Markov-switching model: Regime 1 (Rec.) | Markov-switching model: Regime 2 (Exp.) |
|---|---|---|---|---|---|
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | −0.17** (0.01) | 0.11 (0.81) | −0.37** (0.00) | −0.02 (0.43) | −0.33** (0.00) |
| 10 − 5 | −0.44** (0.00) | −0.33* (0.02) | −0.63** (0.00) | −0.41** (0.00) | −0.47** (0.00) |
| *Portfolios formed on external finance/assets* | | | | | |
| Only Sentiment$^{\perp}$ | | | | | |
| 10 − 1 | −0.52** (0.00) | −0.59*(0.02) | −0.55** (0.00) | −0.70** (0.00) | −0.43** (0.00) |
| 10 − 5 | −0.85** (0.00) | −0.89** (0.00) | −0.93** (0.00) | −1.07(0.07) | −0.70** (0.00) |
| Controlling for RMRF, SMB, and HML | | | | | |
| 10 − 1 | −0.28** (0.00) | −0.20(0.13) | −0.45** (0.00) | −0.34** (0.01) | −0.31** (0.00) |
| 10 − 5 | −0.48** (0.00) | −0.49** (0.01) | −0.69** (0.00) | −0.62** (0.00) | −0.39** (0.00) |

Note: This table contains the results of (i) regressions of value-weighted long-short portfolio returns on the lagged SENTIMENT$^{\perp}$,

$$r_{(i,k_2),t}-r_{(i,k_1),t}=\alpha_i+\gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1}+\varepsilon_{i,t},$$

and (ii) regressions of valued-weighted long-short portfolio returns on the lagged SENTIMENT$^{\perp}$, the market factor (RMRF), and the Fama–French factors (HML and SMB),

$$r_{(i,k_2),t}-r_{(i,k_1),t}=\alpha_i+\gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1}+\beta_{i,\circ}\mathrm{RMRF}_t+\gamma_{i,2}\mathrm{SMB}_t+\gamma_{i,3}\mathrm{HML}_t+\varepsilon_{i,t},$$

where $r_{(i,k_2),t}-r_{(i,k_1),t}$ is the long-short portfolio return that longs portfolio $k_2$ and shorts portfolio $k_1$ with firm characteristic $i$ (including size, book-to-market, dividend yield, earnings/price, age, sigma, R&D expense/assets, fixed assets, sales growth, and external finance/assets) at time $t$, and $k_1,k_2\in\{\leq 0,=0,1,2,\ldots,10\}$. "1", "2", …, and "10" indicate the portfolios in the 1st (the lowest characteristic value), the 2nd, …, and the 10th (the highest characteristic value) deciles, respectively. "≤0" and "= 0" represent the portfolios for non-earnings stocks and non-dividend-paying stocks. This table only reports the parameter estimates of $\gamma_{i,1}$. The sample period for the portfolios formed on size, book-to-market, dividend yield, earnings/price, sigma, and fixed assets is from January 1966 to December 2007 and those of the portfolios formed on R&D expense/assets, age, sales growth, and external finance/assets begin by January 1975, January 1980, July 1967, and July 1967, respectively, and end by December 2007. The portfolio returns are in percentage. SENTIMENT$^{\perp}$ is the Baker and Wurgler's orthogonalized sentiment proxy. The column "All" reports the results without regime sorting. The other columns "Rec." and "Exp." ("Regime $j$", $j=1,2$) show the results based on NBER dummy-sorted (regime-sorted) observations as recessions and expansions (regime $=j$). The bootstrapped $p$-values are in parentheses. * and ** denote significance at 5% and 1% levels, respectively.

the predictive regressions using the regime-sorted (identified by the NBER and the Markov-switching model) to control for the effects of regime shifts,

$$r_{(i,k_2),t}-r_{(i,k_1),t}=\alpha_i+\gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1}+\epsilon_{i,t},\tag{3}$$

where $r_{(i,k_2),t}-r_{(i,k_1),t}$ is the return in month $t$ on a long-short portfolio that is long the $k_2$ portfolio and short the $k_1$ portfolio formed on firm characteristic $i$, and $k_1,k_2\in\{\leq 0,=0,1,2,\ldots,10\}$. We use "1", "2", …, and "10" to denote the portfolios of the 1st (the bottom characteristic decile), 2nd, …, and the top characteristic decile, respectively. We use "≤0" and "= 0" to represent the portfolios of non-earnings stocks and non-dividend-paying stocks, respectively. $\mathrm{SENTIMENT}^{\perp}_{t-1}$ is the Baker and Wurgler's (2006) orthogonalized sentiment index in month $t-1$.

To distinguish the predictability effects from well-known factors, we follow Stambaugh et al. (in press) and control for three factors in the following multivariate predictive regression,

$$r_{(i,k_2),t}-r_{(i,k_1),t}=\alpha_i+\gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1}+\beta_{i,\circ}\mathrm{RMRF}_t+\gamma_{i,2}\mathrm{SMB}_t+\gamma_{i,3}\mathrm{HML}_t+\varepsilon_{i,t},\tag{4}$$

where $\mathrm{RMRF}_t$ is the return on the CRSP value-weighted market index in excess of the one-month T-bill rate. $\mathrm{SMB}_t$ and $\mathrm{HML}_t$ are the Fama–French factors. When performing the predictive regressions, we eliminate observations at the turning points where the economic regime actually switches from one state to another because the portfolio returns of these samples may be affected by both the investor sentiment and regime shifts. This helps clearly identify the predictive ability of sentiment on stock returns.

### 4.3. Results

We consider the long-short portfolios, "10 − 1" or "10 − ≤(or =) 0", that are long the top characteristic decile and short the lowest characteristic decile. We also use the middle decile in the long-short portfolio position to reveal any non-linear relation between the lagged sentiment and portfolio returns. Table 4 reports the coefficient estimate on sentiment of the predictive regression and the Newey–West $p$-value (based on two-tailed test) in parentheses. We follow Andrew (1991) and Stock and Watson (2007) and determine the number of lags by $0.75T^{1/3}$, where $T$ is the sample length.

**Table 6**
Out-of-sample predictability test results using the Clark and West's (2007) MSPE-adjusted statistic.

| | | NBER recession index | | Markov-switching model | |
|---|---|---|---|---|---|
| Long-short | All | Rec. | Exp. | Regime 1 (Rec.) | Regime 2 (Exp.) |
| *Portfolios formed on size* | | | | | |
| 10 – 1 | 2.09** | −0.55 | 3.55** | −0.97 | 2.52** |
| 5 – 1 | 1.99** | −1.84 | 2.01** | −0.76 | 2.03** |
| *Portfolios formed on book-to-market* | | | | | |
| 10 – 1 | −0.39 | −0.85 | −0.16 | 0.31 | −0.64 |
| 5 – 1 | 1.00 | −0.10 | 2.21** | −0.58 | 0.88 |
| *Portfolios formed on dividend yield* | | | | | |
| 10 – = 0 | 4.43** | 4.90** | 3.93** | −0.35 | 2.38** |
| 5 − = 0 | 3.29** | 2.03** | 3.90** | 0.33 | 2.51** |
| *Portfolios formed on earnings/price* | | | | | |
| 10 − ≤0 | 4.28** | 7.06** | 3.93** | 2.39** | 2.46** |
| 5 − ≤0 | 3.76** | 4.32** | 4.98** | 1.30 | 3.13 |
| *Portfolios formed on age* | | | | | |
| 10 − 1 | 5.73** | −0.04 | 11.27** | −1.77 | 4.61** |
| 5 − 1 | 7.49** | −1.83 | 9.56** | 0.35 | 2.76** |
| *Portfolios formed on sigma* | | | | | |
| 10 − 1 | 4.01** | 2.93** | 8.69** | 0.50 | 5.57** |
| 100 − 5 | 3.65** | 3.36** | 7.60** | 1.69* | 4.15** |
| *Portfolios formed on R&D expense/assets* | | | | | |
| 10 − 1 | −0.78 | −0.54 | −0.27 | −3.02 | 0.59 |
| 10 − 5 | −1.10 | 1.41 | −1.83 | −3.15 | 1.93* |
| *Portfolios formed on fixed assets* | | | | | |
| 10 − 1 | −0.07 | 0.89 | −1.52 | 1.09 | −1.83 |
| 5 − 1 | 0.03 | 0.11 | −1.20 | 0.65 | −1.74 |
| *Portfolios formed on sales growth* | | | | | |
| 10 − 1 | 0.51 | −1.29 | 2.64** | −1.92 | −0.05 |
| 10 − 5 | 3.00 | 0.42 | 5.14** | 0.92 | 1.50 |
| *Portfolios formed on external finance/assets* | | | | | |
| 10 − 1 | 1.00 | −0.84 | 3.51** | 0.24 | −0.23 |
| 10 − 5 | 2.61 | 0.96 | 5.35** | 1.45 | 1.29 |

Note: This table reports the results of the out-of-sample tests. The Clark and West's (2007) MSPE-adjusted statistic is computed using the prediction errors of the unrestricted and restricted models of

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1} + \beta_{i,\circ}\mathrm{RMRF}_t + \gamma_{i,2}\mathrm{SMB}_t + \gamma_{i,3}\mathrm{HML}_t + \varepsilon_{i,t},$$

and

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \beta_{i,\circ}\mathrm{RMRF}_t + \gamma_{i,2}\mathrm{SMB}_t + \gamma_{i,3}\mathrm{HML}_t + \varepsilon_{i,t},$$

for the returns of the long-short portfolios of size, book-to-market, dividend yield, earnings/price, age, sigma, and R&D expense/assets. The unrestricted model is the regressions of long-short portfolio returns on the lagged Baker and Wurgler's orthogonalized sentiment proxy, the market factor (RMRF), and the Fama–French factors (HML and SMB). The restricted model is the predictive regression model without the lagged Baker and Wurgler's orthogonalized sentiment proxy. The long-short portfolio return that longs portfolio $k_2$ and shorts portfolio $k_1$ with firm characteristic $i$ (including size, book-to-market, dividend yield, earnings/price, age, sigma, R&D expense/assets, fixed assets, sales growth, and external finance/assets) at time $t$, and $k_1, k_2 \in \{\leq 0, =0, 1, 2, \ldots, 10\}$. "1", "2", …, and "10" indicate the portfolios in the 1st (the lowest characteristic value), the 2nd, …, and the 10th (the highest characteristic value) deciles, respectively. "≤0" and "= 0" represent the portfolios for non-earnings stocks and non-dividend-paying stocks. The sample period for the portfolios formed on size, book-to-market, dividend yield, earnings/price, sigma, and fixed assets is from January 1966 to December 2007 and those of the portfolios formed on R&D expense/assets, age, sales growth, and external finance/assets begin by January 1975, January 1980, July 1967, and July 1967, respectively, and end by December 2007. The portfolio returns are in percentage. SENTIMENT$^{\perp}$ is the Baker and Wurgler's orthogonalized sentiment proxy. The column "All" reports the results without regime sorting. The other columns "Rec." and "Exp." ("Regime $j$", $j = 1,2$) show the results based on NBER dummy-sorted (regime-sorted) observations as recessions and expansions (regime $= j$). The critical value for the one-sided test at the 5% and 1% significance levels are 1.645 and 1.96, respectively. * and ** denote significance at 5% and 1% levels, respectively.

Without sorting observations by regime, the column "All" shows that the coefficient estimates, $\hat{\gamma}_{i,1}$, are positive and significant for the long-short portfolios associated with size, book-to-market ratio, age, earnings and dividend-paying status, and the slope coefficient estimates are negative for the portfolios formed on *SIGMA*, *RD/A*, *GS*, and *EF/A*, without including any control variables in the regressions. The results indicate that, when sentiment is high, future returns on small stocks, young stocks, high volatility stocks, high growth opportunities, non-earnings stocks, and non-dividend-paying stocks are lower than stocks of large size, old, value, low volatility, low intangible asset, low growth opportunities, high dividend yields, and high earnings firms. These patterns are little affected by controlling for RMRF, SMB, and HML. However, the $\hat{\gamma}_{i,1}$ estimates for the portfolios formed on R&D expense-to-assets become insignificant after controlling for RMRF, SMB, and HML. These results are consistent with the findings of Baker and Wurgler (2006).

We report the results using the NBER indicator to identify economic regimes, and include side-by-side the results using the regime-switch model. The columns "Rec." and "Exp." ("Regime $j$", $j=1,2$) show the results based on NBER dummy-sorted (regime-sorted) observations as recessions and expansions (regime $=j$).

Importantly, the results based on regime-sorted data show that the predictability patterns associated with sentiment are regime-dependent once we control for the economic regime. Under a recession state (regime 1), the predictive power of investor sentiment becomes insignificant in most cases after controlling for RMRF, SMB, and HML. In contrast, in an expansion state (regime 2), sentiment shows strong and significant predictive power. High sentiment results in relatively low subsequent returns on small stocks, young stocks, growth stocks, high volatility stocks, high growth opportunities, non-earning stocks, and non-dividend-paying stocks. Moreover, the significance of the predictive ability of sentiment is little affected by controlling for RMRF, SMB, and HML in most cases.[21]

To further confirm our findings, we perform a bootstrap test procedure to draw robust inferences, following Kosowski et al. (2006). This approach is useful for many reasons. For example, Table 3 shows that the distributions of returns on portfolios sorted by stock characteristics may not be normally distributed. The correlation between the endogenous regressors and return innovations may result in biased estimates as shown in Stambaugh (1999). Table 5 reports the bootstrapped $p$-values in parentheses. Comparing the results in Table 5 with those in Table 4, the return predictability of investor sentiment are similar, and do not change our conclusions.

In short, the evidence suggests that controlling for the economic state is crucial for examining the return predictive ability of sentiment.[22]

### 4.4. Out-of-sample test using the MSPE-adjusted statistic of Clark and West (2007)

Since investors only have access to past observations in making forecasts, the in-sample estimation using the full sample might cause a possible look-ahead bias. We thus perform an out-of-sample test to examine the robustness of the regime-dependent predictability patterns associated with sentiment.[23] We test whether the predictive power of an unrestricted model (with the predictor) is better than that of a restricted model (without the predictor) in terms of prediction errors. Clark and West (2007) develop a test statistic using mean squared prediction errors (MSPE) to measure the prediction performance of models. This test statistically assesses the difference in MSPEs between the restricted and the unrestricted models. The unrestricted model is deemed to have better forecasting performance if its MSPE is smaller than that of the restricted model. The unrestricted model in our test of the predictive power of sentiment is Eq. (4), and the restricted model with a constraint of $\gamma_{i,1}=0$, i.e. no predictive power of sentiment, is

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \beta_{i,\circ}\mathrm{RMRF}_t + \gamma_{i,2}\mathrm{SMB}_t + \gamma_{i,3}\mathrm{HML}_t + \epsilon_{i,t}. \tag{5}$$

For a sample of $T$ observations including $R$ in-sample ($t=1,\ldots,R$) and $P$ out-of-sample ($t=R+1,\ldots,R+P$) observations, we calculate the model prediction errors by performing the recursive estimation and prediction procedures as follows. First, for each of the long-short portfolios, we estimate models (4) and (5) using the in-sample data, and obtain the sets of the coefficient estimates, $\tilde{\lambda}_i^{u'}$ and $\tilde{\lambda}_i^{c'}$, for the unrestricted and the restricted models, respectively. Next, for months $t=R+1$, $\ldots,R+P$, we recursively update the coefficient estimates and incorporate the out-of-sample realizations of the explanatory variables $\mathbf{x}_{i,t}^u$ (with $\mathrm{SENTIMENT}_{t-1}^{\perp}$) and $\mathbf{x}_{i,t}^c$ (without $\mathrm{SENTIMENT}_{t-1}^{\perp}$) to compute the return predictions $\tilde{y}_{i,t}^u = \tilde{\lambda}_i^{u'}\mathbf{x}_{i,t}^u$ and $\tilde{y}_{i,t}^c = \tilde{\lambda}_i^{c'}\mathbf{x}_{i,t}^c$ for the unrestricted and restricted models, respectively.[24] The prediction errors, $\epsilon_{i,t}^{\tilde{u}}$ and $\epsilon_{i,t}^{\tilde{c}}$, of the unrestricted

[21] As suggested by the associate editor, we also separately control for macroeconomic variables including the yield spread, default premium, dividend–price ratio, the growth rate of industrial production, and the growth rate of personal consumption expenditures in durables, nondurables and services. The results (not reported but available upon request) are consistent with the findings and do not change our conclusions.

[22] As suggested by one of the referees, we also perform the predictive regressions based on the regimes identified by sentiment. Low (high) sentiment regime is the sentiment value in the previous month below (above) the median of the sentiment series. We find that the predictive ability of sentiment is not significant in both low and high sentiment regimes for most cases. It seems that the two-regime pattern is likely to be driven by fundamentals in this regard.

[23] We are grateful for one of the referees for pointing out that the monthly sentiment index of Baker and Wurgler (2006) itself has a look-ahead bias during construction.

[24] In the recursive estimations the sample size available for estimating the coefficients $\tilde{\lambda}_i^{u'}$ and $\tilde{\lambda}_i^{c'}$ grows as one makes predictions for successive observations. For example, we first estimate the parameters $\tilde{\lambda}_i^{u'}$ and $\tilde{\lambda}_i^{c'}$ using data observed from time $t=1$ to $R$. We then use these estimates to make the prediction for time $t=R+1$.In the next period, we estimate the parameters using data observed from time $t=1$ to $R+1$ to incorporate new information. We then use these estimates to make the prediction for time $t=R+2$. Campbell and Thompson (2008) also use a similar recursive estimation scheme.

**Table 7**
Conditional market betas.

| Long-short | | All | NBER recession index | | Markov-switching model | |
|---|---|---|---|---|---|---|
| | | | Rec. | Exp. | Regime 1 (Rec.) | Regime 2 (Exp.) |
| *Portfolios formed on size* | | | | | | |
| 10 − 1 | $\gamma_{i,1}$ | 0.34** (0.01) | 0.13 (0.48) | 0.53** (0.00) | 0.18(0.44) | 0.36** (0.01) |
| 5 − 1 | $\gamma_{i,1}$ | 0.31*(0.02) | 0.12 (0.66) | 0.36* (0.04) | 0.25(0.18) | 0.27(0.10) |
| 10 − 1 | $\beta_{i,1}$ | −0.01(0.72) | 0.01(0.68) | −0.05 (0.12) | 0.01(0.77) | −0.05(0.08) |
| 5 − 1 | $\beta_{i,1}$ | 0.03(0.26) | 0.06 (0.14) | −0.01 (0.76) | 0.08*(0.04) | −0.03(0.47) |
| *Portfolios formed on book-to-market* | | | | | | |
| 10 − 1 | $\gamma_{i,1}$ | −0.01(0.91) | −0.37(0.11) | 0.15(0.19) | −0.32(0.14) | 0.24*(0.04) |
| 5 − 1 | $\gamma_{i,1}$ | 0.23*(0.03) | 0.09(0.70) | 0.34** (0.01) | 0.13(0.54) | 0.31*(0.04) |
| 10 − 1 | $\beta_{i,1}$ | −0.08(1.00) | −0.12** (0.00) | −0.04 (0.10) | −0.13** (0.00) | −0.02(0.40) |
| 5 − 1 | $\beta_{i,1}$ | −0.04(0.34) | −0.03(0.47) | −0.02 (0.52) | −0.05(0.24) | 0.00(0.96) |
| *Portfolios formed on dividend yield* | | | | | | |
| 10 − = 0 | $\gamma_{i,1}$ | 0.49** (0.00) | 0.50(0.08) | 0.52** (0.01) | 0.25 (0.29) | 0.47 (0.03) |
| 5 − = 0 | $\gamma_{i,1}$ | 0.43** (0.00) | 0.29(0.24) | 0.57** (0.00) | 0.36 (0.15) | 0.40** (0.01) |
| 10 − = 0 | $\beta_{i,1}$ | −0.07*(0.02) | −0.12** (0.00) | −0.01(0.85) | −0.12** (0.00) | −0.02 (0.63) |
| 5 − = 0 | $\beta_{i,1}$ | −0.05(0.11) | −0.07(0.06) | −0.02(0.58) | −0.09*(0.03) | 0.02 (0.50) |
| *Portfolios formed on earnings/price* | | | | | | |
| 10 − ≤ 0 | $\gamma_{i,1}$ | 0.53** (0.00) | 0.84** (0.01) | 0.56*(0.02) | 0.63*(0.02) | 0.42*(0.05) |
| 5 − ≤ 0 | $\gamma_{i,1}$ | 0.57** (0.01) | 0.73(0.07) | 0.70*(0.02) | 0.57(0.13) | 0.47*(0.05) |
| 10 − ≤ 0 | $\beta_{i,1}$ | −0.06(0.09) | −0.05(0.29) | −0.04(0.30) | −0.07(0.13) | −0.05(0.22) |
| 5 − ≤ 0 | $\beta_{i,1}$ | −0.02(0.67) | −0.01(0.92) | −0.01(0.91) | −0.04(0.56) | 0.02(0.62) |
| *Portfolios formed on age* | | | | | | |
| 10 − 1 | $\gamma_{i,1}$ | 1.26** (0.00) | −0.30(0.63) | 1.96** (0.00) | 0.56(0.66) | 1.07** (0.00) |
| 5 − 1 | $\gamma_{i,1}$ | 1.30** (0.00) | 0.52 (0.46) | 1.59*(0.03) | 1.53(0.23) | 0.73(0.07) |
| 10 − 1 | $\beta_{i,1}$ | −0.19(0.08) | −0.36** (0.00) | 0.01(0.97) | −0.46** (0.00) | 0.04(0.65) |
| 5 − 1 | $\beta_{i,1}$ | −0.15(0.10) | −0.29*(0.02) | −0.05(0.73) | −0.29*(0.04) | 0.05(0.62) |
| *Portfolios formed on sigma* | | | | | | |
| 10 − 1 | $\gamma_{i,1}$ | −1.12** (0.00) | −0.65(0.40) | −1.34** (0.00) | −1.48(0.10) | −0.89** (0.01) |
| 10 − 5 | $\gamma_{i,1}$ | −0.98** (0.00) | −0.52(0.44) | −1.14** (0.00) | −1.36(0.07) | −0.73** (0.01) |
| 10 − 1 | $\beta_{i,1}$ | 0.11 (0.25) | 0.34** (0.01) | 0.00(0.98) | 0.29(0.08) | −0.07(0.24) |
| 10 − 5 | $\beta_{i,1}$ | 0.13 (0.17) | 0.40(0.28) | 0.02(0.84) | 0.24(0.08) | −0.03(0.56) |
| *Portfolios formed on R&D expense/assets* | | | | | | |
| 10 − 1 | $\gamma_{i,1}$ | −0.01 (0.98) | −0.77 (0.16) | 0.20 (0.61) | −0.07 (0.92) | −0.20 (0.41) |
| 10 − 5 | $\gamma_{i,1}$ | −0.11 (0.63) | −0.22 (0.65) | −0.06 (0.86) | −0.02 (0.96) | −0.34 (0.10) |
| 10 − 1 | $\beta_{i,1}$ | 0.02 (0.32) | 0.40 (0.08) | 0.07 (0.34) | −0.04 (0.72) | −0.02 (0.83) |
| 10 − 5 | $\beta_{i,1}$ | −0.03 (0.32) | 0.22* (0.03) | 0.04 (0.50) | −0.16 (0.06) | 0.02 (0.67) |
| *Portfolios formed on fixed assets* | | | | | | |
| 10 − 1 | $\gamma_{i,1}$ | 0.11 (0.45) | 0.37 (0.22) | 0.10 (0.62) | 0.33 (0.20) | −0.17 (0.25) |
| 5 − 1 | $\gamma_{i,1}$ | 0.08 (0.50) | 0.16 (0.50) | 0.07 (0.61) | 0.17 (0.39) | −0.12 (0.33) |
| 10 − 1 | $\beta_{i,1}$ | 0.00 (0.95) | 0.03 (0.52) | −0.01 (0.76) | 0.00 (0.94) | 0.01 (0.80) |
| 5 − 1 | $\beta_{i,1}$ | −0.03 (0.25) | −0.01 (0.87) | −0.05* (0.05) | −0.03 (0.38) | −0.03 (0.32) |
| *Portfolios formed on sales growth* | | | | | | |
| 10 − 1 | $\gamma_{i,1}$ | −0.16 (0.31) | 0.28 (0.29) | −0.38* (0.05) | 0.05 (0.83) | −0.32* (0.02) |
| 10 − 5 | $\gamma_{i,1}$ | −0.42** (0.00) | −0.07 (0.79) | −0.65** (0.00) | −0.31 (0.26) | −0.47** (0.00) |
| 10 − 1 | $\beta_{i,1}$ | 0.06* (0.05) | 0.09* (0.05) | 0.04 (0.41) | 0.10* (0.02) | 0.00 (0.89) |
| 10 − 5 | $\beta_{i,1}$ | 0.09 (0.19) | 0.13 (0.16) | 0.05 (0.19) | 0.14 (0.16) | −0.01 (0.83) |
| *Portfolios formed on external finance/assets* | | | | | | |
| 10 − 1 | $\gamma_{i,1}$ | −0.26* (0.05) | 0.01 (0.97) | −0.47*(0.03) | −0.25 (0.37) | −0.30** (0.01) |
| 10 − 5 | $\gamma_{i,1}$ | −0.46** (0.01) | −0.23 (0.57) | −0.70** (0.01) | −0.51 (0.13) | −0.38** (0.00) |
| 10 − 1 | $\beta_{i,1}$ | 0.08 (0.11) | 0.11 (0.11) | 0.03 (0.40) | 0.12 (0.06) | −0.04 (0.29) |
| 10 − 5 | $\beta_{i,1}$ | 0.09 (0.16) | 0.12 (0.09) | 0.03 (0.44) | 0.14 (0.36) | −0.04 (0.38) |

and the restricted models, respectively, are the differences between the realized and the model predicted returns of the long-short portfolios. The MSPE-adjusted statistic of Clark and West (2007) is:

$$\text{MSPE}-\text{adjusted} = \frac{\sqrt{P}\bar{f}}{\sqrt{\tilde{V}_f}}, \tag{6}$$

where $\bar{f} = P^{-1}\sum_{t=R+1}^{R+P}\tilde{f}_t, \tilde{f}_t = \tilde{\varepsilon}_{i,t}^{c2} - \left[\tilde{\varepsilon}_{i,t}^{u2} - \left(\tilde{y}_{i,t}^{c} - \tilde{y}_{i,t}^{u}\right)^2\right]$, and $\tilde{V}_f$ is the sample variance of $\left(\tilde{f}_t - \bar{f}\right)$.

Under the null hypothesis of no difference in the model prediction errors, the MSPE-adjusted statistic is zero. Clark and West (2007) demonstrate that the asymptotic distribution for the MSPE—adjusted statistic can be approximated by a standard normal distribution. Note that since it is a one-sided test, the critical values for 5% and 1% significance level are 1.645 and 1.96, respectively.

We follow the standard convention and set the proportion of in-sample and out-of-sample data (R/P) to 1 in all the tests. Table 6 reports the results of the out-of-sample test. The first column shows that, without segregating the state of the economy, sentiment does not exhibit out-of-sample predictive power for returns firmly since ten out of twenty cases in Table 6 are statistically insignificant. This result is in sharp contrast with the in-sample result reported in the first column of Table 4. Noticeably, the coefficient estimate of $\hat{\gamma}_{i,1}$ changes with the state of the economy, suggesting that the instability of the parameter estimates causes the discrepancy between the in-sample and out-of-sample results (see also, Butler et al. (2005)).

More importantly, after controlling for the economic state identified by the NBER recession index, in the NBER expansion state the MSPE-adjusted statistics are significant in most cases, suggesting that investor sentiment does have an out-of-sample predictive power for returns. In the NBER recession state, by contrast, sentiment does not have significant predictive power. This out-of-sample regime-dependent feature on sentiment remains valid in the results based on the Markov-switching model. The out-of-sample performance of sentiment after controlling for the state of the economic regime is consistent with the in-sample results in Table 4.

### 4.5. Time-variation with sentiment in systematic risk

In the predictive regression model below we allow the market beta to vary with investor sentiment to examine whether the predictability in the cross-section of stock returns is due to time-varying sensitivity in the market risk factor driven by investor sentiment,

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\text{SENTIMENT}_{t-1}^{\perp} + \left(\beta_{i,\circ} + \beta_{i,1}\text{SENTIMENT}_{t-1}^{\perp}\right)\text{RMRF}_t + \gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \epsilon_{i,t}. \tag{7}$$

We test whether the return predictability of investor sentiment is attributable to the time-varying market beta associated with investor sentiment. Under the null hypothesis, the slope coefficient $\beta_{i,1}$ in the specification $(\beta_{i,\circ} + \beta_{i,1}\text{SENTIMENT}_{t-1}^{\perp})$ is zero and statistically insignificant. The results in Table 7 show that, in all cases, the estimate of $\beta_{i,1}$ is indeed close to zero and statistically insignificant for all portfolios. Again, sentiment exhibits predictive power for stock returns under the NBER expansion state but not in the NBER recession state. The results based on the Markov-switching model also perform similar pattern. The evidence suggests that the sentiment-driven variation in the market beta does not capture the predictive ability of sentiment on the cross-section of stock returns.

Notes to Table 7

Note: This table reports the results about regressions of long-short portfolio returns on lagged SENTIMENT$^{\perp}$, the market factor (RMRF), the Fama–French factors (HML and SMB), and interaction of RMRF and SENTIMENT$^{\perp}$,

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\text{SENTIMENT}_{t-1}^{\perp} + \left(\beta_{i,\circ} + \beta_{i,1}\text{SENTIMENT}_{t-1}^{\perp}\right)\text{RMRF}_t$$

$$+\gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\ \text{HML}_t + \varepsilon_{i,t},$$

where $r_{(i,k_2),t} - r_{(i,k_1),t}$ is the long-short portfolio return that longs portfolio $k_2$ and shorts portfolio $k_1$ with firm characteristic $i$ (including size, book-to-market, dividend yield, earnings/price, age, sigma, R&D expense/assets, fixed assets, sales growth, and external finance/assets) at time $t$, and $k_1, k_2 \in \{\leq 0, =0, 1, 2, \ldots, 10\}$. "1", "2", …, and "10" indicate the portfolios in the 1st (the lowest characteristic value), the 2nd, …, and the 10th (the highest characteristic value) deciles, respectively. "≤0" and "=0" represent the portfolios for non-earnings stocks and non-dividend-paying stocks. The sample period for the portfolios formed on size, book-to-market, dividend yield, earnings/price, sigma, and fixed assets is from January 1966 to December 2007 and those of the portfolios formed on R&D expense/assets, age, sales growth, and external finance/assets begin by January 1975, January 1980, July 1967, and July 1967, respectively, and end by December 2007. The portfolio returns are in percentage. SENTIMENT$^{\perp}$ is the Baker and Wurgler's orthogonalized sentiment proxy. The column "All" reports the results without regime sorting. The other columns "Rec." and "Exp." ("Regime $j$", $j=1,2$) show the results based on NBER dummy-sorted (regime-sorted) observations as recessions and expansions (regime $=j$). The Newey–West $p$-values (lagged terms are determined by $0.75T^{1/3}$, where $T$ is the sample length) are in parentheses. $*$ and $**$ denote significance at 5% and 1% levels, respectively.

## 5. Robustness checks

### 5.1. Predictive regressions using consumer confidence

For robustness checks, we use the monthly consumer confidence index provided by the Survey Research Center of the University of Michigan as a proxy for investor sentiment to test the predictive power of sentiment over the period between January 1978, when the index turned into monthly frequency data, and December 2007.[25] Following Baker and Wurgler (2006) and Lemmon and Portniaguina (2006), we orthogonalize consumer sentiment from macroeconomic variables by regressing the consumer sentiment on the growth rate of the industrial production, a dummy variable for NBER recessions, and the growth rate of personal consumption expenditures in durables, nondurables and services.[26] We denote the regression residuals as $\text{ConSENTIMENT}_t^{\perp}$ for the orthogonalized consumer confidence and plot the time-series observations in the bottom panel of Fig. 2. The orthogonalized sentiment dropped to a record low level in 1979 and then moved higher gradually before going back down again to a new low by 1992. It increased afterwards over time and reached a record high by 2000 which is then followed by dramatic decreases until 2003, corresponding to the boom and burst of the internet stock bubble. The overall pattern resembles that in the upper panel which depicts the Baker and Wurgler orthogonalized sentiment.

Next, we replace $\text{SENTIMENT}_{t-1}^{\perp}$ in Eq. (4) by $\text{ConSENTIMENT}_{t-1}^{\perp}$ and run the predictive regression,

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\text{ConSENTIMENT}_{t-1}^{\perp} + \beta_{i,\circ}\text{RMRF}_t + \gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \epsilon_{i,t}. \tag{8}$$

Table 8 reports the coefficient estimates on consumer sentiment. The results show that, in most cases, the orthogonalized sentiment of the consumer confidence index exhibits predictive ability for stock returns under the NBER expansion state, but not in the NBER recession state, suggesting that the predictive power of sentiment is regime-dependent.

The main difference between the results in Tables 4 and 8 is that, without segregating the regime of the economy, the Baker and Wurgler's orthogonalized sentiment is capable of predicting the returns on all portfolios (except the portfolio formed on fixed asset portfolio) while the orthogonalized consumer confidence is only able to predict the returns on the portfolios formed on firm size, fixed assets and growth opportunities. The discrepancy in results may be attributable to the difference in the information content reflected by the different sentiment measures. Consumer confidence mainly reflects the households' outlook about the economic activities related to the macroeconomic conditions, whereas the measure of Baker and Wurgler's (2006) reveals investors' outlook about the aggregate stock market.

### 5.2. Predictive regressions with regime dummies

We also examine the regime-dependency of the predictive power of investor sentiment using regime dummy variables in the regression model. Specifically, we run the predictive regression:

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_{i,\circ} + \alpha_{i,1}D_1 + \alpha_{i,2}D_2 + \left(\delta_{i,1}D_1 + \delta_{i,2}D_2\right)\text{SENTIMENT}_{t-1}^{\perp} + \beta_{i,\circ}\text{RMRF}_t + \gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \epsilon_{i,t}, \tag{9}$$

where $D_j$ is the dummy variable of regime $j$, which is equal to 1 for regime $j$, and 0 otherwise. Note that the specification in Eq. (9) treats the state of the regime as an exogenous variable. In addition to examining the sentiment effect under different regimes, we allow for the intercept to change with the regime dummies to control for the effect of regime shifts on the subsequent returns. We use all observations (including the turning points) to run this regression.

Table 9 reports the estimates of the coefficients on the two regime dummies. In the NBER expansions (and regime 2), investor sentiment displays positive and highly significant predictive power for portfolio returns of all characteristics. In the NBER recessions (and regime 1), however, investor sentiment loses its predictive ability for the cross-section of stock returns. These results are consistent with those presented in Table 4 and confirm our main finding that the predictive power of investor sentiment is regime-dependent.

### 5.3. Predictive regressions for the portfolio returns associated with 11 anomalies

We further examine the predictability of sentiment on the portfolio returns of 11 asset pricing anomalies studied by Stambaugh et al. (in press). These pricing anomalies are associated with financial distress, net stock issues, composite equity

[25] One of the referees noted that the Michigan consumer sentiment index data earlier than 1960 are available. Since the index series started from 1952 at a quarterly frequency until 1977, we linearly interpolate the quarterly index to obtain the monthly data for the period 1966/01–1977/12. We start our analyses from 1966/01 because the returns data for all the 10 portfolios of Baker and Wurgler (2006) are available from 1966/01 or later. We add the interpolated series to the original monthly data (1978/1–2007/12) and run the predictive regressions using this sentiment index. The overall pattern of the results is very similar to Table 7 and do not change our conclusions. For the purpose of maintaining the consistency of data frequency, we only report the results using this sentiment index starting from 1978. The details are available upon request.

[26] The data of personal consumption expenditures in durables, nondurables and services are obtained from the web-page of the Bureau of Economic Analysis: http://www.bea.gov/national/index.htm, National Income Accounts Table 2.8.

**Table 8**
Predictive regressions for long-short portfolio returns using consumer confidence (sample period: 1978:1–2007:12).

| Long-short | All | NBER recession index | | Markov-switching model | |
|---|---|---|---|---|---|
| | | Rec. | Exp. | Regime 1 (Rec.) | Regime 2 (Exp.) |
| *Portfolios formed on size* | | | | | |
| 10 − 1 | 0.06*(0.03) | 0.00(0.94) | 0.03*(0.05) | 0.02(0.59) | 0.01(0.50) |
| 5 − 1 | 0.01(0.87) | 0.00(0.99) | 0.02(0.39) | 0.02(0.80) | 0.01(0.70) |
| *Portfolios formed on book-to-market* | | | | | |
| 10 − 1 | 0.01(0.57) | −0.01(0.73) | 0.01(0.33) | −0.02(0.50) | 0.03*(0.03) |
| 5 − 1 | 0.01(0.26) | 0.01(0.86) | 0.02*(0.05) | 0.01(0.70) | 0.02** (0.00) |
| *Portfolios formed on dividend yield* | | | | | |
| 10 − =0 | 0.03(0.07) | 0.05(0.56) | 0.04*(0.02) | 0.03(0.35) | 0.03(0.29) |
| 5 − =0 | 0.01(0.76) | 0.07(0.25) | 0.01(0.69) | −0.01(0.79) | 0.03(0.21) |
| *Portfolios formed on earnings/price* | | | | | |
| 10 − ≤0 | 0.01(0.71) | 0.05(0.45) | 0.02(0.56) | −0.01(0.92) | 0.03(0.21) |
| 5 − ≤0 | 0.01(0.77) | 0.09(0.21) | 0.02(0.64) | 0.00(0.95) | 0.03(0.21) |
| *Portfolios formed on age* | | | | | |
| 10 − 1 | 0.00(0.56) | 0.00(0.62) | 0.02(0.91) | 0.03(0.70) | 0.03(0.37) |
| 5 − 1 | −0.01(0.41) | −0.03(0.45) | 0.00(0.68) | 0.00(0.97) | 0.01(0.70) |
| *Portfolios formed on sigma* | | | | | |
| 10 − 1 | −0.03(0.33) | −0.13(0.17) | −0.04(0.28) | −0.05(0.53) | −0.05*(0.05) |
| 10 − 5 | −0.02(0.53) | −0.08(0.27) | −0.03(0.37) | −0.06(0.42) | −0.02(0.34) |
| *Portfolios formed on R&D expense/assets* | | | | | |
| 10 − 1 | 0.08(0.09) | −0.05(0.46) | 0.10*(0.05) | 0.16(0.32) | 0.01(0.83) |
| 10 − 5 | 0.06(0.11) | −0.01(0.73) | 0.07(0.07) | 0.12(0.29) | −0.01(0.76) |
| *Portfolios formed on fixed assets* | | | | | |
| 10 − 1 | 0.04*(0.02) | 0.02(0.59) | 0.05**(0.01) | 0.06(0.12) | 0.02(0.42) |
| 5 − 1 | 0.03**(0.01) | 0.02(0.49) | 0.04**(0.01) | 0.04*(0.03) | 0.02(0.25) |
| *Portfolios formed on sales growth* | | | | | |
| 10 − 1 | −0.04(0.06) | −0.06(0.18) | −0.05*(0.04) | −0.09(0.27) | 0.00(0.84) |
| 10 − 5 | −0.03*(0.04) | −0.05(0.34) | −0.04*(0.04) | −0.07(0.26) | −0.01(0.74) |
| *Portfolios formed on external finance/assets* | | | | | |
| 10 − 1 | −0.02(0.21) | −0.08(0.14) | −0.02(0.29) | −0.05(0.12) | −0.01(0.65) |
| 10 − 5 | −0.02(0.38) | −0.07(0.25) | −0.03(0.36) | −0.04(0.28) | −0.01(0.72) |

Note: This table represents the results about regressions of long-short portfolio returns on lagged ConSENTIMENT$^{\perp}$, the market factor (RMRF), and the Fama–French factors (HML and SMB),

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\mathrm{ConSENTIMENT}^{\perp}_{t-1} + \beta_{i,\circ}\mathrm{RMRF}_t + \gamma_{i,2}\mathrm{SMB}_t + \gamma_{i,3}\mathrm{HML}_t + \varepsilon_{i,t},$$

where $r_{(i,k_2),t} - r_{(i,k_1),t}$ is the long-short portfolio return that longs portfolio $k_2$ and shorts portfolio $k_1$ with firm characteristic $i$ (including size, book-to-market, dividend yield, earnings/price, age, sigma, R&D expense/assets, fixed assets, sales growth, and external finance/assets) at time $t$, and $k_1, k_2 \in \{\leq 0, =0, 1, 2, \ldots, 10\}$. "1", "2", …, and "10" indicate the portfolios in the 1st (the lowest characteristic value), the 2nd, …, and the 10th (the highest characteristic value) deciles, respectively. "≤0" and "=0" represent the portfolios for non-earnings stocks and non-dividend-paying stocks. The sample period for the portfolios formed on size, book-to-market, dividend yield, earnings/price, sigma, and fixed assets is from January 1966 to December 2007 and those of the portfolios formed on R&D expense/assets, age, sales growth, and external finance/assets begin by January 1975, January 1980, July 1967, and July 1967, respectively, and end by December 2007. The portfolio returns are in percentage. ConSENTIMENT$^{\perp}$ is the orthogonalized consumer sentiment proxy from January 1978 to December 2007. The column "All" reports the results without regime sorting. The other columns "Rec." and "Exp." ("Regime $j$", $j=1,2$) show the results based on NBER dummy-sorted (regime-sorted) observations as recessions and expansions (regime $=j$). The Newey–West $p$-values (lagged terms are determined by $0.75T^{1/3}$, where $T$ is the sample length) are in parentheses. $*$ and $**$ denote significance at 5% and 1% levels, respectively.

issues, total accruals, net operating assets, momentum, gross profit-to-assets, asset growth, return-on-assets, and investment-to-assets. The anomaly associated with financial distress is advocated by Campbell et al. (2008) that firms with high failure probability have lower returns. To measure financial distress, we follow Chen et al. (2010) and Stambaugh et al. (in press) and adopt Campbell et al. (2008) failure probability (the third column of their Table 4) and Ohlson's (1980) O-score (model one in his Table 4).

The anomaly associated with total accruals is suggested by Sloan (1996) that firms with high accruals earn abnormally lower average returns than those with low accruals. The total accruals are measured by the change in non-cash working

**Table 9**
Predictive regressions with regime dummies.

| Long-short | NBER recession index | | Markov-switching model | |
|---|---|---|---|---|
| | Rec. | Exp. | Regime 1 (Rec.) | Regime 2 (Exp.) |
| *Portfolios formed on size* | | | | |
| 10 − 1 | 0.03(0.44) | 0.49**(0.00) | 0.15(0.23) | 0.40**(0.00) |
| 5 − 1 | −0.04(0.42) | 0.36*(0.03) | 0.19(0.15) | 0.28*(0.05) |
| *Portfolios formed on book-to-market* | | | | |
| 10 − 1 | −0.05(0.41) | 0.14(0.13) | −0.20(0.17) | 0.25**(0.01) |
| 5 − 1 | 0.21(0.15) | 0.34**(0.00) | 0.21(0.11) | 0.27**(0.00) |
| *Portfolios formed on dividend yield* | | | | |
| 10 − =0 | 0.74*(0.02) | 0.53**(0.00) | 0.38(0.08) | 0.54**(0.00) |
| 5 − =0 | 0.41(0.06) | 0.55**(0.00) | 0.47*(0.05) | 0.37**(0.01) |
| *Portfolios formed on earnings/price* | | | | |
| 10 − ≤0 | 0.91**(0.01) | 0.53**(0.00) | 0.75**(0.01) | 0.37*(0.02) |
| 5 − ≤0 | 0.72(0.06) | 0.69**(0.00) | 0.68(0.10) | 0.39*(0.02) |
| *Portfolios formed on age* | | | | |
| 10 − 1 | 0.47(0.21) | 1.87**(0.00) | 1.80*(0.04) | 0.90**(0.01) |
| 5 − 1 | 0.88(0.07) | 1.51**(0.01) | 2.48**(0.01) | 0.62(0.07) |
| *Portfolios formed on sigma* | | | | |
| 10 − 1 | −1.30(0.08) | −1.33**(0.00) | −1.54**(0.01) | −0.89**(0.00) |
| 10 − 5 | −1.18(0.13) | −1.12**(0.00) | −1.46**(0.01) | −0.65**(0.01) |
| *Portfolios formed on R&D expense/assets* | | | | |
| 10 − 1 | −0.93(0.10) | 0.13(0.37) | −0.23(0.34) | 0.01(0.50) |
| 10 − 5 | −0.08(0.45) | −0.11(0.36) | −0.21(0.32) | −0.23(0.14) |
| *Portfolios formed on fixed assets* | | | | |
| 10 − 1 | 0.36(0.07) | 0.06(0.37) | 0.34(0.08) | −0.11(0.27) |
| 5 − 1 | 0.23(0.14) | 0.0(0.41) | 0.20(0.15) | −0.08(0.28) |
| *Portfolios formed on sales growth* | | | | |
| 10 − 1 | 0.03(0.45) | −0.35*(0.04) | −0.05(0.41) | −0.31*(0.02) |
| 10 − 5 | −0.37(0.08) | −0.61**(0.00) | −0.46*(0.04) | −0.43**(0.00) |
| *Portfolios formed on external finance/assets* | | | | |
| 10 − 1 | −0.25(0.20) | −0.44**(0.01) | −0.40(0.07) | −0.21(0.09) |
| 10 − 5 | −0.52(0.09) | −0.67**(0.00) | −0.70*(0.02) | −0.30*(0.02) |

Note: This table reports the results about regressions of long-short portfolio returns on the regime dummies, the interactions of regime dummy variables and lagged SENTIMENT$^{\perp}$, the market factor (RMRF), and the Fama–French factors (HML and SMB),

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_{i,\circ} + \alpha_{i,1}D_1 + \alpha_{i,2}D_2 + (\delta_{i,1}D_1 + \delta_{i,2}D_2)\mathrm{SENTIMENT}^{\perp}_{t-1} + \beta_{i,\circ}\mathrm{RMRF}_t + \gamma_{i,2}\mathrm{SMB}_t + \gamma_{i,3}\mathrm{HML}_t + \varepsilon_{i,t},$$

where $r_{(i,k_2),t} - r_{(i,k_1),t}$ is the long-short portfolio return that longs portfolio $k_2$ and shorts portfolio $k_1$ with firm characteristic $i$ (including size, book-to-market, dividend yield, earnings/price, age, sigma, R&D expense/assets, fixed assets, sales growth, and external finance/assets) at time $t$, $k_1, k_2 \in \{\leq 0, =0, 1, 2, \ldots, 10\}$. "1", "2", …, and "10" indicate the portfolios in the 1st (the lowest characteristic value), the 2nd, …, and the 10th (the highest characteristic value) deciles, respectively. "≤0" and "=0" represent the portfolios for non-earnings stocks and non-dividend-paying stocks. The sample period for the portfolios formed on size, book-to-market, dividend yield, earnings/price, sigma, and fixed assets is from January 1966 to December 2007 and those of the portfolios formed on R&D expense/assets, age, sales growth, and external finance/assets begin by January 1975, January 1980, July 1967, and July 1967, respectively, and end by December 2007. The portfolio returns are in percentage. SENTIMENT$^{\perp}$ is the Baker and Wurgler's orthogonalized sentiment proxy. The columns "Rec." and "Exp." ("Regime $j$", $j = 1, 2$) show the results based on NBER dummy-sorted (regime-sorted) observations as recessions and expansions (regime $= j$). The Newey–West $p$-values (lagged terms are determined by $0.75T^{1/3}$, where $T$ is the sample length) are in parentheses. * and ** denote significance at 5% and 1% levels, respectively.

capital minus depreciation at the fiscal year-end of year $t-1$. Non-cash working capital is current assets minus cash and current liabilities, plus debt in current liabilities, plus taxes payable, scaled by the average of total assets over the fiscal year-end of years $t-1$ and $t-2$.

The anomaly of asset growth is found by Cooper et al. (2008) that firms with high growth in total assets have lower returns. Asset growth is the annual firm asset growth rate, defined as the change in total assets between years $t-1$ and $t-2$, divided by the lagged total assets at the fiscal year-end of year $t-2$. The anomaly of net operating assets is found by Hirshleifer et al. (2004) that firms with high net operating assets earn lower returns. Net operating assets are measured by the operating assets minus operating liabilities at the fiscal year-end of year $t-1$. The operating assets are total assets minus cash. The operating liabilities are total assets minus debt included in current liabilities, minus long-term debt, minority interests, preferred stocks, and common equity, and then divided by lagged total assets at the fiscal year-end of year $t-2$.

Novy-Marx (2010) suggests that, sorting by gross profits-to-assets, firms with higher profits have higher returns than those with lower profits. We measure gross profits-to-assets by net sales minus costs of goods sold at the fiscal year-end of year $t-1$, divided by lagged total asset at the fiscal year-end of year $t-1$. The anomaly of net stock issues and composite equity issues is found by Loughran and Ritter (1995) and Daniel and Titman (2006) that the long-run performance of equity issuers is worse than nonissuers with similar firm characteristics. Following Fama and French (2008), net share issuance is the annual share issuance, defined as the natural log of the ratio of split-adjusted shares outstanding at the end of December of year $t-1$ to split-adjusted shares outstanding at the end of December of year $t-2$. The split-adjusted shares outstanding is equal to Compustat shares outstanding time by the Compustat adjustment factor.[27] Composite equity issuance is the amount of equity a firm issues in exchange for cash and services, i.e. share repurchases and SEOs.

Titman et al. (2004) document that firms with high past investment-to-assets may give rise to lower returns. Following Stambaugh et al. (in press), investment-to-assets is measured by the annual change in gross property, plant, and equipment plus the annual change in inventories scaled by the lagged book value of assets. Fama and French (2006) and Chen et al. (2010) find that firms with high past return-on-assets have abnormally higher returns. Jegadeesh and Titman (1993) find that longing firms that have performed well in the past and shorting stocks that have performed poorly in the past can generate significant positive returns over next 3 to 12 months.[28]

For each anomaly (except for failure probability), at the end of June in year $t$, we sort stocks based on the measures related to anomalies and then allocate them into 10 groups. For the anomaly of failure probability, we use monthly rebalancing and remove all the financial companies in compiling the anomaly pertaining to the failure probability. Following Chen et al. (2010) and Stambaugh et al. (in press), we compute the monthly value-weighted portfolio returns for each group, and then obtain the returns of anomalies that are long in the stocks within the highest-performing decile and short in those within the lowest-performing decile.

Panel A of Table 10 reports the mean returns, the CAPM alphas, and the 3-factor alphas (and $t$-statistics) for portfolio returns related to the 11 anomalies. Consistent with the results of Stambaugh et al. (in press), these 11 anomalies are not captured by either the CAPM or the three factor model of Fama and French (1993) because alphas are significantly different from zero.

To explore the predictive power of investor sentiment across economic regimes for 11 anomalies, we perform the following predictive regressions using regime-sorted data as identified by the NBER as well as the Markov-switching model

$$r_{(i,h),t}-r_{(i,l),t}=\alpha_i+\gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1}+\beta_{i,\circ}\mathrm{RMRF}_t+\gamma_{i,2}\mathrm{SMB}_t+\gamma_{i,3}\mathrm{HML}_t+\epsilon_{i,t}, \tag{10}$$

where $r_{(i,h),t}-r_{(i,l),t}$ is the long-short portfolio return that longs portfolio in the highest-performing decile $h$ and shorts portfolio in the lowest-performing decile $l$ with anomaly $i$. Panel B of Table 10 reports the coefficient estimates on sentiment of the predictive regressions and the Newey–West $p$-values (based on one-tailed test as suggested by Stambaugh et al. (in press)) in parentheses. The first column, without sorting observations by NBER states or regimes, shows that the coefficient estimates, $\hat{\gamma}_{i,1}$, are positive and significant for the anomalies associated with failure probability, Ohlson's O-score, net stock issues, composite equity issues, net operating assets, gross profitability, and return-on-assets. These results are also consistent with the findings of Stambaugh et al. (in press).

Crucially, by controlling for the economic regime identified by either the NBER business cycles or the Markov-switching model, we find that the predictability pattern associated with sentiment is regime-dependent. The overall evidence suggests that investor sentiment has stronger predictive power for long-short portfolio returns of the anomaly strategies in the NBER expansion state (regime 2) than in the recession state (regime 1).

## 6. Conclusion

In this study we examine, across different states of the economy, the asymmetry in the predictive power of investor sentiment on the cross-section of stock returns. In addition to characterizing the economic regime by the NBER recession index, we also implement a multivariate Markov-switching model to characterize two economic regimes, the expansion state and the recession state. We then use the sentiment measures to forecast the returns of stock portfolios formed on firm size, age, return volatility, R&D expense, fixed assets, sales growth, external finance, book-to-market ratio, earnings-to-price ratio, dividend yields, and 11 well-documented anomalies conditional on the identified economic states.

We find strong evidence, both in-sample and out-of-sample, that the predictive power of investor sentiment is regime-dependent. Only under the expansion regime investor sentiment exhibits predictive power for the returns of portfolios formed

[27] It is worth noting that Pontiff and Woodgate (2008) measure the split-adjusted shares outstanding as CRSP shares outstanding divided by cumulative total factor to adjust shares outstanding from CRSP.

[28] The momentum and ROA anomalies are downloaded from Kenneth French's and Long Chen's websites, respectively.

**Table 10**
Mean returns, alphas, and predictive regressions for 11 anomalies.

Panel A: Mean returns and alphas

| | Mean return | CAPM alpha | 3-factor alpha |
|---|---|---|---|
| Failure probability | 1.00(2.28) | 0.96(2.56) | 1.62(4.16) |
| Ohlson's O-Score | 0.71(2.25) | 0.74(2.35) | 0.94(3.15) |
| Net stock issues | 0.65(3.75) | 0.80(4.95) | 0.66(4.57) |
| Comp. equity issues | 0.53(4.00) | 0.64(4.96) | 0.56(4.90) |
| Total accruals | 0.68(3.24) | 0.77(3.68) | 0.70(3.36) |
| Net operating assets | 0.66(3.27) | 0.80(4.03) | 0.90(4.47) |
| Momentum | 1.51(5.26) | 1.58(5.49) | 1.74(5.97) |
| Gross profitability | 0.35(2.08) | 0.43(2.57) | 0.70(4.34) |
| Asset growth | 0.78(4.13) | 0.93(5.19) | 0.45(2.94) |
| Return on assets | 1.00(5.16) | 1.07(5.58) | 1.03(5.59) |
| Investment/assets | 0.78(4.44) | 0.84(4.78) | 0.52(3.23) |

Panel B: Predictive regressions

| Long-short | All | NBER recession index | | Markov-switching model | |
|---|---|---|---|---|---|
| | | Rec. | Exp. | Regime 1 (Rec.) | Regime 2 (Exp.) |
| Failure probability | 0.94*(0.02) | −1.09(0.18) | 1.27**(0.01) | 3.12(0.16) | 1.20**(0.01) |
| Ohlson's O-Score | 1.37**(0.00) | 0.18(0.43) | 1.98**(0.00) | 0.76(0.25) | 1.50**(0.00) |
| Net stock issues | 0.32*(0.02) | 0.06(0.42) | 0.24(0.10) | −0.47(0.27) | 0.29*(0.03) |
| Comp. equity issues | 0.19*(0.04) | −0.15(0.25) | 0.21(0.06) | −0.12(0.42) | 0.16(0.14) |
| Total accruals | −0.02(0.45) | −0.21(0.33) | −0.04(0.45) | −0.10(0.41) | 0.16(0.20) |
| Net operating assets | 0.74**(0.00) | 0.59**(0.00) | 0.78**(0.00) | 1.02**(0.01) | 0.38**(0.01) |
| Momentum | 0.16(0.26) | 0.52(0.16) | 0.27(0.18) | 0.10(0.42) | 0.09(0.38) |
| Gross profitability | 0.30*(0.02) | 0.49(0.17) | 0.36*(0.02) | 0.54(0.08) | 0.19(0.14) |
| Asset growth | 0.17(0.13) | 0.11(0.34) | 0.25(0.11) | 0.15(0.31) | 0.08(0.29) |
| Return on assets | 0.56**(0.01) | 1.14**(0.01) | 0.56*(0.04) | 1.24**(0.00) | 0.22(0.06) |
| Investment/assets | 0.08(0.30) | −0.02(0.52) | 0.12*(0.04) | −0.10(0.67) | 0.21(0.12) |

Note: Panel A reports the mean returns, CAPM's alphas, and 3-factor's alphas for 11 portfolios related to the anomalies and $t$-statistics are in the parentheses. Panel B reports the coefficient estimates on sentiment of the predictive regression and the bootstrapped $p$-values in the parentheses. The regressions of portfolio returns associated with anomalies on the lagged SENTIMENT$^{\perp}$, the market factor (RMRF), and the Fama–French factors (HML and SMB)

$$r_{(i,h),t} - r_{(i,l),t} = \alpha_i + \gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1} + \beta_{i,\circ}\mathrm{RMRF}_t + \gamma_{i,2}\mathrm{SMB}_t + \gamma_{i,3}\mathrm{HML}_t + \varepsilon_{i,t},$$

where $r_{(i,h),t} - r_{(i,l),t}$ is the long-short portfolio return that longs portfolio in the highest-performing decile $h$ and shorts portfolio in the lowest-performing decile $l$ with firm characteristic $i$ (the momentum factor will be excluded when we perform the predictive regression for the anomaly associated with momentum). This table only reports the parameter estimates of $\gamma_{i,1}$. The sample period for the anomalies associated with asset growth, total accruals, net operating assets, and growth profitability is from July 1967 to December 2007 and those associated with net stock issues and composite equity issues begin by July 1975 and end by December 2007. The anomalies of return-on-assets and investment-to-assets begin by January 1972 and end by December 2007. Failure probability and O-score begin by January 1976 and January 1979, respectively, and end by December 2007. Momentum is from January 1966 to December 2007. The portfolio returns are in percentage. SENTIMENT$^{\perp}$ is the Baker and Wurgler's orthogonalized sentiment proxy. The column "All" reports the results without regime sorting. The other columns "Rec." and "Exp." ("Regime $j$", $j=1,2$) show the results based on NBER dummy-sorted (regime-sorted) observations as recessions and expansions (regime$=j$). The Newey–West $p$-values (lagged terms are determined by $0.75T^{1/3}$, where $T$ is the sample length) are in parentheses. $*$ and $**$ denote significance at 5% and 1% levels, respectively.

on firm characteristics and the anomalies. Time-variation in the market beta driven by investor sentiment cannot account for the predictive ability of sentiment.

## Appendix A. The bootstrap test procedure

We use the predictive regression model (4) to illustrate the testing procedure. The implementation is as follows:

- Step 1: We run the following regression model for returns of the $k$th portfolio with firm characteristic $i$:

$$r_{(i,k),t} = \alpha^{\dagger}_{(i,k)} + \beta^{\dagger}_{(i,k),\circ}\mathrm{RMRF}_t + \gamma^{\dagger}_{(i,k),1}\mathrm{SENTIMENT}^{\perp}_{t-1} + \gamma^{\dagger}_{(i,k),2}\mathrm{SMB}_t + \gamma^{\dagger}_{(i,k),3}\mathrm{HML}_t + \epsilon_{(i,k),t}, \tag{A.1}$$

and save all OLS-estimated risk loadings $\left\{\hat{\alpha}^{\dagger}_{(i,k)}, \hat{\beta}^{\dagger}_{(i,k),\circ}, \hat{\gamma}^{\dagger}_{(i,k),1}, \hat{\gamma}^{\dagger}_{(i,k),2}, \hat{\gamma}^{\dagger}_{(i,k),3}\right\}$, residuals $\{\hat{\epsilon}_{(i,k),t}, t = T_0, \ldots, T_n\}$ for all stock portfolios $k = 1,\ldots,N$, where $T_0$ and $T_n$ are the dates of the first and last observations.

- Step 2: Denote the cross-section of residuals at time $t$ by $\hat{\Upsilon}_{i,t} = \left(\hat{\epsilon}_{(i,1),t}, \ldots, \hat{\epsilon}_{(i,N),t}\right)'$. We re-sample a sequence of the time indices $s_{T_0}{}^b, \ldots, s_{T_n}{}^b$ that are drawn randomly from $[T_0, \ldots, T_n]$, where $b$ is the index for the bootstrap sample (for example, $b = 1$ means the re-sample number one). The sequence of the re-sampled residual vectors are given by $\left\{\hat{\Upsilon}^b_{i,t_\varepsilon}, t_\varepsilon = s^b_{T_0}, \ldots, s^b_{T_n}\right\}$, where $\hat{\Upsilon}^b_{i,t_\epsilon} = \left(\hat{\epsilon}^b_{(i,1),t_\epsilon}, \ldots, \hat{\epsilon}^b_{(i,N),t_\epsilon}\right)'$.

- Step 3: The pseudo portfolio returns are constructed under the null hypothesis $\gamma_{(i,k),1} = 0$ by

$$r^b_{(i,k),t_\varepsilon} = \hat{\alpha}^\dagger_{(i,k)} + \hat{\beta}^\dagger_{(i,k),\circ}\mathrm{RMRF}_{t_\varepsilon} + \hat{\gamma}^\dagger_{(i,k),2}\mathrm{SMB}_{t_\varepsilon} + \hat{\gamma}^\dagger_{(i,k),3}\mathrm{HML}_{t_\epsilon} + \hat{\epsilon}_{(i,k),t_\epsilon{}^b}, \tag{A.2}$$

for $k = 1, \ldots, N$ and $t_\varepsilon = s_{T_0}{}^b, \ldots, s_{T_n}{}^b$. So, there are $N$ re-sampled time series of stock portfolio returns in a re-sampling, $\{r_{(i,k),t_\varepsilon}{}^b$, $t_\varepsilon = s_{T_0}{}^b, \ldots, s_{T_n}{}^b\}$, $k = 1, \ldots, N$.

- Step 4: We run the predictive regression model (4) using the re-sampled data generated in Step 3, and compute the corresponding $t$-statistic, $t\left(\hat{\gamma}^b_{i,1}\right)$, for $\gamma_{i,1}$. Repeating Steps 2 and 3 for $M$ times (the largest $b$ is M), a bootstrap distribution of $t$-statistic of $\gamma_{i,1}$, $\left\{t\left(\hat{\gamma}^b_{i,1}\right), b = 1, \ldots, M\right\}$, under the null hypothesis ($\gamma_{i,1} = 0$) is available. We set $M = 20{,}000$ in our empirical studies.
- Step 5: We compute the $p$-value associated with $t$ -statistic by comparing $\sqrt{T}\cdot t\left(\hat{\gamma}_{i,1}\right)$ to the quintiles of $\sqrt{T}\cdot\left[t\left(\hat{\gamma}^b_{i,1}\right) - t\left(\hat{\gamma}_{i,1}\right)\right]$ to obtain the $p$-value. Note that $T$ is the total sample size between $T_0$ and $T_n$, $t\left(\hat{\gamma}_{i,1}\right)$ and $t\left(\hat{\gamma}^b_{i,1}\right)$ are the $t$-statistics computed by real data and the $b$-th re-sampled data, respectively. The bootstrapped $p$-value may be defined as the probability in favor of the null hypothesis

$$\mathrm{Prob}\left(\sqrt{T}\left[t\left(\hat{\gamma}^b_{i,1}\right) - t\left(\hat{\gamma}_{i,1}\right)\right] > \sqrt{T}t\left(\hat{\gamma}_{i,1}\right)\right).$$

The readers can refer to Sullivan et al. (1999) for the details of implementing bootstrap methods.

## References


Andrew, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Ang, A., Bekaert, G., 2002. International asset allocation with regime shifts. Rev. Financ. Stud. 15, 1137–1187.
Ang, A., Chen, J., 2002. Asymmetric correlations of equity portfolios. J. Financ. Econ. 63, 443–494.
Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. J. Finance 61, 1645–1680.
Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. J. Econ. Perspect. 21, 129–151.
Bris, A., Goetzmann, W., Zhu, N., 2007. Efficiency and the bear: short sales and markets around the world. J. Finance 62, 1029–1079.
Brown, G., Cliff, M., 2004. Investor sentiment and the near-term stock market. J. Empir. Finance 11, 1–27.
Brown, G., Cliff, M., 2005. Investor sentiment and asset valuation. J. Bus. 78, 405–440.
Butler, A., Grullon, G., Weston, J., 2005. Can managers forecast aggregate market returns? J. Finance 60, 963–986.
Campbell, J., Thompson, S., 2008. Predicting excess stock returns out of sample: can anything beat the historical average. Rev. Financ. Stud. 21, 1509–1531.
Campbell, J., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. J. Finance 63, 2899–2939.
Chang, E., Cheng, J., Yu, Y., 2007. Short-sale constraint and price discovery: evidence from the Hong Kong market. J. Finance 62, 2097–2121.
Chen, L., Novy-Marx, R., Zhang, L., 2010. An Alternative Three-Factor Model. University of Chicago. Working Paper.
Clark, T., West, K., 2007. Approximately normal tests for equal predictive accuracy in nested models. J. Econom. 138, 291–311.
Cooper, M., Gulen, H., Schill, M., 2008. Asset growth and the cross-section of stock returns. J. Finance 63, 1609–1652.
Daniel, K., Titman, S., 2006. Market reactions to tangible and intangible information. J. Finance 61, 1605–1643.
Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. J. Finance 53, 1839–1886.
De Long, J., Shleifer, A., Summers, L., Waldmann, R., 1990. Noise trader risk in financial markets. J. Polit. Econ. 98, 703–738.
Dechow, P., Hutton, A., Meulbroek, L., Sloan, R., 2001. Short-sellers, fundamental analysis, and stock returns. J. Financ. Econ. 61, 77–106.
Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33, 3–56.
Fama, E., French, K., 2006. Profitability, investment, and average returns. J. Financ. Econ. 82, 491–518.
Fama, E., French, K., 2008. Dissecting anomalies. J. Finance 63, 1653–1678.
Figlewski, S., Webb, G., 1993. Options, short sales, and market completeness. J. Finance 48, 761–777.
Gervais, S., Odean, T., 2001. Learning to be overconfident. Rev. Financ. Stud. 14, 1–28.
Guidolin, M., Timmermann, A., 2008. Size and value anomalies under regime shifts. J. Financ. Econom. 6, 1–48.
Gulen, H., Xing, Y., Zhang, L., 2008. Value Versus Growth: Time-Varying Expected Returns. University of Michigan. Working Paper.
Hamilton, J., 1990. Analysis of time series subject to changes in regimes. J. Econom. 45, 39–70.
Hardouvelis, G., Theodossiou, P., 2002. The asymmetric relation between initial margin requirements and stock market volatility across bull and bear markets. Rev. Financ. Stud. 15, 1525–1559.
Hirshleifer, D., Hou, K., Teoh, S., Zhang, Y., 2004. Do investors overvalue firms with bloated balance sheets? J. Account. Econ. 38, 297–331.
Ho, C., Hung, C., 2009. Investor sentiment as conditioning information in asset pricing. J. Bank. Finance 33, 892–903.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. J. Finance 48, 65–91.
Jones, C., Lamont, O., 2002. Short-sale constraints and stock returns. J. Financ. Econ. 66, 207–239.
Kosowski, R., Timmermann, A., Wermers, R., White, H., 2006. Can mutual fund stars really pick stocks? new evidence from a bootstrap analysis. J. Finance 61, 2551–2595.
Kumar, A., Lee, C., 2006. Retail investor sentiment and return comovements. J. Finance 61, 2451–2486.

Lamont, A., 2005. Short-sale constraints and overpricing. In: Fabozzi, F.J. (Ed.), The Theory and Practice of Short Selling.
Lamont, A., Thaler, R., 2003. Can the market add and subtract? mispricing in tech stock carve-outs. J. Polit. Econ. 111, 227–268.
Lee, C., Shleifer, A., Thaler, R., 1991. Investor sentiment and the closed-end fund puzzle. J. Finance 46, 75–109.
Lemmon, M., Portniaguina, E., 2006. Consumer confidence and asset prices: some empirical evidence. Rev. Financ. Stud. 19, 1499–1529.
Loughran, T., Ritter, J., 1995. The new issues puzzle. J. Finance 50, 23–51.
Mitchell, M., Pulvino, T., Stafford, E., 2002. Limited arbitrage in equity markets. J. Finance 57, 551–584.
Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. J. Financ. Econ. 78, 277–309.
Novy-Marx, R., 2010. The Other Side of Value: Good Growth and the Gross Profitability Premium. University of Chicago. Working Paper.
Ofek, E., Richardson, M., 2003. DotCom mania: the rise and fall of internet stock prices. J. Finance 58, 1113–1137.
Ofek, E., Richardson, M., Whitelaw, R., 2004. Limited arbitrage and short sale restrictions: evidence from the options market. J. Financ. Econ. 74, 305–342.
Ohlson, J., 1980. Financial ratios and the probabilistic prediction of bankruptcy. J. Account. Res. 18, 109–131.
Ozoguz, A., 2009. Good times or bad times? investors' uncertainty and stock returns. Rev. Financ. Stud. 22, 4377–4422.
Perez-Quiros, G., Timmermann, A., 2000. Firm size and cyclical variation in stock returns. J. Finance 55, 1229–1262.
Pontiff, J., Woodgate, A., 2008. Share issuance and cross-sectional returns. J. Finance 63, 921–945.
Schmeling, M., 2009. Investor sentiment and stock returns: some international evidence. J. Empir. Finance 16, 394–408.
Schwert, W., 1990. Stock returns and real activity: a century of evidence. J. Finance 45, 1237–1257.
Shleifer, A., Vishny, R., 1997. The limits of arbitrage. J. Finance 52, 35–55.
Sloan, R., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? Account. Rev. 71, 289–315.
Stambaugh, R., 1999. Predictive regressions. J. Financ. Econ. 54, 375–421.
Stambaugh, R., Yu, J., Yuan, Y., in press. The short of it: Investor sentiment and anomalies. J. Financ. Econ. doi:10.1016/j.jfineco.2011.12.001.
Stock, J., Watson, M., 2007. Introduction to Econometrics, second ed. Pearson Addison-Wesley.
Sullivan, R., Timmermann, A., White, H., 1999. Data snooping, technical trading rule performance and the bootstrap. J. Finance 54, 1647–1692.
Titman, S., Wei, J., Xie, F., 2004. Capital investments and stock returns. J. Financ. Quant. Anal. 39, 677–700.
Whitelaw, R., 2000. Stock market risk and return: an equilibrium approach. Rev. Financ. Stud. 13, 521–574.
Yu, J., Yuan, Y., 2011. Investor sentiment and the mean-variance relation. J. Financ. Econ. 100, 367–381.