IS NEWS SENTIMENT MORE THAN JUST NOISE?

Complete Research

Alfano, Simon, University of Freiburg, Freiburg, Germany, simon.alfano@is.uni-freiburg.de
Feuerriegel, Stefan, University of Freiburg, Freiburg, Germany, stefan.feuerriegel@is.uni-freiburg.de
Neumann, Dirk, University of Freiburg, Freiburg, Germany, dirk.neumann@is.uni-freiburg.de

Abstract

Big Data analytics has recently fostered significant research on the influence of news sentiment in finance. This paper thus examines the effect of news sentiment on crude oil prices for different investor types according to the noise trader approach. The noise trader approach assumes the presence of informed and uninformed investors. Informed investors possess a perfect information horizon, whereas uninformed investors trade upon noise signals, such as sentiment. Methodologically, we decompose the crude oil price with a Kalman filter into a Kalman-smoothed, fundamental price component and a noise residual. We then regress news sentiment on both decomposed oil price components. Our findings suggest that news sentiment not only has a significant positive effect on the noise residual (as suggested by the noise trader approach), but also on the fundamental price. Thus, we find empirical evidence contradicting the noise trader model, which assumes that only uninformed investors trade on sentiment.

Keywords: Information Processing, Sentiment Analysis, Text Mining, Kalman Filter, Noise Trader Approach, Oil Market.

1 Introduction

The ability to process and rapidly internalize newly available information characterizes efficient markets (Fama, 1965, 1998; Friedman, 1953; Malkiel, 2003, 2005; Malkiel and Fama, 1970). This is formulated by the so-called *Efficient Market Hypothesis (EMH)* of Fama (1965), which builds on the ability of financial markets (in form of e. g. traders, buyers, sellers) to reflect new information in an *information efficient* manner at fair market prices. This should hold particularly true for modern electronic markets, where the increasing amount, speed of dissemination and accessibility of information facilitates efficient financial markets. Regardless, turbulence in asset markets have been a recurring theme over the last decades (Carlson, 2007; Easley, López de Prado, and O'Hara, 2011; Sornette, 2003). For example, the S&P 500 declined by nine percent within one week during the October 1987 stock market crash, even though no fundamental information entered the market. Consequently, this crash highlighted the need for a revision of the EMH in its original version (DeLong, Shleifer, Summers, and Waldmann, 1990; Shleifer, 2000; Shleifer and Summers, 1990; Thaler, 2005). In this context, DeLong, Shleifer, Summers, and Waldmann (1990) developed the so-called *noise trader* approach to reflect the fact that not all traders possess the same information. Consequently, noise trader models (and its variations) typically assume that traders or investors are either *informed* or *uninformed*. Uninformed traders do not dispose of the same information horizon as informed traders. Thus, uninformed traders trade on noisy signals (e. g. recent market trends, news sentiment). In contrast, informed traders base their rational decisions on full information availability and make arbitrage trades to exploit the uninformed trading decisions of uninformed noise traders (Shleifer and Summers, 1990).

While previous research has focused on establishing and extending the noise trader theory, we are not aware of any research that conducts an empirical investigation of noise trader theory by utilizing news sentiment. Brown and Cliff (2004) studied the impact of investor sentiment on different investor types. In order to work through large quantities of news releases, we exploit Information Systems (IS) research, since it is perfectly positioned to apply available methodological tools to process *Big Data*, as is the specific case for financial markets. Consequently, Big Data analytics research has examined the processing of news information in many publications. The explored research questions cover several dimensions, in particular volume, variety, velocity and veracity (IBM, 2013). With the advent of Big Data, there exists nowadays more (unstructured) data to analyze sentiment (e. g. Cambria, Schuller, Xia, and Havasi, 2013; Chen, Chiang, and Storey, 2012; Hilbert and López, 2011). Hence, this field of research, which has received significant traction recently, is positioned at the intersection between Information Systems, Big Data analytics and Behavioral Finance (Chen, Chiang, and Storey, 2012; Nassirtoussi, Aghabozorgi, Wah, and Ngo, 2014). IS research has also established robust methods to analyze news sentiment and investigate its influence on financial markets (e. g. Antweiler and Frank, 2004; Mittermayer and Knolmayer, 2006a). In addition, recent research (e. g. Cenesizoglu, 2014) proposes theoretical models to assess the role of news signals on asset prices, as well as behavioral experiments (Bosmana, Kräussl, and Mirgorodskaya, 2015) to verify news reception in practice. Consequently, this paper attempts to shed light on how news sentiment influences noise traders and informed traders. We focus our analysis on the oil market and decompose the oil price with a Kalman filter into a fundamental price component (i. e. the oil price for an informed trader) and a noise residual (i. e. caused by noise traders deviating from the fundamental value of the market price). Wen can then separately measure and compare the impact of news sentiment on both price components. This provides evidence that uninformed investors trade on sentiment signals, opposed to the classical noise trader theory.

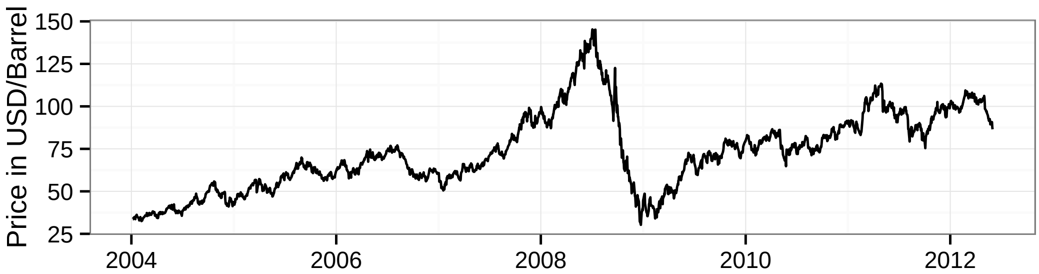

Figure 1. Nominal Western Texas Intermediate (WTI) crude oil price from January 2004 until May 2012.

In this paper, we select the crude oil market for several reasons: first and foremost, crude oil is a pivotal commodity for global economies. As such, increasing chemicals manufacturing and transportation across the globe drive the global demand for oil. In 2007, a total of 138.5 million oil futures contracts, each accounting for 1000 barrels, were traded.[1] In addition, the Organization of the Petroleum Exporting Countries (OPEC) forecasts that demand[2] will increase by 20 million barrels a day until the year 2035. In the last decade, oil prices have been very volatile. As illustrated in Figure 1, the Western Texas Intermediate (WTI) crude oil price gradually increased between 2003 and 2008, exceeding $ 145 per barrel in the year 2008, followed by a price crash of more than a hundred dollars. In the light of increasing

[1] CBS News. *Oil Trading's Powerful "Dark Markets"*. 2011. URL: `http://www.cbsnews.com/2100-18564_162-4188620.html`, accessed September 9, 2014.

[2] OPEC. *2013 World Oil Outlook*. URL: `http://www.opec.org/opec_web/static_files_project/media/downloads/publications/WOO_2013.pdf`, accessed September 9, 2014.

demand and volatile prices, the oil market is subject to broad news coverage. Consequently, crude oil seems to be suitable for investigating the influence of news sentiment on its price.
The remainder of this paper is organized as follows: Section 2 reviews previous research on the noise trader approach and news sentiment, which we combine to derive our research hypotheses. We describe our news sentiment analysis and the Kalman filter for price decomposition in Section 3. This is followed by Section 4, in which we discuss the results of our analysis regarding the impact of news sentiment on the Kalman-smoothed, fundamental price and the noise residual. Finally, Section 5 concludes and provides a research outlook.

2 Related Work

In this section, we present related literature according to two topics: first, we review studies relating to the role of sentiment in *market micro-structures*, with a particular focus on the *noise trader theory*. Second, we review approaches to *measure news sentiment*. The following discussion of related literature motivates the investigation of the impact of sentiment on both noise residuals and the fundamental value of oil prices as an important and relevant research topic to the IS body of knowledge. By combining both research streams, we establish our research hypotheses at the end of this section.

2.1 Theoretical Background: Noise Trader Theory

Standard economic theory assumes efficient markets that rely upon the availability of information (Fama, 1965). Triggered by the stock market crash on October 19, 1987, many research publications focused on extensions of the Efficient Market Hypothesis to account for uninformed traders (DeLong, Shleifer, Summers, and Waldmann, 1990; Shleifer and Summers, 1990). Such uninformed investors, who have no access to inside information, are thus called *noise traders* (Black, 1986; Kyle, 1985). Their rationale is that they interpret *noisy signals* as genuine information, i. e. information resulting in a competitive advantage when forming investment decisions.
Based on the idea of uninformed traders, Shleifer and Summers (1990) developed the *noise trader* approach to provide an explanation for inefficiencies in financial markets not reflected in the EMH. The approach assumes two types of investors as follows:

- *Uninformed (noise) investors*, whose demand for risky assets is (1) partially not justified by fundamentals and (2) subject to sentiment. The sentiment of noise traders refers to the over- and underestimation of expected returns relative to rational expectations. One of the sentiment drivers is news sentiment found in financial disclosures.
- *Informed investors*, who are not subject to sentiment. While they dispose of a full information horizon, Lee, Shleifer, and Thaler (1991) state that noise traders' sentiment is stochastic and cannot be perfectly predicted by rational investors. For instance, while informed investors can differentiate between information and noise, they cannot predict when overvalued or undervalued asset prices will return to their fundamental prices, as a result of sentiment. For this reason, their arbitrage trading is risky and thus limited.

Altogether, the *noise trader* approach acknowledges the role of sentiment in influencing uninformed investors. Other research papers further developed the concept of sentiment as a noisy signal affecting noise traders (Brown, 1999; Brown and Cliff, 2004; Lee, Shleifer, and Thaler, 1991; Sanders, Irwin, and Leuthold, 1997; Shleifer, 2000; Shleifer and Vishny, 1997; Thaler, 2005; Yan, 2010) and tested the impact of noise traders on market characteristics, such as the role of noise traders in the long-term reversal of stock prices (Gerber, Hens, and Vogt, 2002). Recent publications contribute to the research body by improving our understanding of noise traders: Bloomfield, O'Hara, and Saar (2009) find a relationship with market liquidity, while Mendel and Shleifer (2012) show that uniformed traders can amplify sentiment shocks and increase the distance from fundamental prices.

The Efficient Market Hypothesis presumes that market participants quickly react to novel information, so that uninformed traders vanish. In opposition to this hypothesis, several publications (DeLong, Shleifer, Summers, and Waldmann, 1990; Shleifer, 2000) and Thaler (2005) argue that arbitrageurs will not drive all noise traders out of the market, instead claiming that noise traders are presumably integral participants in asset markets. Noise traders can even lead to positive effects, as the novel experimental results of Bloomfield, O'Hara, and Saar (2009) show: higher market liquidity increases market volume, as well as depth, and reduces spreads.

2.2 News Sentiment

Decision Analytics, as a sub-discipline of Information Systems (IS) research, is well-equipped to study how agents in financial markets process information conveyed by textual news (Chen, Chiang, and Storey, 2012). The textual content of news provides relevant information beyond purely quantitative facts, such as earnings or profit forecasts, but also the tone of the language. This subjective tone can be extracted from text documents through a so-called *sentiment analysis*. Sentiment analysis (also known as *opinion mining*) refers to methods that measure how positive or negative the content of text sources is.

News sentiment measures allow for the exploration of the effect of financial news on financial markets. Previous research has focused on how human agents process the sentiment of financial news. For instance, empirical evidence highlights a noticeable relationship between news information and stock market movements (e. g. Antweiler and Frank, 2004; Tetlock, 2007). Beyond the pure establishing of a link, several papers have contributed to the understanding of financial news reception. Examples are as follows: Siering (2013) highlights the differential interpretation of novel information by investors and analysts. According to Feuerriegel, Ratku, and Neumann (2015), considerable differences in news sentiment reception according to the underlying news topic also exist.

Related works have investigated information processing around financial news originating from commodity markets in particular. For example, Feuerriegel and Neumann (2013) argue that negative news sentiment is a stronger driver of oil and gold markets than positive sentiment. There exists further differences in oil news reception, for example, across bullish and bearish market regimes (Feuerriegel, Lampe, and Neumann, 2014), even finding evidence of irrational exuberance (Ratku, Feuerriegel, and Neumann, 2014). Interestingly, Feuerriegel, Heitzmann, and Neumann (2015) also prove that the tone of oil news *causes* price reactions, solving the endogeneity problem. In addition, a link between news tone and commodities can also be found in the gas market (Borovkova and Mahakena, 2015).

Information Systems research has developed various approaches to measure sentiment, since sentiment analysis is deployed across various domains and for different textual sources. For instance, Pang and Lee (2008) provide a comprehensive domain-independent survey. Within the finance domain, recent literature reviews (Minev, Schommer, and Grammatikos, 2012; Mittermayer and Knolmayer, 2006b; Nassirtoussi, Aghabozorgi, Wah, and Ngo, 2014) focus on studies aimed at stock market prediction. Financial text mining research prevalently deploys *dictionary-based methods* (cp. Demers and Vega, 2010; Henry, 2008; Jegadeesh and Wu, 2013; Loughran and McDonald, 2011; Tetlock, Saar-Tsechansky, and Macskassy, 2008). Dictionary-based approaches produce reliable results by counting the frequency of pre-defined negative and positive words from a given dictionary. *Machine learning methods* (e. g. Antweiler and Frank, 2004; Li, 2010; Mittermayer and Knolmayer, 2006a; Schumaker and Chen, 2009) represent a variety of methods, but may be subject to overfitting (Sharma and Dey, 2012). A remedy may originate from regularization methods that utilize variable selection to generate domain-dependent dictionaries. Such a dictionary has been generated for the finance domain by Pröllochs, Feuerriegel, and Neumann (2015b) with the help of Bayesian learning.

2.3 Research Hypotheses

This section combines the previous two concepts of noise trader theory and news sentiment by raising the research question of how news sentiment influences the trading of informed and uninformed investors respectively. We approach this question methodologically by decomposing the WTI crude oil price into two components: the fundamental oil price and a noise residual. This is similar to Schwartz and Smith (2000), who applied the Kalman filter to decompose the oil price into both its fundamental and noise components, and to Brown and Cliff (2004), who decomposed investor sentiment with a Kalman filter to study the relationship between investor sentiment and near-term stock market returns, but investor sentiment is radically different news sentiment. To our best knowledge, no previous research has investigated the different impact of news sentiment on de-noised fundamental asset prices in general, and de-noised oil prices in particular. In this paper, we first decompose the oil price by applying a Kalman filter and then investigate the differential effect of news sentiment on the respective price components. Thus, to extend prior research, we address the following hypotheses:

- ***Hypothesis (H1):*** *News sentiment positively influences the Kalman noise residual.*
- ***Hypothesis (H2):*** *News sentiment does not interfere with the returns of the fundamental oil price.*
- ***Hypothesis (H3):*** *The effect of news sentiment is stronger for Kalman noise residuals than for the returns of the fundamental oil price.*

All three hypotheses immediately follow from the noise trader approach (Brown, 1999; DeLong, Shleifer, Summers, and Waldmann, 1990; Lee, Shleifer, and Thaler, 1991; Shleifer, 2000; Thaler, 2005; Yan, 2010). In order to investigate the three hypotheses, we first de-noise our oil price with a Kalman-filter in order to attain a fundamental oil price and a noise residual. Second, we regress a standardized Net-Optimism news sentiment variable on the noise residual to test hypothesis (H1) and on the real return of the fundamental oil price residual to test hypothesis (H2). Finally, we compare the two results as part of hypothesis (H3).

3 Methodology and Data Sources

This section introduces the applied data set and the methodology. As shown in Figure 2, our approach consists of two steps: first, we process our data by decomposing the WTI crude oil price via a Kalman process *(dependent variables of interest)* and establishing a standardized Net-Optimism news sentiment score *(independent variable of interest)*. Second, we test our hypotheses with a Newey-West corrected OLS regression.

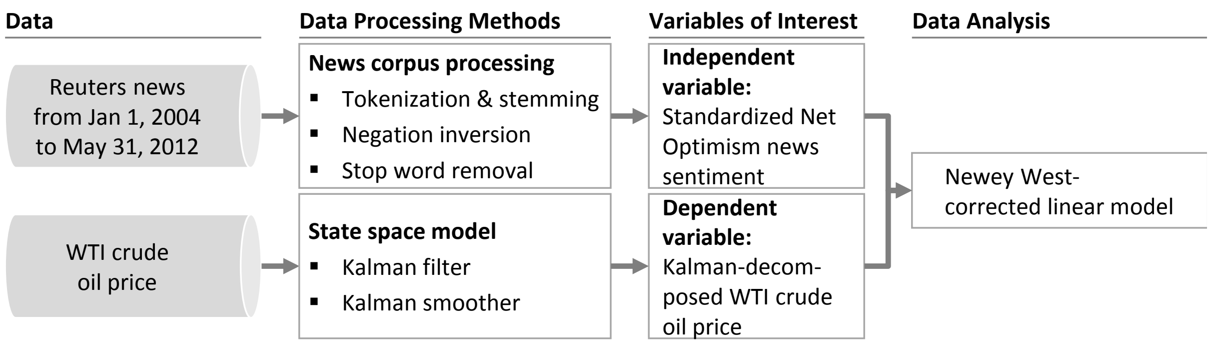

Figure 2. The above research framework shows how this paper evaluates the influence of news sentiment on fundamental oil prices and noise residuals.

We utilize a method named sentiment analysis from the text mining domain and thus present its high-level idea in the following. Sentiment analysis frequently uses this text mining process to derive information from text (Manning and Schütze, 1999; Medhat, Hassan, and Korashy, 2014; Nassirtoussi, Aghabozorgi, Wah, and Ngo, 2014). Since text appears in the form of unstructured data, the first step is to transform it into a machine-readable representation. This data representation is the result of a feature selection, where relevant features (e. g. single terms or bag-of-words) are extracted. The extracted features are then input to form of decision algorithm, such as a machine learning method or a rule-based approach, to calculate the subjective sentiment.

3.1 News Sentiment Analysis

Sentiment analysis refers to methods that measure the positivity or negativity of the content of text sources. Indeed, subjective information can be extracted from text documents through sentiment analysis. In addition, sentiment analysis can also gauge how market participants process and respond to news.
Prior to carrying out our sentiment analysis, we need to pre-process our news dataset by applying the following steps (Feuerriegel and Neumann, 2013; Manning and Schütze, 1999):

1. **Tokenization.** We separate corpus elements into uniquely labeled word *tokens*.
2. **Negations.** We utilize a rule-based approach to detect negation scopes and invert the meaning accordingly (Dadvar, Hauff, and de Jong, 2011; Pröllochs, Feuerriegel, and Neumann, 2015a).
3. **Stop word removal.** We remove words without relevant significance, e. g. articles and pronouns (Lewis, Yang, Rose, and Li, 2004).
4. **Stemming.** During the stemming process step (Manning and Schütze, 1999), we truncate all inflected words to their stem. In this paper, we apply the so-called Porter stemming algorithm.

Upon completion of the pre-processing phase, we can analyze news sentiment and its influence on financial markets. According to a recent study of Feuerriegel and Neumann (2013) on sentiment analysis robustness, the correlation between news sentiment and abnormal returns in oil markets differs across various sentiment metrics. The Net-Optimism metric of Demers and Vega (2010), combined with Henry's Finance-Specific Dictionary (Henry, 2008), is a sentiment approach that yields a robust relationship. The Net-Optimism metric $S(t)$ for day t processes all news on that business day. It is given by the difference between the count of positive $W_{\text{pos}}(A)$ and negative $W_{\text{neg}}(A)$ words divided by the total count of words $W_{\text{tot}}(A)$ in all announcements A on day t. Thus, Net-Optimism is defined by

$$S(t) = \frac{\sum_A W_{\text{pos}}(A) - W_{\text{neg}}(A)}{\sum_A W_{\text{tot}}(A)} \in [-1, +1]. \tag{1}$$

with mean μ and standard deviation σ. To facilitate calculations and later comparisons, we *standardize* this sentiment metric

$$S^*(t) = \frac{S(t) - \mu}{\sigma} \in (-\infty, +\infty), \tag{2}$$

scaled to a zero mean and a standard deviation of one. When we use the standardized Net-Optimism news sentiment in the following analysis, we will refer to it as *news sentiment*.

3.2 Data Sources

Our news corpus originates from the *Thomson Reuters News Archive* for Machine Readable News. We selected this news corpus for several reasons: Thomson Reuters transmits third-party, independent announcements faster than print media (MacGregor, 2013; Paterson, 2007), including online channels of print media. Thus, the news corpus is highly suitable to evaluate stock market reactions. The provided Reuters announcements span the period from January 6, 2004 until May 31, 2012. We only include business days, providing a total of 2108 observation days. Furthermore, the Thomson Reuters news corpus enables us to effectively gather all announcements related to crude oil in the English language, while automatically removing personal opinions or alerts. The information content of opinions and alerts might be limited and potentially difficult to interpret. We also discard announcements communicating changes in prices to avoid simultaneity in our statistical analysis. We apply a further set of filter criteria to omit unsuitable information according to Feuerriegel, Heitzmann, and Neumann (2015) and Feuerriegel and Neumann (2013). Overall, we yield a total of 307,430 crude oil-related announcements. These announcements are aggregated into daily sentiment scores of which 2047 values are positive and 61 negative.

Because the WTI crude oil price serves as the benchmark U. S. crude oil price, it is a common choice in oil market-related research (Bencivenga, D'Ecclesia, and Triulzi, 2012; Chatrath, Miao, and Ramchander, 2012; Kilian and Vega, 2011). Moreover, consistent with prior research (Feuerriegel, Heitzmann, and Neumann, 2015; Kilian, 2009; Lechthaler and Leinert, 2012), we add the following fundamentals as control variables to our econometric model: (a) U. S. interest rate, (b) U. S. Dollar/Euro exchange rate, (c) level of oil imports (in million barrel), (d) total open interest in crude oil future contracts (in million), (e) gold price (London, afternoon fixing) and (f) S&P 500 index (additionally included). All financial data originates from Thomson Reuters Datastream. The corresponding descriptive statistics are provided in Table 1.

Table 1. Descriptive statistics of time series ranging from January 6, 2004 until May 31, 2012.

Variable		Freq.	Mean	Median	Min.	Max.	Std. Dev.	Skew.	Kurt.
$S^*(t)$	News Sentiment (Standardized Net-Optimism)	Daily	0.00	0.01	−3.98	4.17	1.00	−0.10	0.74
$P(t)$	WTI Crude Oil Price	Daily	73.23	71.42	30.28	145.31	22.76	0.44	−0.17
$P_{\text{Kalman}}(t)$	Fundamental Oil Price	Daily	73.23	71.1	33.83	139.04	22.54	0.42	−0.23
$N_{\text{Kalman}}(t)$	Noise Residual	Daily	0.00	0.03	−10.21	19.08	1.81	0.39	7.89
$r(t)$	U. S. Interest Rate	Monthly	1.82	1.30	0.01	5.01	1.85	0.55	−1.29
$FX(t)$	U. S. Dollar/Euro Exchange Rate	Daily	1.34	1.32	1.17	1.60	0.10	0.53	−0.34
$IM(t)$	Level of Oil Imports (in Million Barrel)	Monthly	292.99	297.81	229.14	327.48	22.29	−0.44	−0.67
$OI(t)$	Open Interest in Crude Oil Futures (in Million)	Weekly	1.18	1.23	0.63	1.65	0.28	−0.50	−0.95
$G(t)$	Gold Price	Daily	892.94	827.00	375.00	1895.00	411.87	0.62	−0.72
$SP(t)$	S&P 500 Index	Daily	1220.07	1226.14	676.53	1565.15	168.06	−0.47	0.18

3.3 Price Decomposition

Perturbations and inaccuracies often impair financial time series data (Haven, Liu, and Shen, 2012). The *Kalman filter* represents a recursive approach to linear filtering problems with discrete data (Kalman, 1960). The Kalman filter decomposes discrete datasets, such as time series of prices, into both a *de-noised fundamental price* and a *noise component*. Utilizing the Kalman filter to decompose market prices is a widely-used approach for financial time series (Brogaard, Hendershott, and Riordan, 2014; Haven, Liu, and Shen, 2012; Hendershott and Menkveld, 2014; Hendershott, Menkveld, Li, and Seasholes, 2013; Lopes and Tsay, 2011; Schwartz and Smith, 2000; Wong, 2010).

We describe the mechanisms of the Kalman filter in the following. The Kalman filter uses a feedback control to estimate a process: (1) the filter estimates the process state at the next time step t and then (2) obtains feedback in the form of (noisy) measurements to correct the prediction for the next state (Welch and Bishop, 1997). Thus, the Kalman filter consists of two types of equations:

1. *Time update* or *predictor* equations to attain *a priori* estimates for the next time step by projecting forward the current state and error covariance estimates.

2. *Measurement update* or *corrector* equations to include additional measurements into the *a priori* estimates in order to attain improved *a posteriori* estimates.

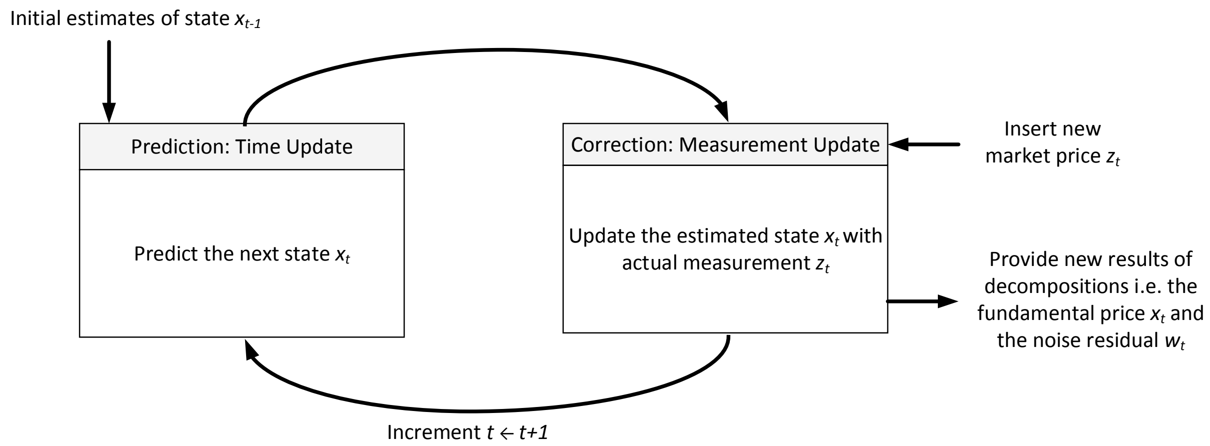

Figure 3. Discrete Kalman filter cycle with the prediction of current state estimate ahead in time and the correction of prediction by an actual measurement update (Welch and Bishop, 1997).

The Kalman filter estimates (Welch and Bishop, 1997) the state $x_t \in \mathbb{R}^n$ at time t of a controlled discrete-time process driven by a linear stochastic difference equation

$$x_{t+1} = F_t x_t + B_t u_t + w_t, \qquad w_t \sim N(0,Q), \tag{3}$$

and a measurement equation

$$z_t = H_t x_t + v_t, \qquad v_t \sim N(0,R), \tag{4}$$

where $F_t \in \mathbb{R}^{n\times n}$ provides a transition matrix, $B_t \in \mathbb{R}^{n\times l}$ a control input model, $z_t \in \mathbb{R}^m$ a measurement and $H_t \in \mathbb{R}^{m\times n}$ a model that maps the true state onto the observed state. Furthermore, the variable $u_t \in \mathbb{R}^l$ denotes a control vector and noise is given by $w_t \in \mathbb{R}^n$ and $v_t \in \mathbb{R}^m$ in the form of i.i.d. random processes $N(0,\cdot)$.

The advantage of the Kalman filter is its property of updating the filter from x_{t-1} to x_t based on a new measurement z_t without the requirement of reproducing the full data set $z_1,\ldots,z_t$ (Welch and Bishop, 1997). With the above definition, we can translate the Kalman filter to our specific setting as a decomposition analysis. In our case, we observe a price sequence z_t given by the true market prices. Then, we can decompose the price series into a fundamental price x_t and a noise residual w_t. The fundamental price is a time series of Kalman-smoothed values, which calculate the dynamic mean of the prices. The estimation error w_t of a state equation represents the noise component, i. e. the gap between fundamental and observed prices.

The Kalman process is repeated with the prior *a posteriori* estimates to predict the new *a priori* estimates. Thus, the Kalman filter recursively conditions all current estimates on prior measurements. This recursive approach facilitates the implementation of the Kalman filter as compared to e. g. the Wiener filter (Welch and Bishop, 1997).

4 Decomposition Analysis

The above sections have introduced the concepts of price decompositions and news sentiment analysis. By combining the two, we perform our decomposition analysis, as well as the interference of news sentiment on the fundamental oil price and its noise residual. For this, we select the oil price model proposed by Kilian (2009). In the domain of oil-related news, Lechthaler and Leinert (2012) have utilized this oil price model to show the influence of news sentiment on monthly crude oil returns, while Feuerriegel, Heitzmann, and Neumann (2015) haven even established a causal relationship between daily news sentiment and the abnormal returns of crude oil. In the following, we study how news sentiment influences both the noise residual and the fundamental oil price.

An initial graphical assessment indicates a positive relationship between news sentiment and the return of the fundamental oil price and the noise residual respectively. Figure 4 highlights so-called LOWESS trend lines with a 95 % confidence interval, i. e. a *locally weighted scatterplot smoothing calculated with local regressions* (Cleveland, 1979; Cleveland and Devlin, 1988). The LOWESS trend line elicits a positive relationship between news sentiment with the noise residual (left) and the return of the fundamental oil price (right) respectively. In addition, the right plot of Figure 4 indicates that the relationship between news sentiment with the return of the fundamental oil price seems to be less pronounced on days with very low and very high returns.

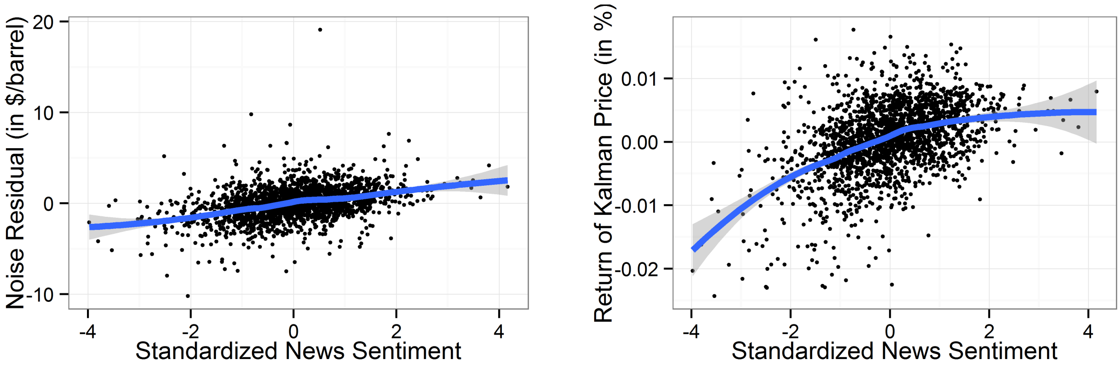

Figure 4. LOWESS (locally weighted scatterplot smoothing with the smoothing parameter f set to 0.5*) calculates a trend line with a 95 % confidence interval and indicates a relationship between news sentiment and the noise residual (left) and the returns of the fundamental oil price (right).*

4.1 Kalman Filtering

This section applies the Kalman filter (as a Kalman smoother) to decompose the WTI crude oil price time series into both a fundamental price component and a noise residual. The noise residuals are plotted Figure 5, revealing a highly volatile pattern. We also observe several spikes in positive and negative directions during the year 2008, which coincides with a drop in oil prices at the same time.

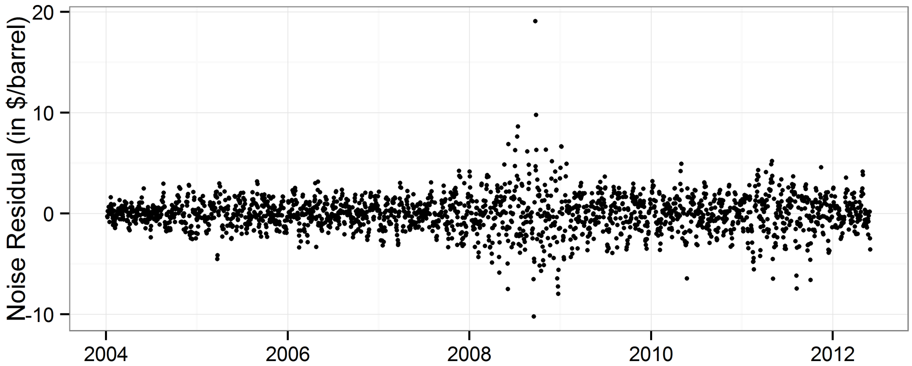

Figure 5. Noise residuals of the (Kalman-smoothed) fundamental oil price.

The previous LOWESS trend lines in Figure 4 show a clearly visible link between news sentiment and both price components. To investigate this characteristic further, we perform a correlation analysis. This analysis confirms, as visualized earlier in Figure 4, a positive relationship between news sentiment and our price components.

From Table 2, we observe a high correlation between news sentiment and and the originally quoted WTI crude oil price (correlation coefficient of 0.186); however, the correlation coefficient is larger for both components of the Kalman-decomposed prices. On the one hand, news sentiment is positively correlated with the daily returns of our fundamental oil price denoted by a correlation coefficient of 0.453. On the other hand, the correlation coefficient between news sentiment and the noise residual accounts for 0.383. Both correlation coefficients are significantly different from zero, as given by a P-value < 0.001 in Table 2. Interestingly, the relationship of news sentiment is larger in combination with the fundamental price than for the noise residual. By observing a higher magnitude of positive correlation between news sentiment and the fundamental price than with the noise residual, we find early evidence that lies in contrast to the noise trader theory of sentiment as only relevant to noise traders. We investigate further the relationship between news sentiment and both price components in the subsequent section of this paper.

Table 2. Correlation analysis of news sentiment and the decomposed price components.

	WTI Crude Oil Price	Return on Fundamental Oil Price (in %)	Noise Residual
News Sentiment	0.186***	0.453***	0.383***
(i. e. Standardized Net-Optimism)	(8.668)	(23.299)	(19.017)

Stated: Correlation Coeff., t-Stat. in Parenthesis Significance: *** 0.001, ** 0.01, * 0.05

4.2 News Sentiment Influence across Decomposition Distributions

This section specifies our regression models, which include relevant oil price control variables according to Kilian (2009) and Lechthaler and Leinert (2012). In order to address our hypotheses, we analyze the influence of news sentiment on the (Kalman) noise residual $N_{\text{Kalman}}(t)$ by the regression model

$$N_{\text{Kalman}}(t) = \beta_0 + \beta_1 S^*(t) + \beta_2 r(t) + \beta_3 FX(t) + \beta_4 IM(t) + \beta_5 OI(t) + \beta_6 G(t) + \beta_7 SP(t) + \varepsilon_t \quad (5)$$

and on returns $R_{\text{Kalman}}(t)$ the (Kalman-smoothed) fundamental crude oil price via

$$R_{\text{Kalman}}(t) = \beta_0 + \beta_1 S^*(t) + \beta_2 r(t) + \beta_3 FX(t) + \beta_4 IM(t) + \beta_5 OI(t) + \beta_6 G(t) + \beta_7 SP(t) + \varepsilon_t, \quad (6)$$

where the standardized news sentiment $S^*(t)$ provides our independent variable of interest. In addition, we include several control variables in line with previous research (Feuerriegel, Heitzmann, and Neumann, 2015; Kilian, 2009; Lechthaler and Leinert, 2012): U. S. interest rate $r(t)$, U. S. Dollar/Euro exchange rate $FX(t)$, level of oil imports $IM(t)$, open interest $OI(t)$, the gold price $G(t)$ and the S&P 500 index $SP(t)$. We control both of the above econometric models for heteroskedasticity (Goldfeld-Quandt test with P-value < 0.001) and autocorrelation (Durbin-Watson test with P-value < 0.001). Consequently, we must amend our OLS regression by calculating standard errors according to the Newey-West estimator that are robust to heteroskedasticity and autocorrelation. When checking Variance Inflation Factors, we see no indication of multicollinearity. In addition, we account for extreme stock price effects and remove outliers at the 0.5 % level on both sides.

Altogether, our Newey-West corrected OLS regressions result in the following key findings:

- ***Confirming Hypothesis (H1):*** *News sentiment positively influences the noise residual of decomposed crude oil prices at a statistically significant level.*
  As seen in the regression column (1) in Table 3, the coefficient of news sentiment on the noise residual in a univariate model is positive and statistically significant with a P-value < 0.001. In addition, the adjusted R^2 of 0.172 suggests that news sentiment accounts for approximately 17 percent of the fluctuations in the noise residual. In a next step, we also extend the linear model by including all control variables, providing the results in regression column (7). The coefficient of news sentiment remains positive and highly significant with a P-value < 0.001. The adjusted R^2 increases to 0.185.

- ***Rejecting Hypothesis (H2):*** *As opposed to the original hypothesis, news sentiment also positively influences the daily returns of the fundamental oil price component at a statistically significant level.* This finding is in stark contrast to the assumption of the noise trader approach that sentiment does not influence informed investors (DeLong, Shleifer, Summers, and Waldmann, 1990; Lee, Shleifer, and Thaler, 1991): our results in Table 4 suggest that news sentiment is also a significant regressor on the fundamental crude oil price with a P-value < 0.001. This relationship remains robust even when adding control variables relevant to the oil market.

- ***Confirming Hypothesis (H3):*** *News sentiment has a stronger impact on the noise residual than the fundamental oil price component.*
  The coefficient of the news sentiment $S^*(t)$ is statistically significant at all common significance levels in both Newey-West corrected regressions on the noise residual and the fundamental price component. The t-statistic of the news sentiment $S^*(t)$ coefficient is higher in a Newey-West corrected regression on the noise residual (t-value of 14.526) than on the fundamental price component (t-value of 8.61). Hence, the comparison of the t-statistics of the two news sentiment $S^*(t)$ coefficients reveals that the effect of news sentiment is stronger on the noise residual. Similar evidence is found when comparing the magnitude of the corresponding coefficients. One standard deviation increase in the sentiment measure correlates positively with a change in fundamental oil prices by 0.2 %, while the same increase in news sentiment relates to an economically significant change in the noise residual by 65.8 %.

Table 3. Results of a Newey-West corrected OLS regression on the noise residual with news from January 6, 2004 until May 31, 2012.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	−0.008	−0.006	−0.005	1.618	2.074	2.110	2.159
	(−0.125)	(−0.061)	(−0.052)	(1.696)	(1.833)	(1.829)	(1.867)
$S^*(t)$	0.678***	0.678***	0.655***	0.645***	0.638***	0.654***	0.658***
	(15.428)	(16.159)	(15.436)	(14.766)	(14.607)	(14.589)	(14.526)
$r(t)$		−0.001	−0.002	0.036	0.035	0.037	0.038
		(−0.032)	(−0.069)	(0.955)	(0.918)	(0.965)	(0.974)
$FX(t)$			0.186***	0.193***	0.197***	0.213***	0.228***
			(3.304)	(3.587)	(3.735)	(3.612)	(3.73)
$IM(t)$				−0.006	−0.007	−0.007	−0.007
				(−1.728)	(−1.841)	(−1.845)	(−1.879)
$OI(t)$					−0.185	−0.179	−0.184
					(−0.859)	(−0.843)	(−0.843)
$G(t)$						−0.019	−0.016
						(−0.419)	(−0.393)
$SP(t)$							−0.025
							(−0.877)
Adj. R^2	0.172	0.172	0.179	0.181	0.178	0.184	0.185
AIC	7563.954	7565.95	7555.048	7547.713	7547.642	7550.104	7551.568
BIC	7580.883	7588.522	7583.263	7581.571	7587.143	7595.248	7602.355

Stated: Newey-West Corrected OLS Coeff., t-Stat. in Parenthesis; Obs.: 2108 Significance: *** 0.001, ** 0.01, * 0.05

Table 4. Results of a Newey-West corrected OLS regression on the daily returns of the fundamental WTI crude oil price, using news from January 6, 2004 until May 31, 2012.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	0.000	0.000	0.000	0.012*	0.018**	0.018**	0.018**
	(1.317)	(0.347)	(0.38)	(2.136)	(2.657)	(2.663)	(2.66)
$S^*(t)$	0.003***	0.003***	0.002***	0.002***	0.002***	0.002***	0.002***
	(8.126)	(7.973)	(8.074)	(8.589)	(8.662)	(8.672)	(8.61)
$r(t)$		0.000	0.000	0.000	0.000	0.000	0.000
		(0.84)	(0.83)	(1.889)	(1.824)	(1.838)	(1.843)
$FX(t)$			0.001*	0.001*	0.001*	0.001*	0.001*
			(2.417)	(2.263)	(2.228)	(2.463)	(2.385)
$IM(t)$				−0.000*	−0.000*	−0.000*	−0.000*
				(−2.08)	(−2.494)	(−2.498)	(−2.497)
$OI(t)$					−0.002	−0.002	−0.002
					(−1.793)	(−1.811)	(−1.792)
$G(t)$						−0.000	−0.000
						(−1.091)	(−0.768)
$SP(t)$							0.000
							(1.879)
Adj. R^2	0.227	0.23	0.234	0.245	0.257	0.258	0.26
AIC	−16457.769	−16463.458	−16474.059	−16519.301	−16551.888	−16551.260	−16552.858
BIC	−16440.841	−16440.888	−16445.846	−16485.446	−16512.390	−16506.120	−16502.075

Stated: Newey-West Corrected OLS Coeff., t-Stat. in Parenthesis; Obs.: 2107 Significance: *** 0.001, ** 0.01, * 0.05

4.3 Discussion

Overall, the above results are only partially consistent with the *noise trader approach*. While some findings support the concept of informed and uninformed investors, other findings reject – to a certain extent – the noise trader assumptions.

- **Findings on hypothesis (H1):** *in confirmation with the noise trader approach*, we observe a positive relationship between news sentiment and the noise residual, i. e. the price component driven by uninformed traders based on noisy signals. This finding corresponds with the noise trader theory that news sentiment represents a noisy signal (Brown, 1999; DeLong, Shleifer, Summers, and Waldmann, 1990; Lee, Shleifer, and Thaler, 1991). In addition, Brown and Cliff (2004) report empirical findings that individual investors (a proxy for uninformed investors) trade based on sentiment.
- **Findings on hypothesis (H2):** *in contrast to the noise trader approach*, we also observe a statistically significant positive effect of news sentiment on the de-noised oil price. This finding opposes the assumption of the noise trader approach that sentiment does not effect informed investors, since it is assumed that they trade based on fundamentals only. While the underlying assumption of the noise trader approach (DeLong, Shleifer, Summers, and Waldmann, 1990; Shleifer and Summers, 1990) does not foresee such an internalization of sentiment into the fundamental price, they acknowledge the potential empirical observation of so-called *positive feedback trading*: *"arbitrageurs optimally buy the stocks that positive feedback investors get interested in when their prices rise. When price increases feed the buying of other investors, arbitrageurs sell out near the top and take their profits."* An empirical study by Brown and Cliff (2004) supports this finding: they found that sentiment is not limited to individual (uninformed) investors. They report a positive relationship between institutional investor sentiment and market returns.
- **Findings on hypothesis (H3):** in addition, the effect of news sentiment is predominantly found on the noise residual. This confirms previous research; for example, the empirical findings of Barber

and Odean (2007) suggest that individual (less informed) investors trade more based on news than institutional (professional) investors. In contrast, Brown and Cliff (2004) suggest that their results indicate an even stronger effect of sentiment on institutional investors than on individual investors.

5 Conclusion and Outlook

The stock market crash in 1987, also known as *Black Monday*, dramatically exposed the limitations of the Efficient Market Hypothesis. The Efficient Market Hypothesis assumes that financial markets process and internalize new information instantly and efficiently. However, in the aftermath of Black Monday, financial market theorists developed models with underlying assumptions different to the Efficient Market Hypothesis with regards to information processing. For instance, DeLong, Shleifer, Summers, and Waldmann (1990) developed the *noise trader* approach. In this model, *uninformed investors*, named *noise traders*, trade upon *noisy signals*, such as news sentiment, and complement informed investors. This theory stipulates that noise traders do not invest based on information, but on sentiment signals. In contrast, informed investors form their trading decisions based on rational information processing only. The noise trader approach has been subject to significant subsequent research, for instance, on different investor types. In this paper, we investigate the influence of news sentiment in the oil market. We assess how news sentiment interferes with noise trading and the fundamental price respectively. For that purpose, we apply a Kalman filter to decompose the WTI crude oil price into both a (Kalman-smoothed) fundamental price component and a noise residual. Consequently, this price component should only reflect the fundamental price movement that is relevant for informed investors. In addition, we extracted sentiment from oil-related news and linked it to the aforementioned decomposed oil price components.

The results of this paper show that news sentiment influences both explanatory variables, namely, the noise residual and the fundamental oil price, at statistically significant levels. In addition, our findings reveal that news sentiment represents not only a relevant factor for noise traders, but also for informed traders. In combination, the findings indicate that the explanatory power of the sentiment measure is higher for the noise residual than for the fundamental price component. In fact, this lies in contrast to the assumption of the noise trader approach, since news sentiment positively impacts the returns of the fundamental oil price component.

This work opens avenues for future research. One evident subsequent research direction would be the investigation of which words from our underlying dictionaries particularly interact with the noise residual and the fundamental price respectively. The method of Bayesian Structural Time Series is a potential methodological toolkit to define such dictionaries relative to the respective decomposed price components. In addition, the differential impact of news sentiment on days with different observed de-noised oil price returns and noise residuals requires further investigations to better understand the role of news sentiment in explaining the movements of the de-noised oil price and noise residuals respectively. Finally, an intriguing approach were the reception of the analysis with a news corpus that also includes the recent years 2013 and 2014, which were not available for our analysis at the time.

References

Antweiler, W. and M. Z. Frank (2004). “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards.” *Journal of Finance* 59 (3), 1259–1294.

Barber, B. M. and T. Odean (2007). “All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors.” *Review of Financial Studies* 21 (2), 785–818.

Bencivenga, C., R. L. D’Ecclesia, and U. Triulzi (2012). “Oil Prices and the Financial Crisis.” *Review of Managerial Science* 6 (3), 227–238.

Black, F. (1986). “Noise.” *Journal of Finance* 41 (3), 529–543.

Bloomfield, R., M. O'Hara, and G. Saar (2009). "How Noise Trading Affects Markets: An Experimental Analysis." *Review of Financial Studies* 22 (6), 2275–2302.

Borovkova, S. and D. Mahakena (2015). "News, Volatility and Jumps: The Case of Natural Gas Futures." *Quantitative Finance*, 1–26.

Bosmana, R., R. Kräussl, and E. Mirgorodskaya (2015). "The "Tone Effect" of News on Investor Beliefs: An Experimental Approach." In: *FMA European Conference*.

Brogaard, J., T. Hendershott, and R. Riordan (2014). "High-Frequency Trading and Price Discovery." *Review of Financial Studies* 27 (8), 2267–2306.

Brown, G. W. (1999). "Volatility, Sentiment, and Noise Traders." *Financial Analysts Journal* 55 (2), 82–90.

Brown, G. W. and M. T. Cliff (2004). "Investor Sentiment and the Near-Term Stock Market." *Journal of Empirical Finance* 11 (1), 1–27.

Cambria, E., B. Schuller, Y. Xia, and C. Havasi (2013). "New Avenues in Opinion Mining and Sentiment Analysis." *IEEE Intelligent Systems* 28 (2), 15–21.

Carlson, M. A. (2007). "A Brief History of the 1987 Stock Market Crash With a Discussion of the Federal Reserve Response." *SSRN Electronic Journal*.

Cenesizoglu, T. (2014). "The Reaction of Stock Returns to News About Fundamentals." *Management Science*.

Chatrath, A., H. Miao, and S. Ramchander (2012). "Does the Price of Crude Oil Respond to Macroeconomic News?" *Journal of Futures Markets* 32 (6), 536–559.

Chen, H., R. R. L. Chiang, and V. C. Storey (2012). "Business Intelligence and Analytics: From Big Data to Big Impact." *MIS Quarterly* 36 (4), 1165–1188.

Cleveland, W. S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74 (368), 829–836.

Cleveland, W. S. and S. J. Devlin (1988). "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association* 83 (403), 596–610.

Dadvar, M., C. Hauff, and F. de Jong (2011). "Scope of Negation Detection in Sentiment Analysis." In: *Proceedings of the Dutch-Belgian Information Retrieval Workshop (DIR 2011)*. Amsterdam, Netherlands, pp. 16–20.

DeLong, B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). "Noise Trader Risk in Financial Markets." *Journal of Political Economy* 98 (4), 703–738.

Demers, E. A. and C. Vega (2010). "Soft Information in Earnings Announcements: News or Noise? INSEAD Working Paper No. 2010/33/AC." *SSRN Electronic Journal*.

Easley, D., M. M. López de Prado, and M. O'Hara (2011). "The Microstructure of the "Flash Crash": Flow Toxicity, Liquidity Crashes, and the Probability of Informed Trading." *The Journal of Portfolio Management* 37 (2), 118–128.

Fama, E. F. (1965). "The Behavior of Stock-Market Prices." *Journal of Business* 38 (1), 34–105.

— (1998). "Market Efficiency, Long-Term Returns, and Behavioral Finance." *Journal of Financial Economics* 49 (3), 283–306.

Feuerriegel, S., S. F. Heitzmann, and D. Neumann (2015). "Do Investors Read Too Much into News? How News Sentiment Causes Price Formation." In: *48th Hawaii International Conference on System Sciences (HICSS)*. IEEE Computer Society.

Feuerriegel, S., M. W. Lampe, and D. Neumann (2014). "News Processing during Speculative Bubbles: Evidence from the Oil Market." In: *47th Hawaii International Conference on System Sciences (HICSS)*, pp. 4103–4112.

Feuerriegel, S. and D. Neumann (2013). "News or Noise? How News Drives Commodity Prices." In: *Proceedings of the International Conference on Information Systems (ICIS 2013)*. Association for Information Systems.

Feuerriegel, S., A. Ratku, and D. Neumann (2015). "Which News Disclosures Matter? News Reception Compared Across Topics Extracted from the Latent Dirichlet Allocation." *SSRN Electronic Journal*.

Friedman, M. (1953). *Essays in Positive Economics*. Chicago, IL: University of Chicago Press.
Gerber, A., T. Hens, and B. Vogt (2002). "Rational Investor Sentiment." *SSRN Electronic Journal*.
Haven, E., X. Liu, and L. Shen (2012). "De-Noising Option Prices with the Wavelet Method." *European Journal of Operational Research* 222 (1), 104–112.
Hendershott, T. and A. J. Menkveld (2014). "Price Pressures." *Journal of Financial Economics* 114 (3), 405–423.
Hendershott, T., A. J. Menkveld, S. X. Li, and M. S. Seasholes (2013). *Asset Price Dynamics with Limited Attention*.
Henry, E. (2008). "Are Investors Influenced By How Earnings Press Releases Are Written?" *Journal of Business Communication* 45 (4), 363–407.
Hilbert, M. and P. López (2011). "The World's Technological Capacity to Store, Communicate, and Compute Information." *Science* 332 (6025), 60–65.
IBM (2013). *The Four V's of Big Data*. URL: `http://www.ibmbigdatahub.com/infographic/four-vs-big-data` (visited on 04/21/2014).
Jegadeesh, N. and D. Wu (2013). "Word Power: A New Approach for Content Analysis." *Journal of Financial Economics* 110 (3), 712–729.
Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering* 82 (1), 35–45.
Kilian, L. (2009). "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market." *American Economic Review* 99 (3), 1053–1069.
Kilian, L. and C. Vega (2011). "Do Energy Prices Respond to U.S. Macroeconomic News? A Test of the Hypothesis of Predetermined Energy Prices." *Review of Economics and Statistics* 93 (2), 660–671.
Kyle, A. (1985). "Continuous Auctions and Insider Trading." *Econometrica* 53 (6), 1315–1335.
Lechthaler, F. and L. Leinert (2012). *Moody Oil: What is Driving the Crude Oil Price?* Zurich.
Lee, C., A. Shleifer, and R. Thaler (1991). "Investor Sentiment and the Closed-End Fund Puzzle." *Journal of Finance* 46 (1), 75–109.
Lewis, D., Y. Yang, T. Rose, and F. Li (2004). "RCV1: A New Benchmark Collection for Text Categorization Research." *Journal of Machine Learning Research* 5, 361–397.
Li, F. (2010). "The Information Content of Forward-Looking Statements in Corporate Filings: A Naïve Bayesian Machine Learning Approach." *Journal of Accounting Research* 48 (5), 1049–1102.
Lopes, H. F. and R. S. Tsay (2011). "Particle Filters and Bayesian Inference in Financial Econometrics." *Journal of Forecasting* 30 (1), 168–209.
Loughran, T. and B. McDonald (2011). "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance* 66 (1), 35–65.
MacGregor, P. (2013). "International News Agencies: Global Eyes that Never Blink." In: *Journalism*. Ed. by K. Fowler-Watt and S. Allan. Centre for Journalism & Communication Research, Bournemouth University, pp. 35–63.
Malkiel, B. G. (2003). "The Efficient Market Hypothesis and Its Critics." *Journal of Economic Perspectives* 17 (1), 59–82.
— (2005). "Reflections on the Efficient Market Hypothesis: 30 Years Later." *The Financial Review* 40 (1), 1–9.
Malkiel, B. G. and E. F. Fama (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25 (2), 383–417.
Manning, C. D. and H. Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
Medhat, W., A. Hassan, and H. Korashy (2014). "Sentiment Analysis Algorithms and Applications: A Survey." *Ain Shams Engineering Journal* 5 (4), 1093–1113.
Mendel, B. and A. Shleifer (2012). "Chasing Noise." *Journal of Financial Economics* 104 (2), 303–320.
Minev, M., C. Schommer, and T. Grammatikos (2012). *News and Stock Markets: A Survey on Abnormal Returns and Prediction Models*. Luxembourg.

Mittermayer, M.-A. and G. F. Knolmayer (2006a). “NewsCATS: A News Categorization and Trading System.” In: *Sixth International Conference on Data Mining (ICDM’06)*, pp. 1002–1007.

— (2006b). *Text Mining Systems for Market Response to News: A Survey*. Bern, Switzerland.

Nassirtoussi, A. K., S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo (2014). “Text Mining for Market Prediction: A Systematic Review.” *Expert Systems with Applications* 41 (16), 7653–7670.

Pang, B. and L. Lee (2008). “Opinion Mining and Sentiment Analysis.” *Foundations and Trends in Information Retrieval* 2 (1–2), 1–135.

Paterson, C. (2007). “International News on the Internet: Why More is Less.” *Ethical Space: The International Journal of Communication Ethics* 4 (1/2), 57–66.

Pröllochs, N., S. Feuerriegel, and D. Neumann (2015a). “Enhancing Sentiment Analysis of Financial News by Detecting Negation Scopes.” In: *48th Hawaii International Conference on System Sciences (HICSS)*. IEEE Computer Society.

— (2015b). “Generating Domain-Specific Dictionaries Using Bayesian Learning.” In: *23rd European Conference on Information Systems (ECIS 2015)*. Münster, Germany.

Ratku, A., S. Feuerriegel, and D. Neumann (2014). “Finding Evidence of Irrational Exuberance in the Oil Market.” In: *"New" Emergent Trends of IT in the Finance Industry (FinanceCom 2014)*. Lecture Notes in Business Information Processing. Springer.

Sanders, D. R., S. H. Irwin, and R. M. Leuthold (1997). “Noise Traders, Market Sentiment, and Futures Price Behavior.” *SSRN Electronic Journal*.

Schumaker, R. P. and H. Chen (2009). “Textual Analysis of Stock Market Prediction using Breaking Financial News.” *ACM Transactions on Information Systems* 27 (2), 1–19.

Schwartz, E. and J. E. Smith (2000). “Short-Term Variations and Long-Term Dynamics in Commodity Prices.” *Management Science* 46 (7), 893–911.

Sharma, A. and S. Dey (2012). “A Comparative Study of Feature Selection and Machine Learning Techniques for Sentiment Analysis.” In: *Proceedings of the 2012 Research in Applied Computation Symposium (RACS 2012)*. New York, NY: ACM, pp. 1–7.

Shleifer, A. (2000). *Inefficient Markets: An introduction to behavioral finance*. Clarendon Lectures in Economics. Oxford: Oxford University Press.

Shleifer, A. and L. H. Summers (1990). “The Noise Trader Approach to Finance.” *Journal of Economic Perspectives* 4 (2), 19–33.

Shleifer, A. and R. W. Vishny (1997). “The Limits of Arbitrage.” *Journal of Finance* 52 (1), 35–55.

Siering, M. (2013). “Investigating the Impact of Media Sentiment and Investor Attention on Financial Markets.” In: *Enterprise Applications and Services in the Finance Industry*. Vol. 135. Lecture Notes in Business Information Processing. Berlin and Heidelberg: Springer, pp. 3–19.

Sornette, D. (2003). *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton, NJ: Princeton University Press.

Tetlock, P. C. (2007). “Giving Content to Investor Sentiment: The Role of Media in the Stock Market.” *Journal of Finance* 62 (3), 1139–1168.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy (2008). “More than Words: Quantifying Language to Measure Firms’ Fundamentals.” *Journal of Finance* 63 (3), 1437–1467.

Thaler, R. H. (2005). *Advances in Behavioral Finance*. Vol. 2. Roundtable Series in Behavioral Economics. Princeton, NJ: Princeton University Press.

Welch, G. and G. Bishop (1997). *An Introduction to the Kalman Filter*. Chapel Hill, NC.

Wong, J. (2010). *Market Predictions using Sentiment Analysis and State-Space Models*. Stanford, CA.

Yan, H. (2010). “Is Noise Trading Cancelled Out by Aggregation?” *Management Science* 56 (7), 1047–1059.