The Long-run Impact of Media Sentiment on Stock Returns∗

Matthias W. Uhl†

January 2011

Abstract

We examine the impact of fundamental and behavioral factors on stock returns in light of the Efficient Market Hypothesis. With two novel sentiment datasets, we find positive and significant correlations between fundamental macroeconomic factors, media sentiment (a combination of sentiment from Reuters and TV news) and stock returns. We show with vector error correction models that media sentiment causes persistent changes in stock returns over months, and that there is a constant and longer-lasting overreaction in stock prices to media sentiment, confirming earlier studies. Reuters sentiment is the more persistent behavioral variable to forecast stock returns than TV sentiment.

Keywords: Behavioral Issues, Equities, Market Efficiency and Anomalies

EFM classifications: 320, 330, 350

∗Thanks go to Thomson Reuters for the sentiment dataset and to MediaTenor for providing the TV sentiment data. I am grateful to Didier Sornette and Jan-Egbert Sturm for helpful comments and suggestions. Thanks go to participants of the internal research seminar at MAN Investments for their feedback. I thank Edward Fridel for his IT support.

†KOF Swiss Economic Institute, ETH Zurich, 8092 Zurich, Switzerland, Tel. +41-44-632-2553, Fax +41-44-632-1218, E-mail: uhl@kof.ethz.ch
The Efficient Market Hypothesis (EMH), first introduced by Fama (1970), has been questioned widely on the grounds of psychological phenomena occurring in financial markets. Financial economists and psychologists alike have devoted time to research that relates sentiment among investors to financial market returns. Kahneman and Tversky (1981) find that subjects overreact to new information in making probabilistic judgments. Based on the same grounds, Shiller (1981) notes that financial markets display excess volatility and overreaction to new information. Summers (1986) then posed the question whether the stock market rationally reflects fundamental values and came to the conclusion that most tests of market efficiency have had little power to solidify the EMH, suggesting that excess volatility and negative autocorrelation can produce a deviation of the price in a rational fundamental market. Further, he elaborates, certain types of inefficiency in market valuations are not likely to be detected using standard methods. Thus, one should not make the erroneous conclusion that market prices represent rational assessments of fundamental valuations based on the grounds that many studies have found that the EMH cannot be rejected.

In this paper, we want to test whether media sentiment\(^1\) has a persistent effect on stock returns, and whether TV and Reuters sentiment, given their different nature of how they are generated, form a complementary basis to explain changes in stock prices. With Reuters and TV sentiment, we mean a (positive or negative) feeling, opinion, or emotion evoked among a reader/TV watcher while reading/watching a certain Reuters news article/TV news show. We further build simple trading strategies with in- and out-of-sample forecasts to test the predictive accuracy of the models.

Section I gives an overview of the existing literature and lays out the motivation. Section II describes the dataset, while section III discusses the econometric modeling approach and the empirical results of the specified models, and section IV the formulation of a simple trading strategy based on in- and out-of-sample forecasting models. Section V concludes.

I. Related Literature and Motivation

Since the late 1980s, when the first studies emerged that postulated irrationality in financial markets, the domain of behavioral finance has introduced ways to explain that irrationality. One of the first studies that attempted to link other

\(^1\)We use the expression of media sentiment for the combination of Reuters and TV sentiment.
exogenous variables to financial market returns was undertaken by De Bondt and Thaler (1985). They show that, based on research in experimental psychology, overreaction occurs mainly when unexpected and dramatic news events happen. A few years later, Cutler et al (1989) were one of the first to identify a link between news coverage and stock prices. Since then, studies have evolved that look at the potential influence that the media has on investor behavior. For example, Klibanoﬀ et al (1998) show that country-speciﬁc news reported on the front page of the New York Times affect the pricing of closed-end country funds. Huberman and Regev (2001) ﬁnd that an article in the Financial Times on a biochemical ﬁrm made prices of that company soar. Antweiler and Frank (2004) consider the inﬂuence of Internet stock message boards. They ﬁnd that stock messages predict market volatility, but their effect on returns is small. In their extensive study on the news media, Mullainathan and Shleifer (2005) identify that there are biases in economic and political news and that these are slanted towards the customers of the media outlet. Mescike and Kim (2011) consider reports on TV. In their recent study, they investigate CEO interviews on CNBC, while documenting signiﬁcant positive abnormal returns accompanied by abnormally high trading volume from two days prior to the event until the day of the event. After the event, there is a negative abnormal return in the following ten trading days. They ﬁnd evidence that enthusiastic individual investors are prone to trading more based on CNBC interviews. Rational utility maximizing investors, they note, take advantage of the regular pricing pattern related to media attention.

Baker and Wurgler (2007) argue that the key nowadays for researchers is to ﬁnd out how to measure investor sentiment and quantify its effects. Owing to the quest for more accuracy in explaining ﬁnancial market returns from a behavioral point of view, studies have been aiming towards the quantiﬁcation of sentiment recently. Baker and Wurgler (2007) describe various possible proxies to measure investor sentiment.\footnote{Such as, e.g., retail investor trades, mutual fund ﬂows, trading volume, dividend premium, closed-end fund discount, option implied volatility, IPO ﬁrst-day returns, IPO volume, equity issues over total new issues, and insider trading.} Although they only use six proxies of the identiﬁed sentiment measures due to data constraints, and, possibly measurement constraints, the obvious question prevails whether these proxies show or drive investor sentiment, or, put differently, whether these measures are rather
outcomes or results of sentiment. In this paper, we do not want to focus on the outcomes of sentiment, we want to consider the input of sentiment, i.e. what drives and shapes sentiment. Thus, we introduce and test new variables that form and influence investor sentiment, which we can measure quantitatively in a systematic way, while trying to avoid subjectivity bias. With the growing importance of the media in the past decades, the obvious publicly available information are news, as De Bondt and Thaler (1985) as well as Cutler et al (1989) noted as early as a few decades ago. Based on these initial findings, we focus on news that relate to the economy and financial markets and are brought to the investor through various channels, such as TV news and Reuters news reports.

The growing evidence in the finance literature about sentiment affecting investors and thus stock returns is key motivator for this study. DeLong et al (1990) are among the first to find that investors are subject to sentiment. In their model, two sets of traders exist: professional arbitrageurs and unsophisticated traders, i.e. noise traders. The prevailing risk in the market, they find, is created by the unpredictability of the noise traders. Professional arbitrageurs respond to the behavior of noise traders rather than acting on fundamentals. In doing so, professional arbitrageurs consider pseudo signals such as volume and price patterns, but also sentiment indices. Barberis et al (1998) show that news can cause both over- and underreaction to stock prices by formulating a parsimonious model of investor sentiment. They claim that news are incorporated only slowly into stock prices, making the case for a lower frequency, i.e. monthly, analysis. Chan (2003) finds evidence of a post-news drift. Investors, he claims, underreact to new information, so that there is a persistent effect of Reuters sentiment on asset prices, which seems strongest after bad news. In a journalistic study, Maier (2005) notes that 61% errors in local news and feature stories in the US, while subjective errors are considered most severe. Maier’s results suggest that how a story is conveyed is at least as important as getting the facts straight. The results of these studies strongly speak for examining media reports, such as TV news shows and Reuters news, for sentiment, and using the sentiment values to explain changes in stock prices.

With his study, Tetlock (2007) is among the first to measure quantitatively the interactions between the media and the stock market using daily content from a Wall Street Journal column. High media pessimism, he finds, predicts falling stock market prices followed by a reversion to fundamentals. Unusually high or low pessimism predicts high trading volume as well. In a follow-up to
Tetlock’s (2007) study, Tetlock et al. (2008) use a simple quantitative measure of language to predict individual firms’ accounting earnings and stock returns. Linguistic media content, they conclude, captures aspects of firms’ fundamentals that are otherwise hard to quantify, which are quickly incorporated into stock prices. Fang and Peres (2009) investigate the cross-sectional relation between media coverage and expected stock returns. They find that stocks with no media coverage earn higher returns than stocks with high media coverage even after controlling for well-known risk factors. Their results are more pronounced among small stocks and stocks with high individual ownership, low analyst following as well as high idiosyncratic volatility. Given their findings, this suggests that the breadth of information dissemination affects stock returns. On a similar note, Livnat and Petrovits (2009) examine whether stock price reactions to earnings surprises and accruals vary systematically with the level of investor sentiment. By formulating a monthly trading strategy, they find evidence that holding extreme good news firms following pessimistic sentiment periods earns significantly higher abnormal returns than holding extreme good news firms following optimistic sentiment periods. These results indicate that investor sentiment influences the source of excess returns from earnings-based trading strategies. In his recent study, Tetlock (2011) tests whether investors distinguish between old and new information about firms, or, what he calls the “staleness of news.” A firm’s return on the day of stale news negatively predicts its return in the following week, which speaks for the fact that individual investors overreact to stale information, leading to temporary movements in firms’ stock prices.

As Baker and Wurgler (2007) point out, it is no longer questionable whether sentiment affects investors and thus stock returns, but rather how to measure sentiment. Many studies have emerged in the past years attempting to tackle the issue of defining sentiment that influences stock markets and, more importantly, measuring it.\(^3\) As previously mentioned, we measure sentiment in Reuters news and TV news shows with the aid of novel techniques. We follow Tetlock’s (2007) methodological approach of measuring sentiment in the media quantitatively. Tetlock uses the General Inquirer (GI), a quantitative content analysis program.\(^4\) As explained in the appendix in Tetlock (2007), the GI has

\(^3\) See, for example, Cao and Wei (2005), Edmans et al. (2007), Hirshleifer (2001), Hirshleifer and Shumway (2003), Kamstra et al. (2003), and Yuan et al. (2006), among others.

one major shortcoming: it is only able to distinguish between positive and negative words, or sentiment categories, but not between contexts. The datasets used in this study correct for this shortcoming and they account for various media sources, such as TV and Reuters news.

II. Dataset

As opposed to Tetlock's (2007) dataset, we are able to account for both positive and negative words, but also to conduct a contextual analysis. Owing to new technological advance in text mining, Thomson Reuters is able to undertake a sentiment analysis that incorporates the contextual sentiment framework to achieve the highest possible scoring accuracy. Based on these findings, we introduce the concept of measuring media sentiment in Reuters news articles and TV news quantitatively that might affect investor sentiment and eventually financial markets. Every Reuters news article/statement of a TV news show is coded as positive \{1\}, neutral \{0\}, or negative \{-1\} sentiment. This concept is fairly new and has only existed for a few years, among the few studies that deal with measuring news sentiment, the technical approaches to measuring news sentiment differ greatly. In the past, most solutions have come from the text mining industry that caters to the financial markets industry, in which news texts can be scanned in great quantities and a short amount of time for sentiment with specific, black-box like, sentiment algorithms. Thomson Reuters is one of the few providers of sentiment classified news.\(^5\) The first dataset at hand consists of high-frequency (tick data) sentiment rated Thomson Reuters news pieces, classified from a wide list of topics for the US market.\(^6\) For this study, we filter all Reuters news items for sentiment from the Equities topic codes section. We then aggregate the sentiment scores to monthly values. Thomson Reuters is able to account for stale news to avoid the issue of staleness as in Tetlock (2011) because the sentiment algorithm is able to tag and distinguish between new and old news items.

The second sentiment dataset that is implemented in this study is from


MediaTenor, a sentiment provider. The two differences between this dataset and the dataset from Thomson Reuters are, first, that the sentiment data were compiled exclusively from US TV news shows, and, second, that the sentiment was coded by humans. Tagged topics are classified in detail and contain possible links to the development of the economy as a whole and the financial markets in particular.\(^7\) Since the dataset is coded by humans, it has the advantage that it is obviously natural for humans to set pre-classified isolated positive and negative words into a larger contextual framework to code the sentiment correctly, while, at the same time, adhering to a strict pre-defined set of classification rules. Important to note is that employees from MediaTenor are trained to adhere to a very specific pre-defined sentiment rating and coding table, to avoid subjectivity bias. Table I shows the number of news pieces that were tagged in both datasets; in total, over 10,000 TV news shows and over 4,200,000 Reuters news items were coded for sentiment from January 2005 to December 2009.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Month & Number of News Pieces \\
\hline
January 2005 & 10,000 \\
December 2009 & 4,200,000 \\
\hline
\end{tabular}
\caption{Number of News Pieces Tagged in Both Datasets}
\end{table}

Monthly price return data for the Dow Jones Industrials stock index were obtained from Thomson Reuters Datastream. The corresponding monthly volume data for the Dow Jones stock index are from MasterData.\(^8\) To capture “hard economic facts,” we use a time series of the Conference Board Leading Economic Indicators Index. This index consists of a combination of leading indices, such as production, employment, monetary, and consumer data.\(^9\) The advantage over using many different indicators is that one variable is easier to handle in our subsequent model than multiple variables. Given that we attempt to explain stock returns with non-conventional measures - inconsistent with the EMH - such as sentiment, we need to include fundamental facts that are consistent with the EMH to capture all possible channels of influence on the stock index, and to compare the fundamental to the non-conventional. The Conference Board Leading Economic Indicators Index appears the most suited for “summarizing” macroeconomic factors in one variable. Monthly data for this indicator were obtained from Thomson Reuters Datastream.

To get a first understanding of the data, we look at the variables graphically...

---

\(^7\)As in the Thomson Reuters dataset, the topics range from financial market to economic and political news that might have an influence on equities markets.

\(^8\)See www.masterdatacsv.com, last accessed 15 October 2010.

in fig. 1. As we will use logged values of the stock index and volume in our model later, we show logged values of these two variables in the graphs, too. For the other variables, we show level data. Later, we use growth rates to account for non-stationarity. The Dow Jones stock index shows a pattern, in which we can make out the bull market from 2005 to 2008 and the subsequent crash when the financial crisis hit global capital markets in 2008. As of March 2009, prices have recovered until the end of the period examined. The volume charts show more or less an inverse pattern to stock prices. This might suggest a negative correlation between stock prices and volume. The TV sentiment graph shows the peak of the bull market in 2006/2007 nicely, with a low at the end of 2008. However, the peaks and troughs are not as clear as in the Reuters sentiment graph, in which we can see a sharp decline in sentiment during the recent financial crisis. Tetlock (2007) finds that a high level of pessimism in the media predicts falling market prices. The Reuters sentiment graph shows that the stock indices follow Reuters sentiment with a certain lag. Most prominently, the trough in Reuters sentiment occurs in December 2008/January 2009, whereas the stock markets bottomed in March 2009. TV sentiment does not appear to be such a clear-cut indicator graphically. The Conference Board Index shows a simultaneous co-movement with the Dow Jones Industrials index. We thus undertake further empirical tests to find out whether a combination of fundamental data, i.e. the Conference Board Index, and behavioral data, i.e. TV and Reuters sentiment, can explain changes in stock prices.

We test all variables for unit roots with the Augmented Dickey-Fuller test according to Dickey and Fuller (1979) and find that the stock index and volume variables, the Conference Board Index and the Reuters sentiment variable have unit roots in the level. To make the data more stationary, we difference and logarithmize Dow Jones stock prices and volume, while we use differenced values for TV and Reuters sentiment as well as for the Conference Board Index. We test the variables further by applying Tetlock's (2007) approach to test whether the variables can explain stock returns.

Figs. 2 to 5 show cross-correlograms of the logged Dow Jones stock index returns, logged differences of stock index volume, as well as differenced values of the Conference Board Index, the TV and Reuters sentiment variables. As graphically anticipated, stock index volume has a negative correlation with the Dow Jones Industrials stock index at most lags. The Conference Board Index

\footnote{See Appendix A.1 for a detailed description of the Augmented Dickey-Fuller Test.}
has a strong correlation with Dow Jones stock returns, greatest at lags zero and one. TV sentiment shows a slight positive correlation with stock returns, which is greatest at a lag of one. The Reuters sentiment variable is positively correlated with stock prices, with the highest correlation at lag 3. Also, Reuters sentiment has the highest correlation with stock returns among all variables, which seems promising for further empirical analyses. Given that we have two sentiment variables at hand, we want to test for possible multicollinearity as well because - although we use two different sources, i.e. TV news shows and Reuters news items - we attempt to measure the same thing: sentiment. Fig. 6 shows a cross-correlogram of Reuters and TV sentiment. For lags zero and one, the correlation appears the greatest, but it is still well below 0.5. For other lags, the correlation converges to zero, so that we can assume that multicollinearity is not an issue between the two sentiment variables. In fig. 7, we construct a scatterplot, which does not indicate multicollinearity either. This might suggest that there is a different set of information in each sentiment variable, which we want to test further in the succeeding sections.

III. Modeling

We test the data more closely by constructing a Vector Autoregression Model (VAR) to tackle possible endogeneity issues. Since we have unit roots in most of the variables, we test for cointegration according to Johansen (1991) first. We find one cointegrating relation, which is most likely between the Conference Board Index and the Dow Jones Industrials Stock Index. Thus, we formulate a Vector Error Correction Model (VECM) according to the reduced rank (RR) estimation procedure as in Johansen (1995) to account for nonstationarity and cointegration in the data as follows:

$$\Delta y_t = \alpha \beta^* [y_{t-1} - D_{t-1}^{co}] + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_p \Delta y_{t-p} + CD_t + u_t,$$ (1)

where $y_t$ refers to the endogenous variables, which are logged values of the Dow Jones Industrials stock price index, Reuters sentiment, TV sentiment, logged Dow Jones stock index volume, and values of the Conference Board index, $D_t$ refers to the deterministic term (here: a constant), $D_{t-1}^{co}$ is the cointegrating relation, $u_t$ is the error term, and the cointegration matrix $\beta^*$ is normalized as followed:
\[
\beta^* = \begin{bmatrix} I_r \\ \beta^*_{(K^*-r)} \end{bmatrix},
\]

where \(\beta^*_{(K^*-r)}\) is a \(((K^* - r) \times r)\) matrix. To find an optimal lag structure of the model, we perform various lag length selection tests, as shown in table II. The results from the Akaike Information Criterion and Final Prediction Error show the most plausible results with lags of three, whereas the Hannan-Quinn and the Schwarz Criterion show an optimal lag length of 0. Given our graphical interpretation as well as the results from the cross-correlograms, which show that sentiment as well as the Conference Board index have leading characteristics over stock returns, it appears more suited to use a lag structure allowing up to three lags. Additionally, Reuters sentiment shows the highest correlation with stock returns in the cross-correlogram in fig. 5.

[insert table II about here]

We empirically test the above model to obtain further clues whether TV and Reuters sentiment as well as other variables have an influence on stock returns. Table III shows the results of the VECM estimation, allowing for up to three lags, as specified with the lag selection tests.

[insert table III about here]

To analyze the dynamic interactions between the endogenous variables of the VEC process, we draw on the impulse response analysis so that we can analyze the dynamic interactions between the endogenous variables of a VEC(p) process.\(^{11}\) Owing to the Cholesky ordering issue in impulse response analyses, a structural vector error correction (SVEC) analysis appears suited in this study for considering impulse response functions, so that we do not have to control for the ordering of the variables in the endogenous vector \(y_t\).

The SVEC model can be used to identify the shocks to be traced in an impulse response analysis by imposing restrictions on the matrix of long-run effects of shocks and the matrix \(B\) of contemporaneous effects of the shocks.

\(^{11}\) See Appendix A.2 for a detailed discussion of Impulse Responses in VEC(p) processes, and the case for a structural vector error correction (SVEC) model.
The matrix $B$ is defined such that $u_t = B\varepsilon_t$ in (1) and the matrix $\Xi$ of long-run effects of the $u_t$ residuals is

$$
\Xi = \beta_\perp \left( \alpha_\perp' \left( I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right)^{-1} \alpha_\perp'.
$$

(3)

Hence, the long-run effects of $\varepsilon$ shocks are given by $\Xi B$. $\text{rk}(\Xi) = K - r$ and, hence, $\Xi B$ has rank $K - r$. Thus, the matrix $\Xi B$ can have at most $r$ columns of zeros. Therefore, there can be at most $r$ shocks with transitory effects (zero long-run impact) and at least $k^* = K - r$ shocks have permanent effects.

Fig. 8 shows the results of the impulse response functions based on the SVEC. We focus on the first row of the impulse response graphs because we want to identify possible impacts of sentiment, volume, and economic factors on stock returns. The graphs show a persistent effect of the Conference Board Index as well as Reuters sentiment on stock returns, while TV sentiment and stock index volume do not have a significant impact on the Dow Jones Industrials stock index returns.

Hong and Stein (1999) make similar findings. They show that prices underreact in the short run, suggesting that this should ultimately lead to overreaction in the long run. In this study, we consider the longer term with our monthly data analysis, in which we find an overreaction to sentiment. In a recent study, Livnat and Petrovits (2009) account for a post-earnings announcement drift among investor sentiment. They find evidence that holding extreme good news firms following pessimistic sentiment periods earns significantly higher abnormal returns than holding extreme good news firms following optimistic sentiment periods. Similarly, they show that holding low accrual firms following pessimistic sentiment periods earns significantly higher abnormal returns than holding low accrual firms following optimistic sentiment periods. Chan (2003) also finds evidence of a post-news drift. This is in line with our findings, as we experience a longer lasting drift in Reuters sentiment that is persistent over months. We do not find this effect in TV sentiment.

We further test how much impact each variable has on stock returns in relation to another. To do this, we draw on the forecast error variance decomposition (FEVD).\textsuperscript{12} The FEVD of the Dow Jones Stock index returns is depicted in fig. 9. Similar to the findings from the impulse responses, the impact of the economic factors, in the form of the Conference Board Index, makes up the

\textsuperscript{12}See Appendix A.3 for a more detailed explanation of the FEVD.
biggest share of the variance of the forecast error of stock returns, increasing in
time and its impulse response. The second largest share has Reuters sentiment,
which is greatest at lag three, making up of around 30% of the variance of the
forecast error of stock returns. Both TV sentiment and volume have negligibly
small shares.

Thus far, we can conclude that both fundamental, i.e. the Conference Board
Index, and behavioral, i.e. Reuters sentiment, factors influence and drive stock
returns. Other factors that we have accounted for, such as stock index vol-
ume and TV sentiment, play negligible roles to explaining and driving stock
returns. In the next section, we test how our model performs in a forecasting
environment.

IV. Forecasting

Tetlock (2007) shows the ability of negative words in news articles to pre-
dict quarterly earnings using ordinary least squares (OLS) regressions. Negative
words, he finds, consistently predict lower earnings, regardless of the measure
and the newspaper. Based on a systematical analysis a measure of media con-
tent specifically tied to either negative investor sentiment or risk aversion, he
constructs a hypothetical zero-cost trading strategy using negative words to pre-
dict returns of the Dow Jones Industrials Stock Index that yields excess returns
(7.3% p.a.). He notes, however, that since this strategy neither accounts for
transaction costs nor for slippage and bid-ask spreads while trading daily, it
is questionable whether this strategy would remain profitable in a real-world
setting. Given these findings, we formulate a simple trading strategy that only
requires to trade once per month, given our low-frequency (monthly) data. We
attempt to formulate a similar strategy by hypothesizing that directive Reuters
and TV sentiment, i.e. positive and negative, can predict both positive and
negative returns.

To practically test the predictive power of our model, we formulate a fore-
casting model as in Lütkepohl (1991). The forecasts are derived from the pre-
viously formulated VECM in (1) based on conditional expectations assuming
independent white noise $u_t$. Hence, an $h$-step forecast at time $T$ is

$$y_{T+h|T} = A_1 y_{T+h-1|T} + \cdots + A_p y_{T+h-p|T} + C D_{T+h},$$

(4)
where $y_t$ is a vector of endogenous variables and $D_T$ refers to the deterministic term. The vector $y_t$, incorporating the endogenous variables Dow Jones Industrials stock index returns and volume (both logged), the Conference Board Index as well as Reuters and TV Sentiment, is altered for the forecasts to test which variables add forecasting power, and which ones do not. The first forecast that we run is without the sentiment variables, so that we only include stock index returns and volume as well as the Conference Board Index in the vector $y_t$. We then add Reuters and TV sentiment to the second and third forecast, respectively, and use all five variables in the fourth forecast. We first run in-sample forecasts and then out-of-sample forecasts. Although in-sample forecasts are not as meaningful as out-of-sample forecasts, we run in-sample forecasts to show how the models perform over a longer time horizon, i.e. the entire dataset horizon from January 2005 to December 2009. The in-sample forecasts are constructed on a static solution basis, i.e. the coefficients are estimated over the entire estimation period and then the forecast starts at time $t = 0$. Then, the forecast is done on a $t + 1$ step-by-step basis, while considering the actual value of month $t$. For the out-of-sample forecast, we estimate the model with values from January 2005 to December 2008. Then, we perform a step-by-step $t + 1$ forecast for each month of 2009, simulating a real-world trading environment.

Based on the predicted values of the model, we formulate a simple long-short strategy. If the forecast is above the month-end closing price of the stock index, the strategy goes long at the beginning of the forecast month. If the forecast is below the month-end closing price of the stock index, the strategy goes short. The position is closed at the end of each month at the closing price and adjusted in the direction if the forecast assumes a reversal. For simplicity reasons, the available equity is always invested in full at the beginning of each month.

Table IV shows the results of the in-sample forecasts. As comparison, the Dow Jones Industrials stock index gained 2% over the entire time horizon from 2005 to 2009, whereas all of the strategies achieve much higher returns with success rates between 72% and 84%. The most successful strategy and thus the most accurate in-sample forecast is the model, which incorporates all variables

\[\text{Success Rate} = \frac{\text{number of correctly forecast trading direction (i.e. up or down) months}}{\text{number of total forecast months}}.\]

\[\text{See Appendix A.4 for a more detailed description of the forecasting model as in Lütkepohl (1991).}\]
at hand. This model includes both TV and Reuters sentiment, and its strategy achieves the highest performance (88% absolute return p.a.). The model that does not incorporate any of the sentiment variables achieves the lowest performance (46% p.a.), yet it outperforms the Dow Jones Industrials stock index. The strategies that include Reuters and TV sentiment separately perform very differently. The strategy with Reuters sentiment performs almost as good as the best strategy that includes all variables (85% p.a.), while the strategy with TV sentiment achieves “only” a performance of 51% p.a. This confirms our previous findings that Reuters sentiment has a higher correlation with and impact on stock returns than TV sentiment, while the Conference Board Index is also an important variable in our model that adds alpha. The combination of both sentiments and the Conference Board Index as well as stock index volume works best in terms of absolute performance, but in terms of success rate, the model that excludes TV sentiment among the endogenous variables performs best. On the one hand, this leads to the assumption that TV sentiment and stock index volume only add, if at all, little value to a forecasting model of stock returns. On the other hand, Reuters sentiment as well as the Conference Board Index add a lot of value to forecast stock returns.

The results of the out-of-sample forecasts are illustrated in table V. The results from the in-sample forecasts appear more pronounced but also justified in the out-of-sample forecasts. In the out-of-sample period examined - from January to December 2009 - the strategies that do not include Reuters sentiment perform equally bad with a success rate just above 50% and an annual performance of under 5%. As a benchmark, the Dow Jones Industrials stock index gained over 18% in that period. The two strategies that include Reuters sentiment, however, outperform the benchmark with over 37% and 44% p.a. The best strategy includes all available endogenous variables, and it is able to forecast the direction of the index in 10 months out of 12 correctly, thus achieving a high success rate (83%).

15Note that the Sharpe Ratio of 1.50 for this strategy is higher than for the same strategy during the in-sample forecast horizon. Naturally, one would expect the Sharpe Ratio to be higher for the same strategy during an in-sample forecast than an out-of-sample forecast. In this case, the lower Sharpe Ratio is due to the higher volatility of the stock market during the collapse of Lehman Brothers, which occurred in 2008. See Appendix A.5 for a detailed calculation of the Sharpe Ratio.
According to the various tests and analyses that we have undertaken, we stress two major findings. First, we confirm the EMH by Fama (1970) to the extent that fundamental factors, such as macroeconomic developments, which we accounted for by the Conference Board Index, have a significant influence on the Dow Jones Industrials stock index. This finding is pronounced in both the impulse responses and the variance decomposition analysis, in which the Conference Board Index makes up as much as 40% of the variance of the stock index. We also find that volume plays a minor role in such a model. Second, we reject the EMH on the grounds that not only fundamental factors influence stock markets, but also behavioral ones. We make the attempt to measure the behavioral effect with sentiment in both TV news shows and Reuters news pieces. Reuters sentiment appears to capture investor sentiment quite well, entailing strong predictive power for stock returns. TV sentiment, however, has only little explanatory power in the model. This makes sense though, as one would assume that Reuters news are much more targeted to the investment society than TV news. These findings are line with the studies by De Bondt and Thaler (1985), DeLong et al (1990) as well as Baker and Wurgler (2007). Tetlock’s (2007) results are confirmed and extended.

V. Conclusion

Based on the EMH by Fama (1970), we examine whether fundamental and/or behavioral factors influence US stock prices. To account for fundamental factors, we use the Conference Board Index that comprises a basket of various macroeconomic variables. We also use stock index volume to control for possible market depth and liquidity constraints. To account for behavioral factors, we use novel datasets that are comprised of media sentiment from US TV news shows and Reuters news pieces. Tetlock’s (2007) approach serves as inspiration for this study, as the use of his textual analysis tool, the General Inquirer (GI), seems limited, given that it is only able to distinguish between positive and negative words, but not between contexts.

We confirm Tetlock’s findings that media sentiment has an impact on stock returns, rejecting the EMH by Fama (1970) on the same grounds, given that we find positive correlations between negative media sentiment and declines in stock returns as well as between positive media sentiment and gains in stock returns. We show with impulse response functions and a variance decomposition analy-
ysis of a Vector Error Correction Model that both fundamental macroeconomic factors, represented by the Conference Board Leading Indicators Index, and behavioral factors, such as Reuters sentiment, have a strong and persistent impact on stock returns. With the aid of in- and out-of-sample forecasts, we infer that Reuters sentiment is the better predictor for changes in stock returns than TV sentiment, manifested in higher success as well as performance rates. Nevertheless, when combined, TV and Reuters sentiment perform best in a model that encapsulates all factors examined, such as fundamental macroeconomic and behavioral factors. We confirm the findings of earlier studies by Hong and Stein (1999) as well as by Livnat and Petrovits (2009) that there is a constant and persistent overreaction of stock returns to sentiment, postulated by a long-term drift in sentiment. Thus, we conclude that although fundamental factors influence stock returns in the long-term, so do behavioral factors, which we measure with two proxies: Reuters and TV sentiment. We find that there exist profound differences in the sentiment datasets with Reuters sentiment having a persistent and strong impact on stock returns, whereas the influence of TV sentiment on stock returns is negligibly small.
Appendix

A.1

The Augmented Dickey-Fuller Unit Root Test according to Dickey and Fuller (1979) takes the following equation

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t,$$

where $\gamma = -\left(1 - \sum_{i=1}^{p} a_i\right)$ and $\beta_i = -\sum_{j=1}^{i} a_j$. The Null hypothesis tests whether $\gamma = 0$, and if so, the equation is entirely in first differences and so has a unit root. If $\gamma \neq 0$, then the equation does not have a unit root.

A.2

In the VECM, we have a vector of endogenous variables, denoted by $y_t$. If the process $y_t$ is stationary, it has a Wald moving average (MA) representation

$$y_t = \Phi_0 u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \cdots,$$

where $\Phi_0 = I_K$ and the $\Phi_s$ can be computed recursively as

$$\Phi_s = \sum_{j=1}^{s} \Phi_{s-j} A_j, \quad s = 1, 2, \ldots,$$

with $\Phi_0 = I_K$ and $A_j = 0$ for $j > p$. The coefficients of this representation may be interpreted as reflecting the responses to impulses hitting the system. The $(i,j)$th elements of the matrices $\Phi_s$, regarded as a function of $s$, trace out the expected response of $y_{t,i+s}$ to a unit change in $y_{j,t}$ holding constant all past values of $y_t$. The elements of $\Phi_s$ represent the impulse responses of the components of $y_t$ with respect to the $u_t$ innovations.

Because the underlying shocks are not likely to occur in isolation if the components of $u_t$ are not instantaneously uncorrelated, that is, if $\sum_u$ is not diagonal, in many applications the innovations of the VAR/VECM are orthogonalized using a Cholesky decomposition of the covariance matrix $\sum_u$. Denoting by $P$ a lower triangular matrix such that $\sum_u = PP'$, the orthogonalized shocks are given by $\varepsilon_t = P^{-1}u_t$. Thus, we obtain
\[ y_t = \Psi_0 \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \cdots, \]

where \( \Psi_i = \Phi_i P \) \((i = 0, 1, 2, \ldots)\). Here \( \Psi_0 = P \) is lower triangular so that an \( \varepsilon \) shock in the first variable may have an instantaneous effect on all the variables, whereas a shock in the second variable cannot have an instantaneous impact on \( y_{1t} \) but only on the other variables and so on.

It is important to notice that if a different ordering of the variables in the vector \( y_t \) is chosen this may produce different impulse responses. Hence, the effects of a shock may depend on the way the variables are arranged in the vector of \( y_t \). Breitung et al. (2004) discuss this issue in detail.

For the impulse responses that are computed from the estimated SVEC coefficients, the confidence intervals (CIs) are constructed with the bootstrap method according to Efron and Tibshirani (1993). The standard percentile interval is determined as

\[
CI_s = \left[ s_{\gamma/2}^*, s_{(1-\gamma/2)}^* \right],
\]

where \( s_{\gamma/2}^\ast \) and \( s_{(1-\gamma/2)}^\ast \) are the \( \gamma/2 \)- and \( (1-\gamma/2) \)-quantiles, respectively, of the bootstrap distribution of the corresponding bootstrap estimator of the impulse response coefficient \( \hat{\Phi}^\ast \).

A.3

The Structural Vector Error Correction (SVEC) Forecast Error Variance Decomposition (FEVD) separates the variation in an endogenous variable into the component shocks to the SVAR, or, in this case, the SVEC. The FEVD provides information about the relative importance of each random innovation in affecting the variables in the SVEC. Denoting the ij-th element of the orthogonalized impulse response coefficient matrix \( \psi_n \), the variance of the forecast error \( y_{k,T+h} - y_{k,T+h}^\ast \) is

\[
\sigma_k^2 (h) = \sum_{n=0}^{h-1} (\psi_{k1,n}^2 + \cdots + \psi_{kK,n}^2) = \sum_{j=1}^{K} (\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2).
\]
A.4

The corresponding forecast errors for the forecasts are

\[ y_{T+h} - y_{T+h'} = u_{T+h} + \phi_1 u_{T+h-1} + \cdots + \phi_{h-1} u_{T+1}, \]

where \( \phi_s = \sum_{j=1}^{s} \phi_{s-j} A_j, \) \( s = 1, 2, \ldots, \) with \( \phi_0 = I_K \) and \( A_j = 0 \) for \( j > p. \) Thus, the forecast errors have zero mean and, hence, the forecasts are unbiased.

A.5

The Sharpe ratio is calculated according to Sharpe (1994):

\[ \frac{R_p - R_f}{\sigma_p}, \]

where \( R_p \) is the annualized return of the portfolio, \( R_f \) the annualized rate of a risk-free asset (in this paper we use the 1-month Treasury Bill rate), and \( \sigma_p \) is the annualized standard deviation of the portfolio returns.
References


Figure 1: Time-Series Charts of the Dow Jones Stock Index and Volume, The Conference Board Index, and TV and Reuters Sentiments
Figure 2: Cross-Correlogram of Dow Jones Industrials Stock Index (log differenced values) and Dow Jones Industrials Stock Index Volume (log differenced values)
Figure 3: Cross-Correlogram of Dow Jones Industrials Stock Index (log differenced values) and the Conference Board Business Cycles Indicator (differenced values)
Figure 4: Cross-Correlogram of Dow Jones Industrials Stock Index (log differenced values) and TV Sentiment (differenced values)
Figure 5: Cross-Correlogram of Dow Jones Industrials Stock Index (log differenced values) and Reuters Sentiment (differenced values)
Figure 6: Cross-Correlogram of TV Sentiment (differenced values) and Reuters Sentiment (differenced values)
Figure 7: Scatterplot of Reuters Sentiment and TV Sentiment (differenced values)
Figure 8: Impulse Responses from the Structural Vector Error Correction Model (SVEC) with 95% Bootstrap Confidence Intervals according to Efron and Tibshirani (1993). Abbreviations used denote the following: logarithmized Dow Jones Industrial Stock Index (log\_dj), logarithmized Dow Jones Industrial Stock Index Volume (log\_dj\_vol), the Conference Board Index (Conf\_Board), TV sentiment (tv\_s) and Reuters sentiment from the equities section (tr\_ns\_eq).
Figure 9. Forecast Error Variance Decomposition of Dow Jones Industrials Stock Index Returns (log)
### Table I

**News Sentiment Sources**

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of TV shows examined for sentiment 2005 - 2009</th>
<th>Number of News Articles examined for sentiment 2005 - 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC World News Tonight*</td>
<td>2'408</td>
<td></td>
</tr>
<tr>
<td>CBS Evening News*</td>
<td>1'981</td>
<td></td>
</tr>
<tr>
<td>FOX: Special Report*</td>
<td>3'306</td>
<td></td>
</tr>
<tr>
<td>NBC Nightly News*</td>
<td>2'734</td>
<td></td>
</tr>
<tr>
<td>Thomson Reuters News Items**</td>
<td></td>
<td>4'235'957</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>10'429</td>
<td>4'235'957</td>
</tr>
</tbody>
</table>

*Sources: MediaTenor* & Thomson Reuters NewsAnalytics**
Table II

Optimal Endogenous Lags from Information Criteria

Endogenous Variables: Dow Jones Industrials Stock Index and Volume (log values), The Conference Business Cycles Indicator, Reuters Sentiment, and TV Sentiment
Deterministic Variables: Constant

<table>
<thead>
<tr>
<th>Sample Range:</th>
<th>[2005 M6, 2009 M12], T = 55</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal number of lags (searched up to 4 lags of 1. differences):</td>
<td></td>
</tr>
<tr>
<td>Akaike Info Criterion:</td>
<td>3</td>
</tr>
<tr>
<td>Final Prediction Error:</td>
<td>3</td>
</tr>
<tr>
<td>Hannan-Quinn Criterion:</td>
<td>0</td>
</tr>
<tr>
<td>Schwarz Criterion:</td>
<td>0</td>
</tr>
</tbody>
</table>
### Table III

**Vector Error Correction Model Coefficient Estimates (monthly values)**

| Lags | log_dj | log_dj_vol | Conf_B | tv_s | r_s | log_dj | log_dj_vol | Conf_B | tv_s | r_s | log_dj | log_dj_vol | Conf_B | tv_s | r_s | log_dj | log_dj_vol | Conf_B | tv_s | r_s |
|------|-------|-----------|-------|-----|----|-------|-----------|-------|-----|----|-------|-----------|-------|-----|----|-------|-----------|-------|-----|----|-------|-----------|-------|-----|----|
| (0,0) | 1.674 | 0.00 | 1.794 | 0.37 | 0.45 | 0.32 | 0.37 | 0.01 | 0.45 | 0.37 | 0.45 | 0.37 | 0.01 |
| (1,0) | -0.173 | -0.023 | -1.337 | -0.573 | -0.642 | 0.081 | -0.441 | -0.132 | 0.37 | 0.081 | -0.441 | -0.132 | 0.37 |
| (0.064) | (0.143) | 0.3724 | (0.777) | 0.7860 | (0.880) | (0.682) | (0.682) | (0.7860) | (0.880) | (0.682) | (0.682) | (0.7860) | (0.880) |
| (0,0) | 0.122 | 0.654 | 0.526 | 0.16 | 0.092 | 0.119 | 0.623 | 1.633 | 0.152 | 0.842 | 0.01 | 0.05 | 0.132 | 0.012 | 0.068 |
| (2,0) | 0.122 | 0.654 | 0.526 | 0.16 | 0.092 | 0.119 | 0.623 | 1.633 | 0.152 | 0.842 | 0.01 | 0.05 | 0.132 | 0.012 | 0.068 |
| (3,0) | 0.122 | 0.654 | 0.526 | 0.16 | 0.092 | 0.119 | 0.623 | 1.633 | 0.152 | 0.842 | 0.01 | 0.05 | 0.132 | 0.012 | 0.068 |
| (4,0) | 0.122 | 0.654 | 0.526 | 0.16 | 0.092 | 0.119 | 0.623 | 1.633 | 0.152 | 0.842 | 0.01 | 0.05 | 0.132 | 0.012 | 0.068 |
| (5,0) | 0.122 | 0.654 | 0.526 | 0.16 | 0.092 | 0.119 | 0.623 | 1.633 | 0.152 | 0.842 | 0.01 | 0.05 | 0.132 | 0.012 | 0.068 |
| (6,0) | 0.122 | 0.654 | 0.526 | 0.16 | 0.092 | 0.119 | 0.623 | 1.633 | 0.152 | 0.842 | 0.01 | 0.05 | 0.132 | 0.012 | 0.068 |

**Endogenous Variables**
- Dow Jones Industrials Stock Index (log_dj)
- Dow Jones Industrials Stock Index Volume (log_dj_vol)
- The Conference Board Index (Conf_B)
- TV Sentiment (tv_s)
- Reuters Sentiment (r_s)

**Exogenous Variables**

**Deterministic Variables**
- Constant (CONST)

**Endogenous Lags (Differences - in months)**
- 3

**Exogenous Lags**
- 0

**Sample Range**
- [2005 M5, 2009 M12], T = 56

**Estimation Procedure**
- One stage Johansen approach according to Johansen (1995)

**Loading coefficients**

- **Estimated cointegration relation**
  - log_dj(t-1) log_dj_vol(t-1) Conf_B(t-1) TV_S(t-1) R_S_ALL(t-1) TV_S(t-1) 0.738 0.002 0.210 0.002 0.685

**Estimated Error Correction Relation**
- log_dj(t-1) log_dj_vol(t-1) Conf_B(t-1) TV_S(t-1) R_S_ALL(t-1) TV_S(t-1) 0.738 0.002 0.210 0.002 0.685

- **Estimated Long-run Relation**
  - log_dj(t-1) log_dj_vol(t-1) Conf_B(t-1) TV_S(t-1) R_S_ALL(t-1) TV_S(t-1) 0.738 0.002 0.210 0.002 0.685

**Estimated Short-run Relation**
- log_dj(t-1) log_dj_vol(t-1) Conf_B(t-1) TV_S(t-1) R_S_ALL(t-1) TV_S(t-1) 0.738 0.002 0.210 0.002 0.685
<table>
<thead>
<tr>
<th>Endogenous Variables</th>
<th>Performance Stock Index - overall %</th>
<th>Performance Stock Index - p.a. %</th>
<th>Performance Strategy - overall %</th>
<th>Performance Strategy - p.a. %</th>
<th>Sharpe Ratio*</th>
<th>Correct Prediction of Direction in Months</th>
<th>Wrong Prediction of Direction in Months</th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dow Jones Industrials Stock Index</td>
<td>2.14</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dow Jones Industrials Stock Index, Dow Jones Industrials Stock Index Volume</td>
<td>222.98</td>
<td>46.94</td>
<td>0.68</td>
<td>40</td>
<td>15</td>
<td>22.71%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dow Jones Industrials Stock Index, Dow Jones Industrials Stock Index Volume, the Conference Board Index</td>
<td>402.63</td>
<td>84.77</td>
<td>1.05</td>
<td>46</td>
<td>9</td>
<td>83.64%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dow Jones Industrials Stock Index, Dow Jones Industrials Stock Index Volume, the Conference Board Index, and Reuters Sentiment</td>
<td>242.36</td>
<td>51.02</td>
<td>0.72</td>
<td>42</td>
<td>13</td>
<td>76.36%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dow Jones Industrials Stock Index, Dow Jones Industrials Stock Index Volume, the Conference Board Index, and TV Sentiment</td>
<td>420.03</td>
<td>88.43</td>
<td>1.08</td>
<td>45</td>
<td>10</td>
<td>81.82%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*For the calculation of the Sharpe Ratio, see Appendix A.5.
Table V
Performance of Trading Strategies based on Vector Error Correction Models Out-of-Sample Prediction Estimates (monthly data)

<table>
<thead>
<tr>
<th>Endogenous Variables</th>
<th>Performance Stock Index - overall % Jan 2009 - Dec 2009</th>
<th>Performance Strategy - overall % Jan 2009 - Dec 2009</th>
<th>Sharpe Ratio*</th>
<th>Correct Prediction of Direction in Months</th>
<th>Wrong Prediction of Direction in Months</th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dow Jones Industrials Stock Index</td>
<td>18.38</td>
<td></td>
<td></td>
<td>7</td>
<td>5</td>
<td>58.33%</td>
</tr>
<tr>
<td>Dow Jones Industrials Stock Index, Dow Jones Industrials Stock Index Volume, and the Conference Board Index</td>
<td>4.97</td>
<td>0.29</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>58.33%</td>
</tr>
<tr>
<td>Dow Jones Industrials Stock Index, Dow Jones Industrials Stock Index Volume, the Conference Board Index, and Reuters Sentiment</td>
<td>37.58</td>
<td>1.29</td>
<td>9</td>
<td>3</td>
<td>5</td>
<td>58.33%</td>
</tr>
<tr>
<td>Dow Jones Industrials Stock Index, Dow Jones Industrials Stock Index Volume, the Conference Board Index, and TV Sentiment</td>
<td>4.97</td>
<td>0.29</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>58.33%</td>
</tr>
<tr>
<td>Dow Jones Industrials Stock Index, Dow Jones Industrials Stock Index Volume, the Conference Board Index, and Reuters and TV Sentiment</td>
<td>44.62</td>
<td>1.50</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>83.33%</td>
</tr>
</tbody>
</table>

*For the calculation of the Sharpe Ratio, see Appendix A.5.