When Does Investor Sentiment Predict Stock Returns?

San-Lin Chung*, Chi-Hsiou Hung†, and Chung-Ying Yeh‡

*Professor of Finance, National Taiwan University, Department of Finance, 85, Section 4, Roosevelt Road, Taipei 106, Taiwan. Tel: 886-2-3366-1084. Email: chungs@management.ntu.edu.tw.

†Hung is with Durham Business School, Durham University, Mill Hill Lane, Durham, DH1 3LB. Tel:44-191-334-5498, E-mail: d.c.hung@durham.ac.uk

‡Yeh is with Department of Finance, Tunghai University, 181, Section 3, Taichung-Kan Road, Taichung, Taiwan. Tel:886-4-2359-0121 ext.35819. Email: cyyeh@thu.edu.tw

When Does Investor Sentiment Predict Stock Returns?

Abstract

We examine the predictive effect of sentiment on the cross-section of stock returns across different economic states. The regime-switching feature on stock returns may cause a problem in identifying the source of the return predictability. In addition, the investors' uncertainty about the state of the economy predicts the presence of asymmetries in the predictive ability of sentiment over different economic states. We implement a multivariate Markov-switching model to characterize the economic states. Conditional on the identified economic states, we use the lagged sentiment proxy to forecast the portfolio returns related to small stocks, non-earning stocks, growth stocks, and non-dividend-paying stocks. We find that sentiment performs both in-sample and out-of-sample predictive power on these categories of stocks only in the expansion state. When a expansion state has high (low) sentiment, these categories of stocks earn relatively low (high) subsequent returns. The predictive ability of sentiment can not be attributed to time-variation in the market beta driven by investor sentiment. The investors' uncertainty about the economy can explain the time-variations on the cross-section of stock return in the recession state.

Keywords: Investor Sentiment; Stock Returns; Return Predictability; Markov-Switching Vector Autoregressive Model; Bootstrap

JEL classification code: E32; G11; G12; G14

1 Introduction

Many studies have shown that investor sentiment impacts stock prices and causes mispricing (e.g., De Long, Shleifer, Summers and Waldman (DSSW), 1990; Brown and Cliff, 2005; and Baker and Wurgler, 2006). The reversal of stock prices from the mispriced levels to their fundamentals implies that sentiment is negatively correlated with future stock returns. As a result, investor sentiment may exhibit predictive ability for stock returns (see, for example, Swaminathan (1996), Brown and Cliff (2004), Qiu and Welch (2004), Baker and Wurgler (2006, 2007), Glushkov (2006), Kumar and Lee (2006), and Lemmon and Portniaguina (2006)).

There is mounting empirical evidence, however, showing that conditional distributions of stock returns are subject to regime shifts. The expected returns, volatilities, and correlations of stock returns vary with regimes. Gray (1996), Maheu and McCurdy (2000), Perez-Quiros and Timmermann (2000), Ang and Chen (2002), and Guidolin and Timmermann (2008) are some examples. Thus, considering a regime-switching feature on stock returns is empirical appealing. We conjecture that the regime-switching feature poses a significant challenge in the literature pertaining to the predictive ability of sentiment on stock returns. This challenge can be understood in two ways.

First of all, we argue that if stock returns exhibit regime shifts, we would be difficult to disentangle the price changes as a correction of a mispricing due to sentiment from an adjustment in relation to the changes in regimes. Consequently, an identification problem may occur in detecting the predictive ability of sentiment. Consider, for example, during a bubble period bullish sentiment may give rise to an overpricing. When the bubble bursts, stock prices fall and a recession regime may be triggered (see, for example, Ofek and Richardson (2003), Abreu and Brunnermeier (2003), Brunnermeier and Nagel (2004) and Pastor and Veronesi (2006)). The decrease in stock prices might reflect a correction for mispricing as sentiment wanes and/or a mean-reverting process as the regime switches to a recession one, characterized by low levels of stock returns with high volatility. Therefore, it would be a problem in identifying the source of the return predictability.

Second, we argue that the investors' uncertainty about the state of the economy predicts the presence of asymmetries in the predictive ability of sentiment over different economic states. Veronesi (1999, 2004) advocated that when there is more uncertainty about the economy, investors' beliefs are prone to react more to news. As a consequence, the expected future volatility and required return for the stock are expected to rise, driving down the stock price. This feedback effects due to the sensitivity of investors' beliefs to news on the stock price might generate asymmetries in the predictive ability of sentiment on stock returns across different economic states. For example, suppose that the current economy is in the recession state and π_t, the conditional probability that the economy is staying in the expansion state, is assumed to be 0.1 in reflecting the idea that investors believe times are bad. Bullish sentiment may give rise to an increase in the stock price. This increase can be regarded as a positive price shock and drive up π_t close to 0.5, the point of maximum uncertainty about the economy. Hence, the overpricing caused by bullish sentiment may be offset by the stock price drop due to the increased uncertainty. In contrast, when the economy is in the expansion state, π_t is assumed to be 0.9 because investors believe times are good. The increase in stock prices caused by bullish sentiment is seen as a positive price shock and then π_t approaches one, the point without uncertainty. This decreased uncertainty may further increase stock prices. The overpricing caused by bullish sentiment is strengthened by the decreased uncertainty. Interestingly, in light of Veronesi (1999, 2004), investors' uncertainty over the economy might cause an overreaction to mispricing in good times and an underreaction in bad times, generating asymmetries in the predictive effect of sentiment across different economic states.

Baker and Wurgler (2006, 2007) and Glushkov (2006) devise a *hard-to-value and difficult-to-arbitrage hypothesis* to explain the cross-sectional effect of sentiment associated with firm characteristics, in particular for small size, non-earnings, growth, and non-dividend-paying stocks. Since these stocks do not have a long enough history of earnings, tangible assets and collateral, their assessments are highly subjective. Therefore, these stocks are strongly affected by fluctuations in the propensity of speculation. In addition, since these stocks tend to have higher idiosyncratic risk and lower liquidity, arbitrage is particularly risky and costly. The mispricing induced by variations in sentiment is hard to be immediately corrected. Moreover, the asymmetry in the predictive effect of sentiment over economic states is particularly so for these stocks. It is intuitive that since these stocks not only have no enough tangible assets and collateral but also perform low liquidity and high idiosyncratic volatility, their prices tend to react more to news when there is more investors' uncertainty to the economy. In contrast, larger stocks which have valuable tangible assets and collateral, high liquidity and low idiosyncratic volatility are expected to be less affected when there is more uncertainty.

The goal of this paper is to explore the predictive ability of investor sentiment on the cross-section of stock returns across different economic states. In the meanwhile, we tackle the identification problem caused by regime shifts when we examine the predictive effect of sentiment on stock returns. These issues are where this paper differs from the previous research and have not, to our best knowledge, received formal research attention. Prior research has studied the relation between investor sentiment and regime-shifts in stock fundamentals via investors' distorted predictions about the fundamentals (Barberis, Shleifer, and Vishny (1998); and Cecchetti, Lam, and Mark (2000)). They do not, however, explore the return predictability of investor sentiment when stock returns are subject to regime shifts.

We follow Perez-Quiros and Timmermann (2000) and Ozoguz (2009) and perform a Markov-switching model with a time-varying state transition matrix to characterize economic regimes. The Markov-switching model has widely been used to characterize occasional, but recurrent, regime shifts related to business cycles. The regime shifts characterized by the Markov-switching model may be attributed to the changes in fundamentals and variations in investor sentiment. For example, the events of the oil shocks or Russian defaults, which led to recessions, are fundamental shifts, whereas the beginning of the Internet bubble in the mid-1990s was sentiment-oriented. In light of the above evidence, we use monthly returns of the market, SMB, and HML portfolios that are orthogonal to the sentiment variation in our empirical implementation in order to remove the regime shifts that are related to the changes in investor sentiment.

Over a sample period of 40 years, from January 1966 to December 2005, we estimate the parameters of the Markov-switching model and identify two economic regimes. Regime 1 is the high-volatility, recession state that captures two oil shocks in the 1970s, the Gulf War in the beginning of the 1990s, the default of Russia sovereign debts and the crash of the Long Term Capital Management (LTCM) in 1998, and the burst of the internet bubble and corporate malfeasance since the beginning of the 2000s. Regime 2 is the low-volatility, expansion state that covers most bull markets since the 1960s, including the run-ups during the 1980s and 1990s.

To examine the predictive effect of sentiment as well as control for the effect of regime shifts, we perform predictive regressions using the regime-sorted data. Our predictive regressions regard the states of regimes as exogenous inputs. We test if the parameter coefficients associated with sentiment are affected by the states of regimes. We employ the lagged investor sentiment, which we proxy by the Baker and Wurgler's (2006) orthogonalized sentiment index, to predict the equally- weighted returns of the long-short portfolios that are long in stocks with high values of a characteristic and short in stocks with low values.[1] We also control for the Fama-French (1993) factors, the momentum factor, and the Pastor and Stambough's (2003) liquidity factor. We use monthly equal-weighted portfolio returns formed on size, book-to-market, dividend yield, and earnings/price. We use the bootstrap testing procedure of Kosowski, Timmermann, Wermers, and White (2006) to calculate the bootstrapped p-values for the empirically computed t-statistic. This bootstrap procedure can enhance the testing power when the sample size is not large.

We find that investor sentiment exhibits predictive power for the cross-section of stock returns, in particular for small size, non-earnings, growth, and non-dividend-paying stocks. Importantly, we find that only under regime 2 (the expansion state), does sentiment show a significant and robust predictive power for the subsequent returns of small stocks, growth stocks, non-earning stocks, and non-dividend-paying stocks. When the economy is in the expansion state, higher sentiment is associated with lower subsequent returns for these

[1] The Baker and Wurgler's theory predicts that large firms will be less affected by sentiment, and hence value-weighted average returns may obscure the identification of relevant patterns.

stocks. Our findings highlight the asymmetry in the predictive ability of sentiment on stock returns over economic states.

For robustness checks, we use an alternative sentiment measure–consumer confidence provided by the Survey Research Center of the University of Michigan. Qiu and Welch (2004) and Lemmon and Portniaguina (2006) document that consumer confidence is an adequate proxy for investor sentiment and is able to predict the subsequent portfolio returns for small stocks. We also conduct an out-of-sample test, proposed by Clark and West (2007), for the predictive effect of sentiment across the states of market regimes. The results suggest that sentiment performs both in-sample and out-of-sample predictive power on the subsequent returns for small stocks, growth stocks, non-earning stocks, and non-dividend-paying stocks under regime 2 (the expansion state). We also consider the predictive regression with regime dummy variables. This predictive regression includes the regime dummies independently in addition to making sentiment loadings conditional on dummies. The results suggest that our main findings continue to hold.

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 due to time variation in systematic risk or mispricing. We show that the predictability patterns associated with sentiment reflect a mispricing/correction pattern, rather than time variation in the market beta, even after conditioning on the state of the economy. We then test whether asymmetries in the predictive effect of sentiment is associated with investors' uncertainty. The results suggest that investors' uncertainty can explain the time-variations on the cross-section of stock returns in the recession state. This simply sustains our conjecture.

The remainder of this paper is organized as follows. Section 2 introduces the multivariate Markov-switching model and characterizes market regimes. The data and the empirical tests are given in Section 3. Two robustness checks are reported in Section 4. The discussions about the source of asymmetries are in section 5. Section 6 concludes the paper.

2 Characterizing Economic Regimes

2.1 A Multivariate Markov-Switching Model for Stock Returns

To characterize economic regimes, we follow Perez-Quiros and Timmermann (2000) and Ozoguz (2009) and estimate a Markov-switching model with time-varying regime transition probabilities. The state of regime will switch at random times but be driven by a latent regime variable following a Markov chain that is assumed to change over time in our setting.

Recent research has advocated that the size and value premiums vary with the state of economic regime; see for example, Perez-Quiros and Timmermann (2000), Cooper, Gutierrez, and Hameed (2004), and Gulen, Xing, and Zhang (2008). In conjunction with this evidence, we follow Guidolin and Timmermann (2008) and characterize economic regimes in the joint process of returns on the portfolios of the market, size and value factors as in Fama and French (1993). These factors are the monthly returns on the market (the CRSP value weighted market index) in excess of the T-bill rate and the monthly returns on the SMB and HML portfolios. The SMB factor is the difference in average returns on the small-size stock portfolios and the big-size stock portfolios. The HML factor is the difference between the average returns on the high book-to-market portfolios and the average returns on the low book-to-market portfolios.

In order to disentangle the regime shifts caused by the shifts in the fundamentals from a potential impact on regimes due to the corresponding shifts in investor sentiment, we remove the sentiment variation from the factor portfolio returns prior to implementing the Markov-switching model. Specifically, we regress the returns of each of three foactor portfolios on the sentiment proxy–the Baker and Wurgler's (2006) orthogonalized sentiment index.[2] The residuals from these regressions, labelled with a superscript $\perp$, are the factor portfolio returns orthogonalized to the sentiment variation. The sample of orthogonalized

[2]Baker and Wurgler (2006) regress each of the six raw proxies on macroeconomic variablels to remove the effects from business cycle variation prior to conducting the principal components analysis. As a result, this sentiment index is orthogonalized to business cycle fluctuations. The descriptions of the sentiment data can refer to section 3.1.

factor portfolios covers the 480-month sample period from January 1966 to December 2005, which is dictated by the availability of the sentiment index. The mean returns for the unorthogonalized market, SMB, and HML portfolios are, respectively, around 0.4%, 0.3%, and 0.4% per month with volatility of 4.5%, 3.3%, and 3.0% per month. The portfolio returns are all skewed and leptokurtic.[3] The orthogonalized factors retain the appealing properties of the unorthogonalized ones.

We model the joint distribution of a 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 $\{1, 2\}$. The specification of the Markov-switching model is:

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

where $\mathbf{X}_{t-1}$ is the vector of publicly available information used to predict stock returns. Following Perez-Quiros and Timmermann (2000) and Ozoguz (2009), $\mathbf{X}_{t-1}$ comprises dividend yield, default premium, and the short-term interest rate. μ_{s_t} is the 3×1 vector of the regime-dependent intercepts, $\mathbf{\Phi}_{s_t}$ is the 3×3 matrix of the regime-dependent coefficients. The vector of return innovations $\varepsilon_t \sim N\left(0, \mathbf{\Omega_{s_t}}\right)$ is assumed to follow a multivariate normal distribution with zero means and a regime-dependent variance-covariance matrix $\mathbf{\Omega_{s_t}}$. The discrete regime variable s_t is assumed to follow a 2−state first-order Markov chain governed by a 2×2 transition probability matrix with time-varying elements

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

where ΔCLI_{t-1} is the one-month lagged value of the change in log composite leading indicator[4] and $N(\cdot)$ is the cumulative density function of a standard normal variable. The regime variable s_t, as a latent variable, can be statistically inferred by the realized observations. The possibilities of the regimes at each time point can be characterized by filtered probabilities $\Pr(s_t = j|\mathbf{Y}^t)$ and smoothed probabilities $\Pr(s_t|\mathbf{Y}^T)$, where $\mathbf{Y}^t$ is

[3] We thank Kenneth French for making the data available at the web-page: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

[4] The composite leading indicator data can be downloaded from the OECD's web-page.

the information set at time t and $\mathbf{Y}^T$ is the complete information set.

Following the standard convention in the literature, the dividend yield is defined as the dividends on the valued-weighted CRSP index over past 12 months. We constructed the dividend payout series using the valued-weighted return including dividends, and the price index series associated with the value-weighted return excluding dividends. We took the dividend series to be the sum of dividend payout over past 12 months. The default premium is defined as the yield spread between Baa and Aaa corporate bonds. The short-term interest rate is defined as the 90-day bill rate. We obtained the bond yield data from the web-page of Federal Reserve at St. Louis.

2.2 Empirical Results of the Multivariate Markov-Switching Model

Before we proceed to implement the Markov-switching model of (1) and (2), an econometric issue needs to be sorted out. We are not able to estimate the Markov-switching model by maximum likelihood directly because the size of the model parameters would cause severe problems in numerical optimization. To tackle this problem, we adopt a pragmatic approach. That is, we estimate $\mu_1, \mu_2, \mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{\Omega_1}$, and $\mathbf{\Omega_2}$ by the EM algorithm devised by Hamilton (1989, 1990) conditional on the prior maximum likelihood estimates of a_1, a_2, b_1 and b_2. We then conduct the maximum likelihood estimation for a_1, a_2, b_1 and b_2 conditional on the prior estimates of the EM algorithm. We iteratively repeat this procedure until convergence. The standard deviations of the parameter estimates are calculated by the standard convention in the maximum likelihood approach.[5]

Table 1 reports the estimation results of the two-regime Markov-switching model. Panel A reveals the parameter estimates in the mean equation. Panel B presents the estimates of volatilities and correlations in the diagonals and the off-diagonals of the correlation matrices, respectively. The parameter estimates associated with the transition probabilities and the steady-state probabilities are shown in Panel C and D. To assist the economic interpretation of the two-state Markov-switching model, Figure 1 plots the historical patterns of the smoothed probabilities for the two regimes. To infer the state

[5]The complete estimation results are available upon request.

of the regime at each time point, the decision criterion is to examine which regime has smoothed probability above 0.7.[6] The interpretations of the regimes are as follows:

Regime 1 is a high-volatility recession state. The monthly volatility of the excess market return is 6.1% in this state. The estimate of the coefficient on the change in the composite leading indicator is significantly negative. The time variation in the transition probabilities highlights that an increase in the leading indicator would decrease the probability of staying in regime 1. Figure 1 shows that this regime 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 cover in the NBER recessions (the shaded areas). In addition, the correlation between the NBER recession indicator and the smoothed probability of regime is about 0.3. Regressing the smoothed probability of regime 1 on the NBER recession indicator comes up with a highly significant positive coefficient. The results related to the NBER recessions are reported in Table 2.

Regime 2 is a low-volatility expansion state. This state covers most of the bull markets with growing stock prices since the mid-1960s, including the run-ups in the 1980s and 1990s. The estimate of the coefficient on the change in the composite leading indicator is significantly positive, suggesting that an increase in the leading indicator brings up the probability of staying in regime 2. The monthly volatility of the excess market return is 3.6% in this state. The levels of volatilities for the factor portfolios of SMB and HML in regime 2 are approximately a half of those in regime 1. Moreover, the correlation between the NBER recession indicator and the smoothed probability of regime is negative. The regression coefficient of the smoothed probability of regime 1 on the NBER recession indicator is significantly negative.

To corroborate the economic interpretation, we sort industrial production growth rates

[6] According to this criterion, there are 114 points classified as regime 1 and 338 points regarded as regime 2. There are 28 points failing to identify the corresponding state of the regime. These points, in realty, involve the turning points of regimes.

based on the identified regimes. We find that the average monthly growth rate of industrial production is 0.05% in regime 1 and 0.3% in regime 2. The volatility of the industrial production growth rate in regime 1 is higher than that in regime 2. It confirms that the regimes characterized by the Markov-switching model are associated with underlying economic fundamentals.

For the mean parameters, the coefficient estimates of the excess market return on the default premium are significant and positive in both regimes. However, the coefficient estimates on the default premium are significantly positive in regime 1 (the recession state) for the SMB factor portfolio. The coefficient estimates on the default premium in the recession state are larger than those in the expansion state. As a result, the default premium is mainly important during the economic recessions and particular so for size premium. These findings are consistent with Perez-Quiros and Timmermann (2000). The coefficient estimates on the lagged interest rate are statistically significant and negative in both regimes for the excess market return. For the SMB factor portfolio, this coefficient estimates on the lagged interest rate are significantly negative in regime 2.

The HML portfolio is negatively correlated with the market portfolio and the SMB portfolio, whereas the correlation between the market portfolio and the SMB portfolio is positive. This may suggest that HML can serve as a hedge against the market portfolio.

3 Predictive Regressions for Long-Short Portfolios

3.1 Data and Sample

To implement the predictability tests, we use the monthly equally weighted returns on portfolios formed on firm characteristics of: (i) size (ME), (ii) book-to-market (BE/ME), (iii) dividend yield (D/P), and (iv) earnings/price (E/P).[7] All portfolios are constructed at the end of each June. In June of year t, all NYSE stocks are sorted by (i) ME, (ii) BE/ME, (iii) D/P, and (iv) E/P, respectively, to determine the decile breakpoints for each firm characteristic. ME is the June market equity of year t. BE/ME is book equity

[7]We obtained the data from Kenneth French's data library.

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 $t-1$. All NYSE, AMEX, and NASDAQ stocks are allocated based on the NYSE's breakpoints of each firm characteristic, forming the portfolios for July of year t to June of year $t+1$.

We use 10 deciles for size and book-to-market portfolios. For dividend yield and earnings/price, there are 11 portfolios in which portfolio returns for non-earning stocks or non-dividend-paying stocks are calculated separately (denoted as "≤ 0"). Table 3 reports the summary statistics for the monthly portfolio returns in percentage for the period from January 1966 to December 2005. The results show that small stocks and value stocks have higher mean returns than large stocks and growth stocks. The size and value premiums in terms of the mean return are approximately 0.60% and 1.28%, respectively. In contrast, the mean return patterns for the portfolios formed on dividend yield and earnings/price do not reveal the cross-sectional effects. The skewness of the middle decile groups is negative and smaller than the skewness of the low- and high-decile groups. The excess kurtosis is positive in all cases. The summary results strongly reject the normality property for the cross-section of stock returns, which is empirically related to the regime-switching feature in the stock returns and this could invalidate the conventional t-statistic in the predictive regressions.

We use the Baker and Wurgler's (2006) orthogonalized sentiment index in the monthly basis as the measure of investor sentiment which has been shown to exhibit the predictive ability for the cross-section of stock returns.[8] The top panel of Figure 2 plots the historical pattern of the Baker and Wurgler's (2006) orthogonal sentiment index over the period from

[8] The sentiment index of Baker and Wurgler (2006) is the first principal component of six major proxies for investor sentiment including the closed-end fund discount, NYSE share turnover, the number of IPOs, the average first-day returns on IPOs, the equity share in new issues, and the dividend premium. Both annual and monthly sentiment data are available from Jeffrey Wurgler's website at: http://pages.stern.nyu.edu/ jwurgler/.

January 1966 to December 2005. The sentiment index shows a spike before 1970's and then turns into negative for a long period during the 1970's which might be attributable to a series of oil crises and recessions. The sentiment index became positive in the 80's until the Gulf War in the early 90's. Investor sentiment reached a big spike before the Dot-Com bubble burst and became negative afterwards until 2003 when the market recovered.

In conjunction with the patterns of smoothed probabilities in Figure 1, we can find that high investor sentiment is followed by the change in the economic regime to a recession state. For example, the recession in the beginning of the 2000s occurs after the spike of the sentiment proxy at the end of the 1990s. Moreover, according to Table 2, the correlations of the lagged sentiment proxy on the smoothed probability of regime 1 (the recession state) and the NBER recession indicator are positive. The regression coefficient of the smoothed probability of regime 1 on the lagged sentiment is significantly positive. These results simply highlight that we will face the identification problem in detecting the source of the price change.

3.2 The Regression Models

Following Baker and Wurgler (2006), we use the lagged sentiment proxy to predict the equally-weighted returns for long-short portfolios that are long on stocks with high values of a characteristic and short on stocks with low values. We run the predictive regressions using the regime-sorted data to control for the effects of regime shifts,

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

where $r_{(i,k_2),t} - r_{(i,k_1),t}$ is the monthly return on a long-short portfolio that are long on portfolio k_2 and short on portfolio k_1 with firm characteristic i at time t, and $k_1, k_2 \in \{\leq 0, 1, 2, \ldots, 10\}$. "1", "2", ..., and "10" indicate the portfolios in the 1st (the smallest), 2nd, ..., and 10th deciles, respectively. "≤ 0" represents the portfolios of non-dividend-paying stocks or non-earnings stocks. $\text{SENTIMENT}^{\perp}_{t-1}$ is the Baker and Wurgler's (2006) orthogonalized sentiment index at time $t-1$ which removes the effects of macroeconomic fluctuations. Baker and Wurgler (2006) documented that, when the beginning-of-period sentiment is high, the subsequent portfolio returns of small stocks, growth stocks, non-dividend-paying stocks and non-earning stocks are lower than those of large size, value, dividend-paying and with-earnings stocks. Thus, sentiment forecasts the returns on a long-short portfolio formed on these characteristics and that the sign of the coefficient on $\hat{\gamma}_{i,1}$ in the regression model is expected to be positive.

We also distinguish predictability effects associated with firm characteristics using the multivariate regression,

$$\begin{aligned} r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\text{SENTIMENT}^{\perp}_{t-1} + \beta_{i,\circ}\text{RMRF}_t \\ + \gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \gamma_{i,4}\text{UMD}_t + \gamma_{i,5}\text{LIQ}_t + \epsilon_{i,t}, \end{aligned} \tag{4}$$

where RMRF_t is the return on the value-weighted market in excess of the risk-free rate. SMB_t and HML_t are the Fama-French factors. UMD_t is the return on high-momentum stocks minus the return on low-momentum stocks, where the measure for momentum is based on the cumulative raw return for the 11 months from 12 through 2 months prior to the return observation.[9] LIQ_t is the Pastor and Stambaugh's (2003) liquidity factor.[10] We include the liquidity factor to mitigate the concern that market liquidity is related to sentiment (see, e.g., Baker and Stein (2004) and Deuskar (2004)).

When we perform predictive regressions, we eliminate the observations at the turning points where regime actually switches from one state to another to completely control for the effect of regime shifts, which helps clearly identify the predictive ability of sentiment on the stock returns.

We perform a bootstrap testing procedure for drawing robust inferences, following Kosowski, Timmermann, Wermers, and White (2006). The Appendix contains details of the bootstrap testing procedure. This approach is useful for many reasons: First, Table 3 shows that return distributions of portfolios sorted by certain stock characteristics such as size, profitability or dividend yield, may not be normally distributed. Second, the finite sample distribution of the standard t-statistic may be shifted to the right, leading to an

[9] We downloaded this momentum factor from the web-page of Kenneth French.

[10] We obtain this liquidity factor from the Fama-French Research Portfolios in Wharton Research Data Services (WRDS). The sample period for the liquidity factor is from Feb. 1962 to Dec. 2006.

over-rejection under the null of no-predictability (see, for example, Hodrick (1992), and Ang and Bekaert (2007)). Finally, the correlation between the endogenous regressors and return innovations may result in biased estimates as shown in Stambaugh (1999).

3.3 Results

For the characteristics of size and book-to-market ratio, we consider the long-short portfolios, "10-1", that are long on the decile with the top characteristic and short on the decile with the lowest characteristic; and, for more detailed examinations, the long-short portfolios, "5-1", that are long on the middle characteristic decile and short on the lowest characteristic decile. For the characteristics of dividend yield and earnings-to-price ratio, we examine two long-short portfolios, "10-≤ 0" and "5-≤ 0", that are, respectively, long on the top characteristic decile and short on the decile of non-earning stocks or non-dividend-paying stocks, and long on the middle characteristic decile and short on the decile of non-earning stocks or non-dividend-paying stocks.

Table 4 reports the estimates of the coefficient on sentiment, $\gamma_{i,1}$, of the predictive regressions for the period from January 1966 to December 2005. The column "All " corresponds to the results without sorting observations by regime. The columns under "Regime j", $j = 1, 2$, report the results using regime-sorted observations with regime $= j$. The bootstrapped p-values are in parentheses. The numbers of observations for the sample period in the two regimes inferred by the smoothed probabilities are 114 and 338, respectively.

The column "All" shows that the coefficient estimates, $\hat{\gamma}_{i,1}$, are positive and significant for the long-short portfolios associated with size, earnings and dividend payment. The results show that when sentiment is high, returns on small stocks, non-earning stocks, and non-dividend-paying stocks are relatively low over the coming year. These patterns are little affected by controlling for RMRF, SMB, HML, UMD, and LIQ. However, the $\hat{\gamma}_{i,1}$ estimates on the portfolios associated with growth stocks become insignificant after controlling for RMRF, SMB, HML, UMD, and LIQ. These results are consistent with those of Baker and Wurgler (2006).

When the economic state is controlled, the results in the columns "Regime j", $j = 1, 2$ show that the predictability patterns associated with sentiment are regime-dependent. In particular, sentiment has strong and significant predictive power in regime 2, i.e. a high sentiment results in relatively low subsequent returns on small stocks, growth stocks, 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, HML, UMD, and LIQ in most cases. Under the other regime, however, sentiment mostly dose not exhibit a significant predictive power. For example, the predictability patterns for non-earning stocks and non-dividend-paying stocks disappear after controlling for RMRF, SMB, HML, UMD, and LIQ under regime 1.

For most parts, the magnitude of the coefficient estimate $\hat{\gamma}_{i,1}$ in regime 2 is larger than that in the case of "All". It suggests that controlling for the economic state has significant influence on the magnitude of the coefficient on sentiment, even after controlling for RMRF, SMB, HML, UMD, and LIQ. Overall results show that sentiment has significant predictive power on the subsequent returns for small stocks, growth stocks, non-earning stocks, and non-dividend-paying stocks only under regime 2.[11]

3.4 Out-of-Sample Test Using the MSPE-adjusted Statistic of Clark-West (2007)

We perform an out-of-sample test to examine the robustness of the regime-dependent predictability patterns associated with sentiment. 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. The unrestricted model is deemed to have better forecasting performance if its MSPE is smaller than that of the restricted model. 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 between the MSPEs of the restricted and the unrestricted models.

[11] In unreported results (available upon request), we run predictive regressions based on the regimes identified by filtered probabilities. The results are consistent with those in Table 4 and do not change our conclusions.

The unrestricted model in our test of the predictive power of sentiment is (4) (as re-written below for clarity),

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\text{SENTIMENT}^{\perp}_{t-1} + \beta_{i,\circ}\text{RMRF}_t + \gamma_{i,2}\text{SMB}_t$$
$$+\gamma_{i,3}\text{HML}_t + \gamma_{i,4}\text{UMD}_t + \gamma_{i,5}\text{LIQ}_t + \epsilon_{i,t},$$

and the restricted model with a constraint of $\gamma_{i,1} = 0$, i.e. no predictive power of sentiment, is

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \beta_{i,\circ}\text{RMRF}_t + \gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \gamma_{i,4}\text{UMD}_t + \gamma_{i,5}\text{LIQ}_t + \epsilon_{i,t}. \quad (5)$$

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 perform the estimation and prediction procedures for calculating the model prediction errors as follows. First, for each of the long-short portfolios, we estimate models (4) and (4) using all the in-sample data to obtain the sets of the estimated coefficients denoted as $\widetilde{\boldsymbol{\lambda}}_i^{u'}$ and $\widetilde{\boldsymbol{\lambda}}_i^{c'}$ for the unrestricted and restricted models, respectively. Next, for month $t = R+1, \ldots, R+P$ we use the estimated coefficients and incorporate the out-of-sample realizations of the explanatory variables $\mathbf{x}^u_{i,t}$ (including $\text{SENTIMENT}^{\perp}_{t-1}$, RMRF_t, SMB_t, HML_t, UMD_t and LIQ_t) and $\mathbf{x}^c_{i,t}$ (without $\text{SENTIMENT}^{\perp}_{t-1}$) to compute the return predictions $\tilde{y}^u_{i,t} = \widetilde{\boldsymbol{\lambda}}_i^{u'}\mathbf{x}^u_{i,t}$ and $\tilde{y}^c_{i,t} = \widetilde{\boldsymbol{\lambda}}_i^{c'}\mathbf{x}^c_{i,t}$ for the unrestricted and restricted models, respectively. The prediction errors, $\tilde{\epsilon}^u_{i,t}$ and $\tilde{\epsilon}^c_{i,t}$, for the unrestricted and restricted models, respectively, are the differences between the realized and the model predicted returns of the long-short portfolio. The MSPE-adjusted statistic of Clark and West (2007) is:

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

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

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.

Table 5 reports the results of the out-of-sample test over the period from January 1966 to December 2005. The first column shows that in most of the cases sentiment doesn't exhibit out-of-sample predictive power for small stocks, growth stocks, non-dividend-paying stocks, and non-earnings stocks. A seemingly strong in-sample result almost breaks down out-of-sample. Butler, Grullon, and Weston (2005) argue that the discrepancy between in-sample and out-of-sample results may be attributed to the instability of parameter estimates. As shown in Table 4, the coefficient estimate for $\hat{\gamma}_{i,1}$ changes with the state of regime, suggesting the existence of parameter instability.

Importantly, after controlling for the economic state, the results in the column of Regime 2 show that the MSPE-adjusted statistics are significant in all cases but the long-short position in the book-to-market deciles, suggesting that, out-of-sample, sentiment does have a significant predictive power. The results of regimes 1, by contrast, show that sentiment does not have significant predictive power. The out-of-sample performance of sentiment after controlling for the state of the market regime is consistent with the in-sample results in Table 4. We find that when the economy is in the expansion state with high sentiment, subsequent returns for small stocks, growth stocks, non-earning stocks, and non-dividend-paying stocks tend to be low.

4 Robustness Checks

4.1 Predictive Regressions Using Consumer Confidence

Qiu and Welch (2004), and Lemmon and Portniaguina (2006) document that consumer confidence is an adequate proxy for investor sentiment and is able to predict the returns for small stocks. For robustness checks, we use monthly consumer confidence index provided by the University of Michigan as a proxy for investor sentiment to test the predictive power of sentiment over the period between January 1978 and December 2007.[12] Following Baker

[12]The survey for consumer sentiment started in 1947 on a quarterly basis for Febuary, May, August and November.The index value on the first quarter of 1966 is 100. After 1978 the sentiment index was

and Wurgler (2006) and Lemmon and Portniaguina (2006), we run the following regression to orthogonalize consumer sentiment from macroeconomic variables,

$$z_t = \mathbf{X}_t'\Upsilon + \varepsilon_t,$$

where z_t is consumer sentiment and $\mathbf{X}_t$ is the vector of macroeconomic variables including: growth in the industrial production index, a dummy variable for NBER recessions, growth in personal consumption expenditures in durables, nondurables and services.[13] 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 Figure 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 to reach a record high by 2000 which is then followed by dramatic decreases until 2003, a period corresponding to the boom and bust of the internet bubble. The overall pattern resembles that in the upper panel which depicts the Baker and Wurgler orthogonalized sentiment.

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

$$\begin{aligned} r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i &+ \gamma_{i,1}\text{ConSENTIMENT}_{t-1}^{\perp} + \beta_{i,\circ}\text{RMRF}_t \\ &+ \gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \gamma_{i,4}\text{UMD}_t + \gamma_{i,5}\text{LIQ}_t + \epsilon_{i,t}. \end{aligned} \tag{7}$$

Table 6 reports the coefficient estimates on sentiment $\gamma_{i,1}$. The column under regime 2 shows that the orthogonalized sentiment of the consumer confidence index exhibits significant ability in predicting stock returns, while the results in regime 1 show that there is no evidence of predictability. These results are in line with those reported in Table 4. The main difference between the results in Tables 4 and 6 is that, in the first column for the whole sample period, the Baker and Wurgler orthogonalized sentiment is capable of predicting stock returns for all portfolios while the orthogonalized consumer confidence is only able to predict stock returns of the size portfolios. Overall, the results

published on a monthly basis.

[13]We use the data of personal consumption expenditures in durables, nondurables and services from BEA National Income Accounts.

suggest that the predictive power of sentiment is regime-dependent. When the economy is in the expansion state, high sentiment is usually associated with low subsequent returns for small stocks, growth stocks, non-earning stocks, and non-dividend-paying stocks.

4.2 Predictive Regressions with Regime Dummies

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

$$\begin{aligned} r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_{i,\circ} + \alpha_{i,1} D_1 + \alpha_{i,2} D_2 + (\delta_{i,1} D_1 + \delta_{i,2} D_2)\mathrm{SENTIMENT}_{t-1}^{\perp} \\ + \beta_{i,\circ}\mathrm{RMRF}_t + \gamma_{i,2}\mathrm{SMB}_t + \gamma_{i,3}\mathrm{HML}_t + \gamma_{i,4}\mathrm{UMD}_t + \gamma_{i,5}\mathrm{LIQ}_t + \epsilon_{i,t}, \quad (8) \end{aligned}$$

where D_j is the dummy variable of regime j, which equals 1 for regime $= j$ and 0 otherwise. Note that the specification in (8) treats the state of regime as an exogenous variable.

Table 7 reports the coefficient estimates of the two regime dummies in the model. In practice we involve the observations at the turning points in this regression. In regime 2 investor sentiment displays positive and highly significant predictive power as indicated by the bootstrapped p-values for all the returns on the long-short portfolios. Importantly for almost all cases, the estimates of the dummy variables in regimes 1 are statistically insignificant, suggesting that investor sentiment loses its predictive ability for the cross-section of stock returns in these economic regimes. These results are consistent with those presented in Table 4 and provide further evidence that the predictive power of investor sentiment is regime-dependent.

5 Sources of the Predictive Ability of Sentiment

5.1 Time-Variation with Sentiment in Systematic Risk

We allow the market beta to vary with investor sentiment in the predictive regression below 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,

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\text{SENTIMENT}^{\perp}_{t-1} + (\beta_{i,\circ} + \beta_{i,1}\text{SENTIMENT}^{\perp}_{t-1})\text{RMRF}_t$$
$$+\gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \gamma_{i,4}\text{UMD}_t + \gamma_{i,5}\text{LIQ}_t\epsilon_{i,t}, \qquad (9)$$

where the specification $(\beta_{i,\circ}+\beta_{i,1}\text{SENTIMENT}^{\perp}_{t-1})$ captures the time-varying market beta associated with sentiment.

We test the null that the return predictability of investor sentiment is attributable to the time-varying market beta associated with investor sentiment, that is, the slope coefficient $\beta_{i,1}$ is significantly different from zero. Table 8 reports the results of the coefficient estimates of $\gamma_{i,1}$ and $\beta_{i,1}$. The results in the first column show that over the period from January 1966 to December 2005 the estimates of $\beta_{i,1}$ are statistically insignificant in all cases as indicated by the bootstrapped p-values. The estimates of $\gamma_{i,1}$, by contrast, are positive and highly significant in all cases but the long-short positions on the book-to-market portfolios. The results for regime 2 are very similar. In most of the cases in regimes 1, the estimates of $\beta_{i,1}$ are statistically insignificant. These results, taken together, show that the sentiment-driven time-variation in the market beta does not capture the predictive ability of sentiment in the cross-section of stock returns.

5.2 Investors' Uncertainty and Sentiment

Broadly speaking, our empirical findings emerges that the regime-dependent return predictability of investment sentiment results from the regime-dependent mispricing caused by the shifts in investor sentiment. In particular, the predictive effect of sentiment on stock returns disappears in the recession state. One possible explanation is based on the investors' uncertainty about the state of the economy that gives rise to a feedback effect from the sensitivity of investor's beliefs to news on the stock price, causing the asymmetry in the predictive ability of sentiment over different economic states.

To verify the effect of investors' uncertainty, we run the following regression:

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\text{SENTIMENT}^{\perp}_{t-1} + \phi_{i,1}\text{UC}_t + \beta_{i,\circ}\text{RMRF}_t$$
$$+\gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \gamma_{i,4}\text{UMD}_t + \gamma_{i,5}\text{LIQ}_t + \epsilon_{i,t}, \qquad (10)$$

where $\text{UC}_t = \pi_t(1 - \pi_t)$ is the uncertainty proxy devised by Ozoguz (2009) and π_t is defined as the filtered probability of regime 2, $\Pr(s_t = 2|\mathbf{Y}^t)$. The coefficients on the uncertainty proxy $\phi_{i,1}$ are expected to be positive because higher uncertainty brings up the expected future volatility and required return, driving down the stock return today. This is particular so for small size, non-earnings, growth, and non-dividend-paying stocks. It is intuitive that these stocks not only have no enough tangible assets and collateral but also perform low liquidity and high idiosyncratic volatility, their prices tend to react more to news when there is more investors' uncertainty to the economy. In contrast, larger stocks which have valuable tangible assets and collateral, high liquidity and low idiosyncratic volatility are expected to be less affected when there is more uncertainty.

Table 9 reports the results of the coefficient estimates of $\hat{\gamma}_{i,1}$ and $\hat{\phi}_{i,1}$. The first column shows that, for most parts, the coefficient estimates on $\hat{\phi}_{i,1}$ are positive and significant for the long-short portfolios associated with size, earnings, growth, and dividend payment. This result shows that higher investors' uncertainty will decrease current returns on small stocks, growth stocks, non-earning stocks, and non-dividend-paying stocks. For the estimates of $\hat{\gamma}_{i,1}$, those patterns are similar to those in Table 4.

When the economic state is controlled, we can find that the coefficients on investors' uncertainty are positive and significant in most cases under regime 1 but three out of eight cases are significant under regime 2. It seems that the effect of investors' uncertainty strongly affect returns on size, earnings, growth, and dividend payment in the recession state, whereas this effect is not stable in the expansion state. Moreover, the predictive patterns of sentiment are little affected after considering investors' uncertainty. Sentiment has significant predictive power on the subsequent returns for small stocks, growth stocks, non-earning stocks, and non-dividend-paying stocks only under the expansion state. Importantly, these findings reveal that investors' uncertainty can explain the time-variation on the cross-section of stocks returns in the recession state. Therefore, by this simply test, we can sustain our explanation that investors' uncertainty may cause the asymmetry in the predictive ability of sentiment on stock returns across different economic states.

6 Conclusion

In this study we examine the predictive effect of investor sentiment on the cross-section of stock returns across different economic states. We first implement a multivariate Markov-switching model to characterize two economic regimes, the expansion state and the recession state. We then use the lagged sentiment proxy to forecast the portfolio returns related to small stocks, non-earning stocks, growth stocks, and non-dividend-paying stocks in the identified economic states. We find that the predictive power of investor sentiment is regime-dependent. Investor sentiment performs both in-sample and out-of-sample predictive power on the portfolio returns related to small stocks, growth stocks, non-earning stocks, and non-dividend-paying stocks only under the bullish regime. When the expansion state has high (low) investor sentiment, these categories of stocks earn relatively low (high) subsequent returns. The predictive ability of sentiment can not be attributed to time-variation in the market beta driven by investor sentiment.

Appendix: A Simple Bootstrap Testing 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 kth portfolio with firm characteristic i:

$$\begin{aligned} r_{(i,k),t} = \alpha^{\dagger}_{(i,k)} + \beta^{\dagger}_{(i,k),\circ}\mathrm{RMRF}_t + \gamma^{\dagger}_{(i,k),1}\mathrm{SENTIMENT}^{\perp}_{t-1} \\ + \gamma^{\dagger}_{(i,k),2}\mathrm{SMB}_t + \gamma^{\dagger}_{(i,k),3}\mathrm{HML}_t + \gamma^{\dagger}_{(i,k),4}\mathrm{UMD}_t + \gamma^{\dagger}_{(i,k),5}\mathrm{LIQ}_t + \epsilon_{(i,k),t}, \quad (11) \end{aligned}$$

  and save all OLS-estimated risk loadings $\{\hat{\alpha}^{\dagger}_{(i,k)}, \hat{\beta}^{\dagger}_{(i,k),\circ}, \hat{\gamma}^{\dagger}_{(i,k),1}, \hat{\gamma}^{\dagger}_{(i,k),2}, \hat{\gamma}^{\dagger}_{(i,k),3}, \hat{\gamma}^{\dagger}_{(i,k),4}, \hat{\gamma}^{\dagger}_{(i,k),5}\}$, 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 $\hat{\mathbf{\Upsilon}}_{i,t} = (\hat{\epsilon}_{(i,1),t}, \ldots, \hat{\epsilon}_{(i,N),t})'$. We resample a sequence of the time indices $s^b_{T_0}, \ldots, s^b_{T_n}$ 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 resample number one). The sequence of the resampled residual vectors are given by $\{\hat{\mathbf{\Upsilon}}^b_{i,t_\epsilon}, t_\epsilon = s^b_{T_0}, \ldots, s^b_{T_n}\}$, where $\hat{\mathbf{\Upsilon}}^b_{i,t_\epsilon} = (\hat{\epsilon}^b_{(i,1),t_\epsilon}, \ldots, \hat{\epsilon}^b_{(i,N),t_\epsilon})'$.

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

$$\begin{aligned} r^b_{(i,k),t_\epsilon} = \hat{\alpha}^{\dagger}_{(i,k)} + \hat{\beta}^{\dagger}_{(i,k),\circ}\mathrm{RMRF}_{t_\varepsilon} + \hat{\gamma}^{\dagger}_{(i,k),2}\mathrm{SMB}_{t_\varepsilon} + \hat{\gamma}^{\dagger}_{(i,k),3}\mathrm{HML}_{t_\epsilon} + \hat{\gamma}^{\dagger}_{(i,k),4}\mathrm{UMD}_{t_\epsilon} \\ + \hat{\gamma}^{\dagger}_{(i,k),5}\mathrm{LIQ}_{t_\epsilon} + \hat{\epsilon}^b_{(i,k),t_\epsilon}, \quad (12) \end{aligned}$$

  for $k = 1, \ldots, N$ and $t_\epsilon = s^b_{T_0}, \ldots, s^b_{T_n}$. So, there are N resampled time series of stock portfolio returns in a resampling, $\{r^b_{(i,k),t_\epsilon},\ t_\epsilon = s^b_{T_0}, \ldots, s^b_{T_n}\},\ k = 1, \ldots, N$.

- **Step 4:** We run the predictive regression model (4) using the resampled data generated in Step 3, and compute the corresponding t-statistic, $t(\hat{\gamma}^b_{i,1})$, for $\gamma_{i,1}$. Repeating Steps 2 and 3 for M times (the largest b is M), a bootstrap distribution of t-statistic of $\gamma_{i,1}$, $\{t(\hat{\gamma}^b_{i,1}), b = 1, \ldots, M\}$, under the null hypothesis ($\gamma_{i,1} = 0$) is available.

- **Step 5:** We compute the p-value associated with t-statistic by comparing $\sqrt{T} \cdot t(\hat{\gamma}_{i,1})$ to the quantiles of $\sqrt{T} \cdot [t(\hat{\gamma}^b_{i,1}) - t(\hat{\gamma}_{i,1})]$ to obtain the p-value. Note that T is the total sample size between T_0 and T_n, $t(\hat{\gamma}_{i,1})$ and $t(\hat{\gamma}^b_{i,1})$ are the t-statistics computed by real data and the b-th resampled data, respectively. The bootstrapped p-value may be defined as the probability in favor of the null hypothesis

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

The readers can refer to Sullivan, Timmermann, and White (1999) and Wang (2005) for the details of implementing bootstrap methods.

References

Ang, A., and G., Bekaert, 2007, “Stock Return Predictability: Is it There,” *Review of Financial Studies*, 20, 651-707.

Ang, A., and J., Chen, 2002, “Asymmetric Correlations of Equity Portfolios,” *Journal of Financial Economics*, 63, 443-393.

Abreu, D., and M., Brunnermeier, 2003, “Bubbles and crashes,” *Econometrica*, 71, 173-204.

Baker, M., and J., Stein, 2004, “Market Liquidity as a Sentiment Indicator,” *Journal of Financial Markets*, 7, 271-299.

Baker, M., and J., Wurgler, 2006, “Investor Sentiment and the Cross-Section of Stock Returns,” *Journal of Finance*, 61, 1645-1680.

Baker, M., and J., Wurgler, 2007, “Investor Sentiment in the Stock Market,” *Journal of Economic Prospective*, 21, 129-152.

Barberis, N., A., Shleifer, and R., Vishny, 1998, “A Model of Investor Sentiment,” *Journal of Financial Economics*, 49, 307-343.

Brown, G., and M., Cliff, 2004, “Investor Sentiment and the Near-Term Stock Market,” *Journal of Empirical Finance*, 11, 1-27.

Brown, G., and M., Cliff, 2005, “Investor Sentiment and asset valuation,” *Journal of Business* , 78, 405-440.

Brunnermeier, M., and S., Nagel, 2004, “Hedge funds and the technology bubble,” *Journal of Finance*, 59, 2013-2040.

Buss, A., C., Schlag, and G., Vilkov, 2009, “CAPM with Option-Implied Betas: Another Rescue Attempt,” *Working Paper*, Goethe University.

Butler, A., G., Grullon, and J., Weston, 2005, “Can Managers Forecast Aggregate Market Returns,” *Journal of Finance*, 60, 963-986.

Cecchetti, S., P., Lam, and N., Mark, 2000, "Asset Pricing with Distorted Beliefs: Are Equity Returns too Good to Be True?," *American Economic Review*, 90, 787-805.

Clark, T., and K., West, 2007, "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics*, 138, 291-311.

Cooper, M., R., Gutierrez, and A., Hameed, 2004, "Market States and Momentum," *Journal of Finance*, 59, 1345-1365.

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

Deuskar, P., 2004, "Aggregate Liquidity and Investor Sentiment: Empirical Connection," *Working Paper*, New York University.

Fama, E., and K. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics*, 33, 3-56.

Glushkov, D., 2006, "Sentiment Beta," *Working Paper*, University of Taxes at Austin.

Gray, S., 1996, "Modelling the Conditional Distribution of Interest Rate as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27-62.

Guidolin, M., and A., Timmermann, 2008, "Size and Value Anomalies under Regime Shifts," *Journal of Financial Econometrics*, 6, 1-48.

Gulen, H., Y., Xing, and L., Zhang, 2008, "Value versus Growth: Time-Varying Expected Returns," *Working Paper*, University of Michigan.

Hamilton, J., 1989, "A New Approach to Economic Analysis to Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 557-590.

Hamilton, J., 1990, "Analysis of Time Series Subject to Changes in Regimes," *Journal of Econometrics*, 45, 39-70.

Hodrick, R., 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5, 357-386.

Kosowski, R., A., Timmermann, R., Wermers, and H., White, 2006, "Can Mutual Fund Stars Really Pick Stocks? New Evidence from a Bootstrap Analysis," *Journal of Finance*, 61, 2551-2595.

Kumar, A., and C., Lee, 2006, "Retail Investor Sentiment and Return Comovements," *Journal of Finance*, 61, 2451-2486.

Lee, C., A., Shleifer, and R., Thaler, 1991, "Investor Sentiment and the closed-end fund puzzle," *Journal of Finance*, 46, 75-109.

Lemmon, M., and E., Portniaguina, 2006, "Consumer Confidence and Asset Prices: Some Empirical Evidence," *Review of Financial Studies*, 19, 1499-1529.

Maheu, J., and T., McCurdy, 2000, "Identifying Bull and Bear Markets in Stock Returns," *Journal of Business and Economic Statistics*, 18, 100-112.

Ofek, E. and M., Richardson, 2003, "DotCom mania: the raise and the fall of internet stock prices," *Journal of Finance*, 58, 1113-1137.

Ozoguz, A., 2009, "Good Times or Bad Times? Investors' Uncertainty and Stock Returns," *Review of Financial Studies*, 22, 4377-4422.

Pastor, L., and P., Veronesi, 2006, "Was there a Nasdaq bubble in the late 1990s?," *Journal of Financial Economics*, 81, 61-100.

Perez-Quiros, G., and A., Timmermann, 2000, "Firm Size and Cyclical Variation in Stock Returns," *Journal of Finance*, 55, 1229-1262.

Qiu, L., and I., Welch, 2004, "Investor Sentiment Measures," NBER Working Series 10794,

Shleifer, A., 2000, *Inefficient Market: An Introduction to Behavior Finance*, Oxford University Press.

Shleifer, A., and R., Vishny, 1997, "The Limits of Arbitrage," *Journal of Finance*, 52, 35-55.

Sullivan, R., A., Timmermann, and H., White, 1999, "Data Snooping, Technical Trading Rule Performance and the Bootstrap," *Journal of Finance*, 54, 1647-1692.

Swaminathan, B., 1996, "Time-Varying Expected Small Firm Returns and Closed-End Fund Discounts," *Review of Financial Studies*, 9, 845-897.

Timmermann, A., 2000, "Moments of Markov Switching Model," *Journal of Econometrics* 96, 75-111.

Wang, K., 2005,"Multifactor Evaluation of Style Rotation," *Journal of Financial and Quantitative Analysis*, 40, 349-372.

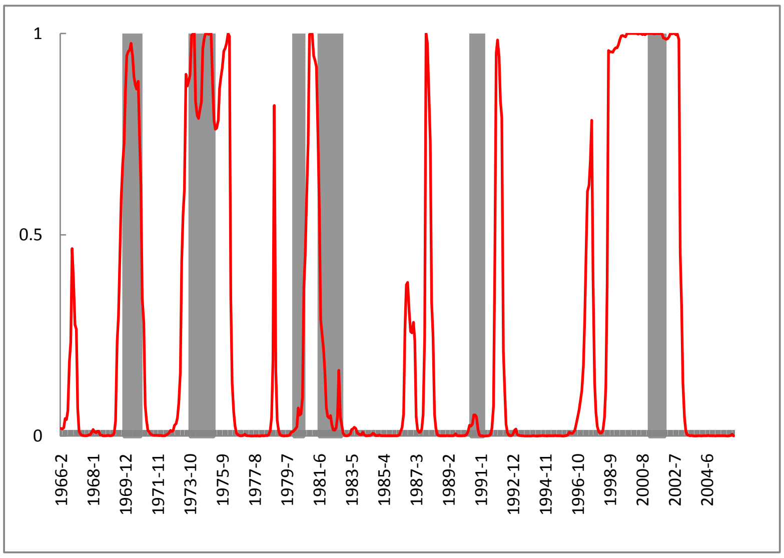
Veronesi, P., 1999, "Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model,"*Review of Financial Studies*, 12, 975-1007.

Veronesi, P., 2004, "The Peso Hypothesis and Stock Market Returns,"*Journal of Economic Dynamics and Control*, 28, 707-725.

Figure 1:

The Smoothed Probabilities of the Two-Sate Markov-Switching Model for the Orthogonalized Market, SMB, and HML Returns

This 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 sample period is from Jan. 1966 to Dec. 2005. 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.



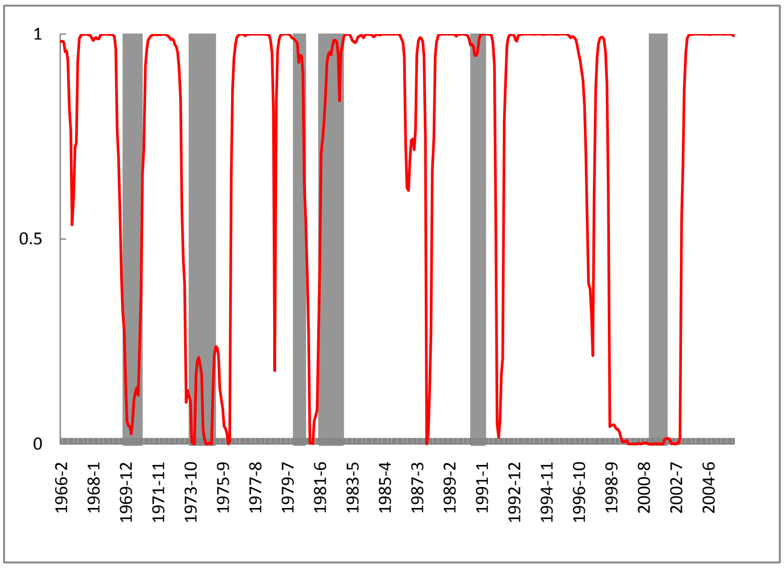
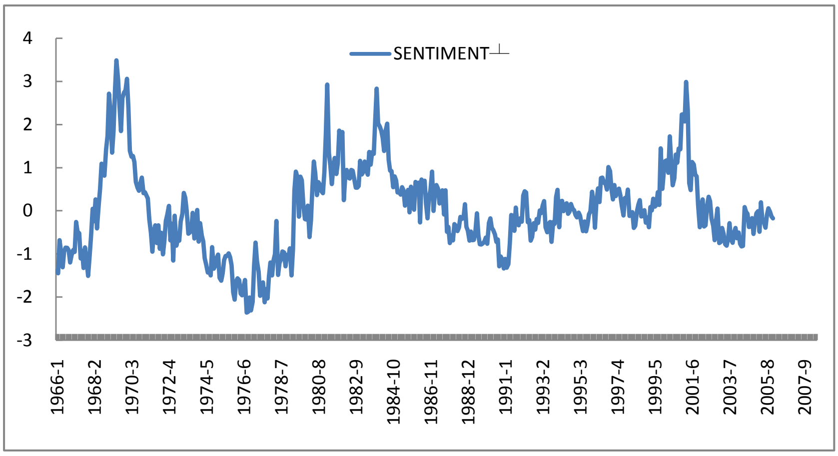


Figure 2:

The Historical Patterns of the Sentiment Proxies

The upper panel plots the Baker and Wurgler's orthogonalized sentiment index for the period from January 1966 to December 2005 and the lower panel displays an orthogonalized consumer sentiment measure for the period from Jan. 1978 to December 2007.



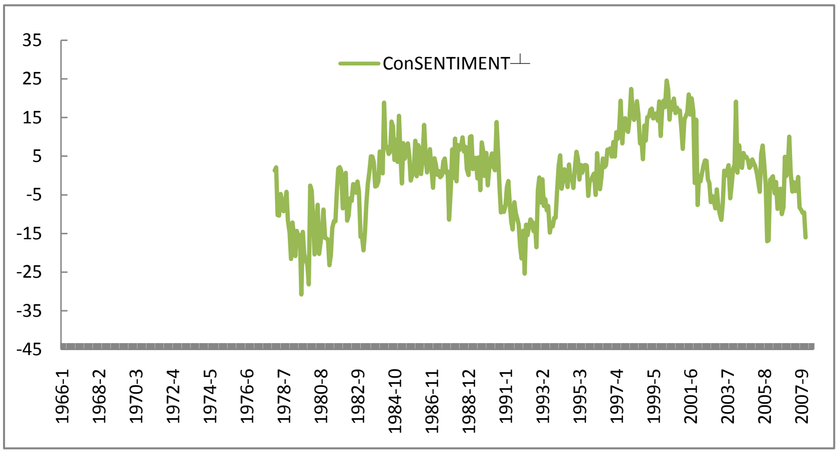


Table 1:
Parameter Estimates of the Markov-Switching Model for the Orthogonalized Market, SMB, and HML Returns

Parameters	Market	SMB	HML
Panel A: Mean Parameters			
Constant, Regime 1	-0.022	-0.004	-0.001
Constant, Regime 2	-0.006	-0.015	0.003
Default Premium(Def_{t-1}), Regime 1	31.127**	20.638**	1.578
Default Premium(Def_{t-1}), Regime 2	23.616**	5.660	-2.454
Interest rate(I_{t-1}), Regime 1	-3.629**	0.279	0.075
Interest rate(I_{t-1}), Regime 2	-4.453**	-4.729**	0.427
Dividend yield(Div_{t-1}), Regime 1	0.172	-0.590	0.071
Dividend yield(Div_{t-1}), Regime 2	0.338	1.056**	-0.137
Panel B: Correlations/Volatilities			
Regime 1			
Market	0.061**		
SMB	0.273*	0.048**	
HML	-0.481**	-0.319**	0.045**
Regime 2			
Market	0.036**		
SMB	0.294*	0.024**	
HML	-0.347**	-0.227*	0.022**
Panel C: Transition Matrix Parameters			
	Regime 1	Regime 2	
Constant	0.757	1.475**	
Leading indicator($\Delta\mathrm{CLI}_{t-1}$)	-18.740**	29.189**	
Panel D: Steady State Probabilities			
	Regime 1	Regime 2	
	0.755**	0.245	

Note: This table reports the results of the parameter estimates for the MSVAR(2,0) model:

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

where μ_{s_t} is the 3×1 intercept vector in regime s_t, $\mathbf{\Phi}_{s_t}$ 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 $\varepsilon_t \sim \mathcal{N}(0, \mathbf{\Omega}_{s_t})$ 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

$$p_{ii,t} = \Pr(s_t = i | s_{t-1} = i, \Delta\mathrm{CLI}_{t-1}) = N(a_i + b_i\Delta\mathrm{CLI}_{t-1}),\ i = 1, 2,$$

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. The three series are excess returns on the values-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 Jan. 1966 to Dec. 2005. Values reported on the diagonals of the correlation matrices are volatilities. All estimates are monthly. $*$ and $**$ denote significance at 5% and 1% level, respectively.

Table 2:
Correlations and Regressions of Smoothed Probabilities with the NBER Indicator and Sentiment

Panel A: Correlations

	$\Pr(s_t = 1\vert\mathbf{Y}^T)$	$\Pr(s_t = 2\vert\mathbf{Y}^T)$	$\text{SENTIMENT}^{\perp}_{t-1}$
$\text{SENTIMENT}^{\perp}_{t-1}$	0.171	-0.171	1
NBER_t	0.289	-0.289	0.129

Panel B: Regressions with NBER Recession Index

	$\Pr(s_t = 1\vert\mathbf{Y}^T)$		$\Pr(s_t = 2\vert\mathbf{Y}^T)$	
Constant	0.226**	0.272**	0.774**	0.728**
	(0.019)	(0.018)	(0.019)	(0.018)
NBER_t	0.330**	-	-0.330**	-
	(0.050)		(0.050)	-
$\text{SENTIMENT}^{\perp}_{t-1}$	-	0.067**	-	-0.067**
	-	(0.018)	-	(0.018)

Note: This table reports (i)the correlations of smoothed probabilities with the NBER recession index NBER_t and lagged $\text{SENTIMENT}^{\perp}$; (ii) regressions of smoothed probabilities on the NBER recession index

$$\Pr(s_t = j|\mathbf{Y}^T) = \rho_\circ + \rho_1 \text{NBER}_t + \epsilon_t,\ j = 1,\ 2,$$

and regressions of smoothed probabilities on lagged $\text{SENTIMENT}^{\perp}$

$$\Pr(s_t = j|\mathbf{Y}^T) = \rho_\circ + \rho_1 \text{SENTIMENT}^{\perp}_{t-1} + \epsilon_t,\ j = 1,\ 2.$$

The standard deviations are in parentheses. $*$ and $**$ denote significance at 5% and 1% level, respectively.

Table 3:
Summary Statistics for the Portfolio Returns

Decile	≤ 0	1	2	3	4	5	6	7	8	9	10
Panel A: Portfolios Formed on Size											
Mean		1.49	1.13	1.16	1.12	1.14	1.08	1.14	1.05	1.02	0.89
Standard Deviation		6.84	6.56	6.33	6.11	5.89	5.58	5.40	5.27	4.85	4.67
Skewness		0.31	-0.07	-0.19	-0.33	-0.35	-0.38	-0.30	-0.30	-0.22	-0.22
Excess Kurtosis		2.70	2.74	2.30	2.17	2.43	1.85	2.13	1.58	1.43	1.74
Maximum		32.88	32.52	29.90	26.41	27.54	23.01	24.81	21.13	19.06	20.92
Minimum		-27.68	-30.31	-28.87	-29.40	-28.12	-25.93	-25.90	-24.07	-22.09	-20.38
Panel B: Portfolios Formed on Book-to-Market											
Mean		0.66	1.02	1.14	1.26	1.32	1.45	1.54	1.57	1.71	1.94
Standard Deviation		7.73	6.59	6.19	5.86	5.53	5.36	5.19	5.25	5.58	6.38
Skewness		0.02	-0.22	-0.30	-0.32	-0.27	-0.18	-0.10	0.12	0.06	0.60
Excess Kurtosis		2.64	1.76	2.41	2.90	3.42	3.55	3.81	4.16	4.30	4.81
Maximum		41.56	25.59	27.50	29.18	29.87	29.70	29.71	32.02	33.08	40.43
Minimum		-32.58	-29.45	-30.12	-29.09	-27.94	-26.11	-25.12	-24.30	-26.83	-25.59
Panel C: Portfolios Formed on Dividend Yield											
Mean	1.34	1.23	1.31	1.28	1.35	1.32	1.41	1.39	1.44	1.34	1.21
Standard Deviation	7.72	5.90	5.44	5.21	5.03	4.79	4.63	4.39	4.26	3.90	4.13
Skewness	0.19	-0.54	-0.53	-0.37	-0.51	-0.40	-0.49	-0.43	-0.24	0.09	1.08
Excess Kurtosis	2.18	2.23	3.11	3.59	3.65	3.70	3.77	4.38	4.48	4.31	8.28
Maximum	35.35	19.92	25.70	27.40	24.63	24.83	23.30	24.24	25.53	25.06	33.20
Minimum	-30.29	-28.64	-26.95	-25.74	-26.24	-25.14	-24.80	-23.98	-22.09	-17.13	-12.40
Panel D: Portfolios Formed on Earnings/Price											
Mean	1.41	0.96	1.17	1.17	1.24	1.29	1.35	1.41	1.48	1.62	1.75
Standard Deviation	8.76	7.02	6.05	5.67	5.33	5.18	5.03	4.88	4.84	5.04	5.81
Skewness	0.59	-0.15	-0.35	-0.36	-0.38	-0.42	-0.33	-0.29	0.00	0.00	0.13
Excess Kurtosis	3.37	1.78	2.21	2.85	4.09	4.04	4.33	4.20	4.77	4.67	4.42
Maximum	46.37	27.30	25.97	27.78	29.62	27.93	28.82	26.99	30.74	30.74	36.32
Minimum	-31.34	-31.58	-29.48	-28.05	-28.66	-28.05	-26.97	-25.58	-22.95	-23.61	-25.43

Note: This table reports the summary statistics of the monthly equal-weighted portfolio returns formed on size, book-to-market, dividend yield, and earnings/price. The data sample is downloaded from the web-page of Kenneth French. All portfolios are constructed at the end of each June. In June of year t, all NYSE stocks are sorted by (i) ME, (ii) BE/ME, (iii) D/P, and (iv) E/P, respectively, to determine the decile breakpoints for each firm characteristic. 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 $t-1$. All NYSE, AMEX, and NASDAQ stocks are allocated based on the NYSE's breakpoints of each firm characteristic, forming the portfolios for July of year t to June of year $t+1$. The monthly equal-weighted returns are calculated for portfolios. For size and book-to-market, 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 is from Jan. 1966 to Dec. 2005.

Table 4:
Predictive Regressions for Long-Short Portfolio Returns

Long-Short	All		Regime 1		Regime 2	
Panel A: Portfolios Formed on Size						
Only Sentiment$^{\perp}$						
10 - 1	0.75**	(0.00)	0.38	(0.16)	1.07**	(0.00)
5 - 1	0.36**	(0.00)	0.21	(0.18)	0.47**	(0.00)
Controlling for RMRF, SMB, HML, UMD, and LIQ						
10 - 1	0.27**	(0.01)	0.04	(0.51)	0.30**	(0.00)
5 - 1	0.24**	(0.00)	0.20	(0.18)	0.19**	(0.01)
Panel B: Portfolios Formed on Book-to-Market						
Only Sentiment$^{\perp}$						
10 - 1	0.24*	(0.03)	-0.12	(0.62)	0.50**	(0.00)
5 - 1	0.47**	(0.00)	0.38	(0.18)	0.57**	(0.00)
Controlling for RMRF, SMB, HML, UMD, and LIQ						
10 - 1	-0.04	(0.77)	-0.61	(1.00)	0.21**	(0.00)
5 - 1	0.16	(0.07)	-0.13	(0.78)	0.25**	(0.00)
Panel C: Portfolios Formed on Dividend Yield						
Only Sentiment$^{\perp}$						
10 - ≤ 0	1.02**	(0.00)	1.02**	(0.00)	1.13**	(0.00)
5 - ≤ 0	0.83**	(0.00)	0.78*	(0.03)	0.90**	(0.00)
Controlling for RMRF, SMB, HML, UMD, and LIQ						
10 - ≤ 0	0.38**	(0.00)	0.27	(0.19)	0.33**	(0.00)
5 - ≤ 0	0.38**	(0.00)	0.19	(0.31)	0.35**	(0.00)

(Continued)

Table 4-Continued

Long-Short	All		Regime 1		Regime 2	
Panel D: Portfolios Formed on Earnings/Price						
Only Sentiment$^{\perp}$						
10 - ≤ 0	0.77**	(0.00)	0.95**	(0.00)	0.75**	(0.00)
5 - ≤ 0	0.85**	(0.00)	0.90*	(0.02)	0.82**	(0.00)
Controlling for RMRF, SMB, HML, UMD, and LIQ						
10 - ≤ 0	0.43**	(0.00)	0.43	(0.10)	0.31**	(0.00)
5 - ≤ 0	0.44**	(0.00)	0.25	(0.29)	0.32**	(0.00)

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

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

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

$$\begin{aligned} r_{(i,k_2),t} - r_{(i,k_1),t} = {} & \alpha_i + \gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1} + \beta_{i,\circ}\mathrm{RMRF}_t + \gamma_{i,2}\mathrm{SMB}_t \\ & + \gamma_{i,3}\mathrm{HML}_t + \gamma_{i,4}\mathrm{UMD}_t + \gamma_{i,5}\mathrm{LIQ}_t + \epsilon_{i,t}, \end{aligned}$$

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, and earnings/price) at time t, and $k_1, k_2 \in \{\leq 0, 1, 2, \ldots, 10\}$. "1", "2", …, and "10" indicate the portfolios in the 1st (the smallest), 2nd, …, and 10th deciles, respectively. "≤ 0" represents the portfolios for non-dividend-paying stocks or non-earning stocks. This table only reports the parameter estimates of $\gamma_{i,1}$. The sample period is from Jan. 1966 to Dec. 2005. The values of portfolio returns are in terms of percentage. SENTIMENT$^{\perp}$ is the Baker and Wurgler's orthogonalized sentiment proxy. The column "All" reports the results without regime sorting. The other columns "Regime j", $j = 1, 2$, show the results based on regime-sorted observations as regime$= j$. The bootstrapped p-values are in parentheses. $*$ and $**$ denote significance at 5% and 1% level, respectively.

Table 5:
Out-of-Sample Predictability Test Results Using the Clark and West's (2007) MSPE-adjusted Statistic

Long-Short	All	Regime 1	Regime 2
Panel A: Portfolios Formed on Size			
10 - 1	0.50	-4.50	2.25**
5 - 1	0.78	-2.39	2.17**
Panel B: Portfolios Formed on Book-to-Market			
10 - 1	-0.69	2.65**	0.98
5 - 1	-0.12	-2.85	1.65*
Panel C: Portfolios Formed on Dividend Yield			
10 - ≤ 0	1.58	-3.68	2.24**
5 - ≤ 0	1.42	-4.12	2.65**
Panel D: Portfolios Formed on Earnings/Price			
10 - ≤ 0	1.82*	-3.30	2.25**
5 - ≤ 0	0.98	-4.41	2.34**

Note: This table reports the results of the out-of-sample tests over the period from Jan. 1966 to Dec. 2005. The Clark and West's (2007) MSPE-adjusted statistic is computed using the prediction errors of the unrestricted and restricted models of (3) and (4) for the returns of the long-short portfolios of size, book-to-market, dividend yield, and earnings/price. 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), the Fama-French factors (HML and SMB), the momentum factor (UMD), the liquidity factor (LIQ). 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) the portfolios for non-dividend-paying stocks or non-earning stocks. Portfolio returns are in percentage. The column "All" reports the results without regime sorting. The other columns "Regime j", $j = 1, 2$, show the results based on regime-sorted observations as 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% level, respectively.

Table 6:
Predictive Regressions for Long-Short Portfolio Returns Using Consumer Confidence

Long-Short	All		Regime 1		Regime 2	
	Panel A: Portfolios Formed on Size					
10 - 1	0.07**	(0.00)	-0.01	(0.61)	0.01*	(0.04)
5 - 1	0.02*	(0.05)	0.01	(0.33)	0.00	(0.37)
	Panel B: Portfolios Formed on Book-to-Market					
10 - 1	0.00	(0.70)	-0.05	(1.00)	0.03**	(0.00)
5 - 1	0.01	(0.19)	-0.01	(0.69)	0.02**	(0.00)
	Panel C: Portfolios Formed on Dividend Yield					
10 - ≤ 0	0.02	(0.10)	-0.01	(0.74)	0.03*	(0.02)
5 - ≤ 0	0.00	(0.48)	-0.03	(0.89)	0.02**	(0.01)
	Panel D: Portfolios Formed on Earnings/Price					
10 - ≤ 0	0.01	(0.21)	-0.01	(0.65)	0.03**	(0.00)
5 - ≤ 0	0.01	(0.20)	-0.01	(0.64)	0.03**	(0.00)

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

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\text{ConSENTIMENT}^{\perp}_{t-1} + \beta_{i,\circ}\text{RMRF}_t + \gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \gamma_{i,4}\text{UMD}_t + \gamma_{i,5}\text{LIQ}_t + \epsilon_{i,t},$$

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, and earnings/price) at time t, and $k_1, k_2 \in \{\leq 0, 1, 2, \ldots, 10\}$. "1", "2", ..., and "10" indicate the portfolios in the 1st (the smallest), 2nd, ..., and 10th deciles, respectively. "≤ 0" represents the portfolios for non-dividend-paying stocks or non-earning stocks. This table only reports the parameter estimates of $\gamma_{i,1}$. The sample period is from Jan. 1978 to Dec. 2005. The values of portfolio returns are in terms of percentage. ConSENTIMENT$^{\perp}$ is the orthogonalized consumer sentiment proxy. The column "All" reports the results without regime sorting. The other columns "Regime j", $j = 1, 2$, show the results based on regime-sorted observations as regime$= j$. The bootstrapped p-values are in parentheses. $*$ and $**$ denote significance at 5% and 1% level, respectively.

Table 7:
Predictive Regressions with Regime Dummies

Long-Short	$\delta_{i,1}$		$\delta_{i,2}$	
Panel A: Portfolios Formed on Size				
10 - 1	-0.09	(0.70)	0.53**	(0.00)
5 - 1	0.14	(0.25)	0.30**	(0.00)
Panel B: Portfolios Formed on Book-to-Market				
10 - 1	-0.63	(1.00)	0.28**	(0.00)
5 - 1	-0.11	(0.75)	0.30**	(0.00)
Panel C: Portfolios Formed on Dividend Yield				
10 - ≤ 0	0.19	(0.31)	0.53**	(0.00)
5 - ≤ 0	0.18	(0.33)	0.48**	(0.00)
Panel D: Portfolios Formed on Earnings/Price				
10 - ≤ 0	0.46	(0.12)	0.43**	(0.00)
5 - ≤ 0	0.33	(0.26)	0.43**	(0.00)

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), the Fama-French factors (HML and SMB), the momentum factor (UMD), and the liquidity factor (LIQ),

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_{i,\circ} + \alpha_{i,1}D_1 + \alpha_{i,2}D_2 + (\delta_{i,1}D_1 + \delta_{i,2}D_2)\text{SENTIMENT}^{\perp}_{t-1}$$
$$+\beta_{i,\circ}\text{RMRF}_t + \gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \gamma_{i,4}\text{UMD}_t + \gamma_{i,5}\text{LIQ}_t + \epsilon_{i,t},$$

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, and earnings/price) at time t, $k_1, k_2 \in \{\leq 0, 1, 2, \ldots, 10\}$, and D_j is the dummy variable of regime j, which equals to 1 when regime= j and equals 0 otherwise. "1", "2", . . . , and "10" indicate the portfolios in the 1st (the smallest), 2nd, . . . , and 10th deciles, respectively. "≤ 0" represents the portfolios for non-dividend-paying stocks or non-earning stocks. The sample period is from Jan. 1966 to Dec. 2005. The values of portfolio returns are in terms of percentage. SENTIMENT$^{\perp}$ is the Baker and Wurgler's orthogonalized sentiment proxy. The bootstrapped p-values are in parentheses. $*$ and $**$ denote significance at 5% and 1% level, respectively.

Table 8:
Conditional Market Betas

Long-Short		All		Regime 1		Regime 2	
		Panel A: Portfolios Formed on Size					
10 - 1	$\gamma_{i,1}$	0.24*	(0.03)	-0.07	(0.67)	0.28**	(0.00)
5 - 1	$\gamma_{i,1}$	0.24**	(0.00)	0.24	(0.17)	0.17**	(0.01)
10 - 1	$\beta_{i,1}$	-0.06	(1.00)	-0.09	(0.98)	-0.05	(1.00)
5 - 1	$\beta_{i,1}$	0.00	(0.58)	0.04	(0.14)	-0.05	(0.99)
		Panel B: Portfolios Formed on Book-to-Market					
10 - 1	$\gamma_{i,1}$	-0.08	(0.90)	-0.82	(1.00)	0.21**	(0.00)
5 - 1	$\gamma_{i,1}$	0.13	(0.12)	-0.31	(0.95)	0.25**	(0.00)
10 - 1	$\beta_{i,1}$	-0.09	(1.00)	-0.18	(1.00)	0.01	(0.25)
5 - 1	$\beta_{i,1}$	-0.07	(1.00)	-0.16	(1.00)	0.01	(0.20)
		Panel C: Portfolios Formed on Dividend Yield					
10 - ≤ 0	$\gamma_{i,1}$	0.34**	(0.00)	0.04	(0.52)	0.34**	(0.00)
5 - ≤ 0	$\gamma_{i,1}$	0.35**	(0.00)	-0.04	(0.64)	0.35**	(0.00)
10 - ≤ 0	$\beta_{i,1}$	-0.10	(1.00)	-0.20	(1.00)	-0.03	(0.86)
5 - ≤ 0	$\beta_{i,1}$	-0.08	(1.00)	-0.20	(1.00)	0.00	(0.27)
		Panel D: Portfolios Formed on Earnings/Price					
10 - ≤ 0	$\gamma_{i,1}$	0.39**	(0.00)	0.21	(0.32)	0.30**	(0.00)
5 - ≤ 0	$\gamma_{i,1}$	0.41**	(0.01)	0.03	(0.56)	0.33**	(0.00)
10 - ≤ 0	$\beta_{i,1}$	-0.10	(1.00)	-0.19	(1.00)	-0.04	(0.98)
5 - ≤ 0	$\beta_{i,1}$	-0.07	(1.00)	-0.20	(1.00)	0.02	(0.23)

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), the momentum factor (UMD), the liquidity factor (LIQ), and interaction of RMRF and SENTIMENT$^{\perp}$,

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\text{SENTIMENT}^{\perp}_{t-1} + (\beta_{i,\circ} + \beta_{i,1}\text{SENTIMENT}^{\perp}_{t-1})\text{RMRF}_t$$
$$+\gamma_{i,2}\text{SMB}_t + \gamma_{i,3}\text{HML}_t + \gamma_{i,4}\text{UMD}_t + \gamma_{i,5}\text{LIQ}_t\epsilon_{i,t},$$

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, and earnings/price) at time t, and $k_1, k_2 \in \{\leq 0, 1, 2, \ldots, 10\}$. "1", "2", …, and "10" indicate the portfolios in the 1st (the smallest), 2nd, …, and 10th deciles, respectively. "≤ 0" represents the portfolios for non-dividend-paying stocks or non-earning stocks. The sample period is from Jan. 1966 to Dec. 2005. The values of portfolio returns are in terms of percentage. SENTIMENT$^{\perp}$ is the Baker and Wurgler's orthogonalized sentiment proxy. The column "All" reports the results without regime sorting. The other columns "Regime j", $j = 1, 2$, show the results based on regime-sorted observations as regime$= j$. The bootstrapped p-values are in parentheses. $*$ and $**$ denote significance at 5% and 1% level, respectively.

Table 9:
Predictive Regressions with Investor Uncertainty Proxy

Long-Short		All		Regime 1		Regime 2	
Panel A: Portfolios Formed on Size							
10 - 1	$\gamma_{i,1}$	0.28**	(0.01)	0.35	(0.20)	0.30**	(0.00)
5 - 1	$\gamma_{i,1}$	0.25**	(0.00)	0.41*	(0.04)	0.19**	(0.01)
10 - 1	$\phi_{i,1}$	5.37**	(0.00)	13.45**	(0.00)	0.84	(0.18)
5 - 1	$\phi_{i,1}$	3.89**	(0.00)	9.30**	(0.00)	0.79	(0.23)
Panel B: Portfolios Formed on Book-to-Market							
10 - 1	$\gamma_{i,1}$	-0.04	(0.75)	-0.58	(1.00)	0.21**	(0.00)
5 - 1	$\gamma_{i,1}$	0.17	(0.06)	0.04	(0.50)	0.25**	(0.00)
10 - 1	$\phi_{i,1}$	0.96	(0.09)	1.27	(0.26)	0.82	(0.12)
5 - 1	$\phi_{i,1}$	2.73**	(0.00)	7.44**	(0.00)	1.21*	(0.03)
Panel C: Portfolios Formed on Dividend Yield							
10 - ≤ 0	$\gamma_{i,1}$	0.40**	(0.00)	0.48	(0.07)	0.32**	(0.00)
5 - ≤ 0	$\gamma_{i,1}$	0.40**	(0.00)	0.45	(0.10)	0.34**	(0.00)
10 - ≤ 0	$\phi_{i,1}$	6.43**	(0.00)	9.16**	(0.00)	5.38**	(0.00)
5 - ≤ 0	$\phi_{i,1}$	4.82**	(0.00)	11.50**	(0.00)	1.64*	(0.05)
Panel D: Portfolios Formed on Earnings/Price							
10 - ≤ 0	$\gamma_{i,1}$	0.44**	(0.00)	0.63	(0.08)	0.31**	(0.00)
5 - ≤ 0	$\gamma_{i,1}$	0.46**	(0.00)	0.50	(0.16)	0.32**	(0.00)
10 - ≤ 0	$\phi_{i,1}$	2.95**	(0.01)	8.94**	(0.00)	0.84	(0.24)
5 - ≤ 0	$\phi_{i,1}$	4.91**	(0.00)	10.81**	(0.00)	1.90	(0.08)

Note: This table contains the results about regressions of long-short portfolio returns on lagged SENTIMENT$^{\perp}$, investors' uncertainty (UC), the market factor (RMRF), the Fama-French factors (HML and SMB), the momentum factor (UMD), and the liquidity factor (LIQ),

$$r_{(i,k_2),t} - r_{(i,k_1),t} = \alpha_i + \gamma_{i,1}\mathrm{SENTIMENT}^{\perp}_{t-1} + \phi_{i,1}\mathrm{UC}_t + \beta_{i,\circ}\mathrm{RMRF}_t + \gamma_{i,2}\mathrm{SMB}_t + \gamma_{i,3}\mathrm{HML}_t + \gamma_{i,4}\mathrm{UMD}_t + \gamma_{i,5}\mathrm{LIQ}_t + \epsilon_{i,t},$$

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, and earnings/price) at time t, and $k_1, k_2 \in \{\leq 0, 1, 2, \ldots, 10\}$. "1", "2", ..., and "10" indicate the portfolios in the 1st (the smallest), 2nd, ..., and 10th deciles, respectively. "≤ 0" represents the portfolios for non-dividend-paying stocks or non-earning stocks. This table only reports the parameter estimates of $\gamma_{i,1}$. The sample period is from Jan. 1966 to Dec. 2005. The values of portfolio returns are in terms of percentage. SENTIMENT$^{\perp}$ is the Baker and Wurgler's orthogonalized sentiment proxy. $\mathrm{UC}_t = \pi_t(1-\pi_t)$ is the investors' uncertainty proxy devised by Ozoguz (2009), where π_t is defined as the filtered probability of regime 2, $\Pr(s_t = 2|\mathbf{Y}^t)$. The column "All" reports the results without regime sorting. The other columns "Regime j", $j = 1, 2$, show the results based on regime-sorted observations as regime= j. The bootstrapped p-values are in parentheses. $*$ and $**$ denote significance at 5% and 1% level, respectively.