Agent Sentiment and Stock Market Predictability

Chandler Lutz[*]

July 15, 2010

Abstract

Using only market return data we create an original index via a dynamic factor model for stock market sentiment. We find that rising sentiment relates to rising returns. Using our index we develop one-step ahead out-of-sample forecasts for market weighted returns. These forecasts outperform a random walk plus drift model and improve on previous out-of-sample exercises. We also employ dynamic Bayesian Model Averaging (BMA) and develop Two-Stage Bayesian Model Averaging (2SBMA). We find that these techniques can improve forecast performance. Lastly, we use our index to quantitatively chronicle sentiment cycles. We find that sentiment cycles often lead bear markets.

Keywords: Behavioral Finance, Agent Sentiment, Financial Forecasting

[*]University of California, Riverside, Department of Economics, Email: chandler.lutz@email.ucr.edu

"I can calculate the motions of the heavenly bodies, but not the madness of people."

-Sir Isaac Newton

Accounts of agent sentiment in asset markets have persisted throughout history: beginning with the Dutch tulip mania in the 1600s through the recent credit crisis of 2008. Developing further insights into how sentiment affects stocks is paramount to our understanding of financial markets.

Classical finance theory posits that investors are fully rational. In this framework, arbitrageurs make a risk-free profit on any security mispricing. This leads to the ideas of the efficient market hypothesis and elusive stock market predictability. The notion of predicting asset returns out-of-sample has intrigued researchers and practitioners for centuries. However, investors outperforming market averages or stock booms and busts have no place in the traditional finance literature. The large failure of the economic and financial literature in forecasting the stock market has added credence to the classical theories. Yet the outstanding performance of investors such as Warren Buffett and George Soros and the many bubbles and crashes throughout history created a dichotomy between researchers the practitioners. Behavioral finance attempts to close this chasm by making two basic assumptions: (1) Market participants are subject to sentiment or beliefs that are not necessarily supported by the facts and (2) there are limits and risks to arbitrage.[1] Using behavioral models, researchers attempt to explain booms and busts and stock market anomalies that escape traditional economic and financial research.

This paper attempts to extend the behavioral literature. Using a dynamic factor model and Bayesian estimation we develop a novel index for agent sentiment in the stock market. We argue that our index closely matches previous documentations of stock market bubbles. Using our index we conduct various in-sample regressions and find a relationship between rising sentiment and rising returns. This supports the idea of sentiment driven bubbles. We use our index to forecast out-of-sample market-weighted returns. We find that conditioning on sentiment allows

[1] Abreu and Brunnermeier (2003) produce a bubble model without this second assumption. Battalio and Schultz (2006) find no evidence of short-sale constraints during the 1990s tech bubble.

us to develop models that outperform the benchmark random walk plus drift under a rolling window and recursive estimation. While conducting our forecasting exercise we also develop an extension of the dynamic Bayesian model averaging (BMA) procedure of Raftery et al. (2005). We call this new technique two-stage Bayesian model averaging (2SBMA). The more flexible 2SBMA approach allows researchers to average forecasts from different models each with a different number of specifications. The technique of Raftery et al. (2005) requires that each forecasting model must have the same number of specifications. We find that both the BMA and 2SBMA procedures exhibit superior out-of-sample performance compared to the benchmark model. Lastly, we modify the Bry-Boschan (1971) algorithm to quantitatively date sentiment cycles. We find that sentiment cycles often lead general stock market cycles.

From a theoretical standpoint, De Long et al. (1990a,b) and Lee, Shleifer and Thaler (1991) implicate noise traders into their models. They find that sentiment affects a broad cross-section of stocks. Abreu and Brunnermeier (2003) construct a bubble model. They contend that sentiment and a dispersion of opinions affect market returns. When a bubble develops arbitrageurs may be limited by the costs or risks for short selling. One could could imagine that these limitations are not constant across firms. D'Avolio (2002) finds that young, small, unprofitable and growth stocks are more costly to buy and sell short. Baker and Wurgler (2006, 2007) find that these types of companies are most affected by sentiment. These firms do not pay dividends, have shorter earnings histories and are more volatile. These characteristics make companies more speculative in nature. This allows investors to plausibly defend a wider range of valuations. All these findings lead to the "Hard-to-Value, Difficult-to-Arbitrage" (HV-DA) hypothesis. This theory posits that the same stocks that are hard to value are also difficult to arbitrage. Baker and Wurgler (2006) develop a sentiment index at the macro level and run in-sample regressions on a cross-section of stock returns. They conclude that the HV-DA stocks are most affected by sentiment. Brown and Cliff (2004,2005) use sentiment driven surveys to study the predictability of asset returns. Lemmon and Portniaguina (2006) and Qui and Welch (2004) use consumer confidence to investigate return predictability. Glushkov (2006) augments Baker and Wurgler's (2006) sentiment index by including sentiment surveys, mutual fund flows, and short-sale and margin data. Using this new index, Glushkov (2006) develops a sentiment beta. This beta measures how much an individual stock is affected by sentiment.[2] Glushkov (2006) also accounts for size and volatility. He finds that stocks with more analyst following, greater institutional ownership and a higher likelihood of S&P500 membership are more affected by sentiment. Chung and Yeh (2009) account for regime changes with a Markov process. They use Baker and Wurgler's sentiment index to study the cross-section of stock returns. They find that sentiment's predictive power only prevails under a 'bullish' regime.

The index of Baker and Wurgler (2006, 2007) is the closest in spirit to the index developed in this paper. They consider six components for their index: The closed-end fund discount, the number and average first day return on IPOs, NYSE share turnover, the equity share of new issues and the dividend premium.[3] They compile their index by using the first principle component as the weight for each series. Baker and Wurgler (2006) develop an annual index for sentiment. This index captures swings in sentiment through the sample. Baker and Wurgler (2007) extend the index to monthly data. We plot Baker and Wurgler's (2007) monthly index (BWsent) against bear markets in figure 3. The sample period is from 1965M07 to 2007M12. BWsent generally follows the overall trends in sentiment. For example, Baker and Wurgler's index captures the low-sentiment times in the mid-1970s. Yet BWsent produces counterintuitive results for other bear episodes. Baker and Wurgler's index peaks in 2001M02. This is nearly a year after the tech crash and the bear market in 2000M03. This is particularly surprising since the tech boom and bust is largely attributed to agent sentiment (see, for example, Shiller, 2006). Additionally, BWsent climbs through the bear market that began in 1968M11 and reaches it global maximum in 1969M12. Baker and Wurgler's index also remains high through the bear market of the early 1980s. These features of BWsent contradict our expectations. Sentiment should peak prior to the end of speculative episodes and crash as bear markets take hold.

[Insert figure 3 here]

[2] This is synonymous to the popular market beta which relates a stock's return and the market average.

[3] Data and further explanations are available from Jeffrey Wurgler's website `http://pages.stern.nyu.edu/~jwurgler/`

Using their index, Baker and Wurgler (2006, 2007) also contend that high sentiment translates into lower future returns for the HV-DA stocks. Baker, Wugler and Yuan (2009) find similar results for other industrialized nations. These results contradict the theoretical findings of Abreu and Brunnermeier (2003). Abreu and Brunnermeier conclude that rational arbitrageurs have a profit incentive to ride sentiment bubbles over some interval. In other words, over a certain interval rising sentiment relates to rising returns. Furthermore, Lutz (2010d) considers the smoothed earnings-price ratio (ep10) as a measure of agent sentiment.[4] He considers in-sample regressions and an out of sample forecasting exercise. Lutz (2010d) finds that rising sentiment (as measured by ep10) relates to rising returns for stocks without dividends and companies without earnings. These results contradict the findings in Baker and Wurgler (2006, 2007). Furthermore, Lutz (2010d) contends that Baker and Wurgler's sentiment index 'mis-times' the conclusion of certain sentiment episodes.

In addition, Baker and Wurgler's index is difficult to calculate in real time. Their index is subject to data revisions. For example, one of their index components is the equity share of new issues. This series is only available monthly. Also, this series is subject to data revisions and only available for past values.

We use only market data to develop our sentiment index. This produces two main advantages: (1) There are no data revisions or measurement errors in our index components and (2) practitioners, firms or policymakers could easily calculate our index in real-time and make decisions based on current information. We use a dynamic factor model to calculate our index. We find that our index more closely corresponds to sentiment bubbles and crashes than the index proposed by Baker and Wurgler (2006, 2007).[5] Our index also closely coincides with anecdotal evidence of sentiment.

Another area of interest in the behavioral literature is stock market predictability. The empirical literature discovered an underreaction of stock prices to earnings announcements or

[4]ep10 is average earnings over the last ten years divided by the price for the S&P500. Averaging earnings over a ten year period mitigates business cycle effects. The result, according to Shiller (2006), is a series that captures agent sentiment in the stock market.

[5]See figures 2 and 3

other news. Researchers have also found an overreaction to a series of good or bad news. Barberis, Shleifer and Vishny (1998) construct a behavioral model to account for these anomalies. In a theoretical model Barberis, Huang and Santos (2001) find predictability in returns in aggregate. There appears to be consensus in the behavioral literature that asset returns are in fact predictable but the degree of this predictability is an open debate. Most of the work is concerned with in-sample predictability. There has been little success in out-of sample forecasting exercises especially when considering market-weighted returns. Timmermann (2008) documents the failure of previous works in forecasting returns out-of-sample and highlights the difficulty of such endeavors in his paper, "Elusive return predictability." Furthermore, Timmermann (2008) uses various models where the regressors are either past return values or standard macro of finance variables. He concludes that stock returns may contain local predictability (for short time periods). Then a wide range of practitioners will learn the model and it loses its power. Timmermann (2008) contends that this makes forecasting returns over long periods of time an extremely difficult task. It should be noted that Timmermann (2008) does not consider any behavioral variables. Campbell and Thompson (2007) find out-of-sample predictability for some variables compared the benchmark random-walk plus drift. However, they restrict the signs of their regression coefficients. In our forecasting exercise we use a variety of models and condition on our sentiment index to construct out-of-sample forecasts. We consider both rolling window and recursive estimation periods. We find that some of these models outperform the benchmark random-walk plus drift over an extended sample period. In particular, we find that linear models and forecast combinations outperform the benchmark.[6] Our forecasting results suggest that sentiment affects a broad cross-section of returns but they do not specify which stocks are most affected by sentiment.

I The Data

Baker and Wurgler (2006, 2007), among others, find that agent sentiment affects small, young, and volatile firms, firms without earnings or dividends, and firms in distress. These results

[6]See tables VI through XI for our forecasting results

help motivate the choice of the series used to construct the sentiment index in this paper.

We use long-short portfolios of equal weighted returns based on dividends, size and earnings. More specifically, we examine the difference between market returns for stocks that do not pay dividends and those that do, companies with earnings less than or equal to zero and those with positive earnings and small and large firms. We define a small (large) firm as one whose market cap is in the bottom (upper) 20 percent. We also consider the equal weighted returns of low momentum firms. We classify firms whose returns are in the bottom ten percent for the previous two to twelve months as having low momentum. We use the returns on low momentum firms to represent companies in distress. The series are all monthly and compiled from Kenneth French's data library.[7] The correlation matrix for these variables is presented in table I. The data for the series comprised from momentum, size, and dividends are available from 1927M07, but the data for earnings begins in 1951M07. Hence, our data contains 698 observations ranging from 1951M07 to 2009M08. We use the returns from the S&P500 to incorporate the findings of Glushkov (2006). He found that after controlling for size and volatility that larger stocks with a higher likelihood of S&P500 membership are more affected by sentiment.[8]

[**Insert table I here**]

Since we derive the data from market returns our index has two main advantages over other sentiment measures and macroeconomic indicators. First, there is no measurement error and no need for the data to be revised. This eliminates the uncertainty related to surveying and sampling that plagues many macro variables. Second, firms and policymakers could compile the data in real time and extend our model to make decisions based on current information.[9]

We abbreviate the series for the long-short portfolios of dividends, earnings and size as div, earn, and size, respectively. The returns for low momentum firms will be abbreviated as lowmom. We represent the S&P500 variable as sp500.

[7]Available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

[8]S&P500 return data is available publicly from Robert Shiller's website at `http://www.irrationalexuberance.com/`

[9]We use market data to capture behavioral biases in aggregate. The nature of the bias is left implicit in the data. Barberis, Huang and Santos (2001), Gervais and Odean (2001) and Grinblatt and Han (2005), among others, use a "micro" level approach. They explicitly model behavioral traits and use them to explain stock market anomalies.

II Model Estimation and Description

We combine the information from the five series and extract a common component using a dynamic factor model. The model takes the following form:

$$y_{it} = \gamma_i C_t + \varepsilon_{it}, i = 1, \ldots, 5 \tag{1}$$

$$\phi(L)C_t = \omega_t, \omega_t \sim N(0,1) \tag{2}$$

$$\psi_i(L)\varepsilon_{it} = v_{it}, v_{it} \sim N(0,\sigma^2) \tag{3}$$

Where y_{it}, $i = 1, ..., 5$, represents one of the five series: div, earn, size, lowmom, and sp500. C_t is the common component, ε_{it} is the idiosyncratic component, γ_i is the factor loading and L is the lag operator. To estimate the model and derive the common component, C_t, we cast it into state-space form. Although a unique representation does not exist we elect a form similar to that of Kim and Nelson (1998). This facilitates computation of the estimation algorithm. We multiply both sides of equation 1 by $\psi_i(L)$ which yields

$$\psi_i(L)y_{it} = \gamma_i \psi_i(L) C_t + v_{it} \tag{4}$$

The Bayesian Information Criterion guides our choice of $\phi(L)$ and $\psi_i(L)$ (given that the number of lags in each equation is greater than or equal to one). We select two lags for both $\phi(L)$ and $\psi_i(L)$. Equations 2 and 3 become

$$C_t = \phi_1 C_{t-1} + \phi_2 C_{t-2} + \omega_t \tag{5}$$

$$\varepsilon_{it} = \psi_{i1}\varepsilon_{i,t-1} + \psi_{i2}\varepsilon_{i,t-2} + v_{it}, i = 1, \ldots, 5 \tag{6}$$

Hence, we utilize the following state-space form:

Measurement Equation:

$$\begin{bmatrix} y_{1t}^* \\ y_{2t}^* \\ y_{3t}^* \\ y_{4t}^* \\ y_{5t}^* \end{bmatrix} = \begin{bmatrix} \gamma_1 & -\gamma_1\psi_{11} & -\gamma_1\psi_{12} \\ \gamma_2 & -\gamma_2\psi_{21} & -\gamma_2\psi_{22} \\ \gamma_3 & -\gamma_3\psi_{31} & -\gamma_3\psi_{32} \\ \gamma_4 & -\gamma_4\psi_{41} & -\gamma_4\psi_{42} \\ \gamma_5 & -\gamma_5\psi_{51} & -\gamma_5\psi_{52} \end{bmatrix} \begin{bmatrix} C_t \\ C_{t-1} \\ C_{t-2} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \\ v_{5t} \end{bmatrix} \tag{7}$$

Which in matrix form becomes

$$(\mathbf{y}_t = \mathbf{H}\beta_t + \mathbf{v}_t)$$

Where $y_{it}^* = y_{it} - \psi_{i1}y_{i,t-1} - \psi_{i2}y_{i,t-2}$ as in the left hand side of equation 4.

$$E(\mathbf{v}_t\mathbf{v}_t') = R = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix} \tag{8}$$

Transition Equation:

$$\begin{bmatrix} C_t \\ C_{t-1} \\ C_{t-2} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} C_{t-1} \\ C_{t-2} \\ C_{t-3} \end{bmatrix} + \begin{bmatrix} \omega_t \\ 0 \\ 0 \end{bmatrix} \tag{9}$$

Equation 9 can be written in matrix form as

$$(\beta_t = \mathbf{F}\beta_{t-1} + \mathbf{e}_t)$$

$$E(\mathbf{e}_t\mathbf{e}_t') = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{10}$$

We estimate the model using the Bayesian multimove Gibbs-sampling approach based on

Carter and Kohn (1994) and Kim and Nelson (1998) instead of the typical maximum likelihood. In the classical maximum likelihood approach inferences about the state vector are based on the estimated parameters. This is critical since the state vector contains the common component representing the sentiment index. In essence, the classical procedure forces the researcher to treat the estimated maximum likelihood parameters as true values when calculating the estimates for the state vector. We circumvent this issue through the Bayesian technique which allows us to jointly estimate the state vector and the model's parameters. To implement the estimation algorithm we use the MCMC Gibbs-sampling method. We run the algorithm 10,000 times and drop the first 2000 iterations. For further explanation of these techniques see Kim and Nelson (1998, 1999).

III Empirical Results

Table II shows the estimated parameters. As mentioned above, the factor loading for variable i is γ_i. The factor loadings are all have the expected positive sign. Furthermore, the loadings for div, earn, size, and lowmom are much higher than the loading for sp500. This results is not surprising. Based on prior research we expect the div, earn, size, and lowmom series to more directly capture sentiment.

[**Insert table II here.**]

We interpret the the common component, C_t, as a sentiment index for the stock market. We plot the series versus market weighted returns in figure 1. The sentiment index is noisy but there appears to be large aberrations in the index around times of high or low sentiment. For example, volatility is high around the tech bubble in 2000. The sentiment index is much more volatile than the market weighted returns during these sentiment episodes. To gain insight into the results of the index we filter the series using the HP filter with the smoothing parameter, $\lambda = 150$.[10] We present the augmented series in figure 2.

[**Insert figures 1 and 2 here.**]

The index appears to be correlated with bear markets and leads NBER recessions. We define

[10] Larger values of λ translate into more smoothing.

a bear market as a 20 percent or more drop in the overall market for at least a two month period. Also, the index levitates after many bear markets. For example, the index jumps following the tech crash in 2000 and the credit crash of 2008. This result is not surprising as firms most affected by sentiment produce their greatest returns immediately after a bear market. Using the filtered series makes the differences between market weighted returns and our index even more pronounced: our index captures major sentiment episodes such at the late 1960s and 1990s while the market weighted returns do not.

Next we conduct an exercise to determine if our index is robust to macroeconomic and business cycle indicators. In this regard, we use a process similar to Baker and Wurgler (2006). For each component of our index we control for (regress out) growth in industrial production, consumer durables and consumer nondurables. We also control for the one month Treasury rate and NBER recessions.[11] We then re-compile our index to develop a "macro-orthogonalized index" using the augmented components. If our original index and the macro-orthogonalized index are similar then we contend that our methodology is robust to business cycle indicators. The correlation between our original index and the macro-orthogonalized index is 0.993. There is little difference between our original index and the macro-orthogonalized index. Hence, for the remainder of the paper we only consider the original index.

Baker and Wurgler (2006, 2007) and Gluskov (2006) among others conclude that smaller stocks are more affected by sentiment. As a robustness check we use our index to calculate the sentiment beta for ten portfolios based on size. The sentiment beta is analogous to the popular market beta. Hence, a higher sentiment beta implies that a portfolio is more affected by sentiment. We follow Baker and Wurgler (2007) and control for market weighted returns. Based on prior research, we expect that portfolios consisting of smaller firms should have higher sentiment betas. Table V shows our results. Our findings correspond to our expectations: portfolios made up of smaller stocks have larger sentiment betas. We discuss the relationship between varying risk and our sentiment index below.

[11] We obtained growth in industrial production, consumer durables and consumer nondurables from the FRED economic database. The series IDs are INDPRO, IPDCONGD, and IPNCONGD, respectively. The one-month Treasury rate is available from Kenneth French's Data library.

[Insert table V here.]

We plot our index versus Baker and Wurgler's (2006, 2007) index in figure 3. In contrast to their index, our sentiment index peaks at the conclusion of speculative episodes and crashes as bear markets take hold. For example, our index peaks before the bear markets beginning in 1966M02, 1968M11, 1981M01, 1987M08, and 2000M03. In comparison, Baker and Wurgler's index peaks or remains high through the bear markets in the late 1960s, early 1980s and early 2000s.

Practitioners often consider the VIX volatility index as a "market fear gage." In other words, the VIX index rises along with investor fear. We expect that sentiment indexes should be negatively correlated with the VIX index: when fear is high sentiment should be low, and vice versa. However, we should not expect strong correlation with the VIX index. As stated by Whaley (2000), "VIX is more a barometer of investors' fear of the downside than it is a barometer of investors' excitement (or greed) in a market rally." We can use the VIX index as a robustness check for the performance of sentiment measures. We opt to use the old formula for the volatility index which is traded under the symbol VXO. VXO contains a longer sample than the VIX index.[12] We compare our raw sentiment index, Baker and Wurgler's sentiment index and VXO volatility for a common sample ranging from 1986M01 to 2007M12. As expected, our sentiment index and the VXO volatility index are negatively correlated. The correlation coefficient between our index and VXO is -0.14. In comparison, the correlation between Baker and Wurgler's sentiment index and the VXO volatility index is positive at 0.29. The positive correlation between Baker and Wurgler's index and the VXO index contradicts our expectations.

Baker and Wurgler (2006) document anecdotal evidence of stock market sentiment comprised from a number of sources between 1961 and 2002. Using the filtered index in figure 2, we find that our sentiment index generally coincides with their findings. First, Baker and Wurgler (2006) find a bubble during 1961. Our index reaches a local maximum during this period

[12] For the calculation of the VIX index and its comparison with VXO see `http://www.cboe.com/micro/vix/vixwhite.pdf`. The correlation coefficient between VIX and VXO over our sample is 0.988

and then falls as the bear market ensues in 1962. Also, Baker and Wurgler (2006) state that sentiment jumps in 1967 and 1968. Our sentiment index rises sharply during the late 1960s before peaking just prior to the bear market beginning in late 1968. Low sentiment ensued as a bear market set in between late 1968 and 1971. Our sentiment index shows a crash during this episode. Baker and Wurgler (2006) document bubbles in 1977 and 1978 and again in the first half of 1983. Our index shows spikes around these bubbles. These periods are associated with the end of bear markets. It may difficult to determine whether the rises in our index are due to the aforementioned bubbles or just a typical jump that follows many bear markets. Baker and Wurgler (2006) conclude that sentiment is high through the mid 1980s and then dissipates in 1987 and 1988. Our index climbs following the end of the bear market in 1982 until the stock market crash in 1987. After this crash however, our index moves upward through the recession and bear market of 1990. Our index shows a dramatic drop at the beginning of 1997 corresponding to the time of the Asian Financial crises. Our sentiment index peaks in the late 1990s as extreme optimism created by the tech bubble gripped the markets and then crashes as the bubble burst. Sentiment falls again prior to the credit crisis in 2008 and rises dramatically at the end of our sample as a bull market took hold beginning in March of 2009.

Using both the HP filtered series with the smoothing parameter, λ, equal to 150 and the unfiltered series we run regressions to determine the in-sample predictive value of the sentiment index. We lag the sentiment index by one period. We also control for the three factors, MKT, HML and SMB of Fama and French (1993) and a momentum factor as is typical in the finance literature. If one of the factors is used as the dependent variable we do not include it in our set of regressors. We run separate regressions using the MKT factor, the HML factor, the SMB factor and the momentum factor as dependent variables. We will also conduct a fifth regression where the dependent variable is the market weighted return, denoted MKT^1. Let us denote the sentiment series as $SENT$ and the momentum factor as UMD. Let the variable Z_t represent any of the five regressands. The regression equation becomes:

$$Z_t = \alpha + \beta_1 SENT_{t-1} + \beta_2 MKT_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_t \qquad (11)$$

We provide the regression results for the filtered and unfiltered sentiment series in tables III and IV, respectively. The true distribution of the errors is likely non-normal. Hence, we calculate the bootstrapped standard errors as in Stambaugh (1999). The size component used to develop our index is similar to the Fama and French (1993) SMB factor.

[Insert tables III and IV here.]

We first consider the results when the sentiment index is filtered. The regression coefficient on $SENT_{t-1}$ is positive and significant when MKT_t or MKT_t^1 is the dependent variable. A positive regressor on sentiment implies that an increase in sentiment at time $t-1$ relates to an increase in the overall market return in time t. Moreover, these coefficients are above one which may suggest a strong relationship between the market returns and the sentiment index. When the dependent variable is HML or UMD the coefficient on $SENT_{t-1}$ is negative and insignificant.

Next we examine the results where the sentiment index is not filtered. When the dependent variable is either MKT or MKT^1, the coefficient corresponding to the sentiment index becomes insignificant but the sign remains positive. It is possible that the noise that we filtered out in the sentiment index is related to the noise in the other Fama-French factors. This may explain the smaller coefficient and lack of significance in this case. Additionally, for the regressions where HML and UMD are the dependent variables the sign for the coefficient on $SENT_{t-1}$ becomes positive and significant.

When SMB is the dependent variable the β for the sentiment index is positive and significant for both the filtered and unfiltered index. It is much larger when the sentiment index is filtered. Baker and Wurgler (2006) also construct a sentiment index and run a similar regression using the SMB factor as the dependent variable. They find the coefficient on their sentiment index to be negative and significant. Their sample contained data from 1963M01 to 2001M12 while our sample draws from 1951M07 to 2009M08. Also, they use significantly different components for their index.

The positive coefficients on the sentiment index when the dependent variable is MKT and MKT^1, among others, may support the idea of sentiment driven bubbles. An increase in sentiment at time $t-1$ relates to an increase in the market averages (or other dependent variables) at time t. Hence, broad increases of sentiment over time translate into higher aggregate returns. These results contradict the empirical results of Baker and Wurgler (2006, 2007) but coincide with the theoretical results of Abreu and Brunnermeier (2003). Baker and Wurgler contend that high sentiment implies lower future returns. Abreu and Brunnermeier conclude that rational arbitrageurs have a profit incentive to ride sentiment bubbles. In other words, they find that rational arbitrageurs may not necessarily sell out if sentiment is high. Lutz (2010d) uses ep10 as a measure of sentiment. For sentiment portfolios based on earnings and dividends he found that rising sentiment relates to rising returns. One must be careful when interpreting these results as correlation does not necessarily imply causation. As noted in the introduction, in the literature the affect of sentiment on a cross-section of stocks is mixed.

Next we examine the relationship between our sentiment index and risk. The rational-based Capital Asset Pricing Model (CAPM) contends that riskier portfolios should always have a higher expected return. The popular beta determines the portfolio's risk. Hence, portfolios with higher betas should earn higher expected returns. From this perspective, one may argue that our sentiment index is just capturing varying degrees of risk rather than the effects of sentiment. To study this issue we consider an exercise similar to that of Baker and Wugler (2007). We form ten decile portfolios based on size.[13] The first row in table V shows the market beta for these portfolios. The results coincide with our expectations: the portfolios made up of smaller stocks are riskier and have larger betas than those consisting of larger stocks. The average returns, also depicted in table V, are higher for portfolios with larger betas. Hence, for average returns the CAPM predictions hold. In addition, we examine portfolio returns when sentiment is high in the previous month and when sentiment is low in the previous month. Sentiment is low during the previous month if it is below the mean and vice versa for high sentiment during the previous month. Table V and figure 4 show the results. When sentiment is high during the previous month the CAPM's effects become more pronounced: portfolios with higher betas earn much higher returns. Furthermore, the high beta portfolios earn better

[13] Size portfolios are from Kenneth French's data library.

than average returns while low beta portfolios earn below average returns. The results reverse when sentiment is low during the previous month. In this case high beta portfolios earn much lower returns than low beta portfolios. These results violate the CAPM predictions. Baker and Wurgler (2007) label the graph in figure 4 as "the sentiment seesaw." Not surprisingly, given the above regression results, our empirical seesaw and Baker and Wurgler's are inverses. However, we both find that the CAPM fails to explain the results: for a given level of sentiment high beta stocks earn lower returns. In Baker and Wurgler's words, "This is a powerful confirmation of the sentiment-driven mispricing view."

[**Insert figure 4 here.**]

IV Out of Sample Forecasting

Now we examine the out of sample forecasting performance of the sentiment index outlined in section II. We follow Timmermann (2008) and attempt to forecast one-step ahead market weighted stock returns. Using market-weighted returns ensures that our results are robust to both large and small firms.[14] To avoid look-ahead bias we will not use the exact sentiment index estimated in section III. Instead we re-estimate the model in section II for each in-sample period considered.

To evaluate the forecasts we employ the mean-absolute error (MAE) and the out-of-sample (OOS) R^2 measure used in Campbell and Thompson (2007) and Goyal and Welch (2008). The formula for the OOS R^2 is

$$\mathrm{R}^2 = 1 - \frac{\mathrm{MSE}_n}{\mathrm{MSE}_{RWD}} \tag{12}$$

Where MSE_n is the mean-squared error for model n and MSE_{RWD} is the mean-squared error for the benchmark random walk plus drift. If the OOS R^2 is positive then model n outperforms the benchmark. We follow the literature and express the OOS R^2 statistic as a percentage. We consider several models to develop our forecasts. We have information available up to time period t and hope to forecast returns at time $t+1$. Let r_{t+1} represent returns for time $t+1$

[14]Some authors also consider equal-weighted returns. However, these averages are biased towards small firms and do not truly represent market averages.

and ε_{t+1} represent an unpredictable white noise process.

The benchmark model we consider is the random walk plus drift:

$$r_{t+1} = \beta_0 + \varepsilon_{t+1} \tag{13}$$

The forecast for this model is the unconditional mean. In out of sample forecasting exercises researchers have struggled to outperform this benchmark on a consistent basis as evinced in Timmermann (2008).

The next model we consider is a linear autoregressive model where lags of the market return and the sentiment index represent the independent variables. The specification for this model is as follows:

$$r_{t+1} = \beta_0 + \sum_{j=1}^{p} \alpha_j r_{t+1-j} + \sum_{j=1}^{q} \beta_j SENT_{t+1-j} + \varepsilon_{t+1} \tag{14}$$

The lags for p and q were chosen using the Bayesian Information Criterion (BIC). We force the model to contain at least one independent variable. This ensures that the linear model differs from the random walk plus drift. We discuss the performance of various autoregressive and sentiment lags in section B.

We also consider two nonparmetric models:

$$r_{t+1} = m(SENT_t) + \varepsilon_{t+1} \tag{15}$$

$$r_{t+1} = m(r_t, SENT_t) + \varepsilon_{t+1} \tag{16}$$

We estimate equation 15 using local liner least squares and equation 16 using local constant least squares.

We also consider a NARX neural network which approximates an equation of the following form in a non-linear fashion:

$$r_{t+1} = f(r_t, r_{t-1}, \ldots, r_{t-p}, SENT_t, SENT_{t-1}, \ldots, SENT_{t-q}) \tag{17}$$

For our NARX forecast we consider a model with one hidden layer and one output layer. We use ten neurons in the hidden layer. To train the network, we opt for Bayesian regulation. This helps avoid overfitting. For more on Bayesian regulation see MacKay (1992) or Foresee and Hagan (1997). We select the lengths of p and q by regressing the fitted values on the actual data. This is standard practice in the neural network literature.

The last model that we entertain is a Generalized Regression Network. These models contain a hidden radial basis layer and special linear layer. One advantage of these models is that they can be trained rather quickly. To use this model we simply specify the market weighted returns as the target vector and the $SENT_{t-1}$ series as the input vector. For more on Radial Basis Networks see Wasserman (1993).

Finally, we average the forecasts over all the models to form a combined forecast.

We consider both rolling window and recursive estimates of the sentiment index and our forecasting models. The data generating process (DGP) for asset returns most likely changes over time. Hence, there is a tradeoff between the recursive and rolling window methods. The recursive procedure uses more information which may yield better estimation results of the model parameters. A rolling window allows the model to adapt to changes in the DGP.

First, we consider a rolling window of ten years or 120 monthly observations. The first forecast corresponds to 1970M01. For each window we estimate the sentiment index as outlined in section II. Then we form the one-step ahead forecast for the various models discussed above. The exact structural form of the sentiment index most likely changes over time. Thus, we consider three different models relating to the different number of lags in for $\phi(L)$ and $\psi(L)$ for every window.[15] We then average the results from these models to develop a single sentiment index. We consider the index as an exogenous independent variable. Then we form the one-step ahead forecast of market weighted returns. We present the results for the MAE and OOS R^2 in the left panel of table VI. RWD is the random walk plus drift model. NP1 and NP2 correspond to the non-parametric models in equations 15 and 16, respectively. Radnet represents the radial basis neural network. Surprisingly, the linear model produces the lowest MAE for the entire sample. This autoregressive model also performs better out of sample than the benchmark random walk plus drift in the 1970 and 2000 subperiods. The NARX model

[15]We used one lag for $\phi(L)$ and one lag for $\psi(L)$ for the first model, two lags for $\phi(L)$ and one lag for $\psi(L)$ for the second, and two lags for both $\phi(L)$ and $\psi(L)$ for the third.

performed well in the 1970s and 2000s but failed to outperform the benchmark model in the overall sample. The rest of the models produced higher MAEs over the entire sample. Also, the OOS R^2 is positive for the linear model. This suggests that the linear model outperforms the benchmark for the overall sample.

[**Insert table VI here.**]

Next, we estimate both the sentiment index and forecasting models recursively. Table VI contains the results. In this case, the linear model outperforms the benchmark model by a wider margin. In contrast to previous results, the NARX model outperforms the random walk plus drift over the entire sample. The NARX model also beats the benchmark in all subperiods except the 1990s. Furthermore, the NARX model produces a lower MAE under recursive estimation. All other models fail to produce lower MAEs than the benchmark for all subperiods. The average forecast outperforms the random walk plus drift for the 1970, 1980 and 2000 subperiods. The OOS R^2 is positive for the NARX model under recursive estimation. Lastly, the R^2 increases significantly for the linear model.

The MAEs are generally equivalent or smaller when we estimate both the sentiment index and the forecasting models using a rolling window. The average linear forecast is equivalent under both approaches. The exceptions are the neural networks. They produce superior forecasts under the recursive approach. The MAE for the benchmark random walk plus drift is smaller under the rolling window approach. This may suggest that using a rolling window allows the benchmark model to adjust to changes in the DGP.

A Bayesian Model Averaging

The true data generating process (DGP) of a series is often not known to researchers. As a result, the researcher must consider a large set of models. Classical regression analysis utilizes specific tools to choose the best model. This model is then used for prediction or forecasting.[16] This approach fails to take into consideration the uncertainty surrounding the choice of the optimal model. Neglecting this uncertainty may lead to errors in inference. Also, only using

[16]Such tools include R^2 measures, Bayesian and Akaike information criterions, log-likelihood values, etc.

one model may disregard information contained in alternative specifications.

In the above forecasting section we combined out-of-sample predictions using a simple average. Now we turn to more advanced mixing techniques. In particular, we employ Bayesian model averaging (BMA). BMA takes into account the uncertainty regarding the true nature of the process. There are two main ways implement BMA. The first method assumes that there is one correct model but the researcher is ignorant to its true form. Raftery (1995) and Hoeting et al. (1999) outline this use of BMA. Raftery et al. (2005) develop a dynamic BMA procedure where the correct model changes from time period to time period. The dynamic BMA technique draws the optimal specification from the weighted average over a given set of models. Furthermore, this technique only requires the model forecasts. This makes it convenient to average over linear and non-linear specifications. In essence, the dynamic BMA procedure produces an optimal forecast by mixing over the forecasts from a given set of models.

In all likelihood, the DGP of stock returns changes over time. Also, we want to combine the linear and non-linear models depicted above. Hence, we select the dynamic BMA approach.

Following the notation of Raftery et al. (2005), the BMA predictive pdf is given by

$$p(y|f_1, \ldots, f_K) = \sum_{k=1}^{K} w_k g_k(y|f_k) \tag{18}$$

Where y is the quantity to be forecast, K is the number of models, w_k is the weight applied to each model, and $g_k(\cdot)$ is an assumed probability distribution. w_k is the posterior probability. We interpret $g_k(y|f_k)$ as the pdf of y conditional on f_k given that model k produced the best forecast over the training period.

We form the log-likelihood function based on the distribution in equation 18. We specify a functional form for $g_k(y|f_k)$. Since this is a finite mixture model we can apply the EM algorithm to maximize the likelihood function. Before we execute the EM algorithm we must also choose the length of the training period. To accommodate the changes in the DGP over time we use a rolling window of forecast data to estimate the likelihood function. Shorter rolling windows allow the model to adjust quickly to changes in the structure of the process. Longer windows use more information. This yields better estimation of model parameters. Researchers widely believe stock returns follow a distribution with thick tales such as the t-distribution. Despite this fact, we consider the normal distribution for $g_k(\cdot)$. As noted by Peel and McLachlan (2000), mixing over the t-distribution simply produces less extreme results than the normal distribution. Hence, using the normal distribution instead of the t-distribution allows us to add larger weights to better forecasts and further punish poor performing models.[17]

Raftery et al. (2005) note that the weights calculated by the BMA technique consider the performance over the training period and the correlation of the forecasts. A model may obtain a larger weight than one might expect if its forecasts are uncorrelated with others. Table VII shows the correlations of the forecasts. The random walk plus drift, linear and NARX models are all highly correlated with a coefficient around 0.65. The non-parametric models are highly correlated with each other but much less so with other models. The Radnet model is almost completely uncorrelated with the others. The non-parametric and the Radnet models perform extremely poorly out-of-sample. Yet the dynamic BMA process will still place significant weights on their forecasts due to their small correlation with the other models. Hence, we drop these specifications and only average over the random walk plus drift, the linear and the NARX models.

We only consider the forecasts from the random walk plus drift, the linear model and the NARX model under a rolling window.[18] We discard the first 60 observations to use as verification. This implies we are eliminating half of the 1970s from our out-of-sample exercise. As noted in table VI, the 1970s were a subperiod where both the linear model and the NARX model outperformed the benchmark. Hence, not using the first half of the 1970s will negatively affect our results. Despite this handicap, we do not change the out-of-sample period to avoid data mining.

The left panel of table VIII shows the results when we average over the random walk plus drift, linear and NARX models. The first column corresponds to the length of the training

[17]Results for the t-distribution are available upon request. Using the t-distribution did not improve our results.

[18]Forecasts under a rolling window produced lower MAEs for the random walk plus drift. Hence, the random walk plus drift under a rolling window is a more difficult benchmark to beat.

period. In columns 2 - 4 we show the posterior weights and their standard deviations in parentheses. We use the OOS R^2 statistic to judge forecast performance. The mean of the posterior weight for all the models hovers around 1/3. However, the weight on the random walk plus drift is usually the highest followed by the linear and NARX models. The BMA forecasts outperform the the benchmark random walk plus drift for all training lengths considered. The best performing model uses 20 training periods and produces an R^2 of 0.52.

[**Insert table VIII here.**]

We find that the standard deviations for the posterior weights are quite large. Hence, the optimal forecasting model changes drastically over time. We plot the posterior weights in figure 5 using a training length of 35 periods. We also shade bear markets. However, this shading can be misleading. Recall that BMA calculates the posterior weights based on prior training lengths. For this case, a posterior weight describes the forecasting performance over the previous 35 training periods. For example, the random walk plus drift model commands a large posterior weight in the early 2000s. This suggests that the random walk plus drift was the top performer in the late 1990s.

[**Insert figure 5 here.**]

As mentioned above, the BMA process will add extra weight to forecasts that are uncorrelated with others. When all the forecasts considered are correlated at the same level, however, we can interpret the BMA posterior weights as out of sample model performance over the training period. This is the case for the random walk plus drift, linear and NARX models. Hence, mean weights for the three models imply similar out of sample forecasting performance for the random walk plus drift and linear frameworks and slightly worse performance for the NARX model.

Next, we apply the BMA technique over just the linear and random walk plus drift models. The right panel in table VIII holds the results. The posterior weight for the random walk plus drift hovers around 0.55. This compares to a weight of about 0.45 for the linear model. These results suggests that random walk plus drift usually provides a slightly better forecast. The standard deviations for posterior weights are quite large. Hence, the optimal model changes over time. The BMA forecast outperforms the benchmark except when we use a training length of ten periods. Furthermore, the largest R^2 is 0.85 in this case. In comparison, the largest R^2 is 0.52 when we average or the random walk plus drift, the linear and the NARX models. Thus, averaging over just the random walk plus drift and the linear model improves performance.

B Two-stage Bayesian Model Averaging

While we mix forecasts in the above BMA procedure from various models we use conventional econometric methods to determine their specifications. This neglects the uncertainty surrounding the choice of model specifications. To account for this uncertainty and combine the forecasts from the various models we employ a method we call two-stage Bayesian model averaging (2SBMA). Before we describe this method let us introduce basic terminology to facilitate the discussion.

Let a *model* be a generic functional representation of the process and a *specification* be a single representation of a model. For example, the models will be the random walk plus drift, linear and NARX. For the linear model a given specification will be a certain number of autoregressive or sentiment lags. The random walk plus drift model only has one specification (corresponding to zero autoregressive and sentiment lags). The linear and NARX models have potentially an infinite number of specifications.

The BMA technique of Raftery et al. (2005) allows for different models but the procedure is difficult to implement if models have different specifications. The 2SBMA approach augments the standard BMA technique to allow models to have different specifications. More explicitly, the 2SMBA approach requires two steps: (1) apply the BMA procedure for each model over all considered specifications; and (2) apply the BMA approach again over all models. In essence, we average over all specifications for each model and then over all models. The 2SBMA technique also allows the researcher to discard specifications that are known to perform poorly or add specifications which may forecast well for some models but not others. Overall, the 2SBMA approach is a simple and flexible framework to combine forecasts. The 2SMBA technique also allows researchers to incorporate linear and non-linear models and their various specifications.[19] In the first stage, the optimal number of training periods may differ across models. Thus, we may want to use different training lengths for different models in the first stage. For example, table IX shows that about 30 training periods are appropriate for the linear model and table X shows that 60 training periods are optimal for the NARX neural network. Ideally, in the second stage we would use the forecast that corresponds to 30 training lengths for the linear model and 60 training lengths for the NARX network. Unfortunately, blindly using these training lengths would lead to look-ahead bias. To circumvent this issue, in stage one we record the OOS R^2 for each model and all training lengths up to time period t. We choose the forecast from each model that produces the highest R^2. Then we conduct the second stage of the process to make the 2SBMA forecast for time $t+1$. This process allows us to use the optimal forecast from stage one in stage two.

Tables IX and X show the results from the first stage of the 2SMBA process for the linear and NARX models, respectively. We do not need to consider the random walk plus drift model in the first stage as it only has one specification. We use the normal distribution for the mixing process. As before, we use the first 60 forecasts as verification. For the linear model, the BMA procedure places the highest weight on the specification where $(p = 0, q = 1)$.[20] This weight is twice as large as the posterior probability for any other specification. The mixing process also places significant weight on the specifications where $(p = 1, q = 0)$, $(p = 2, q = 0)$ and $(p = 0, q = 2)$. The standard deviations for the posterior weights are large. This indicates that the optimal specification changes over time. Additionally, using a training length of 30 periods produces the largest OOS R^2. The BMA forecast for the linear model beat the random walk plus drift by a larger margin than the linear model in section IV. Furthermore, the BMA forecast for the linear model outperforms the benchmark when we use 15-45 or 60 training periods.

[Insert tables IX and X here.]

[19] One could also average the BMA forecasts in some manner over the various training periods. An equal weighted average over different training periods would be easy to implement.

[20] p and q are the autoregressive and sentiment lags, respectively. See equation 14 for the linear model.

The posterior weights for the NARX model are relatively equal across all specifications and training periods. The BMA forecast for the NARX model outperforms the benchmark when 20, 25, 35, 40, 45, 50, 55 and 60 training periods are used.[21]

The left panel in table XI displays the second stage the the 2SBMA process. In this case we average over the random walk plus drift, linear and NARX models. The largest posterior weight belongs to the random walk plus drift model, followed by the linear model. The 2SBMA process places the lowest weight on the NARX model. The R^2 is positive when the training length is set to 15 - 45 or 60 periods.

[**Insert table XI here.**]

We also consider the 2SMBA approach over just the linear and random walk plus drift models. The right panel of XI shows the results. Using only the linear and random walk plus drift models produces even better forecasts. All of the R^2's are extremely large in magnitude and above 1.5 percent. When we use 10 training lengths, the OOS R^2 is 2.319. The 2SBMA forecasts are among the best in our study.

We use 120 observations for verification in the 2SBMA approach. This implies that we do not include the 1970s in our 2SBMA forecasts. Recall that the 1970s were a period where both the linear and NARX models outperformed the benchmark. Hence, our results may be negatively affected. Yet the 2SBMA forecasts still beat the random walk plus drift. We do not adjust the sample to include the 1970s to avoid data mining.

The drawback of the 2SBMA approach compared to the BMA process is that it requires twice as many observations for verification. This cost is minimal, however, when a large time series of forecasts is available (as it is in our case) or when the researcher is mostly concerned with predicting future, unobserved values. Considering shorter training windows also helps substantially reduce this cost.

[21] When only considering autoregressive lags, instead of the NARX framework we use a time delayed neural network. Also, when the independent variables are only lags of the sentiment index we use a standard feed-forward neural network.

C Comparison with other stock market forecasts

There has been a long debate over the predictability of asset returns. Pearson and Timmermann (1995) conclude that return predictability is limited to certain subperiods. Campbell and Thompson (2007) find that some variables exhibit small out-of-sample superiority compared to a recursively estimated random walk plus drift. However, they restrict the signs of the regression coefficients. They also conclude that a variable loses its predictive power over time. Campbell and Thompson also find that returns are least predictable after 1980. Our 2SBMA models produce large and positive R^2's after 1980. Goyal and Welch (2008) contend that returns are not forecastable using typical variables. Furthermore, Timmermann (2008) considers a battery of models and regressors. Using market-weighted returns all of his models fail to outperform the random walk drift under a rolling window. Using recursive estimation he finds that only an equal-weighted average of forecasts outperforms the benchmark random walk plus drift. He concludes models can at best achieve local predictability. Lutz (2010d) considers Baker and Wurgler's sentiment index, ep10, and the VIX volatility index as exogenous variables.[22] He uses a number of unrestricted and restricted models. Lutz (2010d) finds that only the VIX volatility index is able to forecast market weighted using a NARX network. Our results support the proposition that stock returns are forecastable. Controlling for sentiment, we use both a rolling window and recursive estimation. We find individual and average forecasts that outperform the benchmark under both a rolling window and recursive approach.[23]

V Sentiment Cycles

Using the index outlined in sections II and III we date cycles in sentiment. To our knowledge, no other work has quantitatively chronicled sentiment episodes.[24] We apply a modified Bry-Boshan algorithm to date sentiment cycles. The Bry-Boschan algorithm is a set of conditional

[22]See footnote 4 for an explanation of the ep10.

[23]As previously stated, we find that the average MAE under the rolling window is lower than under recursive estimation for the random walk plus drift. Hence, superior performance over the random walk plus drift using a rolling window would be much more useful to practitioners.

[24]Baker and Wurgler (2006) document anecdotal periods of sentiment. Their approach is qualitative. We compare our sentiment index to these accounts in section III.

rules to find local highs and lows in a series. We modify the window lengths and the and cycle durations in the algorithm to fit our noisy data. We also require that the magnitude of each phase be greater than 1.5 standard deviations of the entire index. This ensures that small blips are not recorded as peaks or troughs in sentiment. Appendix A shows the exact set of rules we use for the algorithm. Table XII and figure 6 show the results. The shaded areas in the figure represent peak to trough episodes in agent sentiment. The modified algorithm appears to capture the dynamics of the index. Furthermore, sentiment usually peaks prior to bear markets. This is the case for the bear markets that began in 2008, 2000, 1987, 1981, 1976, 1968, 1966 and 1957. Our algorithm does not record a peak in sentiment prior to the bear market in 1961. Also, sentiment peaks just after the light bear market in the early 1990s. The algorithm also records many false alarms. These usually occur as sentiment levitates following bear markets. For example, the algorithm records peaks following the bear markets in 2002, 1990, 1982 and 1957. As noted above, these results are not that surprising since the stocks most affected by sentiment produce their greatest returns following bear markets. Clearly, there is a correlation between sentiment and bear markets. However, more study is needed to determine if a causal relationship exists and how this affects practitioners and policymakers.

[**Insert table XII and figure 6 here.**]

VI Conclusion and further discussion

Using a dynamic factor model and Bayesian estimation we construct an index for stock sentiment using only market return data. This index generally matches episodes of sentiment bubbles. We run in sample regressions using our index and find a relationship between increasing sentiment and increasing stock returns. This result supports the idea of sentiment driven bubbles and crashes. In an out-of-sample forecasting exercise we find that conditioning on sentiment allows us to outperform a benchmark random walk plus drift model over our sample period. The results are robust for rolling window and recursive estimation. To our knowledge, this result is new to the literature. These findings support the notion that sentiment affects a broad cross section of stocks. Furthermore, we find that sophisticated averaging techniques, such as Bayesian model averaging (BMA), can enhance forecasting performance. In this regard, we develop a simple extension called two-stage Bayesian model averaging (2SBMA). This technique allows researchers to combine forecasts from different models all with different specifications. 2SBMA produces more accurate predictions than the standard BMA approach in our forecasting exercise. Lastly, we date cycles in sentiment. We find that sentiment cycles often lead bear and bull markets. Future research may investigate how sentiment cycles affect bull and bear markets, how sentiment affects stocks in aggregate, and what role does sentiment play in the formulation of bubbles.

A Appendix: Procedure For Determination of Turning Points

1. Remove outliers from the data and replace them using the Spencer curve.
2. Determination of initial turning points in raw data
   (a) Determination of initial turning points in raw data by choosing local peaks (troughs) as occurring when they are the highest (lowest) values in a window 12 months on either side of the date.
   (b) Enforcement of alternation of turns by selecting highest of multiple peaks (or lowest of multiple troughs)
3. Censoring operations (ensure alternation after each)
   (a) Elimination of turns within 6 months of the beginning and end of the series
   (b) Elimination of peaks (or troughs) at both ends of the series which are lower (or higher) than values closer to the end
   (c) Elimination of cycles whose duration is less than 24 months
   (d) Elimination of phases whose duration is less than 4 months or whose magnitude is smaller than 1.5 standard deviations of the sentiment index.
4. Statement of final turning points

B Appendix: Tables and Figures

Table I: The correlation of sentiment components

	Div	Earn	Size	Lowmom
Div	1	0.86	0.8	0.79
Earn	0.86	1	0.73	0.69
Size	0.8	0.73	1	0.62
Lowmom	0.79	0.69	0.62	1

Table II: Bayesian posterior distribution.

Variable	Mean	Med	Std
ϕ_1	0.233	0.234	0.014
ϕ_2	0.009	0.009	0.014
γ_1	3.660	3.657	0.092
ψ_{11}	0.012	0.012	0.055
ψ_{12}	0.006	0.006	0.055
σ_1^2	0.561	0.555	0.121
γ_2	3.778	3.777	0.119
ψ_{21}	-0.168	-0.170	0.043
ψ_{22}	0.025	0.025	0.042
σ_2^2	4.809	4.794	0.287
γ_3	6.282	6.274	0.237
ψ_{31}	0.195	0.195	0.041
ψ_{32}	-0.074	-0.074	0.041
σ_3^2	24.211	24.182	1.358
γ_4	3.233	3.230	0.120
ψ_{41}	0.040	0.040	0.040
ψ_{42}	0.053	0.054	0.041
σ_4^2	6.414	6.401	0.374
γ_5	1.324	1.321	0.118
ψ_{51}	0.220	0.220	0.039
ψ_{52}	-0.055	-0.054	0.037
σ_5^2	9.728	9.683	0.534

Notes: See section II for a description of the model. $i = 1, \ldots, 5$ represent the five series, div, earn, size, lowmom and sp500, respectively. γ_i is the coefficient on the common component in equation 1. ϕ_i are the coefficients on lags of the common component in equation 2. ψ_i are the coefficients on lags of the idiosyncratic component in equation 3.

Table III: Regression Results for filtered sentiment

Dependent Var	α	$SENT_{t-1}$	MKT_t	SMB_t	HML_t	UMD_t
MKT_t	0.7763*	1.3788*	-	0.247*	-0.4141*	-0.1764*
	(0.000)	(0.0008)		(0.000)	(0.000)	(0.000)
HML_t	0.616*	-0.2702	-0.1681*	-0.1453*	-	-0.1486*
	(0.000)	(0.1549)	(0.000)	(0.000)		(0.000)
UMD_t	0.9961*	-0.2314	-0.1704*	-0.0478	-0.3536*	-
	(0.000)	(0.2969)	(0.000)	(0.186)	(0.000)	
SMB	0.2113*	1.6994*	0.1168*	-	-0.1692*	-0.0234
	(0.0283)	(0.000)	(0.000)		(0.000)	(0.1937)
MKT_t^1	1.1706*	1.272*	-	0.2478*	-0.4118*	-0.1742*
	(0.000)	(0.0008)		(0.000)	(0.000)	(0.000)

Notes: Results for the regressions based on equation 11. For these results the sentiment index is filtered using the HP filter ($\lambda = 150$). Bootstrapped p-values are in parentheses. An asterisk means the variable is significant at the five percent level.

Table IV: Regression results for raw sentiment.

Dependent Var	α	$SENT_{t-1}$	MKT_t	SMB_t	HML_t	UMD_t
MKT_t	0.7759*	0.0564	-	0.2866*	-0.4287*	-0.1831*
	(0.000)	(0.3498)		(0.000)	(0.000)	(0.000)
HML_t	0.6242*	0.1686*	-0.171*	-0.1585*	-	-0.1524*
	(0.000)	(0.0358)	(0.000)	(0.000)		(0.000)
UMD_t	1.0053*	0.3986*	-0.1725*	-0.0686*	-0.3599*	-
	(0.000)	(0.0024)	(0.000)	(0.097)	(0.000)	
SMB	0.2135*	0.272*	0.1386*	-	-0.1922*	-0.0352*
	(0.0259)	(0.0045)	(0.000)		(0.000)	(0.093)
MKT_t^1	1.1696*	0.0363	-	0.2849*	-0.4249*	-0.18*
	(0.000)	(0.4016)		(0.000)	(0.000)	(0.000)

Notes: Results based on equation 11. Bootstrapped p-values are in parentheses. An asterisk means the variable is significant at the five percent level.

Table V: 10 Size Portfolios

	Lo 10	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9	Hi 10
Beta	1.0823	1.1625	1.1646	1.1381	1.1283	1.0830	1.0887	1.0746	1.0036	0.9379
Sent-beta	3.2736	2.3234	1.7275	1.3755	1.0572	0.7079	0.4986	0.2850	-0.0728	-0.4899
HighSent_{t-1}	2.0914	1.6304	1.5132	1.3974	1.3280	1.1227	1.2007	1.0585	0.9559	0.7514
Average	1.1664	1.1307	1.1738	1.1243	1.1433	1.0821	1.0916	1.0468	1.0049	0.8787
LowSent_{t-1}	0.3835	0.7022	0.8743	0.8822	0.9756	1.0354	0.9837	1.0196	1.0294	0.9692

Notes: The market beta and sentiment beta for the portfolios are in rows 1 and 2. Row 3 holds the average monthly return when sentiment was high in the previous month, row 4 holds the average monthly return over the whole sample and row five holds the average monthly return when sentiment was low during the previous month.

Table VI: Forecasting Results

	Rolling Window						Recursive					
	Overall	1970s	1980s	1990s	2000s	R^2	Overall	1970s	1980s	1990s	2000s	R^2
RWD	3.536	3.764	3.567	3.020	3.802	0.000	3.539	3.785	3.576	3.038	3.767	0.000
Linear	3.532	3.743	3.576	3.050	3.766	0.091	3.532	3.764	3.561	3.075	3.735	0.611
NARX	3.553	3.747	3.612	3.060	3.800	-0.807	3.534	3.768	3.575	3.050	3.748	0.199
NP1	4.550	4.699	4.688	4.125	4.693	-60.335	4.592	4.946	4.597	4.094	4.737	-67.231
NP2	4.804	5.034	4.956	4.343	4.885	-75.027	4.804	5.034	4.956	4.343	4.885	-80.076
Radnet	4.720	4.703	5.312	3.912	4.960	-84.758	4.111	4.098	4.301	3.435	4.629	-41.587
Average	3.659	3.773	3.818	3.232	3.820	-6.766	3.627	3.770	3.735	3.223	3.787	-5.720

Notes: The mean absolute error and the OOS R^2 for the models outlined in section IV. The left panel shows the results under rolling window estimation. The right panel shows the results under recursive estimation. The first column lists the models. "Overall" is MAE for the entire sample. Columns 2 - 5 and 8 - 11 show the MAE for decade-long subperiods. RWD is the random walk plus drift model. NP1 and NP2 are the nonparemetric models otulined in equations 15 and 16, respectively. Radnet is the radial basis neural network.

Table VII: Forecast correlations.

	RWD	Linear	NARX	NP1	NP2	Radnet
RWD	1.000	0.682	0.631	0.117	0.106	0.036
Linear	0.682	1.000	0.577	0.436	0.465	0.105
NARX	0.631	0.577	1.000	0.240	0.245	0.078
NP1	0.117	0.436	0.240	1.000	0.960	0.062
NP2	0.106	0.465	0.245	0.960	1.000	0.061
Radnet	0.036	0.105	0.078	0.062	0.061	1.000

Notes: RWD is the random walk plus drift model. NP1 and NP2 are the nonparemetric models otulined in equations 15 and 16, respectively. Radnet is the radial basis neural network.

Table VIII: BMA Forecast Results

	RWD, Linear, NARX				RWD, Linear		
Periods	RWD	Linear	NARX	R^2	RWD	Linear	R^2
10	0.389	0.334	0.277	0.017	0.531	0.469	-0.103
	(0.327)	(0.316)	(0.293)		(0.299)	(0.299)	
15	0.382	0.302	0.316	0.479	0.536	0.464	0.502
	(0.279)	(0.258)	(0.264)		(0.254)	(0.254)	
20	0.384	0.297	0.319	0.520	0.545	0.455	0.853
	(0.250)	(0.228)	(0.240)		(0.233)	(0.233)	
25	0.379	0.296	0.325	0.377	0.552	0.448	0.784
	(0.230)	(0.202)	(0.223)		(0.205)	(0.205)	
30	0.372	0.291	0.337	0.337	0.544	0.456	0.680
	(0.214)	(0.183)	(0.223)		(0.185)	(0.185)	
35	0.365	0.286	0.349	0.413	0.545	0.455	0.520
	(0.206)	(0.168)	(0.224)		(0.173)	(0.173)	
40	0.366	0.276	0.358	0.269	0.547	0.453	0.551
	(0.201)	(0.155)	(0.226)		(0.154)	(0.154)	
45	0.372	0.272	0.356	0.097	0.543	0.457	0.323
	(0.204)	(0.148)	(0.223)		(0.147)	(0.147)	
50	0.373	0.272	0.355	0.134	0.537	0.463	0.214
	(0.196)	(0.133)	(0.214)		(0.136)	(0.136)	
55	0.374	0.277	0.349	0.115	0.538	0.462	0.227
	(0.183)	(0.119)	(0.198)		(0.129)	(0.129)	
60	0.375	0.286	0.339	0.002	0.536	0.464	0.239
	(0.166)	(0.110)	(0.178)		(0.115)	(0.115)	

Notes: Column 1 holds the length of the training window. The left panel shows the forecast average over the random walk plus drift, the linear and the NARX models. The right panel shows the forecast average over the random walk plus drift and linear models. The standard deviations of the weights are in parentheses.

Table IX: BMA forecast results for the Linear model

Periods	(1,0)	(0,1)	(1,1)	(1,2)	(2,1)	(2,2)	(2,0)	(0,2)	R^2
10	0.132	0.233	0.076	0.088	0.084	0.100	0.155	0.132	-0.733
	(0.224)	(0.319)	(0.124)	(0.183)	(0.141)	(0.190)	(0.247)	(0.228)	
15	0.120	0.246	0.078	0.081	0.083	0.091	0.168	0.132	0.096
	(0.184)	(0.301)	(0.107)	(0.141)	(0.105)	(0.152)	(0.237)	(0.185)	
20	0.125	0.253	0.086	0.073	0.081	0.090	0.157	0.134	0.266
	(0.168)	(0.277)	(0.109)	(0.095)	(0.087)	(0.141)	(0.197)	(0.173)	
25	0.128	0.277	0.085	0.072	0.078	0.088	0.138	0.134	0.116
	(0.152)	(0.276)	(0.082)	(0.088)	(0.071)	(0.127)	(0.153)	(0.157)	
30	0.130	0.269	0.092	0.067	0.084	0.087	0.141	0.131	0.370
	(0.153)	(0.253)	(0.073)	(0.061)	(0.079)	(0.105)	(0.151)	(0.140)	
35	0.127	0.275	0.090	0.067	0.085	0.085	0.134	0.137	0.279
	(0.142)	(0.248)	(0.062)	(0.054)	(0.081)	(0.098)	(0.144)	(0.142)	
40	0.128	0.266	0.095	0.070	0.082	0.087	0.133	0.139	0.227
	(0.128)	(0.229)	(0.064)	(0.051)	(0.056)	(0.092)	(0.122)	(0.136)	
45	0.123	0.275	0.101	0.068	0.082	0.085	0.129	0.137	0.134
	(0.114)	(0.233)	(0.077)	(0.048)	(0.050)	(0.082)	(0.112)	(0.130)	
50	0.121	0.275	0.107	0.066	0.083	0.082	0.126	0.141	-0.260
	(0.105)	(0.230)	(0.084)	(0.045)	(0.048)	(0.074)	(0.106)	(0.132)	
55	0.120	0.262	0.114	0.069	0.086	0.084	0.127	0.139	-0.120
	(0.101)	(0.223)	(0.088)	(0.045)	(0.046)	(0.072)	(0.105)	(0.126)	
60	0.116	0.256	0.119	0.070	0.089	0.083	0.129	0.138	0.093
	(0.087)	(0.217)	(0.090)	(0.045)	(0.043)	(0.068)	(0.104)	(0.125)	

Notes: The BMA results for all considered specifications of the linear model in equation 14. We conduct the estimation using the normal distribution. The first column shows the number of training periods, columns two through eight contain the weights for the different specifications. The autoregressive and sentiment lags are the first and second number in parentheses, respectively. Columns nine holds the OOS R^2. The OOS R^2 statistic is expressed as a percentage. Standard deviations are in parentheses.

Table X: BMA forecast results for the NARX model.

Periods	(1,0)	(0,1)	(1,1)	(1,2)	(2,1)	(2,2)	(2,0)	(0,2)	R^2
10	0.109	0.141	0.108	0.095	0.124	0.137	0.088	0.198	-2.169
	(0.222)	(0.256)	(0.211)	(0.197)	(0.218)	(0.230)	(0.165)	(0.302)	
15	0.101	0.135	0.112	0.101	0.120	0.162	0.085	0.185	-0.047
	(0.208)	(0.226)	(0.182)	(0.174)	(0.183)	(0.234)	(0.149)	(0.269)	
20	0.095	0.127	0.117	0.126	0.111	0.183	0.097	0.144	0.306
	(0.178)	(0.196)	(0.164)	(0.191)	(0.155)	(0.251)	(0.171)	(0.218)	
25	0.080	0.113	0.111	0.129	0.124	0.202	0.107	0.134	0.296
	(0.146)	(0.173)	(0.142)	(0.195)	(0.167)	(0.238)	(0.178)	(0.188)	
30	0.076	0.106	0.105	0.143	0.128	0.215	0.103	0.123	-0.046
	(0.139)	(0.152)	(0.115)	(0.206)	(0.147)	(0.230)	(0.166)	(0.162)	
35	0.071	0.115	0.105	0.122	0.154	0.214	0.097	0.122	0.062
	(0.118)	(0.153)	(0.104)	(0.165)	(0.176)	(0.229)	(0.141)	(0.154)	
40	0.070	0.116	0.105	0.109	0.157	0.213	0.095	0.136	0.302
	(0.092)	(0.161)	(0.095)	(0.132)	(0.168)	(0.229)	(0.128)	(0.163)	
45	0.079	0.119	0.105	0.097	0.158	0.216	0.091	0.134	0.246
	(0.096)	(0.154)	(0.085)	(0.111)	(0.155)	(0.229)	(0.113)	(0.150)	
50	0.083	0.113	0.106	0.095	0.154	0.223	0.088	0.139	0.265
	(0.089)	(0.148)	(0.076)	(0.110)	(0.154)	(0.222)	(0.105)	(0.162)	
55	0.085	0.112	0.110	0.096	0.150	0.232	0.088	0.127	0.187
	(0.087)	(0.142)	(0.077)	(0.110)	(0.144)	(0.215)	(0.092)	(0.144)	
60	0.079	0.112	0.107	0.094	0.150	0.244	0.086	0.128	0.400
	(0.085)	(0.134)	(0.070)	(0.103)	(0.143)	(0.225)	(0.084)	(0.153)	

Notes: The BMA results for all considered specifications of the NARX model in equation 17. We conduct the estimation using the normal distribution. The first column shows the number of training periods, columns two through eight contain the weights for the different specifications. The autoregressive and sentiment lags are the first and second number in parentheses, respectively. Column nine holds the OOS R^2 statistic. The OOS R^2 statistic is expressed as a percentage. Standard deviations are in parentheses.

Table XI: 2SBMA Forecast Results

	RWD, Linear, NARX				RWD, Linear		
Periods	RWD	Linear	NARX	R^2	RW+Drift	Linear	R^2
10	0.420	0.296	0.284	-0.767	0.539	0.461	2.319
	(0.352)	(0.314)	(0.310)		(0.303)	(0.303)	
15	0.425	0.305	0.269	0.064	0.514	0.486	1.754
	(0.323)	(0.259)	(0.261)		(0.268)	(0.268)	
20	0.439	0.300	0.261	0.283	0.525	0.475	1.596
	(0.292)	(0.236)	(0.234)		(0.237)	(0.237)	
25	0.431	0.310	0.259	0.131	0.522	0.478	1.659
	(0.248)	(0.211)	(0.194)		(0.206)	(0.206)	
30	0.412	0.326	0.262	0.400	0.505	0.495	1.672
	(0.213)	(0.190)	(0.165)		(0.184)	(0.184)	
35	0.397	0.326	0.277	0.309	0.496	0.504	1.778
	(0.188)	(0.171)	(0.166)		(0.165)	(0.165)	
40	0.397	0.320	0.283	0.257	0.482	0.518	1.822
	(0.167)	(0.156)	(0.162)		(0.145)	(0.145)	
45	0.388	0.321	0.291	0.166	0.476	0.524	1.857
	(0.152)	(0.149)	(0.160)		(0.140)	(0.140)	
50	0.380	0.325	0.296	-0.225	0.474	0.526	1.815
	(0.139)	(0.133)	(0.146)		(0.134)	(0.134)	
55	0.377	0.328	0.295	-0.088	0.472	0.528	1.765
	(0.128)	(0.122)	(0.126)		(0.130)	(0.130)	
60	0.376	0.327	0.297	0.101	0.471	0.529	1.835
	(0.124)	(0.118)	(0.119)		(0.130)	(0.130)	

Notes: Column 1 holds the length of the training window. The left panel shows the forecast average over the random walk plus drift, the linear and the NARX models. The right panel shows the forecast average over the random walk plus drift and linear models. The standard deviations of the weights are in parentheses.

Table XII: Peaks and troughs in Sentiment.

Peak	Trough
195412	195707
195808	196004
196512	196608
196707	197003
197201	197212
197504	197608
197804	197902
198001	198110
198302	198406
198703	198711
199107	199207
199601	199701
199911	200011
200306	200408
200511	200806

Notes: The algorithm used to determine the turning points is described in sections V and appendix A.

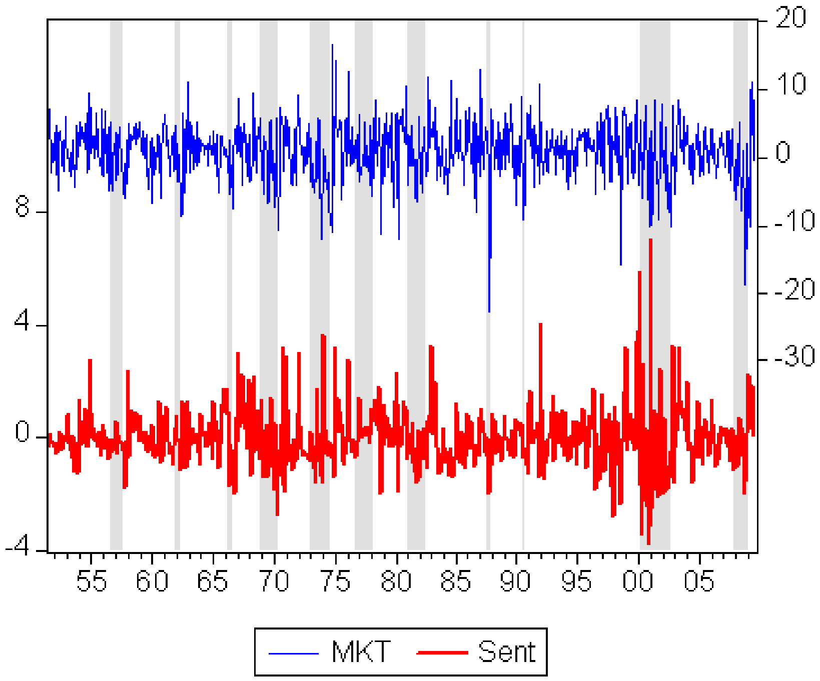

Figure 1: The Sentiment Index vs Market Weighted returns

Notes: Shaded areas are bear markets defined by a 20 percent or drop over a two or month period.

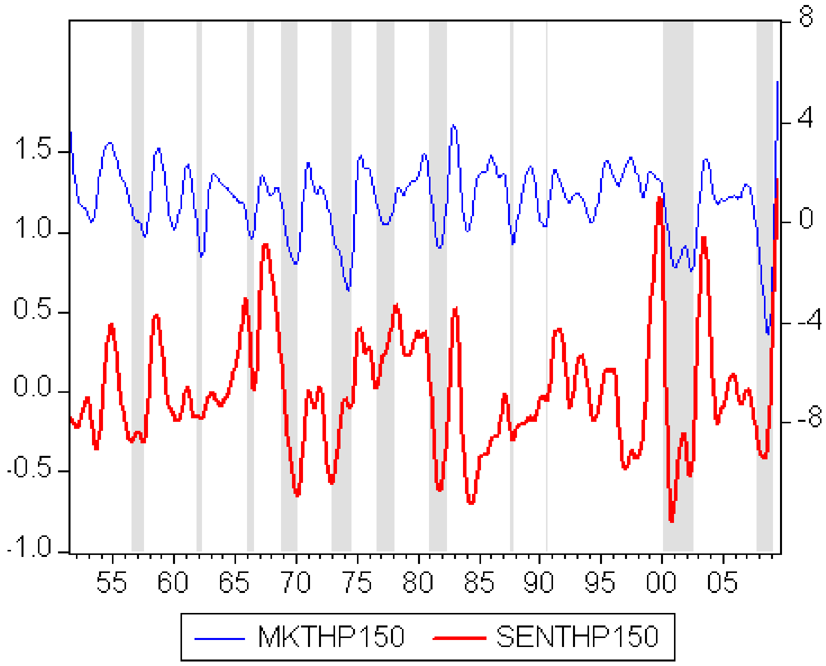

Figure 2: Filtered Sentiment vs. Market Returns

Notes: The sentiment index and market weighted returns filtered using the HP filter ($\lambda = 150$). The shaded areas represent bear markets defined as a 20 percent or more drop over a two or more month period.

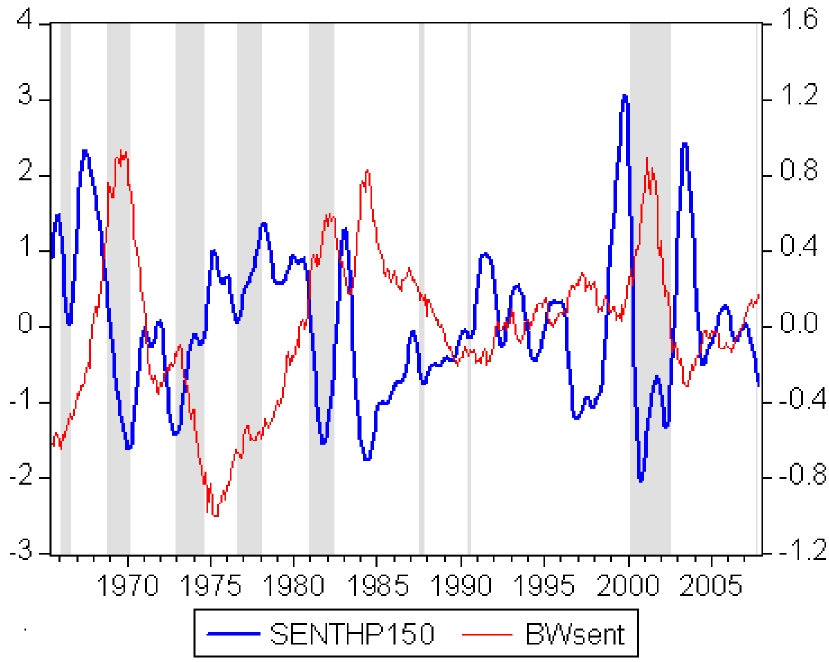

Figure 3: Filtered Sentiment Indexes

Notes: The sentiment index filtered using the HP filter ($\lambda = 150$) (SENTHP150) versus the sentiment index of Baker and Wurgler (2006, 2007) (BWsent). Shaded areas represents the bear markets defined as a 20 percent or more drop over a two or more month period.

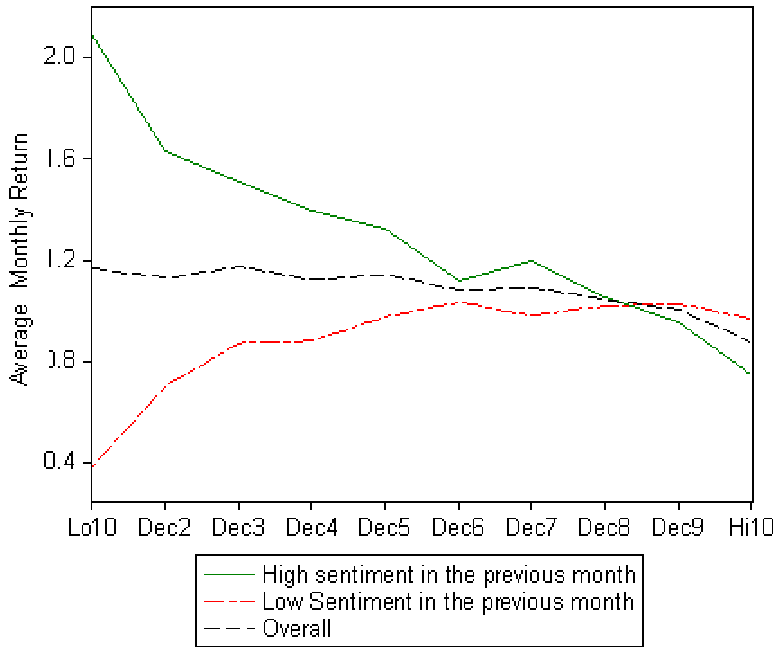

Figure 4: Sentiment, Risk and Returns

Notes: Returns for decile portfolios based on size. The black dashed line represents overall average returns for the portfolio, the green solid line represents the portfolio's returns when sentiment was high in the previous month, and the red dash-dot line represents low sentiment in the previous month.

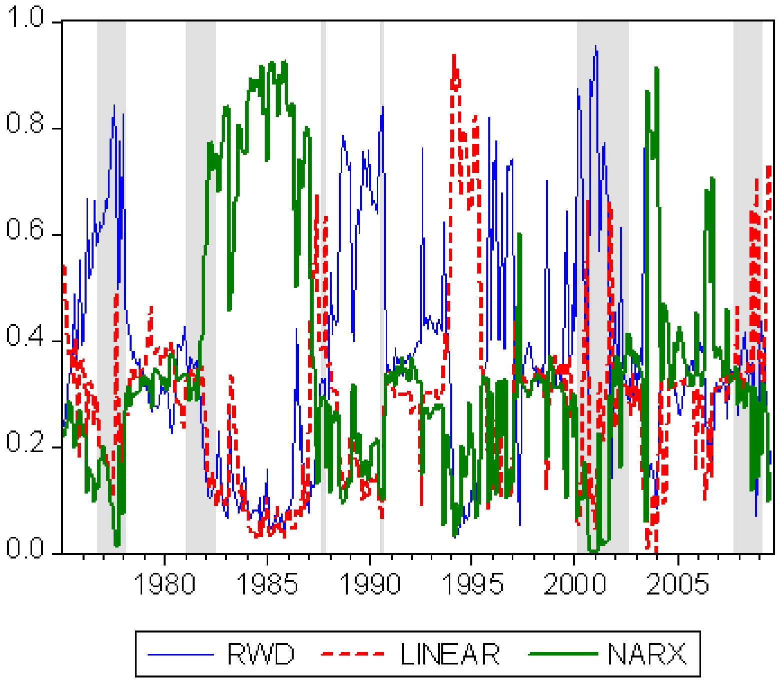

Figure 5: Posterior BMA Weights

Notes: The posterior weights for the BMA process over the random walk plus drift, linear and NARX models. We used 35 training periods and the normal distribution. Shaded areas are bear markets.

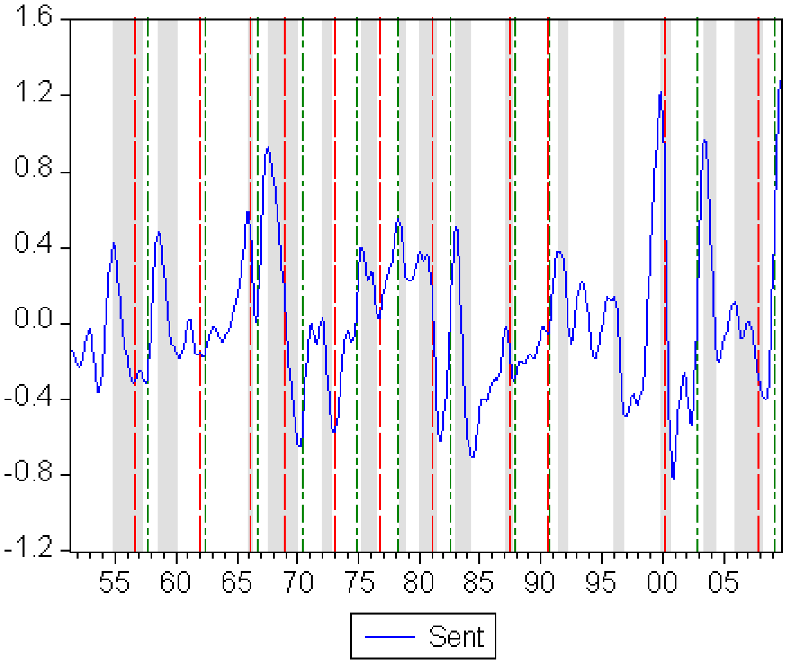

Figure 6: Turning Points in Sentiment

Notes: The filtered sentiment index with $\lambda = 150$. Shaded areas represent peak to troughs in sentiment. Red dashed lines represent the beginning of bear markets. Green dash-dots lines represent the end of bear markets.

References

Abreu, Dilip, and Marcus K. Brunnermeier. 2003. "Bubbles and crashes." *Econometrica*, 71(1): 173–204.

Baker, Malcolm, and Jeffrey Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns." *Journal of Finance*, 61(4): 1645–1680.

Baker, Malcolm, and Jeffrey Wurgler. 2007. "Investor sentiment in the stock market." *Journal of Economic Perspectives*, 21(2): 129–151.

Baker, Malcolm, Jeffrey Wurgler, and Yu Yuan. 2009. "Global, local, and contagious investor sentiment." *Working Papers.*

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A model of investor sentiment." *Journal of Financial Economics*, 49(3): 307–343.

Barberis, Nicholas, Ming Huang, and Tano Santos. 2001. "Prospect Theory and Asset Prices." *Quarterly Journal of Economics*, 116(1): 1–53.

Battalio, Robert, and Paul Schultz. 2006. "Options and the bubble." *Journal Of Finance*, 61(5): 2071.

Brown, Gregory W., and Michael T. Cliff. 2004. "Investor sentiment and the near-term stock market." *Journal of Empirical Finance*, 11(1): 1–27.

Brown, Gregory W., and Michael T. Cliff. 2005. "Investor Sentiment and Asset Valuation*." *The Journal of Business*, 78(2): 405–440.

Bry, Gerhard, and Charlotte Boschan. 1971. "Cyclical analysis of time series: selected procedures and computer programs." *NBER Books.*

Campbell, John Y., and Samuel B. Thompson. 2007. "Predicting excess stock returns out of sample: Can anything beat the historical average?" *Review of Financial Studies.*

Carter, CK, and R. Kohn. 1994. "On Gibbs sampling for state space models." *Biometrika*, 81(3): 541–553.

Chung, San-Lin, and Chung-Ying Yeh. 2009. "Investor Sentiment, Regimes and Stock Returns."

D'avolio, Gene. 2002. "The market for borrowing stock." *Journal of Financial Economics*, 66(2-3): 271–306.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990a. "Noise trader risk in financial markets." *The Journal of Political Economy*, 98(4): 703–738.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990b. "Positive feedback investment strategies and destabilizing rational speculation." *The Journal of Finance*, 45(2): 379–395.

Fama, Eugene F., and Kenneth R. French. 1993. “Common risk factors in the returns on stocks and bonds* 1.” *Journal of financial economics*, 33(1): 3–56.

Foresee, F. Dan, and Martin T. Hagan. 1997. “Gauss-Newton approximation to Bayesian learning.” Vol. 3, 1930–1935, Piscataway: IEEE.

Gervais, Simon, and Terrance Odean. 2001. “Learning to be overconfident.” *Review of Financial Studies*, 14(1): 1.

Glushkov, Denys. 2006. “Sentiment beta.” *Working Paper.*

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

Grinblatt, Mark, and Bing Han. 2005. “Prospect theory, mental accounting, and momentum.” *Journal of Financial Economics*, 78(2): 311–339.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky. 1999. “Bayesian model averaging: A tutorial.” *Statistical science*, 14(4): 382–401.

Kim, Chang-Jin, and Charles R. Nelson. 1998. “Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching.” *Review of Economics and Statistics*, 80(2): 188–201.

Kim, Chang-Jin, and Charles R. Nelson. 1999. *State-space models with regime switching.* MIT press Cambridge, MA.

Lee, Charles M.C., Andrei Shleifer, and Richard H. Thaler. 1991. “Investor sentiment and the closed-end fund puzzle.” *The Journal of Finance*, 46(1): 75–109.

Lemmon, Michael, and Evegenia Portniaguina. 2006. “Consumer confidence and asset prices: Some empirical evidence.” *Review of Financial Studies*, 19(4): 1499.

Lutz, Chandler J. 2010. “Stock Market Forecasting Using Agent Sentiment.” *Working Papers.*

MacKay, David J.C. 1992. “A practical Bayesian framework for backpropagation networks.” *Neural computation*, 4(3): 448–472.

Pesaran, M. Hashem, and Allan Timmermann. 1995. “Predictability of stock returns: Robustness and economic significance.” *The Journal of Finance*, 50(4): 1201–1228.

Qiu, Lily X., and Ivo Welch. 2004. “Investor sentiment measures.” *Working Paper.*

Raftery, Adiran E. 1995. “Bayesian model selection in social research.” *Sociological methodology*, 25: 111–163.

Raftery, Adrian E., Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski. 2005. “Using Bayesian model averaging to calibrate forecast ensembles.” *Monthly Weather Review*, 133(5): 1155–1174.

Shiller, Robert J. 2006. *Irrational exuberance.* Broadway Business.

Stambaugh, Robert F. 1999. "Predictive Regressions." National Bureau of Economic Research, Inc NBER Technical Working Papers 0240.

Timmermann, Allan. 2008. "Elusive return predictability." *International Journal of Forecasting*, 24(1): 1–18.

Wasserman, Phillip D. 1993. *Advanced methods in neural computing.* John Wiley & Sons, Inc. New York, NY, USA.

Whaley, Robert E. 2000. "The investor fear gauge." *The Journal of Portfolio Management*, 26(3): 12–17.