Title
Behavioral Effects in Financial Markets

Author
Lutz, Chandler

Publication Date
2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Behavioral Effects in Financial Markets

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Chandler Joseph Lutz

June 2011

Dissertation Committee:

Dr. Marcelle Chauvet, Chairperson
Dr. Gloria Gonzalez-Rivera
Dr. Jana Grittersova
Dr. Jang-Ting Guo
Dr. Aman Ullah

The Dissertation of Chandler Joseph Lutz is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am grateful to my advisor, without whose help, I would not have been here.

To my parents and family for all the support.

ABSTRACT OF THE DISSERTATION

Behavioral Effects in Financial Markets

by

Chandler Joseph Lutz

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, June 2011
Dr. Marcelle Chauvet, Chairperson

We develop sentiment indexes to study the relationship between sentiment and stock returns. For the first index, we use only market return data; for the second, we use traditional indicators including the closed-end fund discount, NYSE share turnover, and the equity-share of new issues. In sample we find that rising sentiment leads to rising returns and lower risk. We also show that sentiment cycles lead bear markets. In a forecasting exercise we develop two-stage model averaging (2SMA). 2SMA is a flexible framework that allows researchers to incorporate prior economic information into their forecasts. Through 2SMA we develop two-stage Bayesian model averaging (2SBMA) and two-stage equal-weighted averaging (2SEWA). With these techniques we forecast stock returns using investor sentiment. We find that the 2SBMA and 2SEWA forecasts beat their traditional model-averaging counterparts and beat the benchmark random-walk plus drift in a statistically significant manner.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Rational arbitraguers have a profit incentive to ride waves of investor sentiment: Bank stocks rode the liquidity bubble through 2007; hedge funds were overweight tech stocks in the late 1990s; and Sir Issac Newton rode the South Sea Bubble in the 1720s.[1]. Yet classic finance theory fails to explain these historical episodes. In classical financial theory, rational arbitrageurs quickly exploit any mispricing. The failures of the classical financial theory have led to the rise of behavioral finance. In behavioral finance, rational arbitrageurs interact with irrational agents. These irrational agents suffer from one or more behavioral biases that affect their decision making process.

For example, Abreu and Brunnermeier develop a bubble where irrational traders drive asset prices away from equilibrium. The rational arbitrageurs then receive a signal sequentially about the nature of the bubble. This creates a coordination problem. In the model's equilibrium, the rational arbitraguers ride the bubble and try to exit the market just prior to the crash. Hence, rising sentiment leads to rising returns. This theoretical model matches the anecdotal accounts mentioned above. But recent empirical research in behavioral finance, such as Baker and Wurgler (2006, 2007) and Frazzini and Lamont (2008), finds that high sentiment relates to low future returns. Hence, the empirical

[1] See Brunnmermeier (2009), Brunnermeier and Nagel (2004), and Temin and Voth (2004)

research contradicts the theoretical models and the real-world examples.

In this dissertation we develop two new sentiment indexes for the stock market. In chapter 2, we develop a real time stock market sentiment index using only return data. We develop another sentiment index using traditional indicators in chapter 4. With both of these indexes we study the relationship between sentiment and stock returns. We find that rising sentiment leads to rising returns, and vice versa.

In chapter 2 we find that rising sentiment leads lower risk and that cycles in sentiment growth often lead bull and bear markets. Sentiment has its largest on stocks just after bear markets.

We study the time-varying relationship between sentiment and stock returns in chaprter 4. We find that the effects of sentiment have been growing over time. We also find that sentiment sentiment has its largest effect on stocks just after bear markets.

In chapter 3 we develop two-stage model averaging (2SMA). 2SMA is a flexible framework that allows researchers to incorporate prior economic information information into their forecasts. Furthermore, 2SMA handles the 'crowding out effect.' The crowding out effect occurs when a given forecast that performs well over time and doesn't receive a large enough weight in the final forecast.

We combine 2SMA and dynamic Bayesian model averaging to form two-stage Bayesian model averaging (2SBMA). Similarly, we combine 2SMA and equal weighted averaging to develop two-stage equal-weighted averaging (2SEWA). We use 2SBMA and 2SEWA to forecast returns one-step ahead using investor sentiment. We find that 2SBMA and 2SEWA both outperform the benchmark random-walk plus drift and their traditional model averaging counterparts.

Overall, our results correspond with the theoretical predictions of Abreu and Brunnermeier (2003) and anecdotal accounts of investor sentiment. We find that rising sentiment leads to rising returns and lower risk; and vice versa. Whence, rational arbitrageurs have a profit incentive to ride waves of investor sentiment.

Our 2SMA framework is the first forecasting procedure, to our knowledge, that allows researchers to incorporate prior economic in a non-parametric way. Researchers can apply the 2SMA framework to any problem where a given forecast is known to perform well over time.

This dissertation extends the empirical literature in behavioral finance. Unlike previous empirical work, our results match the theoretical and anecdotal literature.

The forecasting work in this dissertation is related to model averaging by clustering. Aiolfi and Timmermann (2006) group forecasts into based on prior forecasting performance, pool within groups, and then shrink the optimal combination weights across clusters to equal weights.

Chapter 2

Real Time Investor Sentiment and the Cross-Correlation of Stock Returns

Abstract: Using only market return data we develop a new index for agent sentiment in the stock market via a dynamic factor model with Bayesian estimation. This index accurately times sentiment episodes. We find that rising sentiment leads to rising returns and lower risk, and vice versa. While sentiment affects most stocks, it has its largest effect on small, young, and volatile stocks, stocks without earnings or dividends, stocks in distress, and high momentum stocks. We also quantitatively chronicle cycles in sentiment growth with a modified Bry-Boschan algorithm. These cycles lead bear and bull markets.

Throughout history sophisticated investors have tried to ride waves of investor sentiment: hedge funds were overweight tech stocks in the late 1990s, banks rode the liquidity bubble through 2007, and Sir Isaac Newton rode the South Sea Bubble in the 1720s.[1] These historical episodes match the theoretical predictions of Abreu and Brunnermeier (2003). Abreu and Brunnermeier find that rational arbitrageurs have a profit incentive to ride waves of investor sentiment and that rising sentiment leads to rising returns. Yet recent empirical research, such as Baker and Wurgler (2006, 2007) and Frazzini and Lamont (2008), contends that high sentiment relates to low future stock returns. Therefore, these empirical papers suggest that rational arbitrageurs should bet against waves of investor sentiment rather than ride them.

In this paper, we develop a novel stock market sentiment index (SENT hereafter) using only market return data. With this index we study the relationship between investor sentiment and stock returns. We report the following findings: (1) SENT accurately times sentiment episodes; (2) rising sentiment leads to above average returns and lower risk, and vice versa; (3) sentiment has its largest effect on small firms, young firms, volatile stocks, firms without earnings or dividends, high momentum stocks, growth stocks, and firms in distress; (4) SENT is robust to the inclusion of macroeconomic indicators and measures of time-varying risk as additional controls; and (5) cycles in sentiment growth often lead bear and bull markets.

Our empirical result—that rising sentiment leads to rising returns and lower risk—corresponds to the theoretical predictions of Abreu and Brunnermeier (2003). Abreu and Brunnermeier build a model where agents face a coordination problem in

[1] See Brunnermeier and Nagal (2004), Brunnermeier (2009) and Abreu and Brunnermeier (2003) for further examples where rational arbitrageurs ride waves of investor sentiment. Temin and Voth (2004) describe the South Sea Bubble. They find that rational arbitrageurs had a profit incentive to ride the bubble. Temin and Voth conclude that the evolution of the South Sea Bubble matches the theoretical model of Abreu and Brunnermeier (2003).

attacking a bubble. In the model's equilibrium, rational arbitrageurs have a profit incentive to ride the bubble. Hence, rising sentiment leads to a higher demand for sentiment-driven stocks. The increase in demand then translates into higher prices.[2]

Our empirical results also match the anecdotal accounts in Temin and Voth (2004), Brunnermeier and Nagel (2004), and Brunnermeier (2009). These papers consider various asset bubbles throughout history. They find that rational arbitrageurs have a profit incentive to ride waves of investor sentiment.

Using only market return data to develop SENT allows researchers, practitioners, and policymakers to compile our index in real time and make decisions based on current information. Market returns are also free from the surveying and sampling issues that plague other macroeconomic and financial indicators. Yet returns are noisy, differenced series. To deal with the noise in returns and construct an index in levels we employ a dynamic factor model with Bayesian estimation. The dynamic factor model (and Kalman filtering it relies on) smooths the large volatility in return data. Furthermore, the dynamic factor model allows us to use the steady-state Kalman Gain, a byproduct from the Kalman filter, to construct an index in levels. Thus, we can compare SENT directly to other stock sentiment aggregates. The most ubiquitous of these aggregates is the stock sentiment index of Baker and Wurgler (2006, 2007) (BWsent hereafter). SENT and BWsent follow the same general pattern, but SENT more precisely times sentiment bubbles and crashes.

With SENT we study the relationship between sentiment and future returns.

[2] Baker and Wurgler (2006) also provide a theoretical foundation that explains the dynamics of the sentiment-driven stocks. In particular, Baker and Wurgler contend that the sentiment-driven stocks are hard-to-value and difficult-to-arbitrage. The sentiment-driven stocks have relatively less information available since they are younger, smaller, without earnings or dividends, etc. This allows investors to plausibly defend a wider range of valuations. D'Avalio (2002) finds that the sentiment-driven stocks are costlier and riskier to arbitrage. Through the hard-to-value and difficult-to-arbitrage channels, sentiment has a larger affect on the set of sentiment-driven stocks.

First, we sort returns by firm characteristics and by the previous month's sentiment growth. We find that rising sentiment leads to above average future returns and lower risk, and vice versa. This result holds for most stocks. Yet sentiment has its largest effect on small, young, and volatile firms; firms without earnings or dividends; growth firms; high momentum stocks; and stocks in distress. We label this set of stocks, the stocks most affected by sentiment, as "the set of sentiment-driven stocks."

Next, we run in-sample regressions. We conclude that rising sentiment relates to rising future returns for the set of sentiment-driven stocks and vice versa. These regression findings match our results based on sorts. Our regression results are robust to the inclusion of macroeconomic indicators, measures of time-varying risk, the Fama-French factors, and the momentum factor as additional controls.

Lastly, we quantitatively date cycles in sentiment growth using a modified Bry-Boschan algorithm. The Bry-Boschan (1971) algorithm is a set of conditional rules to find the highs and lows in a series. Cycles in sentiment growth often lead bear and bull markets. More specifically, sentiment growth peaks prior to all bear markets over the sample period except for the light bear markets that began in 1961 and 1990. Hence, SENT may be useful for investment or policy decisions.

2.1 Related Literature

Baker and Wurgler (2006, 2007) develop a sentiment index from various monthly components (BWsent hereafter). See appendix 4.8 for the description and analysis of Baker and Wurgler's (2006, 2007) data and index construction.

We plot BWsent at the monthly frequency in figure 4.1. Shaded areas represent bear markets defined as a 20 percent or more drop in the S&P 500 over a two or more month period. While BWsent follows general patterns of sentiment, it mis-times the conclusion of certain sentiment episodes. We expect sentiment to rise as bubbles develop and then crash as bubbles burst. Most notably, BWsent peaks nearly a year after the tech bubble crash in 2000. This is surprising since many authors, such as Shiller (2006), attribute the tech boom and bust to agent sentiment. BWsent also peaks well into the bear markets in the late 1960s and the early 1980s.

[Insert figure 4.1 here]

Given figure 4.1, it is not surprising that Baker and Wurgler (2006, 2007) find that high sentiment relates to low future returns. When sentiment is highest (as measured by BWsent), stocks are suffering through bear markets. This result contradicts the theoretical findings of Abreu and Brunnermeier (2003). As noted above, Abreu and Brunnermeier conclude that rational arbitrageurs have a profit incentive to ride waves of investor sentiment.

BWsent is also difficult to calculate in real time. One component of BWsent is the equity-share of new issues (S).[3] This variable is only available from the Federal Reserve at lagged values.[4] Hence, practitioners and policymakers cannot use BWsent to make real time decisions.

Frazzini and Lamont (2008) use mutual fund flows to study the relationship between sentiment and stocks. They also find that high sentiment relates to lower future returns. Other papers use sentiment surveys to study the predictability of asset returns. For example, Lemmon and Portniaguinia (2006) and Qui and Welch (2004) consider consumer confidence. The surveys that these studies employ are only available at lagged values and subject to measurement error. The market returns we use to develop SENT are free from measurement error and are available in real time.

[3] 4.8 for a description of the equity-share of new issues.

[4] See the Federal Reserves' website: `http://www.federalreserve.gov/econresdata/releases/corpsecure/current.htm`

2.2 The Data

Baker and Wurgler (2006, 2007), among others, find that agent sentiment affects small firms, firms without earnings or dividends, and firms in distress. These 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.[5] The correlation matrix for these variables is presented in table 2.1. 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 also use the returns from the S&P500 to incorporate the results from 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.[6]

[Insert table 2.1 here]

Since we use only market return data, our index has two main advantages over other sentiment measures and macroeconomic indicators. First, there is no measurement

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

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

error and no need for the data to be revised. This eliminates the uncertainty related to surveying and sampling found in other sentiment aggregates. Second, firms and policymakers could compile the data in real time and extend our model to make decisions based on current information.

We abbreviate the series for the long-short portfolios based on 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]

2.3 Model Estimation and Description

The return data we use is essentially log-differenced prices. Hence, the data is noisy. If we were to apply principal component analysis (PCA) to the return data the result could be interpreted as a differenced sentiment index. This index would be extremely volatile and not be directly comparable to other sentiment measures. To create an index in levels we use a dynamic factor model with Bayesian estimation. Like PCA, dynamic factor models extract a common component from various series. Unlike PCA, dynamic factor models smooth noisy series via the Kalman Filter. The dynamic factor model also allows us to derive a sentiment series in levels. See Stock and Watson (1991) and Kim and Nelson (1998, 1999) for more on dynamic factor models.

To estimate the model we first take each series in deviations from the mean. The dynamic factor model then takes the following form:

$$\Delta y_{it} = \gamma_i \Delta c_t + \varepsilon_{it}, i = 1, \ldots, 5 \tag{2.1}$$

$$\phi(L)\Delta c_t = \omega_t, \omega_t \sim N(0, 1) \tag{2.2}$$

$$\psi_i(L)\varepsilon_{it} = v_{it}, v_{it} \sim N(0, \sigma^2) \tag{2.3}$$

Where $\Delta y_{it} = \Delta Y_{it} - \Delta \bar{Y}_i$, $i = 1, ..., 5$, represents one of the five series: div,

[7]Baker and Wurgler (2006) test the effects of BWsent on div and earn in their paper. They find a statistically significant relationship between sentiment and the div and earn series.

earn, size lowmom and sp500, respectively. $\Delta c_t = \Delta C_t - \delta$ is the component common to all series, ε_{it} is an idiosyncratic component, γ_i is the factor loading, and L is the lag operator. To estimate the model and derive the common component, Δc_t, we cast it into state-space form. Although a unique representation does not exist we elect a form similar to Kim and Nelson (1998). This facilitates computation of the estimation algorithm. Multiplying both sides of equation 4.1 by $\psi_i(L)$ yields

$$\psi_i(L)\Delta y_{it} = \gamma_i \psi_i(L)\Delta c_t + v_{it} \tag{2.4}$$

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 choose two lags for $\phi(L)$ and two lags for $\psi_i(L)$. Equations 4.2 and 4.3 become

$$\Delta c_t = \phi_1 \Delta c_{t-1} + \phi_2 \Delta c_{t-2} + \omega_t \tag{2.5}$$

$$\varepsilon_{it} = \psi_{i1}\varepsilon_{i,t-1} + \psi_{i2}\varepsilon_{i,t-2} + v_{it}, i = 1, \ldots, 5 \tag{2.6}$$

We employ the following state-space form:

Measurement Equation:

$$\begin{bmatrix} \Delta y^*_{1t} \\ \Delta y^*_{2t} \\ \Delta y^*_{3t} \\ \Delta y^*_{4t} \\ \Delta y^*_{5t} \end{bmatrix} = \begin{bmatrix} \gamma_1 & -\gamma_1\psi_{11} & -\gamma_1\psi_{12} \\ \gamma_2 & -\gamma_2\psi_{21} & -\gamma_2\psi_{22} \\ \gamma_3 & -\gamma_3\psi_{31} & -\gamma_3\psi_{32} \\ \gamma_4 & -\gamma_4\psi_{41} & -\gamma_4\psi_{42} \\ \gamma_5 & -\gamma_5\psi_{51} & -\gamma_5\psi_{52} \end{bmatrix} \begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \\ v_{5t} \end{bmatrix} \tag{2.7}$$

In matrix form this 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 4.4.

$$E(\mathbf{v}_t\mathbf{v}_t') = R = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix} \tag{2.8}$$

Transition Equation:

$$\begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \end{bmatrix} + \begin{bmatrix} \omega_t \\ 0 \\ 0 \end{bmatrix} \tag{2.9}$$

Equation 4.15 can be written in matrix form as

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

$$E(\mathbf{e}_t\mathbf{e}_t') = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.10}$$

Estimating the above model will produce a series, Δc_t. We can interpret Δc_t as the de-meaned first difference in sentiment. Yet our goal is to compile a sentiment index in levels. To derive a common component representing the levels index, we need to identify δ. Our approach follows Stock and Watson (1991). Stock and Watson (1991) notice that ΔC_t is a function of past lags of ΔY_t. Stock and Watson then derive an estimate of δ by taking the expected value of ΔC_t. This method employs the steady-state Kalman gain. Once, we have an estimate for δ we can compile C_t. C_t is the common component or sentiment index in levels. We label C_t as SENT.

We estimate the model using the Bayesian multimove Gibbs-sampling approach based on Carter and Kohn (1994) and Kim and Nelson (1998). The Bayesian method takes into account parameter uncertainty by jointly estimating the state vector and the model parameters. In contrast, traditional maximum likelihood treats the model parameters as true values when estimating the state vector.

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).

2.4 Empirical Results

We interpret the common component, C_t as the sentiment index and label it SENT. Table 2.2 shows the estimated parameters from the dynamic factor model. $i = 1, \ldots, 5$ represent div, earn, size, lowmom, and sp500 respectively. As mentioned above, the factor loading for variable i is γ_i. The factor loadings 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. The parameters ϕ_1 and ϕ_2 represent the autoregressive coefficients on the sentiment index. ϕ_1 and ϕ_2 are positive; this suggests that the sentiment index is persistent. Lastly, the coefficients on the lags of the idiosyncratic components (the ψ_i's) are small in magnitude. Hence, idiosyncratic components are not persistent.

[Insert table 2.2 here.]

We plot SENT in figure 2.2. High values of SENT correspond to high levels of agent sentiment or optimism; low values of SENT correspond to low levels of agent sentiment or pessimism. The gray bands around SENT represent its 95 percent confidence intervals. These confidence intervals are the 5th and 95th percentiles from the posterior distribution of the Gibbs-sampler. The confidence intervals are extremely tight. This shows the robustness of our estimation approach given the data. The vertical bars in figure 2.2 represent bear markets defined as a 20 percent or more drop in the S&P500 over a two or more month period. Overall, SENT rises and falls with agent sentiment: SENT levitates as bubbles develop and crashes as bear markets take hold.

[Insert figure 2.2 here.]

We plot SENT versus BWsent in figure 4.3. SENT is the solid line; BWsent is the dotted line. SENT and BWsent follow the same overall pattern, but SENT more accurately times sentiment episodes. For example, SENT peaks just prior to the bear markets in the mid-1950s, late 1960s, mid 1970s, early 1980s and early 2000s. This matches our expectations as sentiment cycles should coincide with stock booms and busts. BWsent, on the other hand, continues to rise through the bear markets in the late 1960s, early 1980s and early 2000s.

[Insert figure 4.3 here.]

SENT also corresponds with the anecdotal accounts of sentiment episodes outlined in Baker and Wurgler (2006). First, Baker and Wurgler (2006) find a bubble during 1961. SENT reaches a local maximum during this period and then falls as the bear market ensues in 1962. Baker and Wurgler (2006) document bubbles in 1977 and 1978 and again in the first half of 1983. Our index spikes around these bubbles. SENT crashes through the bear markets in 1987 and 1990. SENT then rises through the mid-1990s before peaking prior the 1997 Asian financial crisis and the collapse of Long Term Capital Management. SENT rockets up during the internet bubble in the late 1990s and then crashes in 2000. Lastly, SENT plummets through the credit crisis in 2008.

2.4.1 SENT, Macroeconomic Indicators, and Time-Varying Risk

Next we check the robustness of our index to the inclusion macroeconomic indicators and time-varying risk as additional controls. 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 ($INDPRO$), consumer durables ($IPDCONGD$), consumer nondurables ($IPNCONGD$), the one month Treasury rate ($TREAS$), and NBER recessions ($NBER$). These factors allow us to control for broad macroeconomic effects. We also control for the S&P500 monthly volatility ($sp500vol$) and the BAA–AAA corporate spread ($BAA - AAA$). $sp500vol$ and ($BAA - AAA$) allow us to control for time-varying risk in the stock and bond markets.[8]

We run the following regression where y_{it}, $i = 1, \ldots, 5$ is any of the components our index.

$$\begin{aligned} y_{it} = \alpha &+ \beta_1 INDPRO_t + \beta_2 IPDCONGD_t + \beta_3 IPNCONGD_t \\ &+ \beta_4 TREAS_t + \beta_5 NBER_t + \beta_6 sp500vol_t + \beta_7 (BAA - AAA) + \varepsilon_{it} \end{aligned} \tag{2.11}$$

We then re-compile our index using the residuals from the regression, ε_{it}, to develop an "orthogonalized index." If our original index and the orthogonalized index are similar, we contend that our methodology is robust to business cycle indicators and time-varying risk. The correlation between our original index and the macro-orthogonalized index is 0.80. We plot the orthogonalized index and SENT if figure 2.4. SENT and the orthogonalized index share the same pattern and turning points. The only difference is the magnitude during some periods. The orthogonalized index stays below SENT prior

[8]We obtained growth in industrial production, consumer durables, consumer nondurables, and the BAA and AAA series from the FRED economic database. The series IDs are INDPRO, IPDCONGD, IPNCONGD, and BAA and AAA, respectively. The one-month Treasury rate is available from Kenneth French's Data library. We calculate the monthly volatility by taking the standard deviation of S&P500 daily returns for each month.

to 1980 and above SENT after 1980. After the bear market in 2002, SENT and the orthogonalized index are nearly identical. Overall, our sentiment index appears robust to macro indicators and time-varying risk. As shown below, orthogonalizing SENT to the macro indicators and time-varying risk does not affect the relationship between sentiment and stocks returns.

[Insert figure 2.4 here.]

2.5 Empirical Tests

In this section we study the relationship between sentiment and future returns for various types of stocks. We form decile portfolios based on the following firm characteristics: (1) size (ME); (2) age; (3) volatility (Total Risk, σ); (4) the earnings-price ratio (Earn); (5) the dividend-price ratio (Div); (6) the book-to-market ratio (BE/ME); and (7) momentum (Mom).[9] Portfolios based on ME, Earn, Div, BE/ME, and Mom are from Kenneth French's Data Library. We compile the portfolios based on age and volatility from the CRSP database using share codes 10 and 11. We count a stock's age from the first date it is listed in the CRSP. We then form decile breakpoints using the prior month's age. To compile the portfolios based on volatility, we calculate the prior 2-12 month standard deviation for each stock. We then construct decile based on volatility.

2.5.1 A Stationary Sentiment Index

The sentiment index, SENT, developed sections 4.2 and 4.3 is not stationary. Yet to analyze the relationship between sentiment and stock returns we need a stationary series. First, we test for breaks. We calculate the Quandt (1960) statistic and test the null hypothesis of no structural break based on the Andrews (1993) critical values. Hansen

[9] These portfolios are similar to the portfolios studied in Baker and Wurgler (2006).

(1997) develops p-values based on the Andrews critical values.[10] We find a structural break in 2000M12 with a Hansen (1997) p-value of 0.001. Based on this break date, we split SENT into two subsamples: [1951M09, 2000M11] and [2001M01, 2009M09]. There are no structural breaks in these subsamples. Hence, SENT contains only one break in 2000M12. After controlling for the break in 2000M12, we find that SENT is still not stationary. Thus, we need to use a differenced series.

One benefit of the dynamic factor model is that it produces Δc_t as a byproduct. As noted above, we interpret Δc_t as the de-meaned first difference in SENT (ΔSENT* hereafter).[11] ΔSENT* is a stationary series. Hence, we use it to study the relationship between sentiment and stock returns in the remainder of the paper. We plot ΔSENT* in figure 4.4. As expected, ΔSENT* is most volatile around major episodes of investor sentiment.

[Insert figure 4.4 here.]

2.5.2 Returns for Stocks Sorted by Firm Characteristics and Sentiment

Using the decile portfolios described in section 2.5, we sort returns based on the prior month's sentiment growth, ΔSENT*. More specifically, we compare returns when ΔSENT* is high in the previous month to returns when ΔSENT* is low in the previous month. ΔSENT* is high in the previous month if it is above the mean and vice versa. These conditional sorts allow us to non-parametrically study the relationship between sentiment and future returns across different types of stocks.

Table 2.3 and figure 2.6 show the results. In both table 2.3 and figure 2.6

[10] The Quandt statistic is the maximum Chow (1960) statistic over the sample. We interpret the Quandt statistic as the best candidate for a breakpoint. See Hansen (2001) for an overview of these methods.

[11] We use ΔSENT* rather than the first difference in SENT since ΔSENT* comes directly from the model. Recall that the dynamic factor model derives SENT from ΔSENT*. Hence, ΔSENT* is a more parsimonious than the first difference in SENT. Using the first difference in SENT does not significantly affect our regression results below.

High represents times when $\Delta\text{SENT}^*_{t-1}$ is above the mean; Low represents times when $\Delta\text{SENT}^*_{t-1}$ is below the mean; Ave represents the average across both High and Low; and Dif represents the difference between High and Low. In figure 2.6 we plot Dif times -1 so it is easily visible below the horizontal axis. As the effects of sentiment rise, Dif increases in magnitude.

The top rows in table 2.3 and figure 2.6a show the results based on size (ME). When $\Delta\text{SENT}^*_{t-1}$ is high, returns are above average for all but the largest stocks. In orther words, rising sentiment magnifies the size effect of Banz (1981). Surprisingly, the size effect reverses during times of falling sentiment: when $\Delta\text{SENT}^*_{t-1}$ is Low, smaller stocks earn lower returns on average than larger stocks. For the top two deciles comprising the largest stocks, rising sentiment leads to below average returns and vice versa. This negative relationship between sentiment and future returns is unique to the top two deciles of the size-based portfolios.

Next, we examine the results for the portfolios based on age in the second set of rows in table 2.3 and in figure 4.5b. Sentiment affects the entire cross-section, but the effect is marginal for the oldest stocks. When $\Delta\text{SENT}^*_{t-1}$ is High, returns for all of the deciles are above Ave; and vice versa when $\Delta\text{SENT}^*_{t-1}$ is Low. Thus, rising sentiment leads to above average returns and vice versa.

The results in figure 4.5c for portfolios based on volatility are remarkable. Sentiment has almost no effect on the least volatile stocks, the stocks constituting first decile, but a large effect on the most volatile stocks, the stocks constituting the tenth decile. For the most volatile stocks, Dif is large in magnitude; Dif then converges to zero as volatility abates. Hence, the effect of sentiment is larger for riskier stocks. Furthermore, the effect of $\Delta\text{SENT}^*_{t-1}$ is nearly symmetrical around Ave during periods of rising or falling sentiment: High values of $\Delta\text{SENT}^*_{t-1}$ relate to high returns and vice versa. We discuss the relationship between sentiment and risk further in section 2.5.5.

Figures 4.5e and 4.5d show the sorts based on dividends and earnings. The zero on the horizontal axis in figure 2.6 represents the portfolios comprising stocks without earnings or stocks without dividends. Coinciding with our expectations, sentiment has its largest effect on the stocks without dividends or earnings. The far right column of table 2.3 shows the difference in monthly returns between stocks with positive earnings and those without earnings; and the difference between stocks with positive dividends and those without dividends. When $\Delta\text{SENT}^*_{t-1}$ is High, stocks without earnings or dividends perform best; the opposite occurs when $\Delta\text{SENT}^*_{t-1}$ is Low. This result is not surprising; stocks without earnings or dividends have less information available and should be more affected by sentiment.

Figure 2.6f shows the results based on the book-to-market ratio (BE/ME). Sentiment affects all stocks in the cross-section. Yet as demonstrated by the plot of Dif, sentiment has its largest effect on low book-to-market stocks or growth stocks.

The sorts for the portfolios based on momentum are also noteworthy. The last set of rows in table 2.3 and figure 2.6g show the results. The upward slope in Ave is the well-documented momentum effect: high momentum portfolios earn higher future returns and vice versa for low momentum portfolios. Sentiment affects the entire cross-section of momentum stocks, but the effect is not constant. Sentiment has a larger effect on the extreme momentum portfolios. The U-shaped pattern in Dif highlights the varying effects of sentiment on momentum stocks. Sentiment has a larger effect on the highest and lowest momentum stocks. When $\Delta\text{SENT}^*_{t-1}$ is High, high momentum stock earn extremely high returns. For example, the tenth decile comprising the highest momentum stocks produces monthly returns of 1.82% for Ave compared with monthly returns of 2.52% when $\Delta\text{SENT}^*_{t-1}$ is High. This result corresponds with our expectations: during times of high sentiment, like during the 1990s tech bubble, high momentum stocks perform well. The lowest momentum stocks, the losers constituting the bottom decile, also perform well as sentiment rises. When $\Delta\text{SENT}^*_{t-1}$ is High, the lowest momentum stocks earn 1.96% per month. Interestingly, the momentum effect only persists for Low values of $\Delta\text{SENT}^*_{t-1}$; in figure 2.6g Low slopes upwards.

In summary, for all but the safest stocks, rising sentiment leads to above average returns and vice versa. Yet sentiment has its largest effect on small, young and volatile firms; firms without earnings or dividends; growth stocks; high momentum stocks; and stocks in distress. As noted above, we call these stocks "the set of sentiment driven stocks."

2.5.3 Return Volatility for Stocks Sorted by Firm Characteristics and Sentiment

Next, we study the relationship between sentiment and return volatility for different types of stocks. We sort stocks into bins based on firm characteristics and $\Delta\text{SENT}^*_{t-1}$ using the procedure in sections 2.5 and 2.5.2. Then we calculate standard deviation for each bin. Table 2.4 and figure 2.7 show the results. The format of the table 2.4 and figure 2.7 is similar to table 2.3 and figure 2.6 outlined in section 2.5.2.

For all of the decile portfolios in table 2.4 and figure 2.7, rising sentiment leads to below average return volatility. When $\Delta\text{SENT}^*_{t-1}$ is high, the standard deviation of the returns is below average, and vice versa. Thus, risk falls as sentiment rises. This result holds for portfolios based on size (ME), age, volatility (Total Risk, σ), the earnings-price-ratio, the dividend-price ratio, the book-to-market ratio, and momentum.

Rising sentiment leads to rising returns and lower risk, and vice versa. Hence, our results show that investors should ride waves of agent sentiment and then exit the market before sentiment falls. These findings match the theoretical results of Abreu and Brunnermeier (2003). Abreu and Brunnermeier find that rational arbitrageurs have a profit incentive to ride waves of investor sentiment and then try to exit the market right before the crash.

2.5.4 In-Sample Predictive Regressions

In this section, we run in-sample regressions to determine the predictive power of sentiment on future returns in a model-based framework. We use the stationary ΔSENT^* series outlined in section 4.4

To conduct our regressions, we lag ΔSENT^* by one period. We also control for the three factors, MKT, HML and SMB of Fama and French (1993) and a momentum factor. If one of these factors is used as the dependent variable we do not include it in our set of regressors. We construct long-short portfolios based on the various firm characteristics listed in section 2.5. We build the long-short portfolios by comparing the high, medium and low deciles. The top three deciles are high; the middle four deciles are medium; and the bottom three deciles are low. Let the variable Z_t represent any of the long-short portfolios. The regression equation becomes:

$$Z_t = \alpha + \beta_1 \Delta\text{SENT}^*_{t-1} + \beta_2 MKT_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_t \qquad (2.12)$$

Table 4.3 shows the regression results. The left panel shows the results for ΔSENT^*; the right panel shows the results using the orthogonalized index. We calculate the orthogonalized index by first regressing out the macroeconomic and time-varying risk indicators described in section 2.4.1. The left most column holds the dependent variable used for the regression in equation 4.11. In each panel, the left column holds the beta on the sentiment index and the right column holds its bootstrapped p-value in parentheses.

[Insert table 4.3 here]

The results are similar for the regressions on ΔSENT* and orthogonalized ΔSENT*: controlling for the macroeconomic and time-varying risk indicators only marginally changes the coefficients or their p-values. Thus, the relationship between sentiment and returns is robust to the inclusion of the macroeconomic and time-varying risk indicators as additional controls.

Panel A shows the effects of sentiment on size (ME), age, and risk (σ). We compare small and large firms; young and old firms; and high and low volatility stocks. All of the regression coefficients are positive and significant at the one percent level. Hence, sentiment has a larger effect on small, young and volatile firms. The positive regression coefficients also suggest that rising sentiment relates to rising returns, and vice versa.

In Panel B we study the effect of sentiment on earnings and dividend based portfolios. More specifically, we compare firms without earnings to those with positive earnings, and firms without dividends to those with positive dividends. Clearly, sentiment has a larger effect on stocks without earnings or dividends. The coefficients are positive and significant at the one percent level. So, for stocks without earnings or dividends, rising sentiment leads to rising returns.

Panels C and D show portfolios based on the book-to-market ratio. In Panel C, we examine the effects of sentiment on HML. Stocks with a high book-to-market ratio may be undervalued or in distress; stocks with a low book-to-market ratio are often growth stocks. The positive regression coefficient suggests that sentiment has a larger effect on stocks in distress. In Panel D, we focus on growth stocks. More specifically, we study the Medium – Low long-short portfolio based on the book-to-market ratio. Not surprisingly, the coefficient on the sentiment index is negative. Thus, sentiment has a positive effect on growth stocks.

Panel E shows sentiment's effect on the momentum factor, UMD. The top momentum deciles are obviously high momentum stocks; the bottom deciles are stocks in distress. In this case, the coefficient on the sentiment index is positive and significant. Hence, sentiment has a larger effect on high momentum stocks.

In Panel F we focus soley on high momentum stocks using the long-short portfolio High − Medium. In this regression we control for the momentum factor, UMD. Yet the regression coefficient is still positive and statistically significant at the 10 percent level. This implies that sentiment has an effect on high momentum stocks beyond the momentum factor.

Panel G focuses on stocks in distress. In particular, we study the High − Medium portfolio based on the book-to-market ratio, and the Medium − Low portfolio based on momentum. The regression coefficients have the expected sign: the coefficient is negative for the portfolio based on momentum and positive for the portfolio based on the book-to-market ratio. Hence, rising sentiment lifts stocks in distress.

Overall, the coefficients are positive and significant when we go long the set of sentiment driven stocks, and vice versa. The regressions are robust to the inclusion of macroeconomic and time-varying risk indicators, the Fama-French Factors, the momentum factor as additional controls. These results match our findings based on sorts: rising sentiment leads to rising returns for the set of sentiment driven stocks and vice versa.

2.5.5 Sentiment and Risk

In section 2.4.1 we orthogonolized SENT to time-varying risk. In this section we examine the relationship between our sentiment index and risk in the cross-section. Standard portfolio theory (e.g. the CAPM model) contends that riskier portfolios should always have a higher expected return. The volatility is a measure of the portfolio's risk. Hence, a portfolio with a larger volatility 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 Wurgler (2007). We form ten decile portfolios based on volatility. The third set of rows in table 2.3 and figure 4.5c show the average returns for these portfolios. The results coincide with our expectations: the more volatile portfolios are riskier and earn larger average returns. Hence, on average, the CAPM predictions hold. In addition, we examine portfolio returns when ΔSENT^* is high in the previous month and when ΔSENT^* is low in the previous month. As mentioned above, ΔSENT^* is low during the previous month if it is below the mean and vice versa for high levels of ΔSENT^*. When ΔSENT^* is high during the previous month, the portfolios with higher volatility earn much higher returns. The results reverse when ΔSENT^* is low during the previous month. In this case, high beta portfolios earn much lower returns than low beta portfolios. These results violate CAPM predictions. Baker and Wurgler (2007) label the graph in figure 4.5c as "the sentiment seesaw." Not surprisingly, given the above regressions, our empirical seesaw and Baker and Wurgler's are inverses. We both find, however, that standard portfolio theory fails to explain the results: for a given level of sentiment, riskier stocks earn lower future returns. In Baker and Wurgler's words, "This is a powerful confirmation of the sentiment-driven mispricing view."

2.6 Sentiment Cycles

Using the stationary ΔSENT^* index outlined in sections 4.2 and 4.3 we date cycles in sentiment growth. Recall that we interpret ΔSENT^* as the first difference (or growth) in sentiment. Hence, a peak in ΔSENT^* indicates the point where sentiment grew the fastest and vice versa. Our goal is to determine trends in the growth rates of sentiment. This will allow us to study the relationship between sentiment growth and bear markets. We find that peaks in sentiment growth rates often act as a leading indicator of bear markets. Hence, practitioners and policymakers could use sentiment growth rates to make decisions regarding the state of the market.

To our knowledge, no other work has quantitatively chronicled sentiment cycles.[12] To date the cycles we employ a modified Bry-Boschan algorithm. The Bry-Boschan algorithm is a set of conditional rules to find local highs and lows in a series. We modify the algorithm by first adding an extra layer of smoothing. Then we change the window lengths of the cycle durations in the algorithm to fit our 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 growth. Appendix 2.9 shows the exact set of rules we set for the algorithm. Table 2.7 and figure 2.8 show the results. The gray shaded areas in the figure represent peak to trough episodes in ΔSENT^*; the light blue shaded areas are bear markets. The modified algorithm appears to capture the dynamics in the growth rate of sentiment. Furthermore, sentiment growth peaks prior to most bear markets. This is the case for the bear markets that begin in 2008, 2000, 1987, 1981, 1976, 1968, 1966 and 1957. Our algorithm does not record a peak in sentiment growth prior to the bear market in 1961. Also, sentiment growth peaks just after the light bear market in the early 1990s.

In addition, the algorithm records other sentiment cycles. These usually occur as sentiment growth levitates following bear markets. For example, the algorithm records peaks following the bear markets in 2002, 1990, 1982 and 1957. This result is not

[12] Baker and Wurgler (2006) document anecdotal periods of sentiment. Their approach is qualitative. We compare our sentiment index to these accounts in section 4.3.

surprising since the stocks most affected by sentiment produce their greatest returns following bear markets. Clearly, there is a correlation between sentiment growth and bear markets. More study is needed, however, to determine the nature of the relationship and its causality.

[Insert table 2.7 and figure 2.8 here.]

2.7 Conclusion

In this paper we develop SENT, a market returns based measure of investor sentiment. Using market returns frees SENT from the surveying and sampling issues that plague other sentiment indexes. To compile SENT we use a dynamic factor model with Bayesian estimation. This estimation technique smooths the noise found in return series and compiles an index in levels that we compare directly to other sentiment aggregates. Using conditional sorts we find that rising sentiment leads to rising returns and lower risk. In a regression framework we control for macroeconomic and time-varying risk indicators the Fama-French factors. We again find that rising sentiment leads to rising returns, and vice versa. Furthermore, sentiment has its largest effect on the set of sentiment-driven stocks. The sentiment-driven stocks are small, young and volatile stocks; stock without earnings or dividends; growth stocks; momentum stocks; and stocks in distress. Lastly, we quantitatively chronicle cycles in sentiment growth using a modified Bry-Boschan algorithm. We find that cycles in sentiment growth often lead bear and bull markets.

2.8 Appendix: Baker and Wurgler's (2006, 2007) Sentiment Index

Data for Baker and Wurgler (2006, 2007) is available from Jeffrey Wurgler's website. The following list contains the components of Baker and Wurgler's index:

- The closed end fund discount (CEFD): The average difference between the net-asset-value (NAV) of closed-end stock funds and their market prices.
- NYSE share turnover (TURN): the ratio of reported share volume to the average shares listed for the New York Stock Exchange
- The number of IPOs over a given month (NIPO)
- The average first day return of IPOs for each month (RIPO)
- The equity share of new issues (S): The gross equity issued divided by the gross equity plus debt issued
- The dividend premium (P^{D-ND}): the log difference between the average book-to-market ratios of dividend and non-dividend paying stocks.

See Lee, Shleifer and Thaler (1991), Ibbotson, Sindelar and Ritter (1994), Baker and Wurgler (2000), Baker and Stein (2004) and Baker and Wurgler (2004) for more detailed analysis of these measures.

Baker and Wurgler use the first principal component as the factor loading on each variable. The following equation of shows the construction of BWsent at the monthly frequency as in Baker and Wurgler (2007):

$$\begin{aligned} BWsent = & -0.23CEFD + 0.23TURN + 0.24NIPO \\ & + 0.29RIPO + -0.32P^{D-ND} + 0.23S \end{aligned} \tag{2.13}$$

Baker and Wurgler (2006) develop BWsent at an annual frequency and similar results. Baker and Wurgler (2006) and Baker and Wurgler (2007) find a similar relationship between sentiment and stock prices; high sentiment relates to low returns.

2.9 Appendix: Procedure For Determination of Turning Points

1. Filter the series using the HP Filter with the smoothing parameter set to 150.
2. Remove outliers from the data and replace them using the Spencer curve.
3. 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)
4. 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 that 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.
5. Statement of final turning points

2.10 Appendix: Tables

Table 2.1: 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 2.2: 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 4.2 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 4.1. ϕ_i are the coefficients on lags of the common component in equation 4.2. ψ_i are the coefficients on lags of the idiosyncratic component in equation 4.3.

Table 2.3: Returns for Stocks Sorted by Firms Characteristics and $\Delta\text{SENT}^*_{t-1}$

		Decile											Comparisons			
	$\Delta\text{SENT}^*_{t-1}$	≤ 0	1	2	3	4	5	6	7	8	9	10	$10-1$	$10-5$	$5-1$	$>0-\leq 0$
ME	Ave		1.35	1.09	1.13	1.09	1.10	1.06	1.07	1.01	0.99	0.86	-0.49	-0.24	-0.25	
	High		2.51	1.63	1.49	1.40	1.34	1.12	1.19	1.05	0.95	0.74	-1.77	-0.59	-1.18	
	Low		0.36	0.62	0.82	0.82	0.90	1.01	0.97	0.98	1.03	0.96	0.60	0.06	0.54	
	Dif		2.15	1.01	0.67	0.58	0.43	0.12	0.23	0.07	-0.08	-0.22	-2.37	-0.65	-1.72	
Age	Ave		1.14	1.15	1.19	1.24	1.28	1.30	1.31	1.32	1.29	1.18	0.03	-0.11	0.14	
	High		1.80	1.79	1.80	1.90	1.99	1.91	1.85	1.80	1.69	1.31	-0.49	-0.67	0.19	
	Low		0.59	0.61	0.68	0.69	0.68	0.78	0.84	0.92	0.95	1.06	0.47	0.37	0.09	
	Dif		1.21	1.18	1.12	1.21	1.31	1.13	1.01	0.88	0.74	0.26	-0.95	-1.05	0.09	
σ	Ave		1.10	1.21	1.25	1.27	1.29	1.30	1.30	1.29	1.29	1.34	0.23	0.04	0.19	
	High		1.15	1.30	1.47	1.62	1.76	1.89	2.02	2.13	2.28	2.54	1.39	0.78	0.61	
	Low		1.06	1.13	1.06	0.98	0.89	0.80	0.70	0.58	0.45	0.31	-0.74	-0.58	-0.16	
	Dif		0.09	0.17	0.40	0.64	0.87	1.10	1.32	1.55	1.83	2.23	2.13	1.36	0.77	

Notes: For each month we sort returns by firm characteristics and then by $\Delta\text{SENT}^*_{t-1}$. We form decile portfolios based on firm size (ME), firm age (Age), Total Risk (σ), the earnings-price ration (Earn), the dividend-price ratio (div), the book-to-market ration (BE/ME), and momentum (Mom). We then sort the portfolios based on high and low values of $\Delta\text{SENT}^*_{t-1}$. $\Delta\text{SENT}^*_{t-1}$ is high if it is above the mean and vice versa. Ave is the average monthly return over both high and low sentiment periods. High is the average monthly return when ΔSENT^* is high during the previous month. Low is the average monthly return when ΔSENT^* is low during the previous month. Dif is the difference between High and Low. The left panel shows the returns for various deciles; the right panel shows the difference between certain deciles.

Table 2.3 Continued

		Decile											Comparisons			
	$\Delta\mathrm{SENT}^*_{t-1}$	≤ 0	1	2	3	4	5	6	7	8	9	10	$10-1$	$10-5$	$5-1$	$>0- \leq 0$
Earn	Ave	1.38	0.92	1.05	1.08	1.14	1.20	1.26	1.32	1.40	1.50	1.66	0.74	0.47	0.27	-0.13
	High	2.87	1.64	1.49	1.39	1.43	1.45	1.46	1.55	1.63	1.77	2.16	0.52	0.71	-0.19	-1.28
	Low	0.12	0.31	0.67	0.82	0.89	0.98	1.09	1.13	1.21	1.27	1.24	0.93	0.26	0.67	0.84
	Dif	2.75	1.32	0.82	0.57	0.54	0.47	0.37	0.42	0.42	0.49	0.92	-0.40	0.45	-0.85	-2.12
Div	Ave	1.22	1.10	1.22	1.18	1.24	1.25	1.25	1.28	1.30	1.22	1.11	0.00	-0.14	0.14	0.00
	High	2.36	1.43	1.47	1.38	1.43	1.46	1.35	1.31	1.33	1.24	1.16	-0.27	-0.29	0.02	-1.00
	Low	0.24	0.82	1.00	1.01	1.07	1.07	1.17	1.26	1.27	1.20	1.06	0.24	-0.01	0.24	0.85
	Dif	2.12	0.61	0.48	0.38	0.36	0.39	0.18	0.05	0.06	0.04	0.10	-0.51	-0.29	-0.22	-1.85
BE/ME	Ave		-0.27	0.84	0.70	1.14	1.14	1.22	0.87	0.70	1.00	0.59	0.86	-0.55	1.41	
	High		0.30	1.19	1.05	1.37	1.32	1.38	1.02	1.03	1.19	0.63	0.34	-0.68	1.02	
	Low		-0.75	0.55	0.40	0.95	1.00	1.08	0.74	0.43	0.84	0.55	1.29	-0.45	1.74	
	Dif		1.04	0.64	0.66	0.42	0.32	0.30	0.28	0.60	0.35	0.09	-0.95	-0.23	-0.72	
Mom	Ave		0.79	0.95	1.02	1.07	1.20	1.24	1.34	1.41	1.57	1.82	1.03	0.62	0.41	
	High		1.96	1.66	1.59	1.50	1.59	1.59	1.70	1.84	2.08	2.52	0.56	0.93	-0.37	
	Low		-0.20	0.34	0.54	0.71	0.87	0.93	1.02	1.04	1.13	1.22	1.42	0.35	1.07	
	Dif		2.16	1.32	1.05	0.79	0.72	0.66	0.68	0.80	0.96	1.30	-0.86	0.58	-1.44	

Table 2.4: Return Volatility for Stocks Sorted by Firm Characteristics and $\Delta\text{SENT}^*_{t-1}$

		Decile											Comparisons			
Port	$\Delta\text{SENT}^*_{t-1}$	≤ 0	1	2	3	4	5	6	7	8	9	10	$10-1$	$10-5$	$5-1$	$>0-\leq 0$
ME	Ave		6.35	6.12	5.93	5.73	5.59	5.26	5.14	5.01	4.68	4.45	-1.90	-1.14	-0.76	
	High		5.69	5.41	5.17	5.06	4.84	4.51	4.45	4.25	3.93	3.87	-1.83	-0.97	-0.85	
	Low		6.72	6.64	6.50	6.24	6.16	5.82	5.67	5.57	5.23	4.90	-1.82	-1.26	-0.56	
	Dif		-1.03	-1.24	-1.34	-1.19	-1.32	-1.31	-1.22	-1.33	-1.30	-1.04	-0.01	0.28	-0.29	
Age	Ave		5.58	5.60	5.57	5.65	5.78	5.41	5.39	5.27	5.15	4.59	-0.99	-1.19	0.20	
	High		4.91	4.94	4.92	5.00	5.17	4.84	4.80	4.66	4.58	4.03	-0.88	-1.14	0.26	
	Low		6.04	6.06	6.02	6.09	6.20	5.80	5.81	5.71	5.57	5.02	-1.02	-1.18	0.16	
	Dif		-1.13	-1.12	-1.10	-1.09	-1.03	-0.95	-1.02	-1.05	-0.99	-0.99	0.14	0.04	0.10	
σ	Ave		3.58	4.18	4.65	5.04	5.41	5.78	6.16	6.55	7.01	7.60	4.02	2.19	1.83	
	High		3.14	3.72	4.14	4.48	4.80	5.13	5.46	5.85	6.30	6.85	3.72	2.06	1.66	
	Low		3.92	4.54	5.03	5.47	5.86	6.25	6.64	7.03	7.46	8.04	4.13	2.19	1.94	
	Dif		-0.78	-0.81	-0.89	-0.98	-1.06	-1.13	-1.18	-1.18	-1.16	-1.19	-0.41	-0.13	-0.28	

Notes: This table shows return volatility sorted by firm characteristics and $\Delta\text{SENT}^*_{t-1}$. For each month we sort returns by firm characteristics and then by $\Delta\text{SENT}^*_{t-1}$. We form decile portfolios based on firm size (ME), firm age (Age), Total Risk (σ), the earnings-price ration (Earn), the dividend-price ratio (div), the book-to-market ration (BE/ME), and momentum (Mom). We then sort the portfolios based on high and low values of $\Delta\text{SENT}^*_{t-1}$. $\Delta\text{SENT}^*_{t-1}$ is high if it is above the mean and vice versa. Ave is the average standard deviation of monthly returns over both high and low sentiment periods. High is the standard deviation of monthly returns when $\Delta\text{SENT}^*_{t-1}$ is high in the previous month. Low is the standard deviation of monthly returns when $\Delta\text{SENT}^*_{t-1}$ is low in the previous month. Dif is the difference between High and Low. The left panel shows the standard deviations for various deciles; the right panel shows the difference between certain deciles.

Table 2.5: Table 2.4 Continued

		Decile											Comparisons			
Port	$\Delta \text{SENT}^*_{t-1}$	≤ 0	1	2	3	4	5	6	7	8	9	10	$10-1$	$10-5$	$5-1$	$>0-\leq 0$
Earn	Ave	8.29	6.46	5.58	5.24	4.96	4.87	4.71	4.64	4.72	4.92	5.61	-0.85	0.74	-1.59	-4.44
	High	7.59	5.73	4.93	4.57	4.32	4.22	4.14	4.19	4.32	4.46	5.08	-0.65	0.86	-1.51	-4.15
	Low	8.65	6.97	6.06	5.74	5.44	5.35	5.14	4.98	5.04	5.27	5.99	-0.98	0.64	-1.62	-4.49
	Dif	-1.06	-1.25	-1.13	-1.17	-1.12	-1.13	-1.00	-0.80	-0.72	-0.81	-0.92	0.33	0.22	0.11	0.34
Div	Ave	7.20	5.70	5.16	4.91	4.82	4.62	4.45	4.25	4.14	4.00	4.33	-1.37	-0.29	-1.08	-3.84
	High	6.42	5.09	4.52	4.30	4.38	4.18	4.04	3.83	3.78	3.70	3.92	-1.17	-0.26	-0.91	-3.38
	Low	7.68	6.16	5.65	5.37	5.16	4.97	4.77	4.59	4.44	4.24	4.65	-1.51	-0.31	-1.20	-4.07
	Dif	-1.27	-1.07	-1.13	-1.07	-0.77	-0.79	-0.73	-0.76	-0.66	-0.54	-0.73	0.34	0.05	0.29	0.68
BE/ME	Ave		7.38	5.75	5.64	5.18	4.90	4.83	5.02	5.13	5.26	6.46	-0.92	1.56	-2.47	
	High		6.49	5.01	4.98	4.53	4.27	4.29	4.51	4.72	4.90	6.40	-0.09	2.13	-2.22	
	Low		8.03	6.30	6.13	5.67	5.38	5.24	5.41	5.44	5.55	6.52	-1.51	1.13	-2.65	
	Dif		-1.54	-1.29	-1.15	-1.15	-1.12	-0.95	-0.91	-0.72	-0.66	-0.12	1.42	1.00	0.42	
Mom	Ave		8.52	6.17	5.43	4.92	4.73	4.57	4.60	4.72	5.10	6.34	-2.18	1.61	-3.79	
	High		7.73	5.49	4.70	4.23	4.06	3.94	4.07	4.28	4.66	5.91	-1.82	1.84	-3.66	
	Low		9.03	6.64	5.94	5.41	5.21	5.04	4.98	5.05	5.41	6.63	-2.40	1.42	-3.82	
	Dif		-1.30	-1.15	-1.24	-1.18	-1.14	-1.10	-0.92	-0.76	-0.76	-0.72	0.58	0.42	0.16	

Table 2.6: Regressions—Portfolio Returns and Sentiment

		$\Delta\text{SENT}^*_{t-1}$		Ortho $\Delta\text{SENT}^*_{t-1}$	
		$\boldsymbol{\beta}$	**p-val**	$\boldsymbol{\beta}$	**p-val**
		Panel A: Size, Age, and Risk			
ME	SMB	0.288	(0.004)	0.260	(0.007)
Age	Young-Old	0.189	(0.000)	0.172	(0.000)
σ	High-Low	0.556	(0.000)	0.512	(0.000)
		Panel B: Profitability and Dividends			
DIV	$=0->0$	0.699	(0.000)	0.638	(0.000)
EARN	$\leq 0->0$	0.875	(0.000)	0.875	(0.000)
		Panel C: Growth Opportunities and Distress			
BE/ME	HML	0.180	(0.031)	0.181	(0.026)
		Panel D: Growth Opportunities			
BE/ME	Medium-Low	-0.103	(0.012)	-0.106	(0.007)
		Panel E: Momentum and Distress			
Mom	UMD	0.418	(0.003)	0.440	(0.001)
		Panel F: Momentum			
Mom	High-Medium	0.050	(0.090)	0.055	(0.069)
		Panel G: Distress			
BE/ME	High-Medium	0.098	(0.005)	0.078	(0.007)
Mom	Medium-Low	-0.198	(0.001)	-0.152	(0.011)

Notes: Regressions of long-short portfolios at time t on $\Delta\text{SENT}^*_{t-1}$. See equation 4.11. In each regression we control for the Fama-French factors and the momentum factor. If one of the the Fama-French factors or the momentum factor is the dependent variable, then we eliminate from the set of regressors. Portfolios are formed on size (ME), age, total risk (σ), the dividend-price ratio (Div), the earnings-price ratio (Earn), the book-to-market-ratio (BE/ME), or momentum (Mom). High represents the top three deciles, medium represents the middle four deciles, and low represents the bottom three deciles. The left panel shows the results using the original sentiment index. The right panel shows the results using the sentiment index orthogonalized to the macro indicators listed in section 2.4.1. The left column of each panel lists the regression coefficient (β) on the sentiment index; the right column lists the bootstrapped p-value.

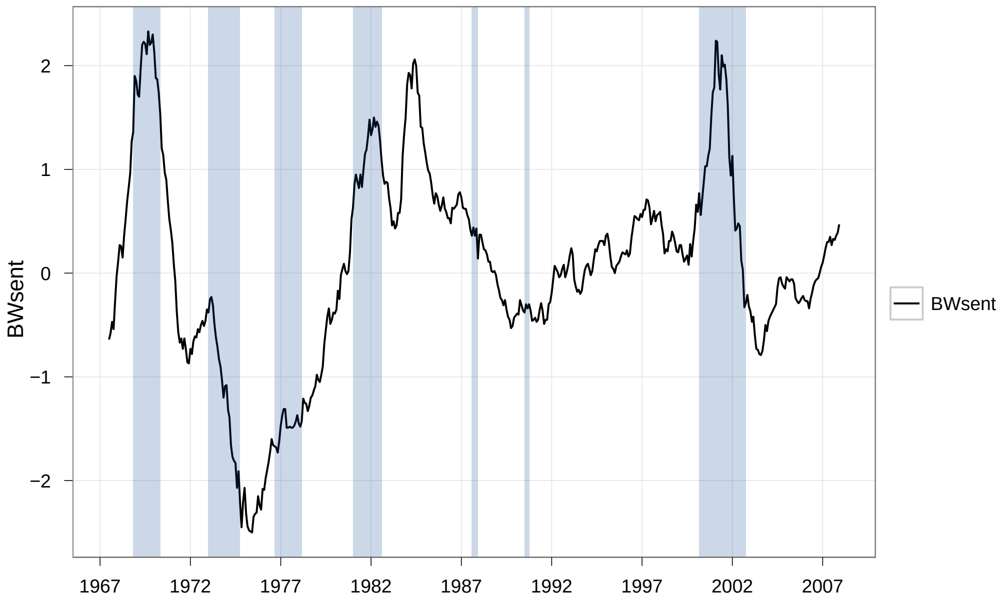
Table 2.7: Peaks and troughs in Sentiment.

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 2.6 and appendix 2.9.

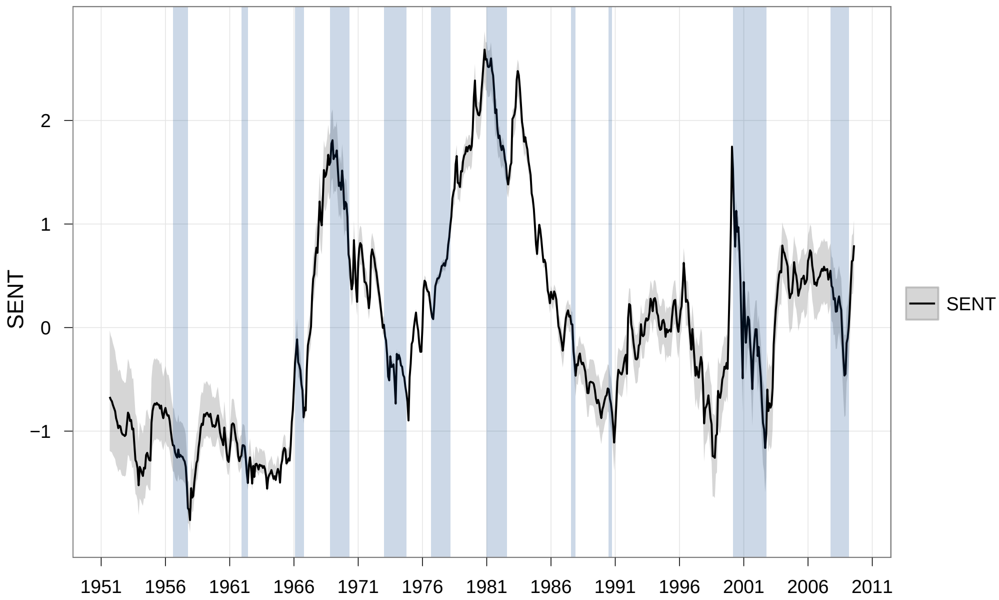
2.11 Appendix: Figures

Figure 2.1: Baker and Wurgler's Sentiment Index (BWsent)



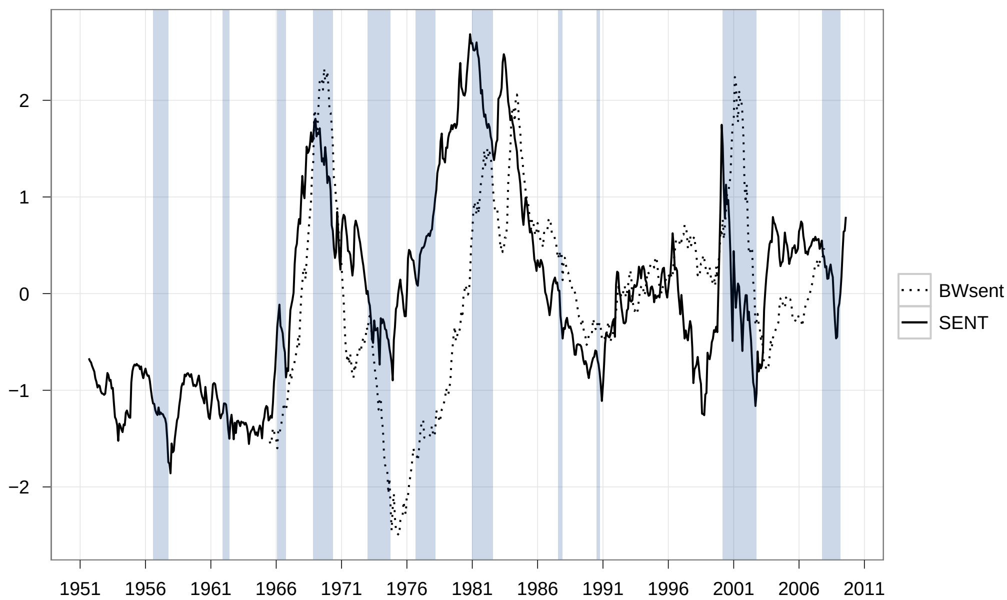
Notes: Shaded areas are bear markets defined by a 20 percent or more drop in the S&P 500 over a two or month period. Data is monthly. The horizontal axis holds the date.

Figure 2.2: SENT



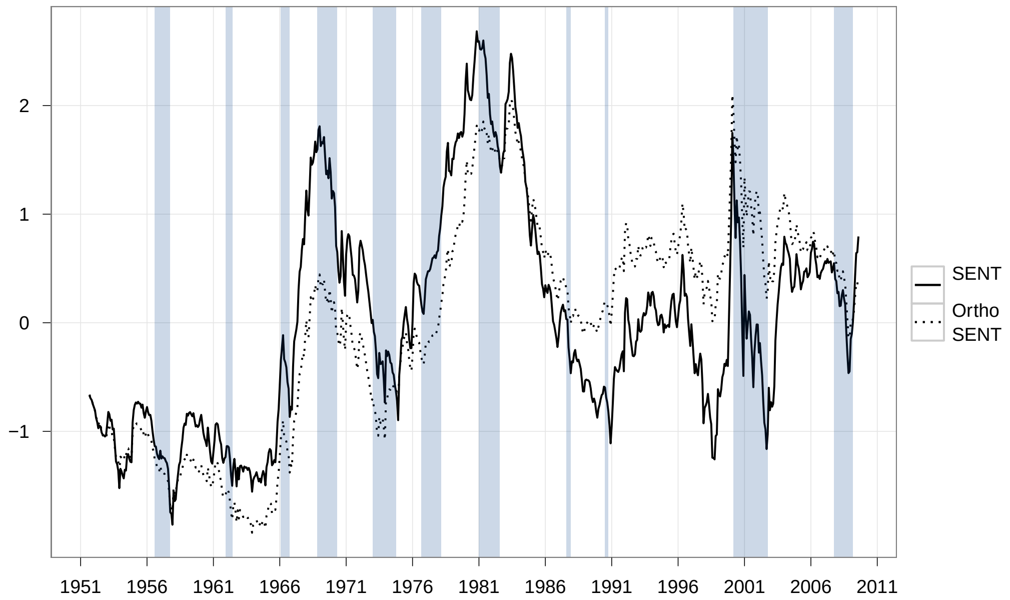
Notes: The sentiment index (SENT) developed in section 4.2. The Date is along the horizontal axis and the level of agent sentiment (SENT) is along the vertical axis. High values of SENT correspond to high levels of agent sentiment. The gray bands represent 95 percent confidence intervals for SENT. The shaded vertical bars represent bear markets defined as 20 percent or more drop in the S&P 500 over a two or more month period.

Figure 2.3: BWsent vs. SENT



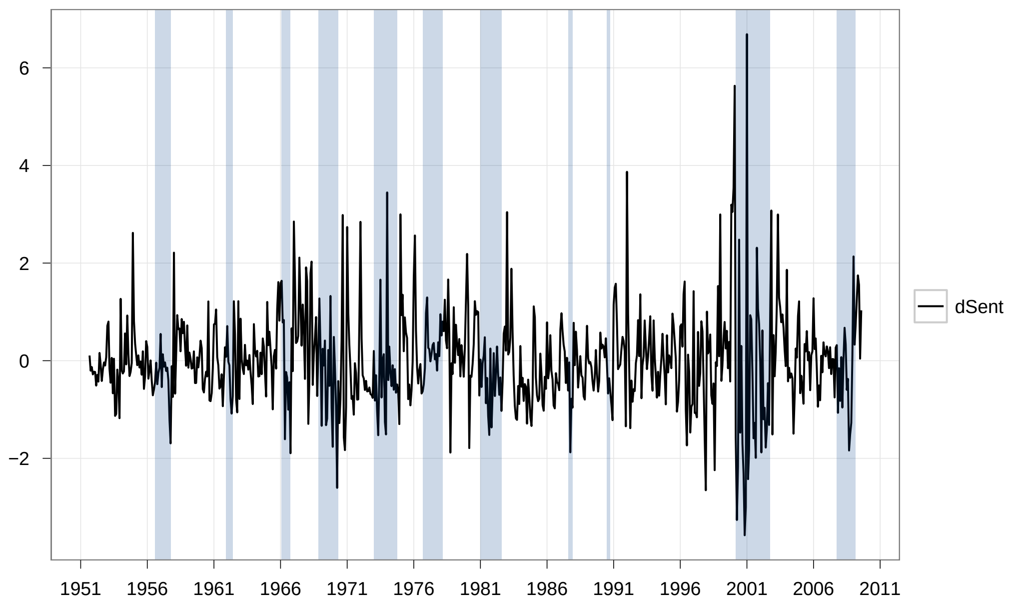
Notes: Shaded areas are bear markets defined by a 20 percent or more drop in the S&P 500 over a two or month period. The black line is SENT. We compile SENT via a dynamic factor model. The red dashed line is BWsent. The horizontal axis holds the date.

Figure 2.4: SENT vs. Orthogonalized SENT



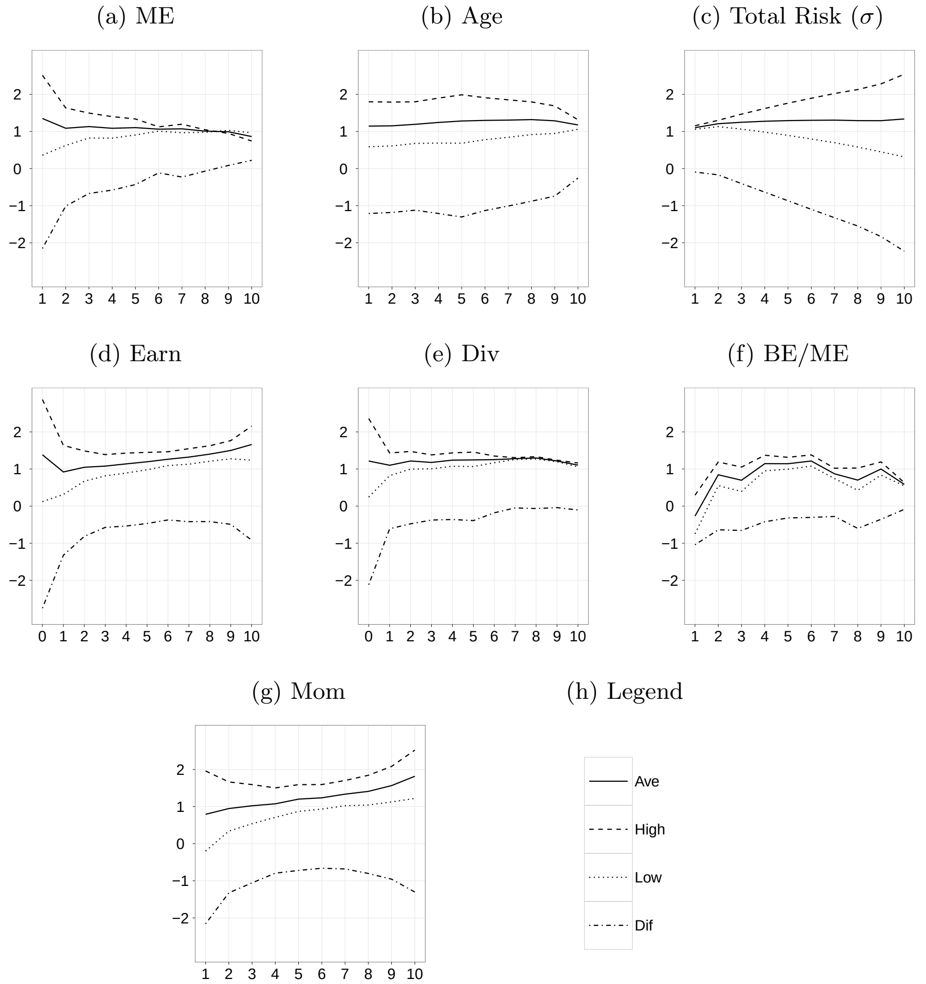
Notes: SENT (the black line) is the sentiment index developed in section 4.2. We calculate Orthogonalized SENT (the red, dashed line) by first regressing the components of SENT on macroeconomic factors and then compiling the index. Orthogonalized SENT accounts for the macroeconomic and time-varying indicators described in 2.4.1. Shaded areas are bear markets defined by a 20 percent or more drop in the S&P 500 over a two or month period. The horizontal axis shows the date.

Figure 2.5: ΔSENT*



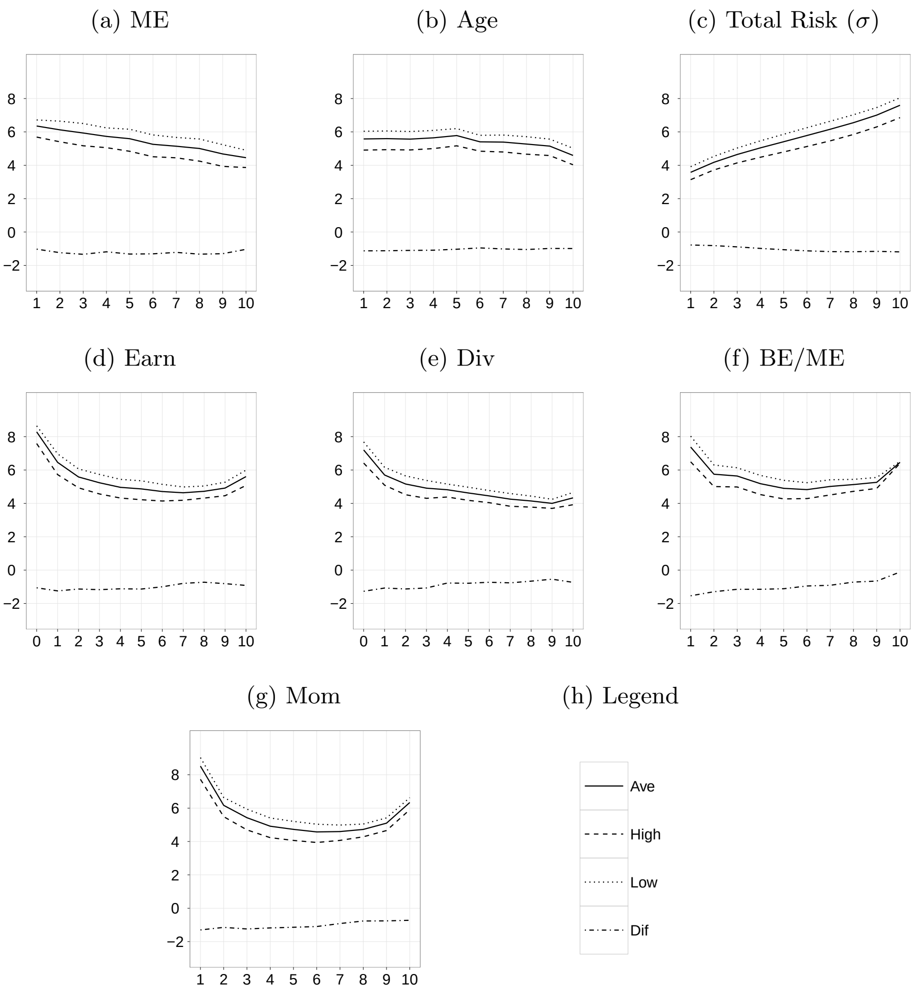
Notes: The graph of ΔSENT*. We interpret ΔSENT* as the first difference in sentiment. The vertical bars represent bear markets defined as a 20 percent of more drop in the S&P 500 over a two or more month period.

Figure 2.6: Returns for Stocks Sorted by Firms Characteristics and ΔSENT*



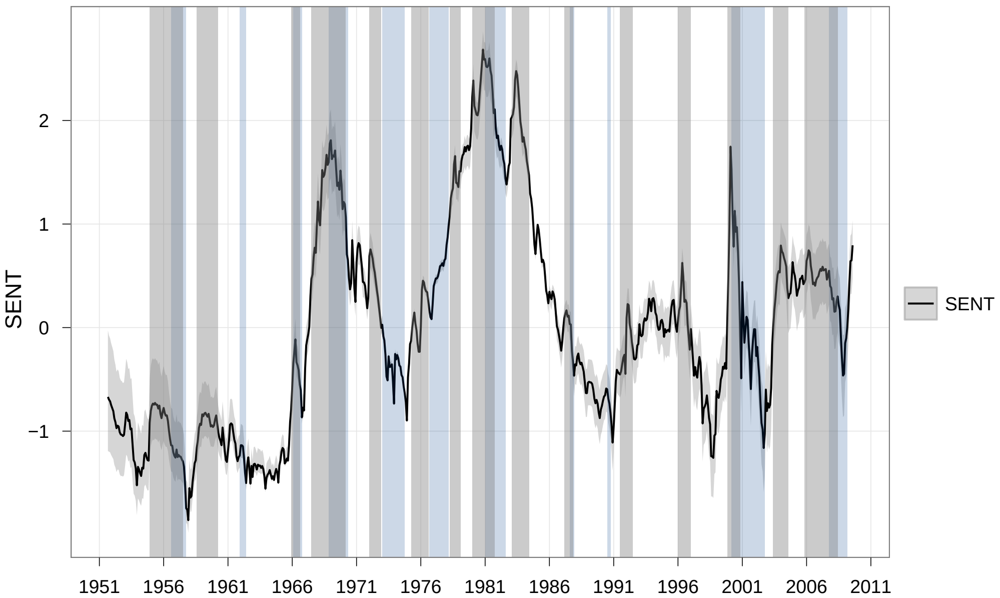
Notes: This figure shows returns sorted by firm characteristics and by $\Delta\text{SENT}^*_{t-1}$. Decile portfolios are on the horizontal axis and monthly returns are on the vertical axis. Panel (h) shows the legend. The solid black line represents average monthly returns (Ave); the dashed line represents returns when ΔSENT* is high in the previous month (High); the dotted line represents returns when ΔSENT* is low in the previous month (Low); and the dash-dot represents -1 times the difference between returns when ΔSENT* is high in the previous month and when ΔSENT* is low on the previous month (Dif). For Panels (a) through (g), portfolios are based on size (ME), age, total risk (σ), the earnings-price ratio (Earn), the dividend-price ratio (Div), the book-to-market ratio (BE/ME), and momentum (Mom). For the portfolios based on earnings and dividends, zero on the horizontal axis represents firms with no earnings or no dividends.

Figure 2.7: Return Volatility for Stocks Sorted by Firms Characteristics and ΔSENT*



Notes: This figure shows return volatility sorted by firm characteristics and by $\Delta\text{SENT}^*_{t-1}$. Decile portfolios are on the horizontal axis and monthly returns are on the vertical axis. Panel (h) shows the legend. The dashed line represents the standard deviation of monthly returns when ΔSENT* is high in the previous month (High); the dotted line represents the standard deviation of monthly returns when ΔSENT* is low in the previous month (Low); the solid black line represents the average standard deviation over both High and Low (Ave); and the dash-dot represents the difference between High and Low (Dif). For Panels (a) through (g), portfolios are based on size (ME), age, total risk (σ), the earnings-price ratio (Earn), the dividend-price ratio (Div), the book-to-market ratio (BE/ME), and momentum (Mom). For the portfolios based on earnings and dividends, zero on the horizontal axis represents firms with no earnings or no dividends.

Figure 2.8: Cycles in sentiment growth



Notes: SENT is the black line. The gray bands around SENT are its 95 percent confidence intervals. The shaded gray areas represent peaks to troughs in sentiment growth, ΔSENT^*. The modified Bry-Boschan algorithm described in section 2.6 and appendix 2.9 determines the turning points in sentiment growth, ΔSENT^*. The shaded light blue areas represent bear markets defined as a 20 percent or more drop in the S&P500 over a two or more month period.

Chapter 3

Two-Stage Model Averaging: An Application to Forecasting Stock Returns using Investor Sentiment

Abstract: We develop two-stage model averaging (2SMA), an extension of model averaging. 2SMA is an easy-to-implement, flexible framework that allows researchers to incorporate prior economic information. 2SMA helps researchers deal with the 'crowding out effect.' Researchers can combine 2SMA with any other model averaging technique. In an application, we develop two-stage Bayesian model averaging (2SBMA), the two-stage model averaging extension of dynamic Bayesian model averaging, and two-stage equal-weighted averaging (2SEWA). Through 2SBMA and 2SEWA we forecast stock returns using investor sentiment. We find that the 2SBMA and 2SEWA forecasts statistically outperform the benchmark random walk plus drift over the sample period. The 2SBMA and 2SEWA forecasts also both beat their traditional model averaging counterparts.

Researchers consider a large number of models that often contain a large number of specifications when developing their forecasts. To circumvent choosing just one model with one specification researchers use model averaging. Model averaging allows researchers to diversify, account for uncertainty, reduce volatility, and improve accuracy. But as the number of models and specifications grow, the weights on consistently top-performing models may fall.[1] We call this the 'the crowding out effect.'

Crowding out is problematic when certain models perform well over time and don't receive a large enough weight in the final forecast.[2] For example, suppose we want to forecast stock returns one step ahead; where y_{t+1} is the return at time $t+1$. Suppose moreover that we use an exogenous variable, x_t to forecast y_{t+1} with a linear model. We may have many specifications to choose from for the linear model. Yet the efficient market hypothesis says that the random-walk plus drift model, whose forecast is an equal-weighted average, should outperform other models over time.[3] So, to create the final forecast we average over the random-walk plus drift and the linear model. In all likelihood, there is uncertainty regarding the choice of the optimal specification of the linear model. If we consider all possible specifications in the combined forecast, the weights on the specifications of the linear model may crowd out the random-walk plus drift. This is especially true if we use equal or relative weights. Hence, the random-walk plus drift may receive a lower weight than we expect. In essence, crowding out occurs when there is unexploited prior information in a model averaging problem.

Previously, researchers dealt with crowding out in one of the following ways:

1. Use equal-weights to average across all specifications of the linear model and the

[1] This result holds for the popular equal-weights and for schemes based relative weighting. Equal weights and relative weights usually perform best in practice. See Timmermann (2006)

[2] Aiolfi and Timmermann (2006) say models that perform well over time are 'persistent.'

[3] The random-walk plus drift also performs well compared to various individual models in practice. See Goyal and Welch (2008), Timmermann (2008).

random-walk plus drift: Even though equal weights often perform well in practice (Timmermann (2006)), as noted above, crowding out will occur.

2. Use a model selection criterion to choose a specification for the linear model: Timmermann (2008) and Aiolfi and Timmermann (2006) use the Bayesian information criterion (BIC) to choose the optimal specification for the linear model. This approach neglects information contained in competing specifications and the uncertainty related to the choice of the optimal specification.

3. Trimming: If we trim poor performing specifications of the linear model there is still no guarantee that crowding out will not occur.

4. Estimate the weights based on past performance: Estimating the weights based on past performance has received much attention in the literature.[4] Yet these methods may not eliminate the crowding out effect.[5] Schemes based on past performance may also place large weights on models that perform well for short periods or randomly just prior to the forecast date. The weights on these forecasts may overpower a different model that is known to perform well over time. This point is alluded to in Timmermann (2008).[6] Furthermore, as noted by Timmermann (2006), more complicated weighting schemes rarely beat simple weighting schemes in practice.[7]

To deal with the crowding out effect and incorporate prior economic information, this paper develops two-stage model averaging (2SMA). Two-stage model averaging is a flexible framework that researchers can combine with any other model averaging

[4]See Timmermann (2006) for an overview of these methods.

[5]We demonstrate this point in the application below using dynamic Bayesian Model Averaging in section 3.2.

[6]Timmermann notes, when forecasting stock returns, that models show local predictability and then lose their predictive power over time.

[7]See also Clemen (1989).

technique. Two-stage model averaging has the following advantages: (1) 2SMA handles the crowding out effect; (2) 2SMA incorporates economic information into a model averaging framework which can improve forecast performance; (3) 2SMA is easy to implement; and (4) 2SMA retains all of benefits of traditional model averaging.

2SMA is a two-stage approach. To implement 2SMA in the above example, we first average over all specifications of the linear model in the first stage, and then average over both the random-walk plus drift and the linear model in the second stage. This approach will place a large weight on the random-walk plus drift model even if we use equal weights. Hence, 2SMA diversifies across specifications of the linear model while retaining the strong performance of the random-walk plus drift. In general, 2SMA incorporates prior information by saving the predetermined top-performing model(s) for the second stage; averaging over other models or specifications in the first stage. In our above example, we saved the random-walk plus drift for the second stage and averaged over the linear model in the first stage. The most obvious use for 2SMA occurs when there is a known benchmark that is difficult to beat. Yet through 2SMA researchers can augment any model that is known to perform well over time with alternative forecasts. 2SMA easily extends to the case where researchers have different exogenous variables or different exogenous variables and different models with different specifications.

2SMA is related to model averaging based on clustering. Aiolfi and Timmermann (2006) group forecasts into clusters based on prior forecasting performance, pool within clusters, and then shrink the optimal combination weights across clusters toward equal weights.[8] Practitioners may prefer 2SMA when they have prior information about a given model that performs well over long time periods even as other models

[8] Aiolfi and Timmermann use the Bayesian information criterion to choose the specification for the linear model, 18 Artificial Neural Network Models, 15 Logistic STAR models, and 21 time-varying autoregressive models.

show strong performance for short intervals. Hence, 2SMA will perform well when predictability is 'elusive' (Timmermann 2008). Furthermore, 2SMA is easy for practitioners to implement.

In an application, we forecast market-weighted returns out-of-sample using the sentiment index of Lutz (2011). We combine 2SMA and Bayesian model averaging (BMA) to develop two-stage Bayesian model averaging (2SBMA). We compare these results to two-stage equal-weighted averaging (2SEWA). We find that both the 2SBMA and 2SEWA techniques beat their traditional model averaging counterparts and outperform the benchmark random-walk plus drift in a statistically significant manner. Lastly, we combine equal-weighted averaging and Bayesian model averaging within the 2SMA framework to form an equal-weight, Bayesian model averaging hybrid. This technique also beats the benchmark in a statistically significant manner. We use both Bayesian model averaging and equal-weighted averaging to demonstrate that the 2SMA approach can be easily work with simple and more complicated averaging schemes. Overall, our 2SMA forecasts are noteworthy since equal-weights are difficult to beat in practice (Timmermann 2006).

We introduce the application in section 3.1. In section 3.2 and appendix 3.8 we describe dynamic Bayesian model averaging and demonstrate the crowding out effect. In section 3.3 we extend dynamic Bayesian model averaging by exploring methods to choose the optimal training length. We also use these methods to in the first stage of 2SBMA.[9] In section 3.4 we outline the two-stage model averaging procedure (2SMA). Lastly, in section 3.5 we develop two-stage Bayesian model averaging (2SBMA) and two-stage equal weighted averaging (2SEWA).

[9]Generally, choosing a training length is necessary for the first stage of 2SMA whenever the weights are estimated based on past performance.

3.1 The Data and Application

We introduce the application to serve as an instructive example. Our goal is to predict market-weighted returns, y_t, one-step ahead using the stock sentiment index of Lutz (2011).[10] Lutz (2011) finds that sentiment cycles lead bear markets. Lutz's sentiment index uses only return data. So, we can use the sentiment index to replicate a real-time forecasting exercise. In appendix 3.7, we describe the construction of the index. Following the notation of Lutz (2011), we represent the sentiment index as ΔSENT^*. To avoid look ahead bias, we re-estimate ΔSENT^* for each in-sample period considered.

We have information available up to time period t and forecast returns at time $t+1$. Let y_{t+1} represent returns for time $t+1$ and ε_{t+1} represent an unpredictable white noise process.

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

$$y_{t+1} = \beta_0 + \varepsilon_{t+1} \tag{3.1}$$

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.[11]

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:

$$y_{t+1} = \beta_0 + \sum_{j=1}^{p} \alpha_j y_{t+1-j} + \sum_{j=1}^{q} \beta_j \Delta\text{SENT}^*_{t+1-j} + \varepsilon_{t+1} \tag{3.2}$$

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 employ a NARX neural network which approximates an equation of the

[10] Market-weighted returns are available from Kenneth French's data library. Other recent sentiment indexes are developed by Baker and Wurgler (2006, 2007) and Frazzani and Lamont (2008). These sentiment measures, however, cant not be calculated in real time.

[11] See Goyal and Welch (2008) and Timmermann (2008).

following form in a non-linear fashion:

$$r_{t+1} = f(r_t, r_{t-1}, \ldots, r_{t-p}, \Delta\mathrm{SENT}^*_t, \Delta\mathrm{SENT}^*_{t-1}, \ldots, \Delta\mathrm{SENT}^*_{t-q}) \tag{3.3}$$

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.[12]

To evaluate the forecasts we employ 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

$$\mathrm{R}^2 = 1 - \frac{\mathrm{MSE}_n}{\mathrm{MSE}_{RWD}} \tag{3.4}$$

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.

To evaluate the statistical significance of a model's performance compared to the benchmark, we employ the Diebold-Mariano test for non-nested models and the Clark and West (2007) test for models that nest the benchmark. These tests compare the loss-differential between the benchmark and the given model. Rejecting the null implies a statistically significant difference in the loss-differential between the two models.

3.1.1 Individual Forecasts

We show the individual model forecasts in table 3.1. We compile the forecasts from 1970M01 to 2009M09 using a rolling window and recursive estimation.[13] We use ten years of data for the rolling window forecasts and all available data for the recursive

[12] For more on Bayesian regulation see MacKay (1992) or Foresee and Hagan (1997).

[13] We estimate the sentiment index using a rolling window (recursively) for the rolling window (recursive) forecasts.

forecasts. In this section we choose the optimal specification for both the linear and NARX models using standard econometric techniques.[14]

We report the mean-absolute error (MAE) for each decade and the OOS R^2 statistic for the overall sample. In table 3.1 RWD is the random-walk plus drift model. Under rolling window estimation, the linear model beats the random-walk plus drift for the overall sample but the loss-differential is not statistically significant. The linear and the NARX models both beat the benchmark in the 1970s and 2000s subperiods.

Under recursive estimation, both the linear model and the NARX neural network beat the random-walk plus drift. The loss-differential is significant at the 15 percent level for the linear model.

The random-walk plus drift is more difficult to beat under rolling window estimation. So, in the remainder of the paper we use the rolling window forecasts for the individual models and their specifications.

3.2 Average Forecasts and the Crowding Out Effect

We average forecasts using a simple equal-weighted average and 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 econometrician is ignorant to its true form. Raftery (1995) and Hoeting et al. (1999) outline this use of BMA. Cremers (2002) and Avramov (2002) apply this type of BMA procedure to stock return data. Raftery et al. (2005) develop a dynamic BMA procedure where the correct model changes from time period to time period. Hence, dynamic BMA estimates the model weights based on past perfor-

[14] We use the Bayesian Information Criterion (BIC) to choose the optimal specification for the linear model. For the NARX model, we regress the fitted values on the actual values. We then use the R^2 to choose the optimal specification for the NARX network. This is standard practice in the neural network literature.

mance. The dynamic BMA technique then creates the optimal forecast from a weighted average over a given set of models. Furthermore, dynamic BMA 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 a given set of models. We describe dynamic Bayesian model averaging in more detail in appendix 3.8.

To implement dynamic Bayesian model averaging we need to choose a training length over which to estimate the posterior weights. We discard the first 60 observations to study the different training lengths. This implies we are eliminating half of the 1970s from our out-of-sample exercise. As noted in table 3.1, 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 dynamic BMA approach combines the random walk plus drift, linear and NARX forecasts. This implies that the BMA forecast nests the random-walk plus drift model. Therefore, we must use the Clark and West (2007) test to evaluate forecast performance.

We consider eight specifications for the linear model (not including the random-walk plus drift) and eight specifications for the NARX neural network. These specifications correspond to the number of lags for the autoregressive and sentiment components.

As noted in the introduction, researchers have traditionally averaged over different models with various specifications in the following ways: (1) using equal weights; (2) estimating the weights based on past performance; or (3) using a model selection criterion to determine the optimal specification. But all of these techniques are susceptible to the 'crowding out effect.' The crowding out effect occurs when the weight on a model known to perform well falls as the number of peripheral forecasts grow.

In table 3.2 we average over all models and all specifications using dynamic Bayesian model averaging and equal-weights.[15] The far left column shows the training length. The middle columns show the average weight applied to each model; standard deviations are in parentheses. The far right column shows the OOS R^2 statistic. For each $AR_{p,q}$, p is the number of autoregressive lags and q is the number of sentiment lags. Similarly for $NARX_{p,q}$. The last row shows the results using equal weights.

The efficient market hypothesis says that stock returns are unforecastable over time. This implies that the random-walk plus drift will be the top model. Furthermore, as noted above, the random-walk plus drift performs well in practice versus other individual models. Yet there are 17 different forecasts when we consider all models and all specifications. Hence, when we use equal-weights, the random-walk plus drift receives a weight of only 1/17 (5.6 percent); the other forecasts crowd out the random-walk plus drift. The equal-weighted forecast also performs poorly with an OOS R^2 statistic of -0.123 percent.

Dynamic BMA partially alleviates the crowding out effect: the average weight on the random-walk plus drift varies from 7.3 percent to 12.6 percent depending on the training length. The top-performing BMA forecast uses a training length of 40 periods. Yet the average weight on the random-walk plus drift for this forecast is only 8.9 percent, a small increase compared to equal-weights. Clearly, the crowding out effect still occurs under dynamic BMA.

Dynamic BMA does improve performance compared to the equal-weighted forecast. The OOS R^2 statistic is positive for 15, 20, 25, 30, 35, 40, 45 and 50 training lengths. The loss differential between the benchmark and the BMA forecast is sta-

[15]Table 3.2 extends over two pages.

tistically significant when we use 25, 40, or 45 training lengths. 40 training lengths produces the largest OOS R^2 statistic at 0.571 percent. Dynamic BMA also beats the equal-weighted forecasts.

In dynamic BMA, we can interpret the average posterior weights as the average forecasting performance over the sample. So, the best model overall is the random-walk plus drift. But the standard deviations of the posterior weights are quite large. This suggests that the optimal model, according to dynamic BMA, changes over time.

As previously mentioned, we can also choose the optimal specification for each model and then average over all models to deal with the crowding out effect. This implies that we average over just three forecasts: one for the random-walk plus drift, one for the linear model, and one for the NARX neural network. We apply this approach to our application and show the results in table 3.3. We choose the optimal specification for the linear model using the Bayesian information criterion (BIC). For the NARX neural network, we regress the fitted values on the actual data and choose the specification with the highest R^2. We then average over the random-walk plus drift, the linear model, and the NARX neural network. As noted above, choosing just one specification for each model neglects information contained in competing specifications.

In table 3.3, the far left column holds the number of training periods for the dynamic BMA process (the last column shows the results using equal-weights). The middle columns hold the average posterior weights with their standard deviations in parentheses. The far right column shows the OOS R^2 statistic expressed as a percentage.

Since we average over just three models, the weight on the random-walk plus drift is 1/3 for equal weights. Yet the OOS R^2 statistic is about 0. The weight on the random-walk plus drift grows when we use dynamic BMA. This leads to an increase in the OOS R^2 statistic.

Under dynamic BMA, the OOS R^2 statistic is positive except when we use 45 or 60 training lengths. Clearly, the drawback to using a model selection criterion is losing information contained in competing specifications. In our case, this loss of information adversely affects performance.

Through our application, we have demonstrated that averaging over all specifications for all models crowds out the top-performing forecasts. Furthermore, choosing the optimal specification via a selection criterion neglects the information contained in competing model specifications. To deal with the crowding out effect while diversifying across all models we develop two-stage model averaging in section 3.4.

3.3 Choosing a Training Length

Model averaging techniques that use past information need a window over which to estimate the weights. For example, we consider various training lengths to estimate the posterior weights in the dynamic BMA process. In this section, we outline methods to choose this training length. Below, we apply these methods to dynamic BMA and two-stage Bayesian model averaging (2SBMA).

We consider the following methods to determine the optimal training length:

1. Pooled: Average across all considered training lengths.

2. Cross-Validation (CV): The following steps describe the cross-validation technique.

   (a) For each model, suppose we have the forecasts, $\hat{y}_1, \ldots, \hat{y}_t$, of $y_1, \ldots, y_t$.

   (b) For each training length k, estimate the weight on each model using $\hat{y}_k, \ldots, \hat{y}_t$ and $y_k, \ldots, y_t$, compute the average forecast $\hat{y}^{\text{ave}}_{t+1,k}$, and compute the forecast error $\hat{\varepsilon}^{\text{ave}}_{t+1,k} = y_{t+1} - \hat{y}^{\text{ave}}_{t+1,k}$.

   (c) Repeat step (b) for $t = m, \ldots, n-1$ where m is the number of observations used for the longest training length. Hence, we use $n-m$ observations in the cross-validation procedure.

(d) For each training length considered, compute the MSE from $\hat{\varepsilon}^{\text{ave}}_{m+1,k},\ldots,\hat{\varepsilon}^{\text{ave}}_{n,k}$

(e) Select the training length that produces the lowest MSE. Using this training length, create the forecast for time $n+1$, $\hat{y}^{\text{ave}}_{n+1}$.

3. Inverse-Error (InvErr): This procedure is equivalent to cross-validation except in step (e) we average across training lengths using the inverse MSE. Thus, we calculate the weight, w_i, on each training length as follows:

$$w_i = \frac{1/\text{MSE}_i}{\sum_{j=1}^{n} 1/\text{MSE}_j} \tag{3.5}$$

Where n is the number of different training lengths.

4. Cross-Validation Window (CVwin): Conduct the cross-validation procedure described above to determine the optimal training length and select a window around this length. Then average around the optimal training length using equal weights. This procedure takes into account the uncertainty related to choosing the training length via cross-validation.

We apply these training length selection techniques to the dynamic BMA forecasts. The above methods to choose a training length extend the dynamic BMA framework of Raftery et al. (2005). Table 3.4 shows the results.

We estimate CV, InvErr and CVwin using windows of 1, 5 or 10 periods. The far left column in table 3.4 shows the training length selection method. CV-1, CV-5, and CV-10 represent the cross-validation technique using 1, 5, and 10 periods respectively. Similarly for InvErr and CVwin. All of the OOS R^2 statistics in table 3.4 are positive, but none of the loss differentials are statistically significant. The top-performing training length criterion is CV-5 with an OOS R^2 statistic of 0.426 percent.

3.4 Two-Stage Model Averaging (2SMA)

In the introduction and in section 3.2 we outlined and defined the crowding out effect. In model averaging, crowding out occurs when the weights on the top-performing models drop as the number of alternative forecasts grow. Crowding out is particularly problematic when we consider equal or relative weights. Furthermore, as demonstrated in section 3.2, crowding out can occur even when we estimate the weights based on past performance.

If researchers use economic information a priori to determine models that will perform well over time, then two-stage model averaging can circumvent the crowding out effect. Two-stage model averaging (2SMA) incorporates prior information to ensure that top-performing models receive a large weight in the final forecast. 2SMA is a two stage approach that requires the following steps:

Stage 1:

1. Use prior economic information to determine the top-performing models. Save these models and their forecasts for the second stage.

2. Average over the other models and specifications to develop the stage 1 forecasts.

Stage 2:

1. Average over the saved models and the stage 1 forecasts to develop the 2SMA forecasts.

Since 2SMA saves the top-performing models for the second stage and averages over other forecasts in the first stage, it will eliminate the crowding out effect while diversifying across a large number of models. Also, researchers can combine 2SMA with any other model averaging technique.

To see why 2SMA improves forecast performance consider the following example: Suppose, for time $t+1$ we want to average over n forecasts, $\hat{y}^1_{t+1}, \ldots, \hat{y}^n_{t+1}$. Suppose moreover that prior economic information tells us that the forecast from model 1, $\hat{y}^1$, performs well over time. As noted above, when we use equal-weights, relative-weights, or estimated weights, the alternative forecasts will crowd out $\hat{y}^1$. In 2SMA, however, we average over $\hat{y}^2, \ldots, \hat{y}^n$ to develop the first stage forecast. Then we average over the first-stage forecast and $\hat{y}^1$ to develop the 2SMA forecast. This ensures that $\hat{y}^1$ receives a large weight in the final forecast.

If researchers use past performance to select the weights for the stage 1 forecasts, then they need to choose a window over which to estimate these weights. This window is commonly called the training window or the training length. In section 3.3 we outline various methods to choose a training length or average over different training lengths. These methods include pooling across training lengths, using cross-validation, weighting training lengths based on past performance, or using a cross-validation window. The window selection techniques from section 3.3 apply to any scheme that estimates weights based on past performance.

3.4.1 Choosing What to average over in the First and Second Stages

An important issue is how to choose what to average over in the first and second stages. Prior economic information will play a large roll in this decision. Researchers should reserve models or specifications that are known to perform well for the second stage. This will ensure that they receive a large weight in the final forecast. For example, the efficient market hypothesis states that the random-walk plus drift will outperform other forecasting models over time. So, in the application in section 3.5 we reserve the random-walk plus drift model for the second stage.

The simplest way to implement 2SMA is to save the models known to perform well, average over all other forecasts in stage 1, and then average over the stage 1 forecast and the saved models in stage 2. This approach will directly incorporate economic information and eliminate the crowding out effect. We demonstrate this approach in section 3.5. Yet there are other ways to choose the first and second stage forecasts.

When considering different models with one or zero exogenous variables, a natural approach is to first average all specifications for each model; then average over all models. This technique first takes into account the uncertainty related to choosing the optimal specification for each model and the uncertainty related to choosing the optimal model. This approach also allows researchers to study the performance of each model in the first stage. We use this intuitive technique for the application in section 3.5. A similar approach applies when researchers are considering multiple exogenous variables and only one model.

The choice becomes trickier when there are multiple models and multiple exogenous variables. In this case, researchers may develop the first-stage forecasts by applying 2SMA to each exogenous variable for the various models and their specifications. Then the researchers could average over the first-stage forecasts and the saved models to develop the 2SMA forecasts. Researchers must be careful, however, as a large number of first-stage forecasts may crowd out the top-performing models originally saved for stage 2.

A more methodical approach involves clustering forecasts. Researchers can cluster forecasts based on performance in the first stage, average within clusters, and then average across clusters in the second stage. Aiolfi and Timmermann (2006) use this approach.[16] But Aiofli and Timmermann do not save models known to perform well for the second stage. Hence, Aiofli and Timmermann's technique is susceptible to the crowding out effect.

Regardless of the methods used in the first stage, the top-performing models, based on prior economic information, must receive large weights in the second stage. This will eliminate the crowding out effect and improve performance.

3.4.2 A Bayesian Analogy

Bayesian estimation combines a prior distribution with the likelihood function to develop a posterior distribution. Researchers then use the posterior distribution for inference. Researchers incorporate economic information through the form prior distribution.

2SMA uses prior economic information to weight some forecasts more than others. More precisely, in 2SMA researchers have some prior belief regarding the performance of a certain forecast. They save this forecast for the second stage. Researchers average over peripheral models in the first stage and then 'update' the pre-selected forecasts in the second stage. In this sense, 2SMA is Bayesian approach.

3.5 Two-Stage Bayesian Model Averaging (2SBMA) and Two-Stage Equal-Weighted Averaging (2SEWA)

Next, we apply 2SMA to dynamic BMA and equal-weighted averaging to develop two-stage Bayesian model averaging (2SBMA) and two-stage equal-weighted averaging (2SEWA). In 2SEWA we use equal weights to average forecasts in stage 1 and stage 2. Similarly, in 2SBMA we use dynamic BMA to average forecasts in both stages. We implement 2SBMA in the following way:

[16] Aiofli and Timmermann use equal weights to average within clusters in the first stage, and then optimal weights based on least squares shrunk to equal weights to average across clusters in the second stage.

Stage 1:

1. Save the predetermined top-performing forecasts for the second stage.

2. Apply the BMA procedure over all other forecasts up to time $t + 1$ using different window lengths.

3. Record the corresponding mean-squared errors.

Stage 2:

1. Select the training length that produces to lowest MSE for the forecast at time period t.

2. Apply the BMA procedure again over the stage 1 forecast and the saved models to develop the 2SBMA forecast for time $t + 1$.

Note that we use leave-one-out cross-validation (CV-1) to determine the optimal training length in Stage 1; this produces a single forecast stage 1 forecast.

We apply 2SBMA and 2SEWA to our application and forecast stock returns using investor sentiment. As noted above, prior economic research and the efficient market hypothesis state that the random-walk plus drift should be the top-performing model over time. Hence, we save the random-walk plus drift for the second stage. Also, since 2SBMA uses dynamic BMA in both stages, we need twice as many forecasts to verify the training length. Thus, we are eliminating the 1970s from our out-of-sample exercise. This will negatively affect our results as the 1970s were a period where both the linear model and the NARX neural network beat the benchmark.[17]

In stage 1, we average over all of the specifications of the linear model and the NARX neural network. Table 3.5 presents the stage 1 forecasts.[18] The far left column

[17]See section 3.1.1.

[18]Table 3.5 extends over two pages.

shows the number of training periods used for the BMA process, the middle columns show the posterior weights, and the far right column shows the OOS R^2 statistic. The last row displays the results based on equal-weights. The OOS R^2 statistic is positive when 25-55 training lengths are used. The loss differential between the stage 1 forecasts and the benchmark is also statistically significant for a training length of 25, 30, 40 or 45 periods. The top-performing specification, according to the BMA weights, is $AR_{0,1}$. This specification corresponds to zero autoregressive lags and one sentiment lag.

Table 3.6 displays the results from the second stage. The second column lists the weight on the random-walk plus drift and the third column lists the weight on the stage 1 forecast. The last column shows the OOS R^2 statistic. The second-to-last row shows the 2SEWA forecast results where we use equal-weights in the first and second stages. Finally, the last row shows the 2SMA results from an equal-weight, dynamic BMA hybrid. In this hybrid, we use dynamic BMA in the first stage and equal-weights in the second stage.

All of the OOS R^2 statistics in table 3.6 are positive and large in magnitude. For all of the stage 2 forecasts, the loss differential relative to the random-walk plus drift is statistically significant at the 5 or 1 percent level. Dynamic BMA using 20 training lengths produces the largest OOS R^2 statistic at 2.096 percent. Furthermore, the 2SBMA and 2SEWA forecasts beat their traditional averaging counterparts, dynamic BMA and equal-weighted averaging, by a large margin.[19]

Not surprisingly, the alternative forecasts do not crowd out the random-walk plus drift in 2SBMA and 2SEWA. The weight on the random-walk plus drift is greater than or equal to 1/2 for all of the forecasts in table 3.6.

To further show the flexibility of the 2SMA framework we augment the above

[19]See section 3.2 for the traditional dynamic BMA and equal-weighted averaging forecasts.

approach and separately average over each model in stage 1.[20] This approach allows us to compare the performance of the linear and the NARX models. So, 2SBMA now involves the following steps.

Stage 1:

1. Save the predetermined top-performing forecasts for the second stage.

2. Apply the BMA procedure for *each model*[21] over all considered specifications up to time $t+1$ using different window lengths.

3. Record the corresponding mean-squared errors.

Stage 2:

1. Select the optimal window length for each model (from Stage 1) by choosing the window length that produces the lowest mean-squared error for the forecast at time period t.

2. Apply the BMA procedure again over all models to develop the 2SBMA forecast for time $t+1$.

Hence, in stage 2, we average over the forecasts from the linear model, the NARX neural network, and the random-walk plus drift.

Table 3.7 shows the results for the linear model. Under dynamic BMA, the specification corresponding to zero autoregressive lags and one sentiment lag, $AR_{0,1}$, receives the largest weight on average. This implies that $AR_{0,1}$ is top performing specification for the linear model on average. Yet the standard deviations for the weights are large; this suggests that the optimal changes over time. For the top-performing

[20]We described this technique in 3.4.1.

[21]In our application the models are the linear model and the NARX neural network.

training length of 30 periods, the OOS R^2 statistic is 0.4 percent. The forecast using equal weights does not beat the benchmark.

Table 3.8 displays the results for the NARX neural network. In this case, under dynamic BMA the OOS R^2 statistic is positive and the loss differential, relative to the benchmark, is statistically significant when we use 20, 25, or 40-60 training lengths. The top-performing forecast under dynamic BMA uses 60 training lengths. Lastly, the OOS R^2 statistic using equal-weights is negative.

An interesting result in tables 3.7 and 3.8 is length of the optimal training window. For the linear model in table 3.7, a training window of 30 periods seems most appropriate. Yet the NARX model in table 3.8 produces its best forecasts using much longer training windows. Step 1 in the second stage of the 2SBMA process selects the optimal window length for each model in stage 1. Hence, step 1 of stage 2 allows us to use cross-validation to choose shorter windows for the linear model and longer windows for the NARX network. This improves performance of the 2SBMA technique.

Table 3.9 shows the second stage forecast results. The OOS R^2 statistics are positive and among the largest in our study. Using dynamic BMA with a training length of 10 periods produces the largest OOS R^2 statistic at 2.018 percent. The loss differential relative to the benchmark is statistically significant at the 5 or 1 percent level for all the forecasts in table 3.9. The weight on the random-walk plus drift is at least 1/3 in all cases. Thus, the linear and NARX models do not crowd out the random walk plus drift.

3.6 Conclusion

This paper develops two-stage model averaging (2SMA). 2SMA is a flexible framework that allows researchers to incorporate prior economic information. To implement 2SMA, researchers use economic information to save certain models for the second stage, and average over peripheral models in the first stage. In the second stage, the researcher then averages over the saved models and the stage 1 forecasts. This approach ensures that the saved models receive a large weight in the final forecast.

In an application, we combine 2SMA with dynamic Bayesian model averaging and equal-weighted averaging to develop two-stage Bayesian model averaging (2SBMA) and two-stage equal-weighted averaging (2SEWA). Through 2SBMA and 2SEWA we forecast stock returns using investor sentiment. We find that these forecasts statistically outperform the benchmark random-walk plus drift and beat traditional forecasts that use equal weights.

3.7 Appendix: The Dynaimc Factor Model with Bayesian Estimation

We follow Lutz (2011) in the estimation of the index. To estimate the model we first take each series in deviations from the mean. The dynamic factor model then takes the following form:

$$\Delta y_{it} = \gamma_i \Delta c_t + \varepsilon_{it}, i = 1, \ldots, 5 \tag{3.6}$$

$$\phi(L)\Delta c_t = \omega_t, \omega_t \sim N(0, 1) \tag{3.7}$$

$$\psi_i(L)\varepsilon_{it} = v_{it}, v_{it} \sim N(0, \sigma^2) \tag{3.8}$$

Where $\Delta y_{it} = \Delta Y_{it} - \Delta \bar{Y}_i,\ i = 1, ..., 5$, represents one of the five series: div, earn, size lowmom and sp500, respectively. $\Delta c_t = \Delta C_t - \delta$ is the component common to all series, ε_{it} is an idiosyncratic component, γ_i is the factor loading and L is the lag operator. To estimate the model and derive the common component, Δc_t, we cast it into state-space form. Lutz (2010) uses the state-space form in Kim and Nelson (1998). This facilitates computation of the estimation algorithm. Multiplying both sides of equation 4.1 by $\psi_i(L)$ yields

$$\psi_i(L)\Delta y_{it} = \gamma_i \psi_i(L)\Delta c_t + v_{it} \tag{3.9}$$

By the Bayesian Information Criterion, Lutz (2010) chooses two lags for $\phi(L)$ and two lags for $\psi_i(L)$. Equations 4.2 and 4.3 become

$$\Delta c_t = \phi_1 \Delta c_{t-1} + \phi_2 \Delta c_{t-2} + \omega_t \tag{3.10}$$

$$\varepsilon_{it} = \psi_{i1}\varepsilon_{i,t-1} + \psi_{i2}\varepsilon_{i,t-2} + v_{it}, i = 1, \ldots, 5 \tag{3.11}$$

Hence, we employ the following state-space form:

Measurement Equation:

$$\begin{bmatrix} \Delta y_{1t}^* \\ \Delta y_{2t}^* \\ \Delta y_{3t}^* \\ \Delta y_{4t}^* \\ \Delta y_{5t}^* \end{bmatrix} = \begin{bmatrix} \gamma_1 & -\gamma_1\psi_{11} & -\gamma_1\psi_{12} \\ \gamma_2 & -\gamma_2\psi_{21} & -\gamma_2\psi_{22} \\ \gamma_3 & -\gamma_3\psi_{31} & -\gamma_3\psi_{32} \\ \gamma_4 & -\gamma_4\psi_{41} & -\gamma_4\psi_{42} \\ \gamma_5 & -\gamma_5\psi_{51} & -\gamma_5\psi_{52} \end{bmatrix} \begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \\ v_{5t} \end{bmatrix} \tag{3.12}$$

In matrix form this 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 4.4.

$$E(\mathbf{v}_t\mathbf{v}_t') = R = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix} \tag{3.13}$$

Transition Equation:

$$\begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \end{bmatrix} + \begin{bmatrix} \omega_t \\ 0 \\ 0 \end{bmatrix} \tag{3.14}$$

Equation 4.15 can be written in matrix form as

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

$$E(\mathbf{e}_t\mathbf{e}_t') = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.15}$$

Lutz (2011) then augments the common component, Δc_t using the steady-state Kalman Gain to develop a sentiment index in levels. The levels index is not stationary. Hence, we use stationary series Δc_t and re-label it ΔSENT*. Lutz (2011) interprets ΔSENT* as the growth or first difference in sentiment.

To avoid look-ahead bias we re-estimate the sentiment index for each in-sample period considered. Over time, the exact specification of the dynamic factor model may change. So, we augment Lutz's approach and average over three specifications of the dynamic factor model. This allows us to take into account some uncertainty regarding the structure of the model.

3.7.1 Appendix: Bayesian Estimation of the Factor Model

We follow Lutz (2010) and estimate the model using the Bayesian multimove Gibbs-sampling approach based on Carter and Kohn (1994) and Kim and Nelson (1998). The Bayesian method takes into account parameter by jointly estimating the state vector and the model parameters. In contrast, traditional maximum likelihood treats the model parameters as true values when estimating the state vector. 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)

3.8 Appendix: Dynamic Bayesian Model Averaging

In this appendix we describe Bayesian model averaging (BMA). In a forecasting exercise dynamic Bayesian model averaging forms posterior weights based on past forecasting performance.

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. Cremers (2002) and Avramov (2002) apply this type of BMA procedure to stock return data. 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

$$p(y|f_1, \ldots, f_K) = \sum_{k=1}^{K} w_k g_k(y|f_k) \tag{3.16}$$

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 3.16. 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.[22]

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 3.1, 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 dynamic BMA approach combines the random walk plus drift, linear and NARX forecasts. This implies that the BMA forecast nests the random-walk plus drift model. Therefore we must use the Clark and West (2007) test to evaluate forecast

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

performance.

3.9 Appendix: Tables

Table 3.1: 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

Notes: The mean absolute error and the OOS R^2 for the models outlined in section 3.1. The OOS R^2 statistic is expressed as a percentage. 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. One, two, three and four asterisks represent a statistically significant difference in the loss-differential between the forecast and the benchmark random-walk plus drift at the 15, 10, 5 and 1 percent levels, respectively. For the linear model we use the Clark and West (2007) test. For the NARX neural network, significance tests are based on Diebold-Mariano Test.

Table 3.2: Average forecasts using all models and specifications

Periods	RWD	$AR_{1,0}$	$AR_{0,1}$	$AR_{1,1}$	$AR_{1,2}$	$AR_{2,1}$	$AR_{2,2}$	$AR_{2,0}$	$AR_{0,2}$
10	0.105	0.043	0.063	0.026	0.037	0.032	0.034	0.042	0.058
	(0.219)	(0.133)	(0.164)	(0.056)	(0.117)	(0.096)	(0.115)	(0.129)	(0.170)
15	0.123	0.039	0.067	0.031	0.028	0.030	0.036	0.062	0.051
	(0.232)	(0.104)	(0.154)	(0.055)	(0.081)	(0.059)	(0.114)	(0.147)	(0.126)
20	0.119	0.038	0.081	0.031	0.018	0.026	0.033	0.070	0.057
	(0.203)	(0.088)	(0.163)	(0.048)	(0.033)	(0.052)	(0.102)	(0.151)	(0.130)
25	0.126	0.039	0.099	0.037	0.020	0.030	0.038	0.072	0.060
	(0.204)	(0.073)	(0.180)	(0.052)	(0.038)	(0.043)	(0.101)	(0.146)	(0.135)
30	0.114	0.046	0.120	0.040	0.018	0.029	0.038	0.068	0.055
	(0.185)	(0.092)	(0.202)	(0.057)	(0.023)	(0.036)	(0.089)	(0.126)	(0.122)
35	0.105	0.044	0.120	0.043	0.020	0.031	0.036	0.063	0.060
	(0.168)	(0.082)	(0.182)	(0.057)	(0.025)	(0.034)	(0.086)	(0.111)	(0.119)
40	0.089	0.045	0.138	0.044	0.020	0.031	0.037	0.058	0.064
	(0.135)	(0.083)	(0.205)	(0.057)	(0.023)	(0.032)	(0.085)	(0.087)	(0.118)
45	0.083	0.044	0.141	0.048	0.020	0.033	0.035	0.057	0.067
	(0.126)	(0.078)	(0.206)	(0.066)	(0.021)	(0.033)	(0.078)	(0.081)	(0.117)
50	0.080	0.043	0.146	0.052	0.022	0.035	0.036	0.056	0.070
	(0.119)	(0.062)	(0.208)	(0.071)	(0.022)	(0.033)	(0.074)	(0.073)	(0.114)
55	0.077	0.045	0.142	0.052	0.025	0.037	0.039	0.062	0.073
	(0.108)	(0.056)	(0.209)	(0.061)	(0.021)	(0.030)	(0.069)	(0.074)	(0.112)
60	0.073	0.047	0.146	0.055	0.025	0.040	0.038	0.064	0.073
	(0.103)	(0.055)	(0.211)	(0.061)	(0.021)	(0.033)	(0.066)	(0.075)	(0.109)
EW	1/17	1/17	1/17	1/17	1/17	1/17	1/17	1/17	1/17
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: **Table continued on next page**. We average over all models and specifications using dynamic Bayesian model averaging and equal weights. The far left column holds the length of the training window. The average posterior weight applied to each model is listed in the middle columns. The standard deviations of the posterior weights are in parentheses. For each $AR_{p,q}$, p is the number of autoregressive lags and q is the number of sentiment lags. Similarly for $NARX_{p,q}$. The far right column holds the OOS R^2 statistic. The OOS R^2 statistic is expressed as a percentage. One, two, three and four asterisks represent a statistically significant difference in the loss-differential between the forecast and the benchmark random-walk plus drift at the 15, 10, 5 and 1 percent levels, respectively. The last row shows the results using equal weights.

Table 3.2 Continued: Average forecasts using all models and specifications

Periods	$NARX_{1,0}$	$NARX_{0,1}$	$NARX_{1,1}$	$NARX_{1,2}$	$NARX_{2,1}$	$NARX_{2,2}$	$NARX_{2,0}$	$NARX_{0,2}$	R^2 (%)
10	0.058	0.085	0.040	0.059	0.066	0.076	0.056	0.121	-1.670
	(0.164)	(0.210)	(0.111)	(0.157)	(0.155)	(0.178)	(0.139)	(0.237)	
15	0.050	0.075	0.044	0.053	0.059	0.088	0.058	0.106	0.066
	(0.135)	(0.180)	(0.108)	(0.116)	(0.123)	(0.190)	(0.134)	(0.199)	
20	0.042	0.066	0.052	0.074	0.053	0.092	0.062	0.085	0.108
	(0.095)	(0.136)	(0.106)	(0.151)	(0.111)	(0.198)	(0.134)	(0.169)	
25	0.038	0.059	0.046	0.083	0.042	0.086	0.056	0.069	0.409*
	(0.078)	(0.115)	(0.078)	(0.157)	(0.082)	(0.165)	(0.111)	(0.124)	
30	0.031	0.047	0.045	0.096	0.045	0.091	0.047	0.067	0.397
	(0.061)	(0.084)	(0.071)	(0.169)	(0.078)	(0.150)	(0.091)	(0.110)	
35	0.026	0.055	0.043	0.084	0.058	0.093	0.046	0.074	0.351
	(0.037)	(0.100)	(0.061)	(0.152)	(0.108)	(0.144)	(0.086)	(0.129)	
40	0.026	0.058	0.043	0.080	0.059	0.094	0.037	0.079	0.571*
	(0.041)	(0.111)	(0.064)	(0.141)	(0.097)	(0.148)	(0.058)	(0.142)	
45	0.031	0.060	0.043	0.065	0.070	0.094	0.033	0.076	0.666*
	(0.057)	(0.115)	(0.065)	(0.113)	(0.116)	(0.149)	(0.048)	(0.137)	
50	0.037	0.062	0.044	0.057	0.066	0.088	0.033	0.074	0.025
	(0.062)	(0.110)	(0.048)	(0.095)	(0.106)	(0.133)	(0.045)	(0.131)	
55	0.042	0.061	0.047	0.049	0.065	0.075	0.034	0.072	-0.018
	(0.059)	(0.104)	(0.045)	(0.076)	(0.089)	(0.097)	(0.037)	(0.129)	
60	0.040	0.062	0.046	0.045	0.066	0.079	0.033	0.070	-0.030
	(0.059)	(0.100)	(0.041)	(0.062)	(0.094)	(0.101)	(0.033)	(0.118)	
EW	1/17	1/17	1/17	1/17	1/17	1/17	1/17	1/17	-0.123
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Table 3.3: Average forecasts using the optimal specifications for each model

Periods	RWD	Linear	NARX	R^2 (%)
10	0.381	0.337	0.282	0.015
	(0.324)	(0.317)	(0.295)	
15	0.376	0.305	0.320	0.432
	(0.277)	(0.260)	(0.264)	
20	0.378	0.301	0.321	0.443
	(0.245)	(0.228)	(0.239)	
25	0.379	0.297	0.325	0.317
	(0.232)	(0.203)	(0.224)	
30	0.371	0.290	0.339	0.262
	(0.216)	(0.184)	(0.225)	
35	0.365	0.283	0.352	0.317
	(0.208)	(0.168)	(0.227)	
40	0.366	0.274	0.360	0.177
	(0.204)	(0.156)	(0.229)	
45	0.372	0.270	0.357	-0.004
	(0.207)	(0.149)	(0.225)	
50	0.374	0.271	0.356	0.051
	(0.198)	(0.134)	(0.217)	
55	0.374	0.276	0.350	0.035
	(0.185)	(0.120)	(0.201)	
60	0.376	0.286	0.338	-0.074
	(0.168)	(0.112)	(0.181)	
EW	1/3	1/3	1/3	0.000
	(0.000)	(0.000)	(0.000)	

Notes: The far left column holds the length of the training window. The average posterior weight applied to each model is listed in columns two through four. The standard deviations of the weights are in parentheses. The far right column holds the OOS R^2 statistic. One, two, three and four asterisks represent a statistically significant difference in the loss-differential between the forecast and the benchmark random-walk plus drift at the 15, 10, 5 and 1 percent levels, respectively. The last row shows the results using equal weights.

Table 3.4: The optimal BMA training length using the optimal specification for each model

	RWD	Linear	NARX	R^2 (%)
Pooled	0.374	0.290	0.336	0.002
	(0.146)	(0.120)	(0.171)	
CV-1	0.374	0.306	0.320	0.245
	(0.280)	(0.255)	(0.265)	
CV-5	0.390	0.289	0.321	0.426
	(0.256)	(0.218)	(0.241)	
CV-10	0.382	0.293	0.325	0.107
	(0.253)	(0.221)	(0.237)	
InvErr-1	0.376	0.290	0.334	0.234
	(0.149)	(0.123)	(0.171)	
InvErr-5	0.374	0.290	0.336	0.203
	(0.146)	(0.120)	(0.171)	
InvErr-10	0.374	0.290	0.336	0.208
	(0.145)	(0.120)	(0.171)	
CVwin-1	0.373	0.296	0.331	0.239
	(0.204)	(0.195)	(0.206)	
CVwin-5	0.382	0.282	0.336	0.383
	(0.205)	(0.177)	(0.203)	
CVwin-10	0.379	0.288	0.333	0.127
	(0.215)	(0.183)	(0.207)	

Notes: The far left column holds the the criterion for the choice of the training window over the dynamic Bayesian model averaging procedure. Pooled is an equal-weighted average across all training lengths. CV-1, CV-5, and CV-10 represent the cross-validation method using 1, 5, and 10 periods, respectively. InvErr-1, InvErr-5 and InvErr-10 represent the Inverse MSE method using 1, 5, and 10 periods, respectively. CVwin-1, CVwin-5, and CVwin-10 represent a window around the optimal training length chosen by cross-validation (We average the forecasts on each side of the optimal training length). The average posterior weight applied to each model is listed in columns two through four. The standard deviations of the weights are in parentheses. The far right column holds the OOS R^2 statistic. One, two, three and four asterisks represent a statistically significant difference in the loss-differential between the forecast and the benchmark random-walk plus drift at the 15, 10, 5 and 1 percent levels, respectively. The last row shows the results using equal weights.

Table 3.5: Two-Stage Model Averaging—Stage 1 Forecasts

Periods	$AR_{1,0}$	$AR_{0,1}$	$AR_{1,1}$	$AR_{1,2}$	$AR_{2,1}$	$AR_{2,2}$	$AR_{2,0}$	$AR_{0,2}$
10	0.048	0.072	0.027	0.039	0.032	0.035	0.048	0.059
	(0.137)	(0.178)	(0.057)	(0.128)	(0.095)	(0.118)	(0.139)	(0.169)
15	0.044	0.086	0.033	0.029	0.032	0.037	0.064	0.053
	(0.109)	(0.177)	(0.057)	(0.080)	(0.062)	(0.114)	(0.149)	(0.126)
20	0.041	0.109	0.033	0.020	0.028	0.034	0.073	0.060
	(0.086)	(0.199)	(0.046)	(0.033)	(0.051)	(0.101)	(0.155)	(0.129)
25	0.047	0.124	0.038	0.021	0.031	0.038	0.073	0.064
	(0.093)	(0.207)	(0.050)	(0.038)	(0.044)	(0.099)	(0.142)	(0.134)
30	0.052	0.147	0.041	0.018	0.030	0.038	0.068	0.060
	(0.102)	(0.226)	(0.056)	(0.023)	(0.038)	(0.091)	(0.120)	(0.123)
35	0.047	0.146	0.045	0.021	0.032	0.036	0.064	0.064
	(0.081)	(0.214)	(0.058)	(0.026)	(0.035)	(0.085)	(0.109)	(0.121)
40	0.047	0.154	0.045	0.021	0.033	0.037	0.059	0.068
	(0.079)	(0.223)	(0.057)	(0.023)	(0.033)	(0.083)	(0.086)	(0.121)
45	0.046	0.163	0.050	0.021	0.034	0.035	0.057	0.071
	(0.075)	(0.225)	(0.067)	(0.021)	(0.034)	(0.077)	(0.080)	(0.121)
50	0.045	0.166	0.053	0.023	0.036	0.036	0.056	0.072
	(0.059)	(0.222)	(0.071)	(0.022)	(0.034)	(0.073)	(0.070)	(0.114)
55	0.048	0.159	0.057	0.027	0.040	0.040	0.062	0.075
	(0.054)	(0.216)	(0.068)	(0.022)	(0.033)	(0.068)	(0.070)	(0.110)
60	0.048	0.163	0.058	0.026	0.040	0.039	0.063	0.077
	(0.054)	(0.218)	(0.070)	(0.022)	(0.033)	(0.065)	(0.072)	(0.110)
EW	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: **Table continued on the next page**. This table shows the stage 1 forecasts. For these stage 1 forecasts we average over all specifications of the Linear model and the NARX neural network. The far lefts column shows the number of training lengths used, the middle columns show the weight applied to each model, and the far left column shows the OOS R^2 statistic. The standard deviations of the weights are in parentheses. $AR_{p,q}$ represents a specification of the linear model with p lags of the sentiment index and q lags of the autoregressive component. Similarly for $NARX_{p,q}$.

Table 3.5 Continued: Two-Stage Model Averaging—Stage 1 Forecasts

Periods	$NARX_{1,0}$	$NARX_{0,1}$	$NARX_{1,1}$	$NARX_{1,2}$	$NARX_{2,1}$	$NARX_{2,2}$	$NARX_{2,0}$	$NARX_{0,2}$	R^2 (%)
10	0.060	0.091	0.060	0.066	0.078	0.091	0.061	0.132	-1.923
	(0.165)	(0.212)	(0.143)	(0.169)	(0.174)	(0.193)	(0.143)	(0.248)	
15	0.052	0.087	0.066	0.057	0.081	0.104	0.064	0.111	-0.242
	(0.134)	(0.196)	(0.139)	(0.123)	(0.157)	(0.199)	(0.139)	(0.204)	
20	0.046	0.074	0.068	0.077	0.069	0.109	0.069	0.092	-0.027
	(0.096)	(0.148)	(0.124)	(0.152)	(0.128)	(0.202)	(0.146)	(0.174)	
25	0.040	0.062	0.062	0.089	0.061	0.107	0.066	0.076	0.354*
	(0.083)	(0.122)	(0.107)	(0.161)	(0.107)	(0.180)	(0.136)	(0.130)	
30	0.034	0.055	0.055	0.100	0.066	0.109	0.055	0.071	0.488*
	(0.069)	(0.099)	(0.080)	(0.172)	(0.106)	(0.169)	(0.110)	(0.117)	
35	0.030	0.063	0.052	0.089	0.076	0.106	0.048	0.080	0.281
	(0.049)	(0.114)	(0.068)	(0.154)	(0.120)	(0.156)	(0.091)	(0.133)	
40	0.029	0.069	0.052	0.082	0.078	0.098	0.040	0.087	0.599*
	(0.043)	(0.133)	(0.068)	(0.138)	(0.116)	(0.148)	(0.060)	(0.144)	
45	0.033	0.072	0.051	0.066	0.086	0.097	0.035	0.083	0.516*
	(0.058)	(0.138)	(0.067)	(0.112)	(0.127)	(0.152)	(0.049)	(0.144)	
50	0.039	0.070	0.053	0.058	0.082	0.090	0.034	0.087	0.193
	(0.061)	(0.131)	(0.054)	(0.094)	(0.117)	(0.133)	(0.045)	(0.150)	
55	0.043	0.067	0.056	0.052	0.077	0.080	0.037	0.082	0.293
	(0.058)	(0.122)	(0.050)	(0.080)	(0.096)	(0.104)	(0.039)	(0.138)	
60	0.040	0.067	0.054	0.047	0.078	0.081	0.036	0.083	0.067
	(0.057)	(0.119)	(0.043)	(0.065)	(0.100)	(0.102)	(0.035)	(0.144)	
EW	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	-0.063
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Table 3.6: Two-stage Model Averaging—Stage 2 Forecasts

Periods	RWD	Stage 1	R^2 (%)
10	0.524	0.476	1.833****
	(0.318)	(0.318)	
15	0.539	0.461	1.874****
	(0.286)	(0.286)	
20	0.552	0.448	2.096****
	(0.258)	(0.258)	
25	0.536	0.464	1.775****
	(0.229)	(0.229)	
30	0.530	0.470	1.901****
	(0.211)	(0.211)	
35	0.524	0.476	1.945****
	(0.189)	(0.189)	
40	0.519	0.481	1.806****
	(0.172)	(0.172)	
45	0.517	0.483	1.981****
	(0.147)	(0.147)	
50	0.531	0.469	1.891****
	(0.116)	(0.116)	
55	0.535	0.465	1.982****
	(0.103)	(0.103)	
60	0.532	0.468	1.988****
	(0.093)	(0.093)	
$EW_{1\text{ and }2}$	1/2	1/2	1.205****
	(0.000)	(0.000)	
$EW_{2\text{ only}}$	1/2	1/2	1.699****
	(0.000)	(0.000)	

Notes: The second stage of the 2SMA process. The far left column holds the length of the training window. The average posterior weight applied to each model is listed in columns two and three. The standard deviations of the weights are in parentheses. The far right column holds the OOS R^2 statistic. One, two, three and four asterisks represent a statistically significant difference in the loss-differential between the forecast and the benchmark random-walk plus drift at the 15, 10, 5 and 1 percent levels, respectively. The second-to-last row shows the results using using equal-weights in the first and the second stage. The last row shows the results using BMA in the first stage and equal-weights in the second stage.

Table 3.7: Average Forecasts over the Linear Model

Periods	$AR_{1,0}$	$AR_{0,1}$	$AR_{1,1}$	$AR_{1,2}$	$AR_{2,1}$	$AR_{2,2}$	$AR_{2,0}$	$AR_{0,2}$	R^2 (%)
10	0.132	0.234	0.076	0.088	0.084	0.099	0.154	0.132	-0.767
	(0.223)	(0.321)	(0.124)	(0.183)	(0.141)	(0.189)	(0.247)	(0.228)	
15	0.120	0.248	0.078	0.081	0.083	0.091	0.168	0.132	0.064
	(0.183)	(0.302)	(0.107)	(0.140)	(0.105)	(0.152)	(0.237)	(0.185)	
20	0.125	0.253	0.086	0.073	0.081	0.090	0.157	0.135	0.283
	(0.168)	(0.276)	(0.109)	(0.095)	(0.087)	(0.141)	(0.197)	(0.173)	
25	0.128	0.276	0.085	0.072	0.078	0.088	0.138	0.134	0.131
	(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	0.400
	(0.153)	(0.253)	(0.072)	(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	0.309
	(0.142)	(0.248)	(0.062)	(0.054)	(0.081)	(0.098)	(0.144)	(0.141)	
40	0.128	0.266	0.095	0.070	0.082	0.087	0.133	0.139	0.257
	(0.128)	(0.229)	(0.064)	(0.051)	(0.056)	(0.092)	(0.121)	(0.136)	
45	0.123	0.275	0.101	0.068	0.082	0.085	0.129	0.137	0.166
	(0.114)	(0.233)	(0.077)	(0.048)	(0.050)	(0.082)	(0.112)	(0.130)	
50	0.121	0.274	0.107	0.066	0.083	0.082	0.126	0.141	-0.225
	(0.104)	(0.230)	(0.084)	(0.045)	(0.048)	(0.074)	(0.106)	(0.132)	
55	0.120	0.261	0.114	0.069	0.086	0.084	0.127	0.139	-0.088
	(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	0.139	0.101
	(0.087)	(0.217)	(0.090)	(0.045)	(0.043)	(0.068)	(0.104)	(0.125)	
EW	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8	-0.453
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Notes: The average forecast results for all considered specifications of the linear model. For all but the last row, the first column shows the number of training periods. Columns two through eight contain the weights for the different specifications. In the first row, the autoregressive and sentiment lags are the first and second in parentheses, respectively. The far right column holds the OOS R^2 statistic. The OOS R^2 statistic is expressed as a percentage. One, two, three and four asterisks represent a statistically significant difference in the loss-differential between the forecast and the benchmark random-walk plus drift at the 15, 10, 5 and 1 percent levels, respectively. The last row shows the results using equal weights.

Table 3.8: Average forecasts over the NARX model.

Periods	$NARX_{1,0}$	$NARX_{0,1}$	$NARX_{1,1}$	$NARX_{1,2}$	$NARX_{2,1}$	$NARX_{2,2}$	$NARX_{2,0}$	$NARX_{0,2}$	R^2 (%)
10	0.109	0.141	0.108	0.095	0.124	0.138	0.088	0.198	-2.151
	(0.222)	(0.255)	(0.211)	(0.197)	(0.218)	(0.230)	(0.165)	(0.302)	
15	0.101	0.135	0.112	0.101	0.121	0.162	0.085	0.184	-0.043
	(0.208)	(0.226)	(0.181)	(0.174)	(0.185)	(0.234)	(0.149)	(0.269)	
20	0.096	0.127	0.117	0.126	0.111	0.183	0.097	0.144	0.299**
	(0.178)	(0.196)	(0.164)	(0.191)	(0.155)	(0.250)	(0.171)	(0.218)	
25	0.081	0.113	0.111	0.129	0.124	0.201	0.107	0.134	0.287*
	(0.147)	(0.172)	(0.142)	(0.195)	(0.167)	(0.238)	(0.178)	(0.188)	
30	0.077	0.107	0.105	0.143	0.128	0.215	0.103	0.123	-0.053
	(0.140)	(0.152)	(0.115)	(0.206)	(0.147)	(0.230)	(0.166)	(0.162)	
35	0.072	0.115	0.105	0.122	0.154	0.213	0.097	0.122	0.058
	(0.120)	(0.153)	(0.103)	(0.164)	(0.176)	(0.228)	(0.141)	(0.153)	
40	0.070	0.116	0.105	0.109	0.157	0.212	0.095	0.137	0.306*
	(0.092)	(0.161)	(0.095)	(0.132)	(0.168)	(0.229)	(0.128)	(0.163)	
45	0.079	0.120	0.105	0.097	0.157	0.216	0.091	0.135	0.253*
	(0.096)	(0.154)	(0.085)	(0.111)	(0.155)	(0.229)	(0.113)	(0.152)	
50	0.083	0.113	0.106	0.095	0.153	0.223	0.088	0.139	0.265*
	(0.089)	(0.148)	(0.077)	(0.110)	(0.154)	(0.222)	(0.105)	(0.162)	
55	0.085	0.112	0.110	0.096	0.150	0.231	0.088	0.128	0.189*
	(0.087)	(0.142)	(0.077)	(0.110)	(0.144)	(0.215)	(0.092)	(0.144)	
60	0.078	0.112	0.107	0.094	0.150	0.244	0.086	0.129	0.403**
	(0.085)	(0.134)	(0.070)	(0.103)	(0.143)	(0.225)	(0.084)	(0.156)	
EW	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8	-0.229
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Notes: The average forecast results for all considered specifications of the NARX neural network. For all but the last row, the far left column shows the number of training periods. Columns two through eight contain the weights for the different specifications. In the first row, the autoregressive and sentiment lags are the first and second in parentheses, respectively. The far right column holds the OOS R^2 statistic. The OOS R^2 statistic is expressed as a percentage. One, two, three and four asterisks represent a statistically significant difference in the loss-differential between the forecast and the benchmark random-walk plus drift at the 15, 10, 5 and 1 percent levels, respectively. The last row shows the results using equal weights.

Table 3.9: Stage 2 forecasts over the RWD, Linear and NARX models

Periods	RWD	Linear	NARX	R^2 (%)
10	0.420 (0.352)	0.296 (0.314)	0.284 (0.310)	2.018****
15	0.425 (0.323)	0.305 (0.259)	0.269 (0.261)	1.585****
20	0.439 (0.292)	0.300 (0.236)	0.261 (0.234)	1.365***
25	0.431 (0.248)	0.310 (0.211)	0.259 (0.194)	1.470***
30	0.412 (0.213)	0.326 (0.190)	0.262 (0.165)	1.506****
35	0.397 (0.188)	0.326 (0.171)	0.277 (0.166)	1.552****
40	0.397 (0.167)	0.320 (0.156)	0.283 (0.162)	1.441***
45	0.388 (0.152)	0.321 (0.149)	0.291 (0.160)	1.530***
50	0.380 (0.139)	0.325 (0.133)	0.296 (0.146)	1.573***
55	0.377 (0.128)	0.328 (0.122)	0.295 (0.126)	1.635****
60	0.376 (0.124)	0.327 (0.118)	0.297 (0.119)	1.609****
$EW_{1 \text{ and } 2}$	1/3 (0.000)	1/3 (0.000)	1/3 (0.000)	1.474****
$EW_{2 \text{ only}}$	1/3 (0.000)	1/3 (0.000)	1/3 (0.000)	1.910****

Notes: The second stage of the 2SMA process. The far left column holds the length of the training window. The average posterior weight applied to each model is listed in columns two through four. The standard deviations of the weights are in parentheses. The far right column holds the OOS R^2 statistic. One, two, three and four asterisks represent a statistically significant difference in the loss-differential between the forecast and the benchmark random-walk plus drift at the 15, 10, 5 and 1 percent levels, respectively. The last row shows the results using equal weights.

Chapter 4

The time-varying cross-sectional effects of investor sentiment

Abstract: We develop a new sentiment index for the stock market that accurately times sentiment bubbles and crashes. We compile this index from the closed end fund discount, NYSE share turnover, the dividend premium and the equity-share of new issues. To estimate the index we use a dynamic factor model with Bayesian techniques. In static regressions we find that rising sentiment relates to rising returns for portfolios based on size, dividends, earnings and momentum. We then use our index to study the time-varying effects of agent sentiment on stock returns. We find that the relationship between sentiment and returns has increased dramatically since the late 1990s. We also find a strong relationship between relationship between sentiment and future returns following bear markets.

Since the tech bubble burst in 2000 and the credit crisis in 2008 there has been a growing interest in stock market sentiment. Baker and Wurgler (2006, 2007) develop an index for agent sentiment in the stock market (BWsent hereafter). This index generally captures trends in stock market sentiment. Baker and Wurgler conclude that high sentiment relates to low future returns. Lutz (2011) notes, however, that BWsent 'mis-times' the conclusion of many bear markets. Furthermore, Baker and Wurgler's empirical findings contradict the theoretical predictions of Abreu and Brunnermeier (2003). Abreu and Brunnermeier (2003) develop a theoretical bubble model. They find that rational arbitrageurs have a profit incentive to ride sentiment bubbles. In other words, rising sentiment relates to rising future returns.

In this paper we develop a new sentiment index (Sent hereafter) using the data from Baker and Wurgler (2006, 2007). Baker and Wurgler (2006, 2007) use principal component analysis to develop their sentiment index. We use a dynamic factor model with Bayesian estimation. Sent and BWsent both capture the same general patterns in sentiment, but Sent more accurately times sentiment and bear market episodes. In in-sample predictive regressions we find that rising sentiment relates to rising returns for portfolios based on size, dividends, earnings, and momentum. These regression results match the theoretical predictions of Abreu and Brunnermeier (2003) and previous anecdotal accounts of investor sentiment. Brunnermeier (2009), Nagel and Brunnermeier (2004), and Temin and Voth (2004) consider various asset bubbles throughout history. They find that rational arbitrageurs have a profit incentive to ride the bubble and that rising sentiment relates to rising returns.

We also consider time-varying parameter regressions. In these models the beta on the sentiment index can change over time. We find that the relationship between sentiment and future returns has increased since 2000. This result corresponds to the anecdotal accounts in Shiller (2006). Furthermore, the betas increase along with sentiment and returns after bear markets. Yet the magnitudes of the betas drop during the 1990s tech bubble. Hence, sentiment-driven agents were not the only investors taking part in the boom. This finding matches Abreu and Brunnermeier's theoretical results. As noted above, Abreu and Brunnermeier find that rational arbitrageurs also participate in sentiment-driven bubbles.

Other papers have examined the relationship between sentiment and stock returns, but to our knowledge this is the first study to do so in a dynamic framework. Frazzini and Lamont (2008) use mutual fund flows to construct a sentiment indicator. They find that high sentiment relates to low future returns. Lemmon and Portniaguinia (2006) and Qui and Welch (2004) examine the relationship between consumer confidence and stock returns. Brown and Cliff (2004, 2005) use sentiment driven surveys to study return predictability.

4.1 Data

We develop a sentiment index using the data compiled in Baker and Wurgler (2006, 2007).[1] Baker and Wurgler use six components to compile their index. These components are the closed end fund discount (CEFD), NYSE share turnover (TURN), the number of IPOs over a given month (NIPO), the average first day return on these IPOs (RIPO), the equity-share of new issues (S), and the dividend premium (P^{D-ND}). CEFD is the the average difference between the net-asset-value (NAV) of closed-end stock funds and their market prices. TURN is the ratio of reported share volume to the average shares listed for the New York Stock Exchange. The equity-share of new issues (S) is gross equity issued divided by the gross equity plus gross long-term debt issued. The

[1] The data is available on Jeffrey Wurgler's website: `http://pages.stern.nyu.edu/~jwurgler/`

dividend premium (P^{D-ND}) is the log difference of the average market-to-book ratio of dividend and non-dividend paying stocks. Previous research found all of these measures to be proxies of agent sentiment in the the stock market. See Lee, Shleifer and Thaler (1991), Ibbotson, Sindelar and Ritter (1994), Baker and Wurgler (2000), Baker and Stein (2004) and Baker and Wurgler (2004) for more detailed analysis of these measures. We describe BWsent further in appendix 4.8. The data is monthly from 1965M07 to 2007M12. The RIPO series contains some missing observations. These missing observations correspond to months with no IPOs. RIPO is missing during the mid 1970s and after the tech crash in 2000. These episodes correspond to times of low sentiment. There is no clear way to deal with the missing data in RIPO. Data imputation may provide one solution. Imputation methods, however, typically assume that data is missing completely at random (MCAR) or missing at random (MAR). This is not the case with RIPO. RIPO is missing during times of low sentiment. Furthermore, in many months only a few IPOs occur. For example, in 99 out of 510 months there were five or fewer IPOs. Hence, RIPO draws from a small sample almost 20 percent of the time. In these cases, RIPO may reflect idiosyncrasies in the companies going public rather than the nature of the IPO market. For these reasons we choose to drop the RIPO variable. This still leaves five series that represent agent sentiment. From these five series we compile a sentiment index.

4.2 Model Estimation and description

We combine the information from the five series and extract a common component using a dynamic factor model. Baker and Wurgler (2006, 2007) note that TURN follows an exponential trend. Hence TURN is not stationary. To use the dynamic factor model we require that all the series be stationary.[2] Therefore, we take the first difference in TURN. In this case, Baker and Wurgler (2007) recommend to take the first difference of all of the index components. Table 4.1 shows the correlation matrix between the first differenced series. Unfortunately, the differenced series are largely uncorrelated. Below we discuss how the small correlation between the variables affects our results. Using the first difference in each series will lead to the first difference of the sentiment index. To derive a levels index we employ the method outlined in Stock and Watson (1991) and Kim and Nelson (1998, 1999). We take each series in deviation from from the mean. Hence, the model takes the following form:

$$\Delta y_{it} = \gamma_i \Delta c_t + \varepsilon_{it}, i = 1, \ldots, 5 \tag{4.1}$$

$$\phi(L)\Delta c_t = \omega_t, \omega_t \sim N(0, 1) \tag{4.2}$$

$$\psi_i(L)\varepsilon_{it} = v_{it}, v_{it} \sim N(0, \sigma^2) \tag{4.3}$$

Where $\Delta y_{it} = \Delta Y_{it} - \Delta \bar{Y}_i$, $i = 1, ..., 5$, represents one of the five series: P^{D-ND}, NIPO, TURN, CEFD, and S, respectively. $\Delta c_t = \Delta C_t - \delta$ 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, Δc_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 4.1 by $\psi_i(L)$ which yields

$$\psi_i(L)\Delta y_{it} = \gamma_i \psi_i(L)\Delta c_t + v_{it} \tag{4.4}$$

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 choose two lags for $\phi(L)$ and two lags for $\psi_i(L)$. Equations 4.2 and 4.3 become

[2] Bakera and Wurgler (2006) consider the natural log of TURN minus it five year moving average for annual data. Subtracting moving averages did not make TURN stationary for our monthly data

$$\Delta c_t = \phi_1 \Delta c_{t-1} + \phi_2 \Delta c_{t-2} + \omega_t \tag{4.5}$$

$$\varepsilon_{it} = \psi_{i1}\varepsilon_{i,t-1} + \psi_{i2}\varepsilon_{i,t-2} + v_{it}, i = 1, \ldots, 5 \tag{4.6}$$

Hence, we employ the following state-space form:

Measurement Equation:

$$\begin{bmatrix} \Delta y^*_{1t} \\ \Delta y^*_{2t} \\ \Delta y^*_{3t} \\ \Delta y^*_{4t} \\ \Delta y^*_{5t} \end{bmatrix} = \begin{bmatrix} \gamma_1 & -\gamma_1\psi_{11} & -\gamma_1\psi_{12} \\ \gamma_2 & -\gamma_2\psi_{21} & -\gamma_2\psi_{22} \\ \gamma_3 & -\gamma_3\psi_{31} & -\gamma_3\psi_{32} \\ \gamma_4 & -\gamma_4\psi_{41} & -\gamma_4\psi_{42} \\ \gamma_5 & -\gamma_5\psi_{51} & -\gamma_5\psi_{52} \end{bmatrix} \begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \\ v_{5t} \end{bmatrix} \tag{4.7}$$

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 4.4.

$$E(\mathbf{v}_t\mathbf{v}'_t) = R = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix} \tag{4.8}$$

Transition Equation:

$$\begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \end{bmatrix} + \begin{bmatrix} \omega_t \\ 0 \\ 0 \end{bmatrix} \tag{4.9}$$

Equation 4.15 can be written in matrix form as

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

$$E(\mathbf{e}_t\mathbf{e}_t') = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.10}$$

Estimating the above model will produce the first difference in sentiment. The first difference in sentiment is a noisy series. We want to compile sentiment index in levels to compare it to Baker and Wurgler (2006, 2007). To derive the common component representing the sentiment index in levels, we identify δ using the procedure outlined in Stock and Watson (1991). Stock and Watson (1991) notice that ΔC_t is a function of past lags of ΔY_t. From there they derive an estimate of δ by taking the expected value of ΔC_t. This method employs the steady-state Kalman gain to derive an estimate of δ. Once we have an estimate for δ we can compile C_t. C_t is the common component or sentiment index in levels.

We follow Lutz (2011) and estimate the model using the Bayesian multimove Gibbs-sampling approach based on Carter and Kohn (1994) and Kim and Nelson (1998). This methodology allows us just to jointly estimate the state vector and the model 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).

4.3 Empirical Results

Table 4.2 shows the results. γ_i, $i = 1, \ldots, 5$ represent the factor loadings for ΔP^{D-ND}, ΔNIPO, ΔTURN, ΔCEFD and ΔS, respectively. As noted in table 4.1, the correlation between the variables is very small. Hence, the factor loadings are quite small. Also, the variance on the error for ΔNIPO is very large. This is not surprising since the changes in ΔNIPO from one month to the next can be extreme. We do not make further alterations to the sentiment components to avoid data mining.[3] Despite these drawbacks, the common component representing the sentiment index captures trends in sentiment. Figure 4.2 shows our sentiment index (Sent). Baker and Wurgler (2006) document anecdotal accounts of agent sentiment. Sent generally coincides with their findings. For example, Sent peaks just prior to the bear market in the late 1960s and crashes through the bear market in the early 1970s. Furthermore, Sent reaches its nadir during the 1970s. Baker and Wurgler characterize the mid-1970s as a time of very low sentiment. The index pauses for bear markets as it climbs through the rest of the sample. Most notably, the index crashes prior to the bear markets in 1987 and 2000.

Figure 4.3 compares our sentiment index with Baker and Wurgler's sentiment index (BWsent). The overall pattern and trend of the sentiment indexes are the same. This validates our estimation methodology. Yet our sentiment index more accurately times sentiment episodes. Most notably, our index peaks just prior to the bear episodes in the late 1960s and early 2000s. BWsent continues to rise through these down markets. Lastly, employing the Kalman filter in the estimation procedure makes our sentiment index smoother than BWsent.

4.4 A Stationary Sentiment Index

The sentiment index, Sent, developed sections 4.2 and 4.3 is not stationary. Yet to analyze the relationship between sentiment and stock returns we need a stationary series. First, we test for breaks. We calculate the Quandt (1960) statistic and test the null hypothesis of no structural break based on the Andrews (1993) critical values.

[3] We also want our data to coincide as closely as possible to Baker and Wurgler (2006, 2007).

Hansen (1997) develops p-values based on the Andrews critical values.[4] We find no structural breaks in the series. Hence, we need to use a stationary series.

One benefit of the dynamic factor model is that it produces Δc_t as a byproduct. As noted above, we interpret Δc_t as the de-meaned first difference in Sent (ΔSent* hereafter).[5] ΔSent* is a stationary series. Hence, we use it to study the relationship between sentiment and stock returns in the remainder of the paper. We plot ΔSent* in figure 4.4. As expected, ΔSent* is most volatile around major episodes of investor sentiment. We use ΔSent* in the for the in-sample and time-varying regressions below.

[Insert figure 4.4 here.]

ΔSent* captures the changes in sentiment over the sample period: ΔSent* peaks prior at the of bubbles and crashes as bear markets take hold. Interestingly, ΔSent* falls dramatically just prior to the 1987 stock market crash. ΔSent* then rises through the mid-1990s before crashing prior to the 1997s Asian financial crisis. Lastly, ΔSent* crashes prior to the beginning of the bear market in 2000. ΔSent* then rises once the bear market concludes in 2003.

4.5 In-sample predictive regressions

We study the relationship between sentiment and stock returns for various types of stocks based on the following characteristics: (1) size (ME); (2) age; (3) volatility (Total Risk, σ); (4) the earnings-price ratio (Earn); (5) the dividend-price ratio (Div); (6) the book-to-market ratio (BE/ME); and (7) momentum (Mom).[6] Portfolios based on ME, Earn, Div, BE/ME, and Mom are from Kenneth French's Data Library. We compile

[4]The Quandt statistic is the maximum Chow (1960) statistic over the sample. We interpret the Quandt statistic as the best candidate for a breakpoint. See Hansen (2001) for an overview of these methods.

[5]We use ΔSent* rather than the first difference in Sent since ΔSent* comes directly from the model. Recall that the dynamic factor model derives Sent from ΔSent*. Hence, ΔSent* is a more parsimonious than the first difference in Sent. Using the first difference in Sent does not significantly affect our regression results below.

[6]These portfolios are similar to the portfolios studied in Baker and Wurgler (2006).

the portfolios based on age and volatility from the CRSP database using share codes 10 and 11. We count a stock's age from the first date it is listed in the CRSP. We then form decile breakpoints using the prior month's age. To compile the portfolios based on volatility, we calculate the prior 2-12 month standard deviation for each stock and then form the decile breakpoints. Through the return portfolios listed above, we examine the relationship between sentiment and the cross-section of stock returns.

We run in-sample regressions to determine the predictive power of sentiment on future returns in a model-based framework. We use the stationary ΔSent^* outlined in section 4.4

To conduct our regressions, we lag ΔSent^* by one period. We also control for the three factors, MKT, HML and SMB of Fama and French (1993) and a momentum factor. If one of these factors is used as the dependent variable we do not include it in our set of regressors. We construct long-short portfolios based on the various firm characteristics listed above. We build the long-short portfolios by comparing the high, medium and low deciles. The top three deciles represent the set of high deciles; the middle four deciles represent the set of medium deciles; and the bottom three deciles represent the set low deciles. Let the variable Z_t represent any of the long-short portfolios. The regression equation becomes:

$$Z_t = \alpha + \beta_1 \Delta\text{Sent}^*_{t-1} + \beta_2 MKT_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_t \qquad (4.11)$$

Table 4.3 shows the regression results. The left panel shows the results based on the raw ΔSENT^* series. In the right panel, we show the results after not controlling for the Fama-French and Momentum factors. The left most column holds the dependent variable used for the regression in equation 4.11. In each panel, the left column holds the beta on the sentiment index and the right column holds its bootstrapped p-value in parentheses.

[Insert table 4.3 here]

First, the results are similar for the left and the right panels. Thus, controlling for the Fama-French factors does significantly affect our results.

Panel A shows the effects of sentiment on size (ME), age, and risk. We compare small and large firms (SMB); young and old firms; and high volatility and low volatility stocks. All of the regression coefficients are positive, but the results are only significant when size (SMB) is the dependent variable. This implies that sentiment has a larger effect on small firms and that rising sentiment relates to rising returns.

In Panel B we study the effect of sentiment on earnings and dividend based portfolios. More specifically, we compare firms without earnings to those with positive earnings (Earn); and firms without dividends to those with positive dividends (Div). Clearly, sentiment has a larger effect on stocks without earnings or dividends. The coefficients are positive and significant at the one percent level; suggesting that rising sentiment relates to rising returns.

Panels C and D show portfolios based on the book-to-market ratio. In Panel C, we examine the effects of sentiment on HML. Stocks with a high book-to-market ratio may be undervalued or in distress; stocks with a low book-to-market ratio are often growth stocks. The regression coefficient in Panel C is negative, but not statistically significant. In Panel D, we focus on growth stocks. More specifically, we study the Medium-Low long-short portfolio based on the book-to-market ratio. Not surprisingly, the coefficient on the sentiment index is negative. This suggests that sentiment has a positive effect on growth stocks. The regression coefficient, however, is not statistically significant.

Panel E shows sentiment's effect on the momentum factor, UMD. The top momentum deciles are high momentum stocks; the bottom deciles are stocks in distress. In this case, the coefficient on the sentiment index is positive and significant. Hence, sentiment has a larger effect on high momentum stocks.

In Panel F we focus on high momentum stocks using the long-short portfolio High-Medium. In the left panel we control for the momentum factor, UMD. In this case the regression coefficient is positive but not statistically significant. In the left panel, we do not control for the momentum factor. The p-value then falls to 0.107.

Panel G focuses on stocks in distress. In particular, we study the High-Medium portfolio based on the book-to-market ratio and the Medium-Low portfolio for the momentum stocks. The regression coefficients are negative but not statistically significant.

4.6 Time-varying betas

In section 4.5 we discussed the relationship between sentiment and future returns. In this section we study the time-varying relationship between investor sentiment and stock returns.

Shiller (2006) posits that sentiment's effect on stock returns has increased since the early 1980s. Also, Gonzalez-Rivera (1997) finds evidence of time-varying betas in Arbitrage Pricing Theory models. Huang and Hueng (2008) consider a time-varying parameter beta in the capital asset pricing model (CAPM). They find that the beta risk-return relationship changes over time and depends the state of the market. Hence, time-varying betas appear to apply to stock returns.

We augment the regression in equation 4.11 to allow the beta on Sent to vary over time:

$$Z_t = \alpha + \beta_{1t}\Delta Sent^*_{t-1} + \beta_2 MKT\_RF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_t \quad (4.12)$$

$$\beta_{1t} = \beta_{1,t-1} + v_t, v_t \sim N(0, \sigma_1^2) \quad (4.13)$$

To our knowledge, there is no previous theory that describes the evolution of β_{1t}. Hence, we model β_{1t} as a random-walk process. The model specification in equations 4.12 and 4.13 allows for easy comparison to the regressions in section 4.5. To estimate the model we need to cast equations 4.12 and 4.13 into state-space form and apply the Kalman Filter. Equations 4.12 and 4.13 are nearly in state-space form. All we need to do is group the non-dynamic factors in a matrix. Hence we can write the measurement and transition equations:

Measurement Equation

$$Z_t = \beta_{1t} Sent_{t-1} + Az_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2) \tag{4.14}$$

Transition Equation

$$\beta_{1t} = \beta_{1,t-1} + v_t, v_t \sim N(0, \sigma_1^2) \tag{4.15}$$

Where $A = [\alpha\ \beta_2\ \beta_3\ \beta_4\ \beta_5]$ and $z_t = [1\ MKT\_RF_t\ SMB_t\ HML_t\ UMD_t]'$. Now we can estimate the model's parameters using maximum likelihood. We calculate the state vector, β_{1t}, using the Kalman filter.

If MKT, SMB, HML, or UMD is the regressand then we do not include it in the set of regressors. Figure 4.5 shows the results.

It is important to recall that the time-varying beta describes the relationship between sentiment and the given long-short portfolio over time. When the beta is positive, there is positive correlation between sentiment and the long portfolio and negative correlation between sentiment and the short portfolio; and vice versa. Larger magnitudes indicate a stronger relationship.

In general, the magnitudes of the time-varying betas increase over time. This is especially true after the bear market ended in 2003. This results corresponds to the findings of Shiller (2006). Shiller contends that the proliferation of the internet and low trading fees through online brokers have increased the effect of sentiment on the markets.

The relationship between ΔSent$*$ strengthens after most bear markets. Most notably, the magnitudes of the betas rise dramatically following the end of the bear market in 2003. This result is not surprising as the stocks most affected by sentiment generate their largest returns following bear markets. Furthermore, figure 4.4 shows that the change in sentiment spikes after bear markets.

The magnitudes of the betas shrink, however, during the 1990s tech bubble. During the 1990s, the sentiment index did not rise as fast as returns. Hence, sentiment-driven agents were not the only investors participating in stock market boom. This result corresponds to the theoretical model of Abreu and Brunnermeier (2003). Abreu and Brunnermeier find that rational arbitrageurs ride waves of investor sentiment.

Panel (a) shows the time-varying beta when SMB is the dependent variable. The beta is high during the first half of the sample. Hence, sentiment is strongly related to size effect of Benz (1981) through the early 1980s. The relationship between sentiment and SMB falls after the stock market crash in 1987 through the 1997 Asian financial crisis. The beta then rises after the bear market in 2003.

Panels (b), (c), (d), and (e) show the time-varying betas when age, volatility, Div, or Earn is the dependent variable. For all these cases, the beta increases over time and sharply rises after the end of the bear market in 2003. We see the similar results in panel (f) for HML. For HML, however, the time-varying beta is mostly negative as sentiment has a larger effect on growth stocks (low book-to-market ratio) than value stocks (high book-to-market ratio). We see a similar result in panel (g).

Panel (h) shows the time-varying result when the dependent variable is UMD. The beta switches sign over time. This is not surprising since sentiment affects both high momentum and low momentum stocks (Lutz 2011). When the beta is positive there is a stronger relationship between sentiment and high momentum stocks, and vice versa.

Panel (i) shows the results for high momentum stocks. Not surprisingly, the beta is positive for most of the sample. Hence, there is a positive relationship between sentiment and momentum. Moreover, the beta grows over time.

Panels (j) and (k) show the results for stocks in distress. The betas are positive except around the crisis periods in 1987, 1997 and 2000. This suggests a positive relationship between sentiment and distressed stocks in normal times, and a negative relationship during times of crisis.

4.7 Conclusion and further discussion

Using the closed-end fund discount, NYSE share turnover, the dividend premium, and the equity-share of new issues, we create an index for stock market sentiment. Through static regressions we find that rising sentiment relates to rising returns. This result corresponds to the theoretical predictions of Abreu and Brunnermeier (2003). We also employ time-varying parameter regressions. We find that the relationship between sentiment and future returns has increased dramatically since the mid-1990s. We also find that the relationship between sentiment and future returns strengthens following bear markets.

Future research may develop other measure for aggregating agent sentiment in the stock market. Using these measures researchers could study the relationship between sentiment and bear markets. Future research may also study sentiment in different industries.

4.8 Appendix: Baker and Wurgler's (2006, 2007) Sentiment Index

Data for Baker and Wurgler (2006, 2007) is available from Jeffrey Wurgler's website. The following list contains the components of Baker and Wurgler's index:

- The closed end fund discount (CEFD): The average difference between the net-asset-value (NAV) of closed-end stock funds and their market prices.
- NYSE share turnover (TURN): the ratio of reported share volume to the average shares listed for the New York Stock Exchange
- The number of IPOs over a given month (NIPO)
- The average first day return of IPOs for each month (RIPO)
- The equity share of new issues (S): The gross equity issued divided by the gross equity plus debt issued
- The dividend premium (P^{D-ND}): the log difference between the average book-to-market ratios of dividend and non-dividend paying stocks.

See Lee, Shleifer and Thaler (1991), Ibbotson, Sindelar and Ritter (1994), Baker and Wurgler (2000), Baker and Stein (2004) and Baker and Wurgler (2004) for more detailed analysis of these measures.

Baker and Wurgler use the first principal component as the factor loading on each variable. The following equation of shows the construction of BWsent at the monthly frequency as in Baker and Wurgler (2007):

$$\begin{aligned} BWsent = & -0.23CEFD + 0.23TURN + 0.24NIPO \\ & + 0.29RIPO + -0.32P^{D-ND} + 0.23S \end{aligned} \tag{4.16}$$

Baker and Wurgler (2006) develop BWsent at an annual frequency and similar results. Baker and Wurgler (2006) and Baker and Wurgler (2007) find a similar relationship between sentiment and stock prices; high sentiment relates to low returns.

4.9 Appendix: Tables

Table 4.1: Correlation of Sentiment Components

	ΔP^{D-ND}	ΔNIPO	ΔTURN	ΔCEFD	ΔS
ΔP^{D-ND}	1.000	-0.124	-0.060	-0.090	-0.030
ΔNIPO	-0.124	1.000	-0.049	0.038	0.192
ΔTURN	-0.060	-0.049	1.000	0.068	-0.013
ΔCEFD	-0.090	0.038	0.068	1.000	0.060
ΔS	-0.030	0.192	-0.013	0.060	1.000

Notes: The correlation of the first difference of the sentiment components. The data is described in section 4.1.

Table 4.2: Parameter estimates

Variable	Mean	Std	Med
ϕ_1	0.866	0.087	0.878
ϕ_2	0.010	0.045	0.012
γ_1	-0.001	0.087	-0.001
ψ_{11}	0.103	0.045	0.103
ψ_{12}	-0.032	0.045	-0.032
σ_1^2	10.551	0.674	10.499
γ_2	0.010	0.198	0.010
ψ_{21}	-0.330	0.047	-0.329
ψ_{22}	-0.205	0.043	-0.204
σ_2^2	151.160	9.331	150.865
γ_3	0.000	0.001	0.000
ψ_{31}	-0.504	0.048	-0.504
ψ_{32}	-0.366	0.043	-0.367
σ_3^2	0.008	0.000	0.008
γ_4	-0.001	0.040	0.001
$\psi_4 1$	-0.156	0.045	-0.154
$\psi_4 2$	-0.079	0.044	-0.079
σ_4^2	3.574	0.225	3.564
γ_5	0.000	0.001	0.000
ψ_{51}	-0.426	0.047	-0.424
ψ_{52}	-0.330	0.042	-0.330
σ_{25}	0.004	0.000	0.004

Notes: Parameter estimates for the model in section 4.2. $i = 1, \ldots, 5$ correspond to ΔP^{D-ND}, ΔNIPO, ΔTURN, ΔCEFD and ΔS, respectively.

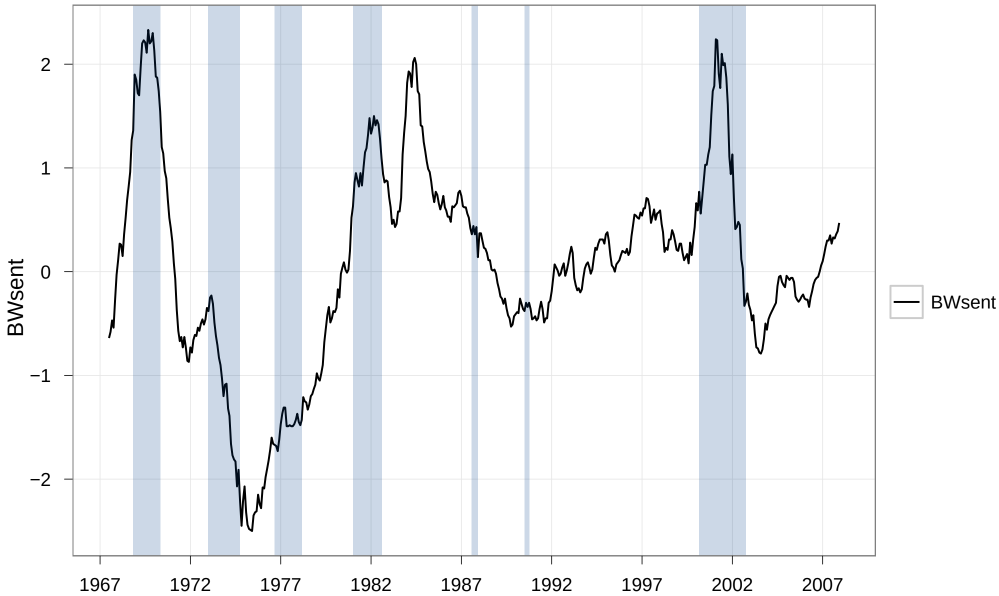
Table 4.3: Regresssions—Portfolio returns and sentiment

		$\Delta\text{Sent}^*_{t-1}$		$\Delta\text{Sent}^*_{t-1}$ No FF	
		$\boldsymbol{\beta}$	**p-val**	$\boldsymbol{\beta}$	**p-val**
		Panel A: Size, Age, and Risk			
ME	SMB	0.289	(0.021)	0.341	(0.007)
Age	Young-Old	0.009	(0.444)	0.118	(0.444)
σ	High-Low	0.444	(0.257)	0.444	(0.630)
		Panel B: Profitability and Dividends			
DIV	$=0->0$	0.315	(0.006)	0.654	(0.000)
EARN	$\leq 0->0$	0.253	(0.053)	0.554	(0.003)
		Panel C: Growth Opportunities and Distress			
BE/ME	HML	-0.016	(0.445)	-0.113	(0.197)
		Panel D: Growth Opportunities			
BE/ME	Medium-Low	-0.018	(0.376)	-0.018	(0.375)
		Panel E: Momentum and Distress			
Mom	UMD	0.445	(0.022)	0.197	(0.031)
		Panel F: Momentum			
Mom	High-Medium	0.420	(0.408)	0.382	(0.107)
		Panel G: Distress			
BE/ME	High-Medium	-0.009	(0.420)	-0.017	(0.382)
Mom	Medium-Low	-0.012	(0.446)	0.116	(0.290)

Notes: Regressions of long-short portfolios at time t on $\Delta\text{Sent}^*_{t-1}$. See equation 4.11. In the left panel, we control for the Fama-French factors and the momentum factor. If one of the the Fama-French factors or the momentum factor is the dependent variable, then we eliminate from the set of regressors. In the right panel, we do not control for the Fama-French Factors or the momentum factors. Portfolios are formed on size (ME), age, total risk (σ), the dividend-price ratio (Div), the earnings-price ratio (Earn), the book-to-market-ratio (BE/ME), or momentum (Mom). High represents the top three deciles, medium represents the middle four deciles, and low represents the bottom three deciles. The left panel shows the results using the original sentiment index. The left column of each panel lists the regression coefficient (β) on the sentiment index; the right column lists the bootstrapped p-value.

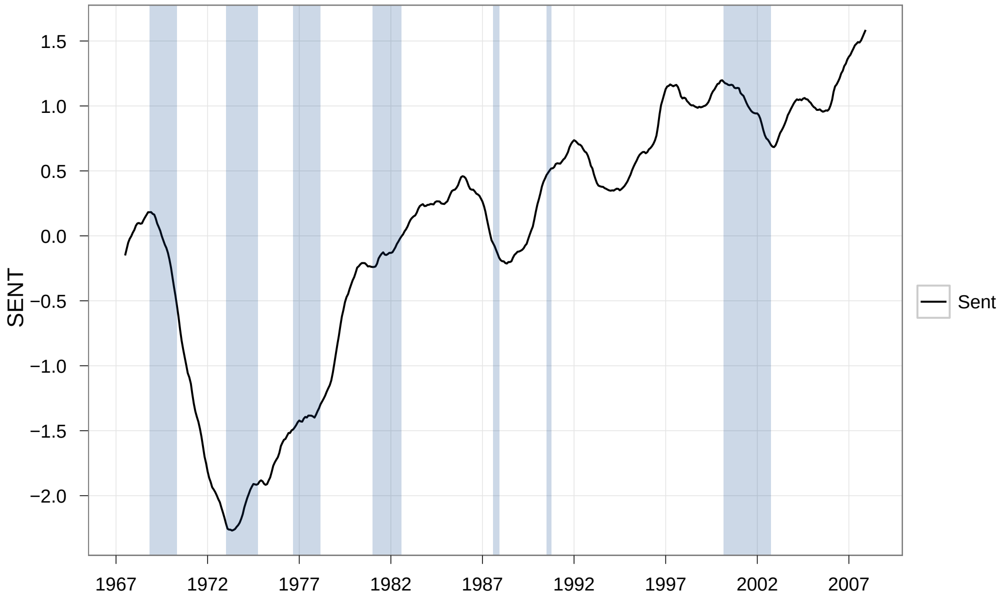
4.10 Appendix: Figures

Figure 4.1: Baker and Wurgler's Sentiment Index (BWsent)



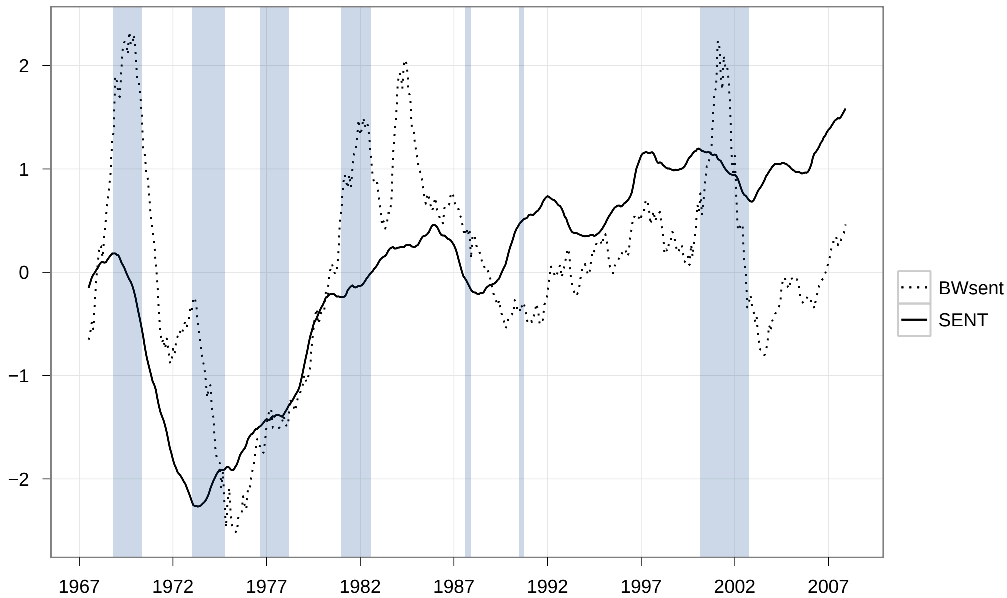
Notes: Shaded areas are bear markets defined by a 20 percent or more drop in the S&P 500 over a two or month period. Data is monthly. The horizontal axis holds the date.

Figure 4.2: Agent Sentiment (Sent)



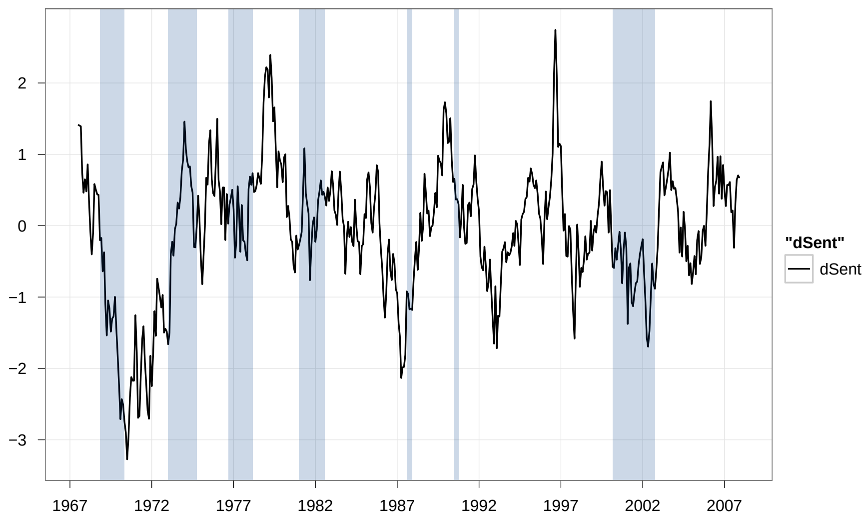
Notes: The sentiment index (Sent) in levels from the model described in section 4.2. Shaded areas are bear markets defined as a 20 percent or more drop over a two or more month period.

Figure 4.3: Sent and BWsent



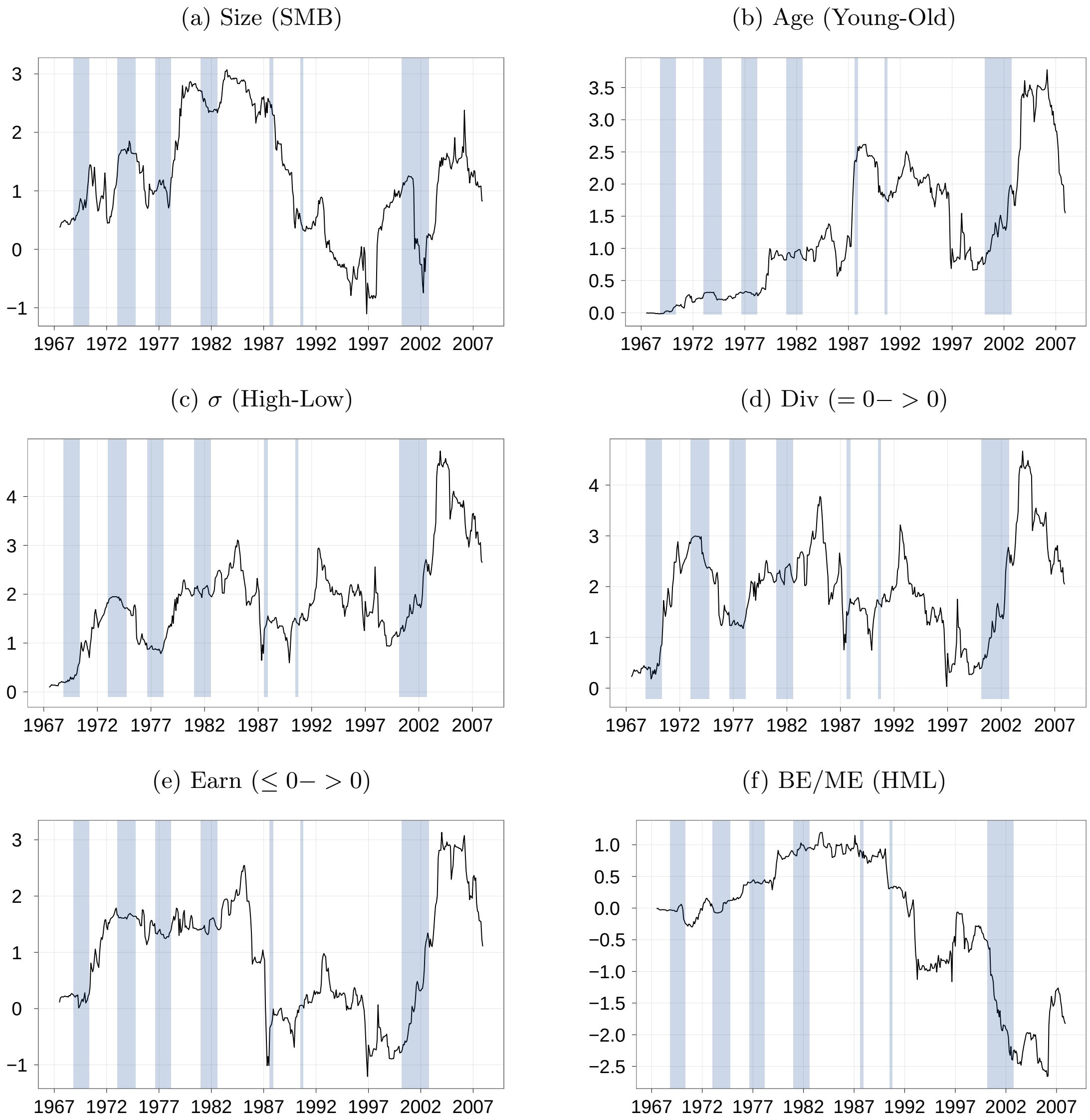
Notes: The sentiment index (Sent) and BWsent in levels from the model described in section 4.2. Shaded areas are bear markets defined as a 20 percent or more drop over a two or more month period.

Figure 4.4: ΔSent*



Notes: The graph of ΔSent*. We interpret ΔSent* as the first difference in sentiment. The vertical bars represent bear markets defined as a 20 percent of more drop in the S&P 500 over a two or more month period.

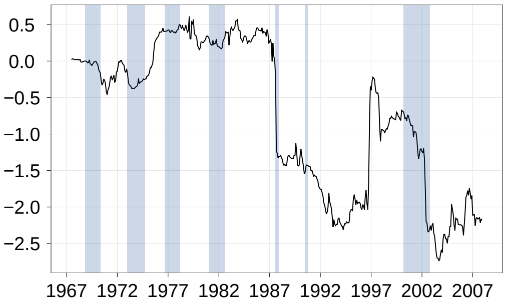
Figure 4.5: Time-Varying Betas



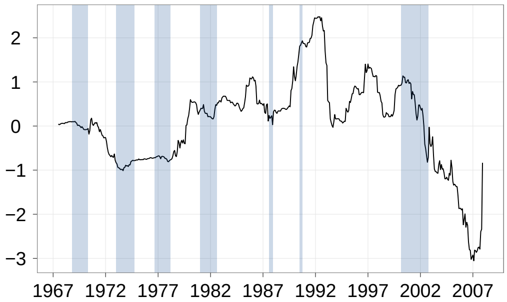
Notes: **Figure continued on next page.** These figures show the time-varying betas for different portfolio characteristics. See section 4.6.

Figure 4.5 Continued: Time-Varying Betas

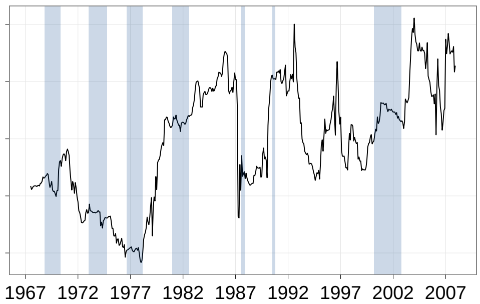
(g) BE/ME (Medium-Low)



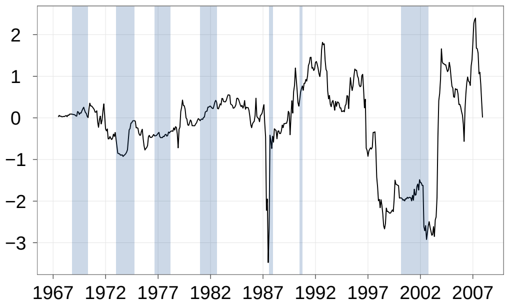
(h) Mom (UMD)



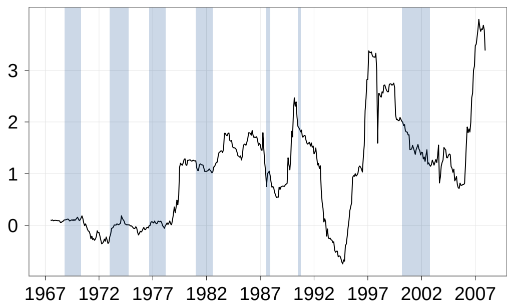
(i) Mom (High-Medium)



(j) BE/ME (High-Medium)



(k) Mom (Medium-Low)



Chapter 5

Conclusion

In this dissertation we develop two new indexes for agent sentiment in the stock market. Using both of these indexes we find that rising sentiment leads to rising returns; and vice versa. In chapter 2 we also find that rising sentiment leads to lower risk. Our results in chapter 2 are robust to the inclusion of macroeconomic indicators, risk factors and the Fama-French factors as additional controls. Furthermore, we find that cycles in sentiment growth often lead bear and bull markets.

In chapter 3 we forecast stock returns using investor sentiment. More specifically, we use the index outlined in chapter 2 for out-of-sample prediction. We also develop two-stage model averaging (2SMA). 2SMA is a flexible framework that allows researchers to incorporate prior economic information. We combine 2SMA with dynamic Bayesian model averaging to develop two-stage Bayesian model averaging (2SBMA). We also combine 2SMA and equal-weighted averaging to for two-stage equal-weighted averaging (2SEWA). The 2SBMA and 2SEWA forecasts statistically outperform the random-walk plus drift and their traditional model averaging counterparts. Our 2SMA results are noteworthy since forecasts using equal-weights are difficult to beat in practice (Timmermann 2006).

In chapter 4 we develop another new sentiment index using traditional sentiment indicators. These indicators include the closed-end fund discount, the equity-share of new issues and the number of IPOs in a given month. Using this index we find that rising sentiment leads to rising returns for stocks based on earnings, dividends, and momentum. We also find that the effect of sentiment has been increasing over time. Lastly, we show that sentiment has its largest effect on stocks just after bear markets.

Future research should study the relationship between sentiment and volatility. Also, the causal relationship between sentiment and bear markets is not clear.

Bibliography

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

[2] Marco Aiolfi and Allan Timmermann. Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1-2):31–53, 2006.

[3] Donald W.K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, 61(4):821–856, 1993.

[4] D. Avramov. Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3):423–458, 2002.

[5] Malcolm Baker and Jeremy C. Stein. Market liquidity as a sentiment indicator. *Journal of Financial Markets*, 7(3):271–299, 2004.

[6] Malcolm Baker and Jeffrey Wurgler. The equity share in new issues and aggregate stock returns. *The Journal of Finance*, 55(5):2219–2257, 2000.

[7] Malcolm Baker and Jeffrey Wurgler. A catering theory of dividends. *The Journal of Finance*, 59(3):1125–1165, 2004.

[8] Malcolm Baker and Jeffrey Wurgler. Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4):1645–1680, 2006.

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

[10] Rolf W. Banz. The relationship between return and market value of common stocks* 1. *Journal of Financial Economics*, 9(1):3–18, 1981.

[11] Marcus K. Brunnermeier. Deciphering the Liquidity and Credit Crunch 2007-2008. *Journal of Economic Perspectives*, 23(1):77–100, 2009.

[12] Marcus K. Brunnermeier and Stefan Nagel. Hedge Funds and the Technology Bubble. *The Journal of Finance*, 59(5):2013–2040, 2004.

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

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

[15] Gregrory C. Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, 28(3):591–605, 1960.

[16] T.E. Clark and K.D. West. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311, 2007.

[17] Robert T. Clemen. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583, 1989.

[18] KJ Cremers. Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies*, 15(4):1223, 2002.

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

[20] F.X. Diebold and R.S. Mariano. Comparing predictive accuracy. *Journal of business and economic statistics*, 20(1):134–144, 2002.

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

[22] F. Dan Foresee and Martin T. Hagan. Gauss-Newton approximation to Bayesian learning. In *Proceedings of the 1997 International Joint Conference on Neural Networks*, volume 3, pages 1930–1935. Piscataway: IEEE, 1997.

[23] Andrea Frazzini and Owen A. Lamont. Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics*, 88(2):299–322, 2008.

[24] Denys Glushkov. Sentiment beta. *Working Paper*, 2006.

[25] Gloria Gonzalez-Rivera. The pricing of time-varying beta. *Empirical Economics*, 22(3):345–363, 1997.

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

[27] Bruce E. Hansen. Approximate asymptotic p values for structural-change tests. *Journal of Business & Economic Statistics*, 15(1):60–67, 1997.

[28] Bruce E. Hansen. The new econometrics of structural change: Dating breaks in US labor productivity. *Journal of Economic perspectives*, 15(4):117–128, 2001.

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

[30] Peng Huang and C. James Hueng. Conditional risk–return relationship in a time-varying beta model. *Quantitative Finance*, 8(4):381–390, 2008.

[31] Roger G. Ibbotson, Jody Sindelar, and Jay R. Ritter. The market's problems with the pricing of initial public offerings. *Journal of Applied Corporate Finance*, 7:66–74, 1994.

[32] Chang-Jin Kim and Charles R. Nelson. 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, 1998.

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

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

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

[36] Chandler J. Lutz. Market Based Investor Sentiment and the Cross-Correlation of Stock Returns. *Working Papers*, 2010a.

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

[38] Lily X. Qiu and Ivo Welch. Investor sentiment measures. *Working Paper*, 2004.

[39] Richard E. Quandt. Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American statistical Association*, 55(290):324–330, 1960.

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

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

[42] Robert J. Shiller. *Irrational exuberance*. Broadway Business, 2006.

[43] James H. Stock and Mark W. Watson. A probability model of the coincident economic indicators. *Leading Economic indicators: new approaches and forecasting records*, 66, 1991.

[44] Peter Temin and Hans-Joachim Voth. Riding the south sea bubble. *The American Economic Review*, 94(5):1654–1668, 2004.

[45] Allan Timmermann. Forecast combinations. *Handbook of economic forecasting*, 1:135–196, 2006.

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