DO PESSIMISTS MOVE ASSET PRICES? EVIDENCE FROM APPLYING PROSPECT THEORY TO NEWS SENTIMENT

Submitted to EFMA 2016

Alfano, Simon, University of Freiburg, Platz der alten Synagoge, 79098 Freiburg, Germany, simon.alfano@is.uni-freiburg.de

Feuerriegel, Stefan, University of Freiburg, Platz der alten Synagoge, 79098 Freiburg, Germany, stefan.feuerriegel@is.uni-freiburg.de

Neumann, Dirk, University of Freiburg, Platz der alten Synagoge, 79098 Freiburg, Germany, dirk.neumann@is.uni-freiburg.de

Abstract

Behavioral finance research increasingly endorses Information Systems (IS) methodologies, such as news sentiment analysis, to empirically study information processing. One intriguing application is the asymmetric processing of information as described by the negativity bias and its economic derivative, prospect theory, according to which negative information outweighs positive. We test for asymmetric information processing by measuring the influence of news sentiment on the WTI crude oil price and trading volume. We use quantile regressions to gauge the differential sentiment response subject to the rate of change in prices and trading volumes. Additionally, we study the differential role of news sentiment for different investor types as stipulated by the noise trader theory. Methodologically, we perform a Kalman decomposition of prices into a fundamental price (representing informed investors) and a noise residual (representing uninformed investors). Our findings provide strong evidence for the presence of the negativity bias in the price formation, but not for trading volumes. Furthermore, our evaluation suggests that both informed and uninformed traders process information asymmetrically.

Keywords: Behavioral finance, information processing, negativity bias, noise trader theory, prospect theory

Introduction

Behavioral finance builds on the assumption that not all human agents in financial markets behave fully rational (Constantinides et al. 2003; Thaler 2005). Research from cognitive psychology sheds light on the central assumption of the restricted rationality of human behavior (Allais 1953; Edwards 1954; Kahneman & Tversky 1979; Tversky & Kahneman 1974). Behavioral finance consists of two building blocks that describe deviations from full rationality: psychology and limits to arbitrage (Constantinides et al. 2003; Thaler 2005). Behavioral finance studies psychological phenomena to explain human behavior that deviates from full rationality, such as asymmetric risk preferences, in which losses are weighed differently from gains (Kahneman & Tversky 1979, 1984; Tversky & Kahneman 1986, 1992), or the asymmetric evaluation of negative and positive information (Kanouse & Hanson 1972; Tversky & Kahneman 1981).

Tversky and Kahneman (1974) introduced the term *cognitive bias* to describe such psychological phenomena that cause irrational human behavior. One such irrational behavioral trait is that humans weigh negative information more heavily than positive information, which Kanouse and Hanson (1972) formalize as *negativity bias*. Kahneman and Tversky (1979) describe the influence of the negativity bias on economic decision-making in the so-called prospect theory. Prospect theory stipulates that human agents have asymmetric risk preferences. The core concept of prospect theory is that agents make decisions based on potential losses and gains rather than the final outcome and also tend to weigh losses more heavily than gains.

To account for different information horizons of investors (e. g. professional versus private investors), DeLong et al. (1990) developed the so-called *noise trader*. The noise trader approach reflects the fact that not all agents in financial markets have the same information horizon. In this model, traders that do not have complete information and do not behave in a wholly rational manner are called noise traders. Noise traders are not fully rational and trade on noisy signals such as news sentiment. Extensions of the noise trader model (DeLong et al. 1990; Shleifer 2000; Thaler 2005) argue that arbitrage is limited; i. e. the trading of rational investors, who are not subject to sentiment, does not drive all noise traders out of the market. In addition, Shleifer and Summers (1990) highlight that some rational investors may mimic noise traders as so-called positive feedback traders in order to take a risky position in a sentiment-driven market rally and try to sell their risky position earlier than noise traders in anticipation of a price shock. Presumably, trader types with different information availability react with a different magnitude to new information.

Consequently, this paper aims to shed light on the differential impact of positive and negative news sentiment on financial markets. First, we evaluate the differential impact of news sentiment (a measure of the tone in financial news) on oil price returns and trading volumes. Second, we study the effect of news sentiment separately for informed and uninformed investors. To study the effect of news sentiment on different investor types, we study how investors in the oil market process new oil-market-relevant information. We select the oil price for this purpose due to the high economic relevance of oil and the high liquidity of the oil market. To model these two different investor types, we decompose the oil price with a Kalman filter, widely used in finance research (e. g. Alfano et al. 2015; Johnson & Sakoulis 2008; Owens & Steigerwald 2006; Zhang & Semmler 2009), into a fundamental price component (i. e. for an informed investor) and a noise residual (i. e. for an uninformed investor; caused by noise traders deviating from the fundamental market price). This approach follows the noise trader approach in disaggregating the underlying price curve into a fundamental price and a noise component for informed and uninformed investors, respectively (DeLong et al. 1990; Shleifer & Summers 1990).

Following the noise trader approach, news sentiment should influence both investor types. Uninformed investors should react more strongly to news sentiment than informed traders based on the noise trader approach. Furthermore, both investor types should react more strongly to negative than to positive news when accounting for asymmetric information processing. We then investigate with a quantile regression the differential impact of (positive vs. negative) news sentiment on the fundamental price and the noise residual. We are not aware of any research that empirically investigates prospect theory for different investor types, as defined by the noise trader approach, by utilizing news sentiment. We focus our analysis on the oil price, since oil represents a pivotal commodity in global markets. In addition, previous research on the influence of news (sentiment) on oil prices has opened avenues for further research (e. g. Alfano et al. 2015; Narayan & Narayan 2016).

The remainder of this paper is organized as follows: Section 2 reviews literature on asymmetric information processing and the noise trader approach. In this theoretical context, we derive our research hypotheses. We review our data sources and our methodological approach in Section 3. We present the results of our analysis on the differential impact of news sentiment in Section 4. Section 5 discusses the findings and outlines their managerial implications. Section 6 concludes and provides a research outlook.

Related Work and Hypotheses Development

This section introduces related behavioral research on asymmetric information processing and noise trader theory. Taken together, this body of information motivates the derivation of our research hypotheses.

Asymmetric Information Processing

Behavioral finance is an alternative concept to classical economic theory, which holds expected utility theory as its central paradigm (Fama 1965; Friedman 1953). Concepts that broke with classical economic theory, and its firmly established paradigm of fully rational decision-making, co-existed concurrently. For instance, Simon (1955, 1962) states that human cognitive capacities are bounded and thus decision-making rather follows heuristics based on a limited sub-set of overall information and choices than based on the full universe of available information. However, it was only after the stock market crash of October 19, 1987, that economics started to broadly recognize and adopt behavioral economics approaches (Thaler 2005). Within behavioral economics, cognitive biases are a class of observed, economic decision-making patterns (Tversky & Kahneman 1974) that do not follow purely rational decision-making.

In this paper, we focus on the class of cognitive biases that describes asymmetric processing of positive and negative information. Peeters (1971) formulates the basic logic of asymmetric information processing. Therein, the negativity bias (Kanouse 1984; Kanouse & Hanson 1972) describes the phenomena whereby human agents assign greater emphasis on average to negative than positive stimuli, as in regard to news. Peeters and Czapinski (1990) summarize the evidence for the negativity bias presented by Kanouse and Hanson (1972) in two observations: (1) in decision-making under risk, potential costs are more heavily weighted than potential gains and (2) in overall evaluations, negative information is weighted more heavily than positive information. Rozin and Royzman (2001) formalize the negativity bias along four dimensions: *(i) negative potency*, that is negative events are experienced more strongly than corresponding positive events; *(ii) steeper negative gradients*, that is the perceived negativity of negative events increases more quickly when approaching the event in time than does the positivity of positive events; *(iii) negativity dominance*, which describes the tendency to negatively skew the interpretation of a combination of negative and positive items, such as words, and *(iv) negative differentiation*, which acknowledges that we have more concepts and a larger vocabulary for negative entities and events.

Behavioral economics explores the negativity bias in the context of economic decision-making. When human agents have to make a choice in which they face the possibility of an equivalent gain or loss subject to a risky outcome, they tend to weigh potential costs more heavily than potential gains as formalized in the so-called prospect theory (Kahneman & Tversky 1979, 1984; Tversky & Kahneman 1986, 1992). According to prospect theory (Kahneman & Tversky 1979), *"the value function is normally concave for gains, convex for* *losses and generally steeper for losses than for gains".* Previous research has focused on extending prospect theory (Barberis 2012; Camerer et al. 2004; Grinblatt & Han 2005; Wakker 2010).

When applying the concept of differential processing of negative and positive information to financial markets, Brown and Cliff (2005) argue that this effect originates from limits to arbitrage: *"practical limitations to short-selling activity may make it difficult for rational investors to prevent market prices from being pushed above their intrinsic value during periods of excessive optimism. On the other hand, when some investors are especially pessimistic, no similar frictions prevent arbitrageurs from taking the necessary long position."* In addition, Akhtar et al. (2011) take a portfolio view to explain why investors weigh negative news more heavily than positive news. They assume that investors can allocate funds into risky (stocks) and safe (bonds) assets. The assumption is that investors perceive news as *good* in the default position and do not change their portfolio if they do not perceive any novel news as *bad*. However, if new information qualifies as *bad*, investors quickly re-balance their portfolio and increase their position in bonds, i. e. reduce their exposure to stocks. When facing uncertainty, this screening strategy is safest since it favors false negative decisions over false positive decisions. If news is truly *bad* the investor correctly shifts assets. Conversely, while investors miss out on the stock risk premium if the information is not actually *bad*, they still can re-balance their portfolio towards riskier stocks.

Prospect theory as a concept of behavioral finance, as well as empirical evidence on reversals and momentum, e. g. Hong et al. (2000), suggests that the influence of sentiment appears to be asymmetric. In other words, negative news may have a stronger impact on stock market returns than positive news. Within the context of stock market announcements, Tetlock (2007) measures the influence of news sentiment on stock returns, which he reports are highly reactive to negative words. Feuerriegel and Neumann (2013) indicate that this also holds true for the oil and gold commodity markets. Their results show a significantly higher impact of negative than positive words on abnormal returns. Furthermore, Akhtar et al. (2011, 2013) report a significant negative announcement effect in stock prices in response to a decrease in consumer sentiment. However, they do not observe a significant effect of an increase in consumer sentiment. Loughran and McDonald (2013) find evidence supporting prospect theory in the context of IPO filings. They state that higher numbers of negative words generate predictable upward offer price revisions. Thus, investors may utilize language with a negative tone in order to mitigate day-1 losses.

We are not aware of a theory that formalizes the differential impact of news sentiment for varying return levels. Liebmann et al. (2012), Schumaker et al. (2012) indicate a positive relationship between news sentiment and returns. In addition, Feuerriegel and Neumann (2013) for commodities, along with Schumaker et al. (2012), Tetlock (2007) for stocks, report a stronger effect of negative news on asset prices. Akhtar et al. (2011, 2013) find evidence that negative consumer sentiment exerts a downward effect on stock prices, while there is no significant market reaction to positive consumer sentiment. Stieglitz and Dang-Xuan (2013) find that negative sentiment induces Twitter users to re-tweet messages in greater quantity than those with a positive sentiment. Prospect theory suggests that agents dislike a negative prospect more than they value a positive prospect (Kahneman & Tversky 1979, 1984). This, in turn, implies that news sentiment has a stronger effect on negative than on positive returns.

Thus, in order to extend prior research on asymmetric information processing by investors, we stipulate the following two first research hypotheses:

Hypothesis (H1a): The effect of news sentiment on the (non-Kalman-decomposed) real return of the oil price is stronger on negative return days (percentiles) than on positive return days (percentiles), given that negative information outweighs positive information according to the negativity bias, as reflected in the findings of Akhtar et al. (2011, 2013), Stieglitz and Dang-Xuan (2013).

Hypothesis (H1b): The effect of news sentiment on the rate of change in the oil trading volume is stronger on days with a positive change in oil trading volume than on days with a negative change in oil trading volume (Akhtar et al. 2011, 2013; Stieglitz & Dang-Xuan 2013).

Noise Trader Theory

Triggered by the stock market crash of October 19, 1987, finance research has increasingly focused on extensions of the Efficient Market Hypothesis to account for uninformed traders (DeLong et al. 1990; Shleifer & Summers 1990), who follow irrational trading patterns. Since uninformed investors trade upon incomplete information signals, they are called *noise traders* (Black 1986; Kyle 1985). Based on the idea of uninformed traders, Shleifer and Summers (1990) developed the *noise trader* approach, which consists of two investor types. **Uninformed (noise) investors** are those whose trading of risky assets is (1) not fully justified by fundamentals and (2) influenced by sentiment. In this model, sentiment describes the over- and underestimation of expected returns as compared to rational expectations. The news sentiment of financial disclosures is one sentiment driver. **Informed investors** are those who are not influenced by sentiment. They have a full information horizon at their disposal and build fully rational expectations. Yet, Lee et al. (1991) state that the sentiment of noise traders is stochastic. Thus, rational investors cannot perfectly predict sentiment. 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 have further developed the concept of sentiment as a noisy signal affecting noise traders (Bloomfield et al. 2009; Brown 1999; Brown & Cliff 2004; Lee et al. 1991; Mendel & Shleifer 2012; Sanders et al. 1997; Shleifer 2000; Shleifer & 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 et al. 2002). Recent publications contribute to the body of research by improving our understanding of noise traders: Bloomfield et al. (2009) report that noise traders can actually have positive effects on financial markets, as their experimental results show that higher market liquidity increases market volume, as well as depth, and reduces spreads. Mendel and Shleifer (2012) show that uniformed traders can amplify sentiment shocks and increase the distance from fundamental prices. Thus, we stipulate the following hypothesis:

Hypothesis (H2a): The effect of news sentiment on the noise residual percentiles is stronger for negative return days (percentiles) as predicted by the negativity bias, in line with Feuerriegel and Neumann (2013).

DeLong et al. (1990), Shleifer and Vishny (1997) introduced the notion of noise trader risk. Noise trader risk describes the possibility that the mispricing of an asset may worsen in the short run due to the potential of sentiment signals to reinforce the beliefs of noise traders. Thus, in the short run, arbitrageurs (informed traders) face the possibility of asset prices further deviating from their fundamental price. Therefore, arbitrageurs are a) risk-averse and have a limited appetite for taking positions against noise traders and may b) even follow sentiment trading in the short run and try to sell before sentiment reverts (the so-called bandwagon effect) (DeLong et al. 1990; Shleifer & Summers 1990). An empirical study confirms a significant positive relationship of news sentiment both with informed and uninformed investors (Alfano et al. 2015). Hence:

Hypothesis (H2b): The effect of news sentiment on the (Kalman-filtered) fundamental oil price is only marginally significant and stronger on negative return days (percentiles). This effect reflects the bandwagon effect whereby rational investors may follow noise investors in the short run (DeLong et al. 1990; Shleifer & Summers 1990).

Overall, we assume the following relative influence of news sentiment on different investor types:

Hypothesis (H3): The effect of news sentiment is stronger on noise residual than on the fundamental price component across all return percentiles. This reflects the fact that informed investors trading on fundamental information are less prone to sentiment than uninformed noise traders (DeLong et al. 1990; Shleifer & Summers 1990).

Methodology and Data Sources

This section introduces the applied methodology and our dataset. As shown in Figure 1, our approach consists of two steps. *First*, we decompose the WTI crude oil price into two components: the fundamental oil price and a noise residual. This follows Schwartz and Smith (2000), who used the Kalman filter to decompose the oil price into fundamental and noise components. Similarly, Neumann et al. (2006) studied gas prices using the Kalman filter. Furthermore, Brown and Cliff (2004) Kalman-filtered investor sentiment to study the link between investor sentiment and near-term stock market returns. Appendix A provides more details on the Kalman filter. We are aware of an extensive body of research that investigates the influence of fundamental variables on oil prices (Baumeister & Kilian 2012; Feuerriegel et al. 2015; Kilian 2009; Lechthaler & Leinert 2012; Zivot & Andrews 2002). *Second*, we regress the return of the oil price and trading volume as well as the two Kalman-filtered oil price components with quantile regressions (see Appendix B for a detailed discussion) on news sentiment and a set of standard control variables in oil-market-related research. In order to establish causality and address the endogeneity problem, we further run a two-stage least square (2SLS) regression.

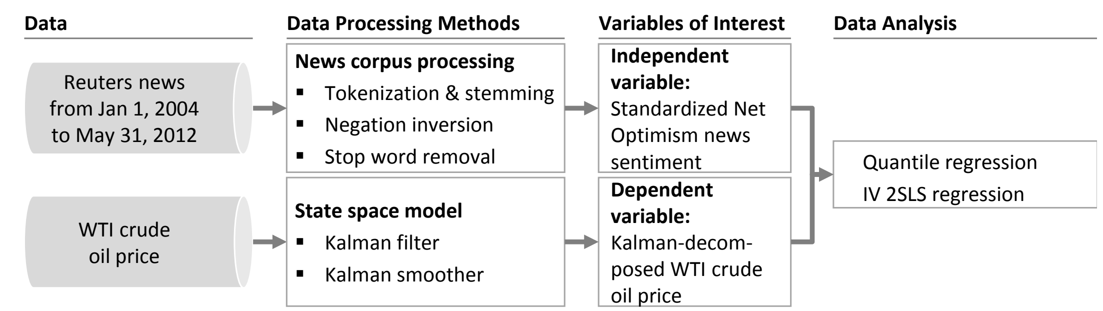

Figure 1. Research Framework Shows How this Paper Evaluates the Differential Influence of News Sentiment on Fundamental Oil Prices and Noise Residuals via Quantile Regressions

In order to aggregate the tone in financial news as a measure of news sentiment, we build on a method, drawn from the text mining domain, referred to as sentiment analysis. In the following, we introduce its high-level idea. Sentiment analysis often utilizes a text mining process to derive information from textual information (Manning & Schütze 1999; Medhat et al. 2014; Nassirtoussi et al. 2014). Since text appears in the form of unstructured data, the first step is to transform it into a machine-readable representation. A feature selection leads to this data representation, where relevant features (e. g. single terms or bag-of-words) are extracted. The extracted features are then input to form a decision algorithm, such as a machine learning method or a rule-based approach, to calculate the subjective news sentiment. For a detailed description of our sentiment analysis approach, please refer to Appendix C.

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 suited to the purpose of evaluating stock market reactions. The provided Reuters announcements span a period from January 6, 2004 to May 31, 2012. We only include business days, providing a total of 2,112 observation days. Furthermore, the Thomson Reuters news corpus enables us to effectively gather all announcements related to crude oil in English language, while automatically removing personal opinions or alerts. The information content of opinions and alerts may 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 et al. (2015), 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 2,051 values are positive and 61 negative. We standardize the daily sentiment scores as described in Appendix C.

In this paper, we select crude oil for a deliberate purpose. In light of increasing 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.

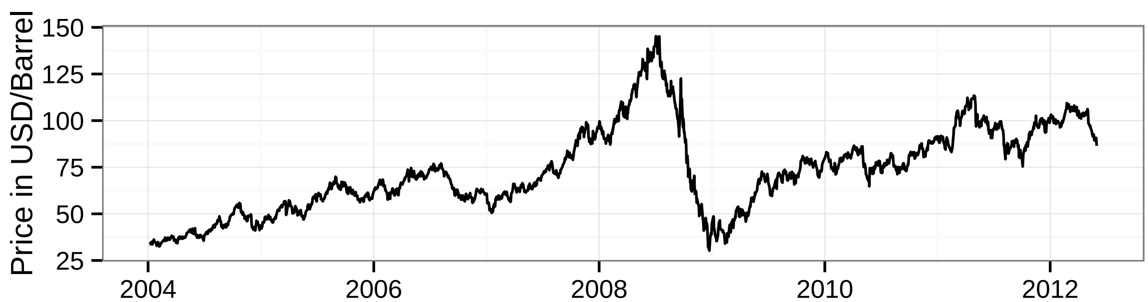

Figure 2. Nominal Western Texas Intermediate (WTI) crude oil price from January 2004 to May 2012.

Following prior oil-related research, we select the Western Texas Intermediate (WTI) crude oil price for our analysis (Bencivenga et al. 2012; Chatrath et al. 2012; Kilian & Vega 2011). Figure 2 depicts the WTI crude oil price curve for our study period. Moreover, consistent with prior research (Feuerriegel et al. 2015; Kilian 2009; Lechthaler & Leinert 2012), we add the following fundamentals as control variables to our econometric model: (a) the U. S. interest rate, (b) the U. S. dollar/euro exchange rate, (c) level of oil imports (in million barrels), (d) total open interest in crude oil future contracts (in million), (e) the gold price (London, afternoon fixing) and (f) the S&P 500 index (included additionally). All financial data originates from Thomson Reuters Datastream.

The corresponding descriptive statistics are provided in Table 1.

Differential Information Processing Analysis

In this section, we investigate whether investors process news sentiment asymmetrically in line with behavioral economics theories. Behavioral economics theory suggests that the influence of news sentiment varies, for example, when comparing days with positive/negative returns. To investigate the influence of news sentiment on different oil price return days, we methodologically deploy quantile regressions. Quantile regressions perform separate regressions in which the bandwidth of point estimates can shift across the full percentile continuum (e. g. the conditional median or any other quantile of the dependent variable) (Koenker 2005). This differs (Koenker 2005) from conventional OLS methods that fit a model to a dataset by minimizing squared deviations from a conditional mean; a quantile regression minimizes squared deviations within a defined quantile. We run quantile regressions at the $25th$, $50th$ and $75th$ percentiles. In each quantile regression, the optimization minimizes the distance of our observation values to the corresponding quantile.

Table 1. Descriptive Statistics of the Dependent and Independent Variables Ranging From January 6, 2004 Until May 31, 2012

Variable		Freq.	Mean	Median	Min.	Max.	Std. Dev.	Skew.	Kurt.
$P(t)$	WTI Oil Price	Daily	73.155	71.39	22.806	30.280	145.310	0.434	-0.178
$p(t)$	WTI Oil Price Return	Daily	0.084	0.091	2.513	-12.038	17.838	0.229	4.952
$P_{\text{Kalman}}(t)$	Fundamental Oil Price (Return)	Daily	0.049	0.086	0.558	-2.433	1.766	-0.758	1.789
$N_{\text{Kalman}}(t)$	Noise Residual	Daily	0.001	0.032	1.811	-10.214	19.082	0.395	7.908
$v(t)$	WTI Oil Trading Volume (Return)	Daily	36.597	-1.578	-99.602	22,736.400	517.600	40.102	1,749.700
$v(t)_{log}$	WTI Oil Trading Volume (Log-Return)	Daily	0.001	-0.016	-5.526	5.431	0.591	0.896	13.214
$S^*(t)$	News Sentiment (Standardized Net-Optimism)	Daily	0.001	0.011	1	-3.978	4.171	-0.101	0.745
$r(t)$	U. S. Interest Rate	Monthly	1.818	1.3	1.847	0.01	5.01	0.556	-1.284
$FX(t)_{log}$	U. S. Dollar/Euro Exchange Rate (Log-Return)	Daily	0.007	0.02	0.66	-4.625	4.12	-0.073	3.166
$IM(t)$	Oil Imports (in Million Barrels)	Monthly	292.999	297.814	22.266	229.14	327.476	-0.443	-0.669
$OI(t)$	Open Interest in Crude Oil Futures (in Million)	Weekly	1.177	1.23	0.279	0.598	1.654	-0.495	-0.951
$G(t)_{log}$	Gold Price (Log-Return)	Daily	0.113	0.107	1.037	-5.132	6.304	0.101	3.936
$SP(t)_{log}$	S&P 500 Index (Log-Return)	Daily	0.017	0.075	1.363	-9.035	11.580	-0.053	10.402

The Influence of Different News Sentiment Levels on Oil Prices and Trading Volumes

To analyze the differential influence of news sentiment on oil prices and oil trading volumes, we utilize the oil price model developed by Kilian (2009), Kilian and Park (2009). In the domain of oil-related news, Lechthaler and Leinert (2012) analyzed the influence of news sentiment on monthly crude oil returns with this oil price model. Thus, we define our regression model including relevant oil price control variables as described by Kilian (2009), Lechthaler and Leinert (2012). In order to address our hypotheses regarding differential information processing, we analyze the influence of news sentiment on the return of the WTI oil price $p_{\text{Return}}(t)$ in *Hypothesis 1a* and on the WTI oil trading volume $v_{\text{Return}}(t)$ based on the following regression model

$$\left.\begin{matrix} p_{\text{Return}}(t) \\ v_{\text{Return}}(t) \end{matrix}\right\} = \beta_0 + \beta_1 S^*(t) + \beta_2 r(t) + \beta_3 FX(t)_{log} + \beta_4 IM(t) + \beta_5 OI(t) + \beta_6 G(t)_{log} + \beta_7 SP(t)_{log} + \varepsilon_t \quad (1)$$

where the standardized news sentiment $S^*(t)$ is our independent variable of interest. In addition, we include further control variables in line with previous research (Feuerriegel et al. 2015; Kilian 2009; Lechthaler & Leinert 2012): the U. S. interest rate $r(t)$, the log-return of the U. S. dollar/euro exchange rate $FX(t)_{log}$, the level of oil imports $IM(t)$, the open interest $OI(t)$, the log-return of the gold price $G(t)_{log}$ and the log-return of the S&P 500 index $SP(t)_{log}$.

An initial visualization indicates a positive relationship between news sentiment and the real return of the WTI oil price. Figure 3 features a so-called LOWESS trend line with a 95 % confidence band, i. e. a *locally weighted scatterplot smoothing calculated with local regressions* (Cleveland 1979; Cleveland & Devlin 1988). The LOWESS trend line highlights a positive relationship between news sentiment and the the real return of the WTI oil price.

Our quantile regressions, as defined in Equation (1), yield the following key findings on information processing:

- ***Confirming Hypothesis (H1a):*** *News sentiment has a positive effect on the real return of the WTI oil price.*
  The quantile regression coefficients are positive on all quantiles as represented in Table 2. In line with

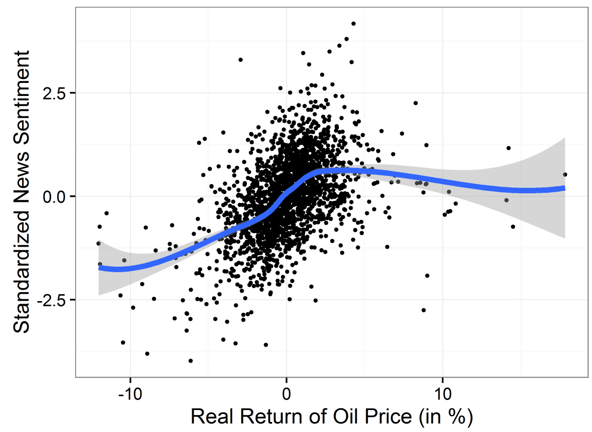

Figure 3. News Sentiment Plotted Against and the Real Returns of the WTI Oil Price Inclusing a LOWESS Trend Line with Smoothing Parameter $f = 0.5$ and a 95 % Confidence Band.

the negativity bias and our hypothesis, the effect size of news sentiment is 22 percent larger on the $25th$ than on the $75th$ quantile (coefficient of 1.094 vs. 0.896), i. e., the effect of news sentiment is greater on days with negative returns than on days with positive returns. The regression coefficients of news sentiment are all statistically significant (p-value < 0.001). The *left* graph of Figure 4 shows graphically the regression coefficients of news sentiment from quantile regressions on every $10th$ quantile of the real returns of the WTI oil price. Additionally, we run a Wald test to gauge the equality of the respective news sentiment coefficients as computed for the $25th$, $50th$ and $75th$ quantile regressions, as suggested by Koenker (2005). The Wald test confirms that the quantile regression coefficients are significantly different (F-test 3.987; P-value < 0.001) for a regression on the WTI oil price real return across the quantiles and thus indicates that the difference in coefficients is robust.

- ***Partly rejecting Hypothesis (H1b):*** *While the coefficients are directionally in line with our hypothesis, we do not find a statistically significant effect of news sentiment on the change in the WTI oil trading volume.* The results in Table 3 show that we cannot confirm our hypothesis due to insignificant effect sizes. The relative effect size, i. e. the coefficients of our quantile regressions, are keeping with our hypothesis, whereas more negative sentiment has a positive effect on an increase in trading volume, i. e. negative sentiment multiplied with the negative coefficient in the $75th$ percentile leads to an increase in trading volume, as similarly observed by Stieglitz and Dang-Xuan (2013) with regard to Twitter news. While the regression coefficients in the WTI oil trading volume are not statistically significant, a Wald test reveals that their levels are significantly different (F-test 2.220; P-value < 0.001) across the different quantiles, i. e. the coefficient on the $25th$ quantile is significantly larger than for the $50th$ and $75th$ quantiles. The *right* graph of Figure 4 shows graphically the regression coefficients of news sentiment from quantile regressions on every $10th$ quantile of the change in WTI oil trading volume.

Kalman Filtering

In this section, we discuss the results of our price decomposition analysis, as well as the influence of *news sentiment* on the fundamental oil price and its noise residual at different return quantiles. The analysis at different return quantiles allows us to identify whether different investor types process sentiment asymmetrically. We proceed as follows:

1. We decompose the oil price into the fundamental oil price and its noise residual

Table 2. Hypothesis (1a): Coefficients of Quantile Regressions on the Return of WTI Oil Price

	(a) OLS	(b) 25 % Quant.	(c) 50 % Quant.	(d) 75 % Quant.
Intercept	1.935^{*} (2.221)	-1.013 (-1.02)	0.441 (0.448)	2.446^{*} (2.138)
News Sentiment $S^{*}(t)$	1.015^{***} (22.38)	1.094^{***} (19.56)	0.983^{***} (20.113)	0.896^{***} (14.895)
U. S. Interest Rate $r(t)$	0.078 (0.732)	-0.001 (-0.423)	-0.001 (-1.578)	-0.001 (-0.799)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$	0.146^{*} (2.000)	0.142 (1.575)	0.147^{*} (1.938)	0.197^{*} (2.101)
Level of Oil Imports $IM(t)$ (in Million Barrel)	-0.002 (-0.84)	0.002 (0.519)	0.002 (0.521)	-0.001 (-0.371)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)	-1.636^{**} (-3.034)	-0.807 (-1.251)	-1.07^{*} (-1.993)	-1.025 (-1.546)
Gold Price $G(t)_{log}$	0.033 (0.733)	-0.036 (-0.617)	0.009 (0.208)	0.040 (0.701)
S&P 500 Index $SP(t)_{log}$	0.386^{***} (12.148)	0.361^{***} (7.479)	0.322^{***} (8.977)	0.385^{***} (8.756)
Stated: Coeff., t-Stat. in Parentheses	Dummies: Yearly	Significance: *** 0.001, ** 0.01, * 0.05		

Table 3. Hypothesis (1b): Coefficients of Quantile Regressions on the Change in WTI Oil Trading Volume

	(a) OLS	(b) 25 % Quant.	(c) 50 % Quant.	(d) 75 % Quant.
Intercept	20.745 (0.357)	-29.647 (-1.816)	2.192 (0.160)	32.006 (1.640)
News Sentiment $S^{*}(t)$	-2.904 (-0.956)	0.582 (0.740)	-0.471 (-0.730)	-1.418 (-1.560)
U. S. Interest Rate $r(t)$	7.971 (1.116)	0.007 (0.670)	-0.003 (-0.298)	-0.009 (-0.515)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$	2.058 (0.429)	-2.522^{*} (-2.050)	-2.200^{*} (-2.119)	2.211 (1.902)
Level of Oil Imports $IM(t)$ (in Million Barrels)	0.087 (0.483)	0.041 (0.838)	0.004 (0.110)	0.011 (0.199)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)	-72.797^{*} (-1.989)	-4.815 (-0.521)	-3.731 (-0.508)	-20.859 (-1.652)
Gold Price $G(t)_{log}$	0.630 (0.733)	1.221 (1.591)	0.553 (0.878)	-1.697 (-1.804)
S&P 500 Index $SP(t)_{log}$	2.416 (1.142)	0.358 (0.753)	-0.700 (-1.517)	-1.103^{*} (-2.019)
Stated: Coeff., t-Stat. in Parentheses	Dummies: Yearly	Significance: *** 0.001, ** 0.01, * 0.05		

2. We investigate the interference of news sentiment with our decomposed oil price components at different return levels to study the cognitive negativity bias
3. We assess whether the effects are significantly different across our modeled investor types

We decompose the WTI crude oil price time series into a fundamental price component and a noise residual with the Kalman filter as a Kalman smoother (Appendix A provides details on the Kalman filter). As outlined by Shleifer and Summers (1990), informed investors build their valuation decisions upon fully rational expectations, i. e. the fundamental value of an asset. In contrast, noise traders trade in diffuse signals such as sentiment, which leads to deviations of asset prices from fundamental prices. To attain such a fundamental price for an asset, the Kalman filter allows one to compute the de-noised fundamental price and separate the noise residual.

In our application case, we estimate the oil price with a Kalman filter at the next time iteration t. In a next step, we receive feedback of the observed oil price as (noisy) measurements to correct the prediction for the next state, i. e. time iteration $t+1$ (Welch & Bishop 1997).

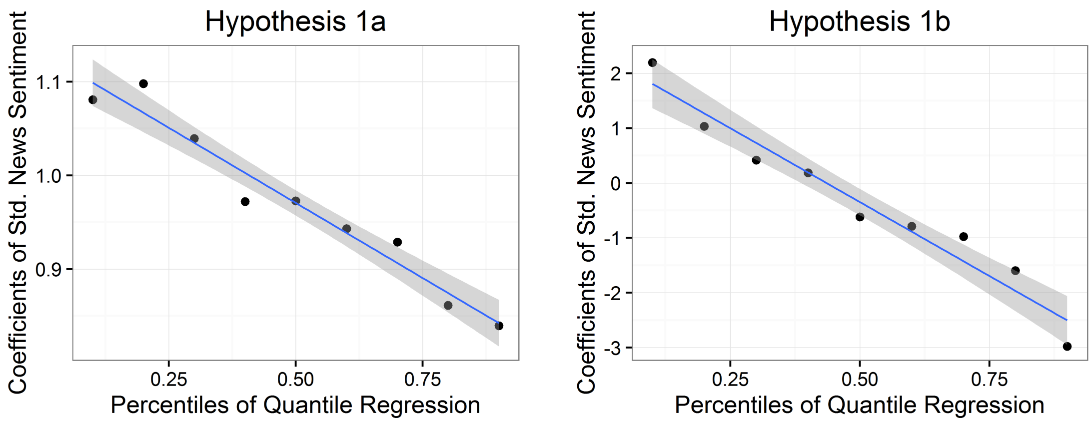

Figure 4. Coefficients of News Sentiment from Quantile Regressions on Every $10th$ Percentile of the WTI Oil Price Real Return (Left) and of the Change in WTI Oil Trading Volume (Right). Plots Include a Trend Line Computed with a Generalized Linear Model.

The plotted noise residuals in Section 4.2 reveal a highly volatile noise pattern. Furthermore, the residual plot illustrates several high positive and negative spikes during the year 2008, coinciding with an oil price drop in the same year.

Table 4 represents a high positive correlation between news sentiment and the real return of the originally quoted WTI crude oil price (correlation coefficient of 0.453). Similarly, we observe high correlation coefficients for both components of the Kalman-decomposed prices: news sentiment is positively correlated with the returns of the fundamental oil price (correlation coefficient of 0.453) and the noise residual (correlation coefficient of 0.383). All three correlation coefficients are significantly different from zero with a P-value < 0.001. Surprisingly, the correlation of news sentiment with the fundamental oil price is stronger than with the noise residual. This early evidence contrasts the noise trader theory of sentiment as being primarily relevant for noise traders. We focus on the relationship between news sentiment and the different price components at different return quantiles in the subsequent section of this paper.

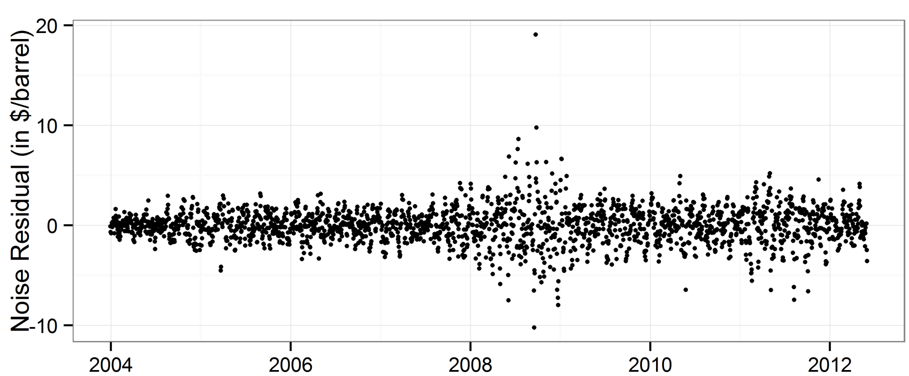

Figure 5. Noise Residuals of the (Kalman-Smoothed) Fundamental WTI Oil Price.

Table 4. Correlation Analysis of News Sentiment and the Decomposed Price Components

	(a) WTI Oil Real Return	(b) Fundamental (Kalman) Return	(c) WTI Oil Price Noise Residual
News Sentiment	0.453^{***} (23.330)	0.453^{***} (23.299)	0.383^{***} (19.017)
Stated: Correlation Coeff., t-Stat. in Parenthesis		Significance: *** 0.001, ** 0.01, * 0.05	

The Influence of News Sentiment on Different Investor Types

Next, we investigate the corresponding effect of news sentiment on both the noise residual and the fundamental oil price at different relative values of the dependent variable. We gain insights into their relationships at different intensities of our de-composed WTI oil price components by applying quantilre regressions. Quantile regression differs (Koenker 2005) from conventional OLS methods that fit a model to a dataset by minimizing squared deviations from a conditional mean; a quantile regression minimizes squared deviations within a defined quantile. We run quantile regressions at the $25th$, $50th$ and $75th$ percentiles. In each quantile regression, the optimization minimizes the distance of our observation values to the corresponding quantile.

In order to address our hypotheses regarding differential information processing, we analyze the influence of news sentiment on the noise residual $N_{\text{Kalman}}(t)$ in *Hypothesis 2a* and on the return of the de-noised fundamental crude oil price $R_{\text{Kalman}}(t)$ in *Hypothesis 2b* by the same regression model as specified in equation Equation (1):

$$\left.\begin{matrix} R_{\text{Kalman}}(t) \\ N_{\text{Kalman}}(t) \end{matrix}\right\} = \beta_0 + \beta_1 S^*(t) + \beta_2 r(t) + \beta_3 FX(t) + \beta_4 IM(t) + \beta_5 OI(t) + \beta_6 G(t) + \beta_7 SP(t) + \varepsilon_t \quad (2)$$

We test *Hypothesis 3*, that the regression coefficients are significantly different between the two models (representing different investor types), using a Kolmogorov-Smirnov test.

Our quantile regressions yield the following key findings on differential information processing:

- ***Confirming Hypothesis (H2a):*** *News sentiment has a stronger effect on negative noise residuals than on positive noise residuals.*
  The quantile regression results in Table 5 confirm, in line with the negativity bias, that the effect of news sentiment on the noise residual is more accentuated on the lower $25th$ quantile (coefficient 0.721, p-value < 0.001) (i. e. when weighing days with negative noise residuals more due to the properties of quantile regression) than on the higher $75th$ quantile (coefficient 0.6125, p-value < 0.001). The effect is 18 percent larger on negative return days than on positive return days.

  As suggested by Koenker (2005), we perform a Wald test to gauge the equality of slope parameters of the respective regression coefficients on the $25th$, $50th$ and $75th$ quantiles. The Wald test confirms that the regression coefficients for news sentiment are significantly different (F-test 4.939 P-value < 0.001) across the different quantiles, i. e. the coefficient on the $25th$ quantile is significantly larger than for the $50th$ and $75th$ quantiles. The *left* graph of Figure 6 shows graphically larger regression coefficients of news sentiment on days with a negative noise effect (i. e. negative noise residuals) from quantile regressions on every $10th$ quantile of the noise residual.

- ***Confirming Hypothesis (H2b):*** *News sentiment has a positive and varying effect on the fundamental oil price return.*
  Our results in Table 6 reveal that news sentiment also has a statistically significant effect on the fundamental, de-noised oil price (all p-values < 0.001). The effect of news sentiment is stronger on negative oil price return days than on positive oil price return days. The magnitude of the effect of

one standard deviation of news sentiment on returns is about 28 percent larger when regressing on the $25th$ (low-return) quantile than on the $75th$ (high-return) quantile. Thus, the relative impact of news sentiment is stronger on low-return days in line with the negativity bias. In further support of our hypothesis, a Wald test confirms that the regression coefficients for news sentiment are significantly different (F-test 6.043; P-value < 0.001) across the different quantiles. The *right* graph of Figure 6 shows larger regression coefficients of news sentiment on days with negative returns (smaller quantiles) on the fundamental oil price, as compared to days with more positive returns (larger quantiles).

- ***Partly Confirming Hypothesis (H3):*** *News sentiment has a similarly strong effect on the fundamental oil price return and on the noise residual. The relative effect size for the noise residual, however, is stronger for negative return days.*

  We evaluate whether the effect of news sentiment on fundamental prices (reflecting informed investors) is different from the effect on the noise residual (reflecting uninformed investors). We test this with a one-sided Kolmogorov-Smirnov test, which is applied under the assumption that noise traders react more strongly to negative sentiment. The Kolmogorov-Smirnov test shows that the quantile coefficients from our two quantile regression models (regressing on the fundamental de-noised oil price and on the noise residual, respectively) are different (P-value < 0.05) across the models, whereas the cumulative distribution function (CDF) of the coefficients for news sentiment in a quantile regression on the noise residual is above the CDF of the respective coefficients for the news sentiment measure in quantile regressions on the fundamental oil price return. Additionally, a student's t-test confirms that the coefficients are different across the two models at statistically significant levels (P-value < 0.01).

  In contrast, we do not find supporting evidence that the effect of news sentiment is stronger on the noise residual than on the fundamental oil price. As already indicated by the correlations in Table 4, the correlation of news sentiment is even stronger with the return on the fundamental oil price than on the noise residual. Also, the t-values for the news sentiment coefficient are slightly larger in the regression model regressing on the $25th$, $50th$ and $75th$ quantiles of the fundamental WTI oil price return (t-statistics between 16 and 19) than in the regression model for the noise residual (t-statistics between 14 and 18).

Table 5. Hypothesis (2a): Coefficients of Quantile Regressions on the Noise Residuals of WTI Oil Price

	(a) OLS	(b) 25 % Quant.	(c) 50 % Quant.	(d) 75 % Quant.
Intercept	2.007^{**}	1.085	2.312^{**}	2.056^{*}
	(2.945)	(1.283)	(3.233)	(2.493)
News Sentiment $S^*(t)$	0.669^{***}	0.721^{***}	0.603^{***}	0.613^{***}
	(18.875)	(17.444)	(16.196)	(14.100)
U. S. Interest Rate $r(t)$	-0.052	-0.034	-0.139	-0.058
	(-0.621)	(-0.339)	(-1.602)	(-0.568)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$	0.217^{***}	0.185^{**}	0.227^{***}	0.14^{*}
	(3.888)	(2.770)	(3.694)	(1.967)
Level of Oil Imports $IM(t)$ (in Million Barrels)	-0.006^{**}	-0.006^{*}	-0.008^{***}	-0.006^{*}
	(-3.024)	(-2.343)	(-3.467)	(-2.250)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)	0.029	0.262	0.229	0.418
	(0.068)	(0.508)	(0.468)	(0.770)
Gold Price $G(t)_{log}$	-0.018	-0.017	-0.02	0.019
	(-0.511)	(-0.391)	(-0.536)	(0.418)
S&P 500 Index $SP(t)_{log}$	-0.02	-0.030	0.022	-0.061
	(-0.829)	(-0.821)	(0.745)	(-1.672)
Stated: Coeff., t-Stat. in Parentheses	Dummies: Yearly		Significance: *** 0.001, ** 0.01, * 0.05	

Comparative Analysis and Robustness Checks

To ensure the stability of our findings, we perform various robustness checks. We conduct a twofold analysis by first inspecting the influence of model and variable variations and then that of dataset variations.

Table 6. Hypothesis (2b): Coefficients of Quantile Regressions on the Return of the Kalman-Filtered WTI Oil Price

	(a) OLS	(b) 25 % Quant.	(c) 50 % Quant.	(d) 75 % Quant.
Intercept	1.275*** (6.228)	0.598* (2.048)	1.182*** (4.951)	1.783*** (7.022)
News Sentiment $S^*(t)$	0.229*** (21.549)	0.238*** (16.520)	0.217*** (18.392)	0.186*** (18.132)
U. S. Interest Rate $r(t)$	0.132*** (5.252)	0.174*** (4.318)	0.089** (2.698)	0.002 (0.072)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$	0.058*** (3.417)	0.063** (2.773)	0.029 (1.386)	0.028 (1.519)
Level of Oil Imports $IM(t)$ (in Million Barrels)	−0.004*** (−7.013)	−0.003*** (−3.555)	−0.003*** (−4.818)	−0.003*** (−4.177)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)	0.02 (0.157)	−0.048 (−0.275)	−0.122 (−0.916)	−0.553*** (−4.085)
Gold Price $G(t)_{log}$	−0.011 (−1.006)	−0.027 (−1.900)	0.007 (0.580)	0.010 (0.809)
S&P 500 Index $SP(t)_{log}$	0.011 (1.467)	0.019 (1.557)	0.019 (1.639)	0.005 (0.460)
Stated: Coeff., t-Stat. in Parentheses	Dummies: Yearly	Significance: *** 0.001, ** 0.01, * 0.05		

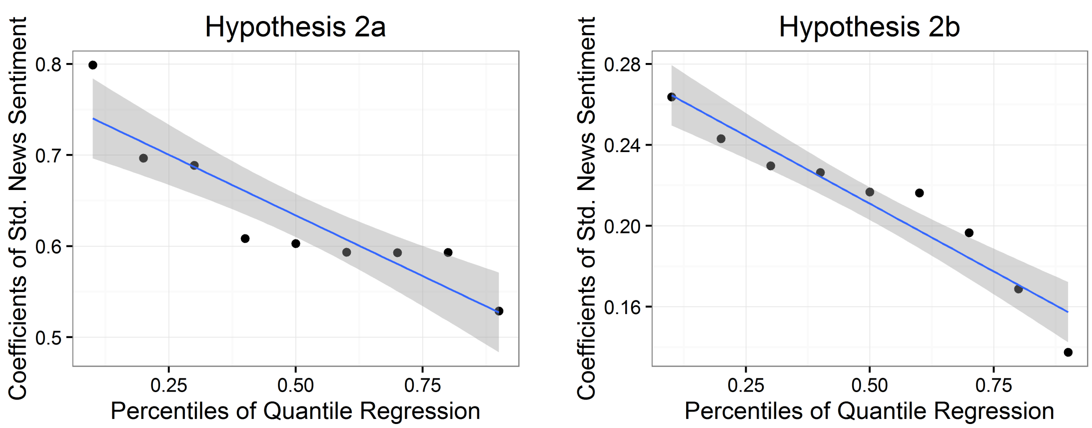

Figure 6. Coefficients of News Sentiment from Quantile Regressions on Every $10th$ Percentile of the Noise Residual of the Kalman-Filtered WTI Oil Price (Left) and of the Return of the Kalman-Filtered WTI Oil Price (Right). Plots Include a Trend Line Computed with a Generalized Linear Model.

Model and Variable Variations

One concern to be addressed is a potential multicollinearity among the control variables. In a first step, we examine the pairwise correlations across the control variables. We observe the highest pairwise correlations between the U. S. interest rate and oil imports (correlation = 0.584), the U. S. dollar/euro exchange rate and the log-return of the gold price (correlation = 0.430) and the open interest in crude oil futures and the level of oil imports (correlation = -0.426). The condition number of the correlation matrix of the control variables is 4.484 and thus not indication of multicollinearity. The variation inflation factors for OLS regressions on the real return of the WTI oil price, the Kalman-smoothed fundamental oil price and the noise residual confirm that the control variables do not suffer from multicollinearity. All observed variance inflation factors are below 2 and thus much smaller than the conventional threshold of 10 (Kutner et al. 2004).

In addition, we assess the robustness of the effect of news sentiment beyond the $25th$, $50th$ and $75th$ quantiles. We run quantile regressions on all 99 percentiles for our four dependent variables of interest (real return of oil price, Kalman-filtered oil price, noise residual and oil trading volume). In line with our results for the quantile regressions at the $25th$, $50th$ and $75th$ quantiles, the coefficients of news sentiment are statistically significant (P-value < 0.001) on all 99 percentiles. The effect of news sentiment on the change in oil trading volume is only marginally significant when strongly weighing negative return days with P-values consistently below 0.05 between the $5th$ and $12th$ percentiles. The only further statistically significant effects (P-value < 0.05) of news sentiment on trading volume shifts can be observed on the $97th$ and $99th$ percentiles. In summary, these findings support our hypotheses and confirm a consistently strong effect of news sentiment on oil price movements, in line with the negativity bias. In contrast, news sentiment seems to only weakly influence oil price trading on days with extreme changes in oil trading. Table 7 shows the coefficients of news sentiment on every $10th$ quantile of the respective regressions.

Table 7. News Sentiment Coefficients of Quantile Regressions on Every 10th Percentile

Quantile	(a) WTI Oil Real Return	(b) Change in Trading Volume	(c) Fundamental (Kalman) Return	(d) WTI Oil Price Noise Residual
0.1	1.081*** (16.838)	2.196* (2.054)	0.799*** (18.215)	0.264*** (13.918)
0.2	1.098*** (18.304)	1.031 (1.263)	0.697*** (19.65)	0.243*** (17.678)
0.3	1.039*** (24.921)	0.417 (0.55)	0.689*** (18.131)	0.23*** (16.243)
0.4	0.972*** (21.46)	0.184 (0.271)	0.608*** (15.428)	0.226*** (17.624)
0.5	0.973*** (19.712)	−0.618 (−1.012)	0.603*** (16.196)	0.217*** (18.392)
0.6	0.943*** (19.871)	−0.787 (−1.09)	0.593*** (14.374)	0.216*** (20.8)
0.7	0.929*** (−16.856)	−0.976 (−1.254)	0.593*** (14.539)	0.197*** (18.202)
0.8	0.861*** (13.976)	−1.595 (−1.231)	0.593*** (15.338)	0.169*** (15.124)
0.9	0.840*** (13.594)	−2.977 (−1.442)	0.529*** (17.990)	0.137*** (15.188)

Stated: Coeff., t-Stat. in Parentheses Dummies: Yearly Significance: *** 0.001, ** 0.01, * 0.05

To test the sensitivity of the news sentiment measure for the dictionary with which we compute the news sentiment, we vary the underlying dictionary. We replace Henry's Finance dictionary (Henry 2008) with the finance-domain-specific Loughran-McDonald's dictionary (Loughran & McDonald 2011) to compute an alternative Net-Optimism metric $S_{\text{NO}}(S)$. The correlation of the sentiment scores computed by the two concurrent approaches is 0.379 (t-test $= 18.798$; P-value < 0.001). Overall, our findings are robust to the variation of the dictionary as the results in Table 8 confirm. As for the above results using Henry's Finance dictionary, the effect size of news sentiment $25th$, $50th$ and $75th$ quantile regressions for the real oil price return and the noise residual is more accentuated on negative return days than on positive return days. For the return on the Kalman-filtered fundamental price, the effect size is 7 percent larger on the $50th$ than on the $25th$ quantile, which is different from the results when using Henry's Finance dictionary. A Wald test also confirms that the slopes between these two quantile regressions are significantly different (P-value < 0.001). In line with the negativity bias and our previous findings, the effect is smaller (and only marginally statistically significant) at the $75th$ quantile. As with the above, we find no statistically significant influence of news sentiment on the oil trading volume. Overall, the t-statistics are smaller in the Loughran-McDonald's dictionary variation. In line with the negativity bias, the robustness of the news sentiment effect (measured by t-test statistics) is less accentuated on positive event days ($75th$ percentile) than on negative event days when computing sentiment scores with Loughran-McDonald's dictionary.

Table 8. Coefficients of News Sentiment from Quantile Regressions with an Alternative Sentiment Measure

Quantile	(a) WTI Oil Real Return	(b) Change in Trading Volume	(c) Fundamental (Kalman) Return	(d) WTI Oil Price Noise Residual
$0.25th$	0.378***	−0.200	0.187***	0.072***
	(5.981)	(−0.229)	(3.777)	(4.408)
$0.5th$	0.280***	−0.333	0.138**	0.077***
	(5.171)	(−0.481)	(3.173)	(6.002)
$0.75th$	0.106	−0.269	0.074	0.039**
	(1.954)	(−0.282)	(1.667)	(3.159)

Stated: Coeff., t-Stat. in Parentheses   Dummies: Yearly   Significance: *** 0.001, ** 0.01, * 0.05

We study the asymmetric effect of news sentiment on the real returns of the WTI oil price, since we model the Kalman-filtered price components based on the observed oil prices and thus select the real return in order to compare the Kalman-filtered fundamental WTI oil price and its noise residual with its originating dataset. However, we also validate the robustness of our results by alternating our dependent variable for the WTI oil price. Specifically, we repeat the quantile regressions by regressing the abnormal return of the WTI oil price on our model specified in Equation (1). Abnormal returns are particulary favored in event studies where the aim is to gauge the impact of a specific event.

In an event study approach, one computes a *predicted* return without accounting for an event X_ρ. The difference between the observed *actual* and *predicted* return represents the *abnormal* return (MacKinlay 1997). In our research, the news sentiment of all daily oil-relevant announcements in our news corpus reflects the event of interest. Following Feuerriegel et al. (2015), we define the *event window* as the (single) day of the announcement given that we have daily financial market data. Hence, the abnormal return is defined by

$$AR(\tau) = R(\rho) - ER(\rho)|\neg X_\rho \tag{3}$$

where $R(\rho)$ and $ER(\rho)|\neg X_\rho$ are the actual and predicted returns in during the event window ρ. The abnormal return is the observed actual return (during the event window) minus the predicted return to measure the impact of the specific event on the returns.

The *predicted* return is estimated based on a pre-event *estimation window*. We compute the *predicted* return by the so-called *market model* MacKinlay 1997. We use a commodity index, the Dow Jones-UBS Commodity Index (Demirer & Kutan 2010), to model the market portfolio and define our event window of 10 trading days prior to the event (Feuerriegel et al. 2015).

The results of estimating our quantile regression model with the abnormal return of the WTI oil price confirms our results. Table 9 shows the coefficients of quantile regressions on the $25th$, $50th$ and $75th$ quantiles. Consistent with our prior findings, we find a statistically significant (P-values < 0.001) stronger effect of news sentiment on days with negative abnormal returns than on days with positive abnormal returns.

Dataset Variations

Furthermore, we validate the robustness of our findings by estimating the effect of news sentiment on oil prices for different subsets of our data. As such, we split our dataset according to our three WTI oil price components (divided by real return on oil price) and, to reflect the change in WTI oil trading volume, also split it into two datasets, one representing the bottom 50 percent *(low)* return days and one showing the top 50 percent *(high)* return days. We estimate OLS regressions with a 0.5% outlier trimming of the lower and upper bound and with yearly dummies on each of these two datasets to compare the regression coefficients of the standardized news sentiment. The results in Table 10 confirm the results from the above quantile regressions. For the *low* and *high* price data subsets, the effect size of news sentiment is consistently larger and of statistically higher significance. Specifically, we always observe larger t-statistics for the *low* than for

Table 9. Robustness Check: Coefficients of Quantile Regressions on the Abnormal Return of WTI Oil Price

	(a) OLS	(b) 25 % Quant.	(c) 50 % Quant.	(d) 75 % Quant.
Intercept	1.251 (0.965)	−2.522 (−2.839)	−1.034 (−1.032)	1.876 (1.757)
News Sentiment $S^*(t)$	0.77*** (12.150)	0.788*** (16.156)	0.700*** (12.581)	0.674*** (12.518)
U. S. Interest Rate $r(t)$	−0.145 (−0.789)	0.28* (2.001)	0.092 (0.675)	−0.219 (−1.517)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$	0.138 (1.122)	0.197** (2.798)	0.053 (0.580)	0.214* (2.350)
Level of Oil Imports $IM(t)$ (in Million Barrel)	0.002 (0.488)	0.005* (1.972)	0.008* (2.508)	0.003 (0.928)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)	−2.321*** (−3.518)	−1.170* (−2.199)	−2.086** (−3.246)	−1.687** (−2.690)
Gold Price $G(t)_{log}$	0.113 (1.306)	0.036 (0.76)	0.067 (1.165)	0.052 (0.906)
S&P 500 Index $SP(t)_{log}$	0.357*** (6.651)	0.302*** (7.798)	0.353*** (8.090)	0.429*** (9.501)
Stated: Coeff., t-Stat. in Parentheses	Dummies: Yearly	Significance: *** 0.001, ** 0.01, * 0.05		

the *high* subset. Furthermore, the results support the weak evidence from our quantile regressions on all percentiles (refer also to Table 7, which shows that news sentiment is of marginal statistical significance on days with the 50 percent bottom changes in trading volume). Separate OLS regressions on the bottom 25 and top 25 percent oil price return data subsets replicate these findings. Not surprisingly, we find that these effects remain robust when evaluating our model with OLS regressions on subsets of the bottom 25 and top 25 percent news sentiment observations.

Next, we transform our continuous standardized news sentiment metric into five discrete quintiles in order to estimate the effect size of news sentiment for each respective quintile by an OLS regression. We substitute the standardized Net-Optimism news sentiment metric $S^*(t)$ with the dummy variables for the $1st$, $2nd$, $3rd$, $4th$ and $5th$ quintiles. We omit the dummy variable for the $3rd$ quintile to mitigate multicollinearity and study. We thus define our OLS regression model as follows:

$$\left.\begin{array}{r} O_{\text{Return}}(t) \\ R_{\text{Kalman}}(t) \\ N_{\text{Kalman}}(t) \end{array}\right\} = \beta_0 + \beta_i S^*(t)_{iqnt.} + \beta_5 r(t) + \beta_6 FX(t) + \beta_7 IM(t) + \beta_8 OI(t) + \beta_9 G(t) + \beta_{10} SP(t) + \varepsilon_t, \quad (4)$$

where β_i represents the regression coefficient for the dummy variable $S^*(t)_{iqnt.}$ for each quintile i with $i \in [1,2,4,5]$. We consistently observe the largest effects size in magnitude and statistical significance (always largest t-statistics) for the bottom quintile sentiment dummy, as the results in Table 11 reveal. This observation holds true for the real return of the WTI oil price, as well as its Kalman-filtered noise and fundamental components. In each case, the magnitude of the effect of news sentiment is largest for the $1st$ (bottom) quintile of sentiment. In the $2nd$ quintile, the effect of news sentiment is still negative, but at a smaller magnitude. This effect is essentially mirrored for the $4th$ and $5th$ quintiles where we observe a positive effect of news sentiment with a larger effect size in the $5th$ quintile. Yet, the absolute effect size of negative sentiment in the bottom news sentiment quintile is between 61 percent (for the noise residual) and 79 percent (for the real return of the WTI oil price) larger than in the top news sentiment quintile.

While the ability of quantile regression to minimize the absolute error terms of a dataset from an estimated regression line (other than OLS, which minimizes the squared error terms) makes quantile regression already more robust with regard to outliers as compared to conventional least square regression approaches, we still want to rule out extreme outlier effects responsible for the previously described asymmetric information

Table 10. Coefficients from OLS Regressions on the Bottom and Top 50 Percent Return-Days Data Subsets

	Hypothesis (1a): Reported (Unfiltered) Oil Price (as Returns)		Hypothesis (1b): Change in Oil Trading Volume		Hypothesis (2a): Noise Residual		Hypothesis (2b): Fundamental Oil Price Component (as Returns)	
	(a) bottom 50%	(b) top 50%	(c) bottom 50%	(d) top 50%	(e) bottom 50%	(f) top 50%	(g) bottom 50%	(h) top 50%
Intercept	-1.361	4.251^{***}	-29.762^{*}	19.230	1.416	1.694	0.867^{**}	1.637^{***}
	(-1.571)	(4.396)	(-2.433)	(0.164)	(1.416)	(1.86)	(2.983)	(6.025)
News Sentiment $S^{*}(t)$	0.547^{***}	0.116^{*}	1.571^{*}	-3.805	0.550^{***}	0.498^{***}	0.208^{***}	0.166^{***}
	(11.114)	(2.104)	(2.369)	(-0.640)	(9.734)	(9.708)	(12.610)	(10.876)
U. S. Interest Rate $r(t)$	0.381^{***}	-0.504^{***}	0.993	13.540	-0.087	-0.091	0.157^{***}	0.057
	(3.640)	(-4.090)	(0.918)	(0.917)	(-0.726)	(-0.802)	(4.481)	(1.646)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$	0.076	0.055	-0.805	7.828	0.288^{***}	0.051	0.044	0.015
	(1.105)	(0.650)	(-0.872)	(0.728)	(3.750)	(0.643)	(1.956)	(0.634)
Level of Oil Imports $IM(t)$ (in Million Barrel)	-0.002	0.000	0.024	0.477	-0.004	-0.005	-0.004^{***}	-0.004^{***}
	(-0.857)	(0.004)	(0.654)	(1.288)	(-1.295)	(-1.923)	(-4.039)	(-5.28)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)	0.213	-2.384^{***}	-1.997	-213.476^{**}	-0.242	0.302	0.094	-0.230
	(0.397)	(-3.918)	(-0.266)	(-2.835)	(-0.395)	(0.535)	(0.522)	(-1.351)
Gold Price $G(t)_{log}$	-0.022	0.062	1.227^{**}	-2.050	0.006	-0.021	0.000	-0.019
	(-0.505)	(1.160)	(2.854)	(-0.329)	(0.121)	(-0.428)	(0.027)	(-1.254)
S&P 500 Index $SP(t)_{log}$	0.246^{***}	0.158^{***}	0.994	4.914	-0.077^{*}	-0.049	0.022^{*}	-0.043^{***}
	(7.590)	(4.059)	(0.315)	(1.117)	(-2.073)	(-1.390)	(1.995)	(-3.872)
AIC	3581	3756	9076	13840	3862	3610	1291	1085
BIC	3665	3840	9155	13925	3946	3694	1375	1170
R2	0.298	0.138	0.030	0.027	0.140	0.123	0.317	0.223
Adj R2	0.288	0.125	0.017	0.013	0.127	0.110	0.307	0.211

Stated: Coef. and t-Stat. in Parentheses   Dummies: Year   Obs.: 2112   Signif.: *** 0.001, ** 0.01, * 0.05

Table 11. Coefficients of OLS Regressions with Quintile Dummies for News Sentiment

	(a) WTI Oil Real Return	(b) Fundamental (Kalman) Return	(c) WTI Oil Price Noise Residual
Intercept	2.000^{*} (2.235)	2.036^{**} (2.933)	1.233^{***} (5.883)
News Sentiment $S^{*}(t)_{1stqnt.}$	-1.667^{***} (-12.118)	-1.103^{***} (-10.308)	-0.376^{***} (-11.693)
News Sentiment $S^{*}(t)_{2ndqnt.}$	-0.666^{***} (-4.977)	-0.568^{***} (-5.484)	-0.157^{***} (-5.054)
News Sentiment $S^{*}(t)_{4thqnt.}$	0.512^{***} (3.836)	0.208^{*} (2.012)	0.104^{***} (3.359)
News Sentiment $S^{*}(t)_{5thqnt.}$	0.931^{***} (6.859)	0.675^{***} (6.409)	0.233^{***} (7.376)
U. S. Interest Rate $r(t)$	0.106 (0.972)	-0.033 (-0.392)	0.142^{***} (5.590)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$	0.201^{**} (2.666)	0.22^{***} (3.900)	0.057^{***} (3.359)
Level of Oil Imports $IM(t)$ (in Million Barrel)	-0.003 (-1.175)	-0.007^{**} (-3.216)	-0.004^{***} (-6.945)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)	-1.104^{*} (-2.017)	0.396 (0.930)	0.129 (1.008)
Gold Price $G(t)_{log}$	0.059 (1.285)	0.006 (0.174)	-0.002 (-0.148)
S&P 500 Index $SP(t)_{log}$	0.392^{***} (12.138)	-0.007 (-0.292)	0.012 (1.585)
AIC	8668.246	7603.026	2573.609
BIC	8781.135	7715.915	2686.498
R2	0.287	0.174	0.302
Adj R2	0.281	0.167	0.296
Stated: Coeff., t-Stat. in Parentheses	Dummies: Yearly		Significance: *** 0.001, ** 0.01, * 0.05

processing whereby negative sentiment outweighs positive sentiment. Therefore, we run $25th$, $50th$ and $75th$ quantile regressions on trimmed subsets of our data. We trim our dataset by removing the 1 (5) percentiles of negative and positive outliers of extreme oil price return days on either side of the dataset, i. e. we trim our dataset by 2 (10) percent of extreme data, respectively. The results in Table 12 reveal that our standardized news sentiment measure remains robust at statistically highly significant levels. All t-statistics are above 10 for the coefficients for regressions on the real return of the WTI oil price and its two Kalman-filtered price components. Additionally, the effect size is consistently larger at the $25th$ percentile as compared to the $75th$ percentile.

Table 12. Coefficients of Quantile Regressions for Trimmed Datasets

	Reported (Unfiltered) WTI Oil Price (as Returns)		WTI Oil Price Noise Residual		Fundamental (Kalman-Filtered) Oil Price (as Returns)	
	(a) 2% **trimming**	(b) 10% **trimming**	(c) 2% **trimming**	(d) 10% **trimming**	(e) 2% **trimming**	(f) 10% **trimming**
0.25 quantile	1.032^{***} (19.220)	0.856^{***} (16.056)	0.689^{***} (16.813)	0.673^{***} (17.160)	0.234^{***} (18.378)	0.211^{***} (13.865)
0.5 quantile	0.927^{***} (19.559)	0.830^{***} (17.353)	0.595^{***} (15.680)	0.554^{***} (13.369)	0.215^{***} (18.036)	0.207^{***} (15.541)
0.75 quantile	0.888^{***} (15.532)	0.751^{***} (13.340)	0.613^{***} (15.495)	0.571^{***} (13.717)	0.186^{***} (16.262)	0.175^{***} (15.862)
Stated: Coeff., t-Stat. in Parentheses			Dummies: Yearly		Significance: *** 0.001, ** 0.01, * 0.05	

Causal Influence of News Sentiment on Oil Prices

One limitation of the above findings on asymmetric information processing of news sentiment is arguably causality. As Loughran and McDonald (2011) argue *"the existing literature on financial text does not actually determine the causal link between tone and returns."* Feuerriegel et al. (2015) provide evidence of a causal relationship between daily news sentiment and abnormal returns of crude oil. In their oil-related paper, they pursue an instrumental variable (IV) design with a two-stage least squares procedure as e. g. described by (Angrist & Evans 1998; Angrist & Krueger 2001; Antonakis et al. 2010). They choose the number of terror attacks as an instrument to account for the endogeneity problem implicit in news releases. Their analysis validates the instrument as relevant, such that they can correct for endogeneity and infer a causal link between news sentiment and oil returns. Similarly, Engelberg and Parsons (2011) find that the presence of individual companies' announcements in local media coverage has a causal effect on the trading volume among retail investors in the coverage area of local press. Bosman et al. (2014) find in a controlled experimental setting that investors predict higher future stock returns and are more likely to buy stocks subsequently when they are exposed to a more positive tone when holding the informational content constant. We address the endogeneity problem by following the approach described by Feuerriegel et al. (2015), since their research examines the causal effect of news sentiment on oil prices and is thus closely related to our specific research domain. We use the same model as defined in Equation (1) and Equation (2). Following Feuerriegel et al. (2015), we define the instrument as the number of *daily reported terrorist attacks* $T^(t)$ in the top 10 crude-oil-producing countries and Iraq. First, we validate whether the instrument meets the necessary conditions of relevance and exogeneity (Stock & Watson 2011).

- **Relevance.** An instrument z_i is relevant if it is correlated with our endogenous explanatory variable x_i, here standardized news sentiment, conditional on the other covariates, i. e.

$$\mathrm{cor}(z_i, x_i) \neq 0. \tag{5}$$

- **Exogeneity.** An instrument z_i is exogenous if it is not correlated to the error term ε_i when regressing a dependent variable y_i on x_i conditional on the other covariates, i. e.

$$\mathrm{cor}(z_i, \epsilon_i) = 0 \tag{6}$$

Subject to fulfilling these two necessary conditions, a two-stage least squares estimator (2SLS) can estimate the desired β_i. In the first stage, the endogenous explanatory variable is regressed on the instrument and assesses its relevance (Antonakis et al. 2010). We define our first stage regression by

$$S^*(t) = \beta_0 + \beta_1 T^(t) + \beta_2 r(t) + \beta_3 FX(t) + \beta_4 IM(t) + \beta_5 OI(t) + \beta_6 G(t) + \beta_7 SP(t) + \varepsilon_t, \tag{7}$$

where $T^(t)$ reflects the number of reported terror attacks on day t.

In a second step, we can inject the predicted values $\hat{S}_t$ into the second-stage regression. We define the second-stage regression in a fashion similar to that of Feuerriegel et al. (2015). However, in contrast to Feuerriegel et al. (2015), we use the real return, not the abnormal return, of the WTI oil price as a dependent variable, and define the model by

$$\left.\begin{array}{r} O_{\text{Return}}(t) \\ R_{\text{Kalman}}(t) \\ N_{\text{Kalman}}(t) \end{array}\right\} = \beta_0 + \beta_1 \hat{S}_t + \beta_2 r(t) + \beta_3 FX(t) + \beta_4 IM(t) + \beta_5 OI(t) + \beta_6 G(t) + \beta_7 SP(t) + \varepsilon_t. \tag{8}$$

Table 13 shows the results of the first stage least square regression. The regression results provide evidence that the correlation of the instrument terror attacks $T(t)$ with the endogenous explanatory variable $S^*(t)$ is statistically significant (t-statistics of 2.916, P-Value < 0.01). Pearson's two product-moment correlation of the two factors is 0.084 (P-value < 0.001). Thus, the instrument qualifies as relevant. The interpretation of this finding is that terrorist attacks correlate with news sentiment at a statistically significant level. This relationship seems plausible. As we cover all oil-related news, including general market conditions, terror attacks likely also affect the overall sentiment of news related to the general market setting.

Furthermore, we validate the second necessary condition for our instrument, its exogeneity. Exogeneity requires that the instrument not be correlated with the error terms of a regression of our model on the real returns of the WTI crude oil price. In fact, the correlation $\mathrm{cor}(z_i, \epsilon_i) \cong 0$ and is statistically insignificant (t-statistics of 0.000 and P-value < 0.001).

In the case of a weak correlation $\mathrm{cor}(z_i, x_i)$, the IV is prone to a larger bias. The F-statistic against the null hypothesis that the excluded instrument is irrelevant in the first-stage regression is a valid measure for the so-called *strength* of instruments. As stipulated by Staiger and Stock (1997), strong instruments should have a test statistic larger than 10 in models with one endogenous regressor. We achieve a first-stage F-statistic of 76.61 for our instrument, which is evidence of a strong instrument. Thus, the necessary conditions for our instrument are met and we can compute the second-stage least square regression models as defined in Equation (8). The results of the second-stage IV regressions are represented in Table 14 for the real return of the WTI oil price, in Table 15 for the noise residual and in Table 16 for the Kalman-filtered fundamental price. The results remain statistically robust for the standardized news sentiment coefficient. Additionally, we observe that an increase of our sentiment measure by one standard deviation increases the real return by an economically significant 0.344 percentage points.

The downside of the IV regression approach is a loss in efficiency, while its purpose is to assure consistency and mitigate the endogeneity problem. Thus, we apply a Hausman test as an endogeneity evaluation. The Hausman test (Hausman 1978; Wooldridge 2009) compares the respective OLS and 2SLS estimates to test for statistically significant differences. Under the null hypothesis, all variables are exogenous and both OLS and 2SLS are consistent. If 2SLS and OLS differ significantly, the Hausman test concludes that the endogenous explanatory variable is indeed endogenous. As such, the Hausman test consists of two steps.

1. The Hausman test estimates the residuals of the first-stage least square equation $\hat{\varepsilon}_t$.
2. Add the estimated residuals $\hat{\varepsilon}_t$ to the first-stage least square equation. Run an OLS regression to test whether the coefficient on $\hat{\varepsilon}_t$ is statistically significant, i. e. different from zero. Under the null hypothesis, we can assume that our endogenous explanatory variable is actually exogenous, else for a statistically significant non-zero coefficient of $\hat{\varepsilon}_t$, we conclude the presence of endogeneity and need to favor IV regression over OLS regression.

We conduct the Hausman test for our three respective WTI oil price variables: the real return of the WTI oil price, the noise residual and the fundamental oil price. Following Wooldridge (2009), we use a heteroskedasticity-robust Newey-West estimator. The coefficients on the fitted residuals $\hat{\varepsilon}_t$ when adding these to our first-stage least square model are as follows: the Hausman test does not produce significant coefficients for the real return of the WTI oil price (t-statistics of -0.814, P-value of 0.416), or the noise residual (t-statistics of -0.152, P-value of 0.879) or for the Kalman-filtered fundamental WTI oil price (t-statistics of -0.152, P-value of 0.879). Thus, we observe that the influence of a possible omitted variable bias is very small. Hence, we can conclude that our news sentiment measure fulfills the criteria of exogeneity and thus represents a valid explanatory variable. As a result, an IV regression approach is not favored.

Table 13. Coefficients of First-Stage Least Square Regressions for our Respective WTI Oil Price Components

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Intercept	-0.032 (-0.512)	-0.404^{***} (-4.163)	-0.409^{***} (-4.351)	0.324 (0.784)	-0.441 (-1.032)	-0.525 (-1.242)	-0.427 (-1.029)
Terror Attacks $T(t)$	0.028^{***} (4.177)	0.029^{***} (4.343)	0.028^{***} (4.319)	0.029^{***} (4.362)	0.029^{***} (4.527)	0.029^{***} (4.471)	0.028^{***} (4.384)
U. S. Interest Rate $r(t)$		0.263^{***} (5.002)	0.256^{***} (5.022)	0.263^{***} (5.155)	0.172^{**} (3.261)	0.174^{***} (3.343)	0.164^{**} (3.206)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$			0.367^{***} (11.691)	0.365^{***} (11.658)	0.363^{***} (11.683)	0.257^{***} (7.561)	0.208^{***} (6.151)
Level of Oil Imports $IM(t)$ (in Million Barrel)				-0.002 (-1.822)	-0.003^{*} (-2.441)	-0.003^{*} (-2.25)	-0.003^{*} (-2.573)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)					1.661^{***} (6.255)	1.641^{***} (6.253)	1.679^{***} (6.517)
Gold Price $G(t)_{log}$						0.157^{***} (7.253)	0.166^{***} (7.802)
S&P 500 Index $SP(t)_{log}$							0.131^{***} (8.874)
AIC	5940	5917	5786	5784	5747	5697	5621
BIC	6002	5985	5859	5864	5832	5788	5717
R2	0.034	0.046	0.104	0.106	0.122	0.143	0.174
Adj R2	0.03	0.041	0.099	0.1	0.117	0.138	0.169

Stated: Coeff., t-Stat. in Parentheses   Dummies: Yearly   Significance: *** 0.001, ** 0.01, * 0.05

Table 14. Coefficients of Second-Stage Least Square Regression for the Real Return of the (Unfiltered) WTI Oil Price

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Intercept	0.174 (1.136)	0.178 (1.283)	-0.035 (-0.283)	-0.03 (-0.241)	0.417^{*} -2.088	0.356 (1.735)	-1.2 (-1.392)
News Sentiment $S^{*}(t)$	0.52^{***} (13.511)	0.472^{***} (13.455)	0.387^{***} (12.231)	0.355^{***} (9.576)	0.339^{***} (9.051)	0.354^{***} (8.978)	0.344^{***} (8.647)
U. S. Interest Rate $r(t)$		0.554^{***} (20.976)	0.491^{***} (20.633)	0.486^{***} (20.249)	0.483^{***} (20.158)	0.481^{***} (20.017)	0.483^{***} (20.091)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$			1.867^{***} (22.926)	1.838^{***} (22.084)	1.857^{***} (22.278)	1.858^{***} (22.289)	1.872^{***} (22.378)
Level of Oil Imports $IM(t)$ (in Million Barrels)				0.224 (1.648)	0.34^{*} (2.405)	0.178 (0.927)	0.238 (1.224)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)					-0.317^{**} (-2.865)	-0.279^{*} (-2.427)	-0.312^{**} (-2.684)
Gold Price $G(t)_{log}$						0.113 (1.250)	0.094 (1.034)
S&P 500 Index $SP(t)_{log}$							0.005 (1.858)
AIC	9724	9324	8854	8853	8847	8848	8846
BIC	9786	9392	8928	8933	8932	8938	8942
R2	0.083	0.242	0.394	0.394	0.397	0.397	0.398
Adj R2	0.079	0.238	0.39	0.391	0.393	0.393	0.394

Stated: Coeff., t-Stat. in Parentheses   Dummies: Yearly   Significance: *** 0.001, ** 0.01, * 0.05

Table 15. Coefficients of Second-Stage Least Square Regression for the Noise Residual of the Fundamental WTI Oil Price

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Intercept	0.688^{***} (134.095)	0.688^{***} (134.114)	0.686^{***} (134.028)	0.689^{***} (151.021)	0.644^{***} (89.443)	0.677^{***} (100.002)	0.37^{***} (13.405)
News Sentiment $S^*(t)$	-0.000 (-0.274)	-0.000 (-0.354)	-0.001 (-0.892)	-0.018^{***} (-13.073)	-0.016^{***} (-11.924)	-0.025^{***} (-18.900)	-0.027^{***} (-20.855)
U. S. Interest Rate $r(t)$		0.001 (1.232)	0.001 (0.700)	-0.002^{*} (-2.279)	-0.002^{*} (-2.006)	-0.001 (-0.712)	0 (-0.232)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$			0.015^{***} (4.565)	0.000 (0.119)	-0.001 (-0.487)	-0.002 (-0.687)	0.001 (0.372)
Level of Oil Imports $IM(t)$ (in Million Barrels)				0.116^{***} (23.409)	0.105^{***} (20.482)	0.192^{***} (30.357)	0.204^{***} (32.751)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)					0.031^{***} (7.863)	0.011^{**} (2.832)	0.004 (1.134)
Gold Price $G(t)_{log}$						-0.061^{***} (-20.458)	-0.065^{***} (-22.235)
S&P 500 Index $SP(t)_{log}$							0.001^{***} (11.447)
AIC	-4614.24	-4613.765	-4632.621	-5120.535	-5179.875	-5562.221	-5688.3
BIC	-4552.031	-4545.9	-4559.101	-5041.359	-5095.044	-5471.734	-5592.158
R2	0.916	0.916	0.917	0.934	0.936	0.947	0.95
Adj R2	0.916	0.916	0.917	0.934	0.936	0.946	0.95
Stated: Coeff., t-Stat. in Parentheses				Dummies: Yearly	Significance: *** 0.001, ** 0.01, * 0.05		

Table 16. Coefficients of Second-Stage Least Square Regression for the Return of the Fundamental WTI Oil Price

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Intercept	0.072 (1.095)	0.072 (1.107)	0.047 (0.728)	0.102^{*} (2.243)	0.977^{***} (14.237)	1.124^{***} (16.211)	-0.718^{*} (-2.487)
News Sentiment $S^*(t)$	0.018 (1.115)	0.014 (0.839)	0.004 (0.234)	-0.329^{***} (-24.329)	-0.36^{***} (-27.986)	-0.399^{***} (-29.886)	-0.411^{***} (-30.799)
U. S. Interest Rate $r(t)$		0.053^{***} (4.277)	0.046^{***} (3.681)	-0.008 (-0.926)	-0.013 (-1.614)	-0.008 (-0.991)	-0.006 (-0.712)
U. S. Dollar/Euro Exchange Rate $FX(t)_{log}$			0.218^{***} (5.114)	-0.081^{**} (-2.681)	-0.045 (-1.576)	-0.047 (-1.673)	-0.03 (-1.064)
Level of Oil Imports $IM(t)$ (in Million Barrels)				2.335^{***} (47.161)	2.564^{***} (52.63)	2.957^{***} (45.575)	3.029^{***} (46.483)
Open Interest in Crude Oil Futures $OI(t)$ (in Million)					-0.622^{***} (-16.317)	-0.715^{***} (-18.418)	-0.755^{***} (-19.388)
Gold Price $G(t)_{log}$						-0.275^{***} (-8.978)	-0.297^{***} (-9.748)
S&P 500 Index $SP(t)_{log}$							0.006^{***} (6.572)
AIC	6157	6140	6116	4592	4342	4264	4223
BIC	6219	6208	6190	4671	4427	4355	4319
R2	0.003	0.011	0.023	0.526	0.579	0.595	0.603
Adj R2	-0.002	0.006	0.018	0.523	0.577	0.592	0.6
Stated: Coeff., t-Stat. in Parentheses				Dummies: Yearly	Significance: *** 0.001, ** 0.01, * 0.05		

Discussion and Managerial Contribution

This section discusses our results in the context of behavioral economics theory. Our approach uses quantile regressions to evaluate the differential effect of news sentiment on varying return levels for different investor types (informed vs. uninformed investors). For the different investor types, we assumed that we would find a stronger effect of negative than positive sentiment, in line with the negativity bias.

While our findings support the concept of asymmetric information processing, some findings reject – to a certain extent – the asymmetric direction of information processing under the negativity bias.

In keeping with the negativity bias, we find evidence that news sentiment has a stronger effect on investors on negative than on positive return days. The results of our quantile regressions produce a consistent outcome: the effect of news sentiment is larger on negative return days for our oil price components. Wald tests confirm this result. This provides strong evidence that negative sentiment, associated with negative market movements (Tetlock 2007; Tetlock et al. 2008), has a stronger influence on investors than positive sentiment.

News sentiment plays a similar role for different investor types. As confirmed by the results of Wald tests for the equality of quantile coefficients, our news sentiment coefficients are significantly different for different return percentiles in all three regression models for the unfiltered oil price return, the Kalman-filtered fundamental oil price and the noise residual. While a Kolmogorov-Smirnov test suggests that the reaction of the noise residual (representing uninformed investors) to news sentiment is more accentuated than the reaction of the fundamental WTI oil price on more negative return days, the overall significance levels of the effect of news sentiment are similar across different investor types.

Overall, our results are consistent with the negativity bias' assumption that negative information outweighs positive information. These findings on asymmetric information processing by different investor types entail several implications for capital market stakeholders such as regulators, investors and communicators.

Today, these stakeholders lack evidence-based decision-support systems which indicate how financial markets process the qualitative, textual content of the information contained in disclosures directed at capital markets. Based on the aforementioned findings, we can draw the following three managerial implications for stakeholders in the financial news domain.

Implication 1: *Brokers need to familiarize traders with information processing asymmetry.* To mitigate high risk exposure, financial institutions need to educate their traders about cognitive biases and the asymmetry of information processing. Asymmetric information processing can (a) influence the genuine decision-making of traders via the bandwagon effect (Shleifer & Summers 1990) and (b) lead to structural market risks, e. g. from sentiment price bubbles.

Implication 2: *Media can amplify sentiment.* News companies need to train journalists to avoid cognitive biases. Otherwise, media may serve as an amplifier of news sentiment and thus contribute to increased price volatility from exaggerated sentiment levels ,as shown by Bajo and Raimondo (2015).

Implication 3: *Avoid overly negative language.* Departments, such as investor relations, which communicate a company's performance to capital markets should avoid using overly negative language in order to prevent negative sentiment shocks. These departments could, for instance, rely on decision-support software that guides them in avoiding very negatively phrased wording.

Based on our findings, the theoretical contribution of this research is twofold: first, we open new avenues to empirically validate information processing of additional behavioral finance theories. Second, the properties of quantile regression suggest further applications of this methodology to other asset classes to shed further light on the robustness of our findings.

Conclusion and Outlook

Behavioral finance models account for deviations of economic agents from rational behavior. Such deviations from expected decision-making are often referred to as cognitive biases. Among other things, cognitive biases describe asymmetric information processing. Asymmetric information processing in which negative information dominates positive information is referred to as negativity bias. The negativity bias is applied to economic decision-making in the so-called prospect theory. According to prospect theory, negative outcomes outweigh positive outcomes in the decision-making process.

As its main contribution, this paper utilizes quantile regressions to investigate whether news sentiment has a differential effect on different return days. We study this with the WTI oil price and trading volume and a dataset of oil-related news. We analyze the effect of news sentiment across different oil price (trading volume) return quantiles. In addition, we decompose the oil price using a Kalman filter into a fundamental oil price component (representing informed investors) and a noise residual (representing uninformed investors). Next, we extract the news sentiment from oil-related news and study how investors integrate different occurrences of news sentiment (negative vs. positive) into the observed oil pricing and trading. Furthermore, we show with an instrumental variable approach and two stage least square regressions that our news sentiment measure is exogenous, a strong indication for a causal relationship of news sentiment with the studied WTI oil price components.

The results of this paper confirm that financial markets process news sentiment asymmetrically. Our findings with regard to the influence of news sentiment on the oil price return and its Kalman-filtered fundamental and noise components provide strong evidence for the negativity bias. In contrast, we do not find a statistically significant relationship between news sentiment and trading volume except at very negative trading volume change quantiles. In accordance with the noise trader theory, the effect of news sentiment on negative return days is stronger for the noise residual (representing uninformed investors) than for the fundamental oil price component (representing informed investors). However, the effect size of news sentiment is of a similar magnitude for both investor types.

Our work opens an avenue for research on the asymmetric processing of news sentiment by financial markets. Thus far, researchers have commonly studied the role of news sentiment in financial markets using more static regression models estimated on the conditional mean. We apply quantile regressions for different return quantiles to investigate the relationship of news sentiment with oil prices more comprehensively. Furthermore, quantile regressions are more robust against outliers than conventional OLS regressions. In further work, we will investigate the observed effects for other asset classes in order to determine whether broader inferences regarding the validity of the negativity bias in financial markets can be made. Finally, an intriguing approach would be the application of the analysis to a news corpus that also includes the more recent years since 2012, which were not available for analysis at the time of our study.

Appendix A: Price Decomposition

Inaccuracies and perturbations frequently affect financial time series data (Harvey 1990; Haven et al. 2012). The *Kalman process* is a recursive approach to linear filtering problems with discrete data (Harvey 1990; Kalman 1960). The Kalman filter decomposes discrete datasets, such as price time series, into a *de-noised fundamental price* and a *noise residual*. Decomposing market prices with the Kalman filter is a widely-used approach for financial time series (Brogaard et al. 2014; Haven et al. 2012; Hendershott & Menkveld 2014; Hendershott et al. 2013; Lopes & Tsay 2011; Manoliu & Tompaidis 2002; Schwartz & Smith 2000; Wong 2010).

In the following, we introduce the mechanisms of the Kalman filter. The Kalman filter utilizes a feedback control to estimate a process: (1) the filter estimates the process state at the next time iteration t and then (2) receives feedback as (noisy) measurements to correct the prediction for the next state (Welch & Bishop 1997). Hence, the Kalman filter consists of two types of equations:

1. *Predictor* or *time update* equations to obtain *a priori* predictions for the next time step by projecting forward the current state and error covariance estimates.
2. *Corrector* or *measurement update* equations to include additional measurements into the *a priori* predictions in order to attain improved *a posteriori* estimates.

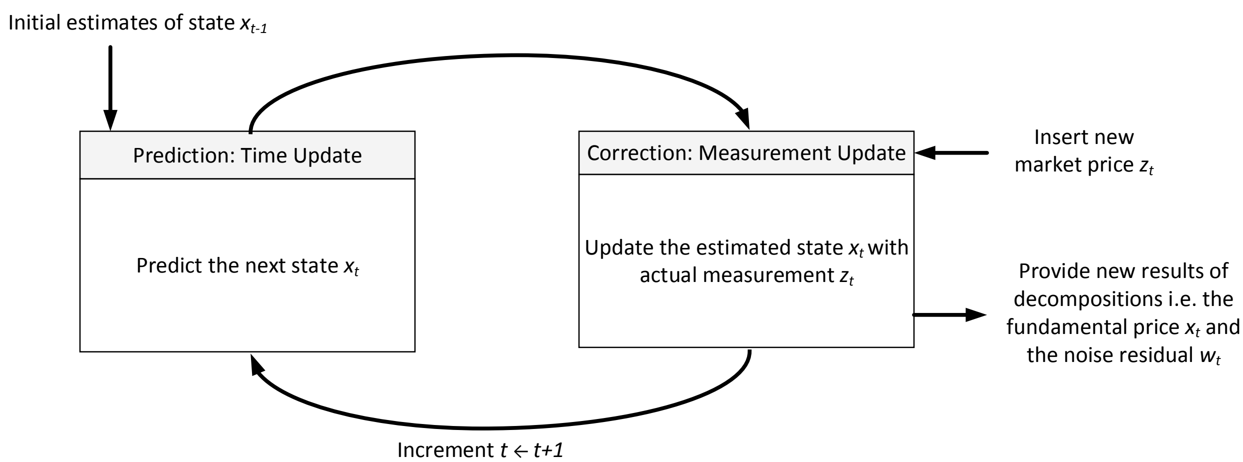

Figure 7. Discrete Kalman filter cycle with the prediction of current state estimate ahead in time and the correction of prediction by an actual measurement update (Welch & Bishop 1997).

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

$$x_{t+1} = F_t x_t + B_t u_t + w_t, \qquad w_t \sim N(0, Q), \tag{9}$$

and a measurement equation

$$z_t = H_t x_t + v_t, \qquad v_t \sim N(0, R), \tag{10}$$

where $B_t \in \mathbb{R}^{n \times l}$ provides a control input model, $F_t \in \mathbb{R}^{n \times n}$ a transition matrix, $z_t \in \mathbb{R}^m$ 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 $u_t \in \mathbb{R}^l$ represents a control vector and $w_t \in \mathbb{R}^n$ and $v_t \in \mathbb{R}^m$ provide the noise residuals in the form of i.i.d. random processes $N(0, \cdot)$.

Based on the above definition, we can apply the Kalman filter as a decomposition analysis to our specific setting. We observe a price sequence z_t given by the observed market prices. Then, we 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 calculates the dynamic mean of the prices. The estimation error w_t of a state equation provides the noise component, i. e. the gap between fundamental and observed prices.

We repeat the Kalman process with the prior *a posteriori* estimates to predict the new *a priori* estimates. Hence, the Kalman filter recursively conditions all current estimates on prior measurements.

Appendix B: Quantile Regression

The increased computation capacity of modern computer systems has resolved computation limitations of numerous statistical methods. This has opened new research avenues for applying statistical methodologies with the potential to address relevant research questions in the context of Data Analytics. This development has triggered a renaissance of statistical methods outside of the academic spotlight. For instance, Advances in Data Analytics and IS methodologies have rendered the computation-intense quantile regressions an attractive statistical analysis tool with applications in e. g. finance and labor markets (e. g. (Buchinsky 1994; Chernozhukov & Hansen 2004; Fitzenberger, Hujer, et al. 2002; Fitzenberger, Koenker, & Machado 2002; Tsai 2012)).

While conventional OLS methods fit a model to a dataset by minimizing squared deviations from a conditional mean, quantile regression minimizes squared deviations within a defined quantile (Buchinsky 1998; Fitzenberger, Koenker, & Machado 2002; Koenker 2005; Koenker & Bassett 1978). Thus, quantile regression allows for the estimation of quantiles (e. g. the median) of the dependent variable. Formally, the cumulative distribution function $F_X(x) = P(X \leq x)$ may characterize any real-valued random variable X, where

$$F_X^{-1}(\tau) = \inf_x \{x \mid F_X(x) \geq \tau\} \tag{11}$$

is the τ-th quantile of X for any $0 \geq \tau \geq 1$. We denote the penalty $\rho_\tau(x)$ for deviations within our quantile τ as

$$\rho_\tau(x) = |x\,(\tau - I_{x<0})| \tag{12}$$

where I is an indicator variable. Following this, a quantile regression searches the $\hat{x}$ that minimizes the absolute deviations within our quantile τ, i. e.

$$\hat{x} = \min_u \; E(\rho_\tau(X-u) = \min_u \,(\tau - 1) \int_{-\infty}^{8} (x-u)\;\mathrm{d}F_X(x) + \tau \int_{u}^{\infty} (x-u)\;\mathrm{d}F_X(x). \tag{13}$$

In order to solve the problem, we differentiate with respect to u and yield

$$0 = (1-\tau) \int_{-\infty}^{\hat{x}} \mathrm{d}F_X(x) - \tau \int_{\hat{x}}^{\infty} \mathrm{d}F_X(x) = F_X(\hat{x}) - \tau. \tag{14}$$

Since F is monotonic, a value $\hat{x}$ which fulfills $F_X(\hat{x}) = \tau$ minimizes the absolute deviation within our quantile τ.

Appendix C: Sentiment Analysis for Financial News

Decision Analytics, as a research area of Information Systems (IS) research, is well-equipped to study how agents process information conveyed by textual news in financial markets (Chen et al. 2012). The textual content of news provides relevant facts beyond quantitative information, e. g. profits or earning forecasts, through the tone of the language. The subjective tone of text documents can be assessed using so-called *sentiment analysis*. Sentiment analysis (also known as *opinion mining*) includes methods that measure how positive or negative the content of textual sources is.

News sentiment measures enable the investigation of the effect of financial news on financial markets. Prior research has explored how human agents process the sentiment of financial news. For instance, empirical evidence highlights a statistically significant relationship between financial news and stock market movements (e. g. Antweiler & Frank 2004; Tetlock 2007).

Beyond the pure description of a relationship, several papers have explored the reception of financial news in stock market transactions. For example, Groth and Muntermann (2011) applied text-mining approaches to risk management to identify volatility-enhancing corporate disclosures that increase risk exposure in particular. Muntermann and Guettler (2007) show that abnormal price effects following an announcement are smaller the larger the disclosing company is. Bajo and Raimondo (2015) report that positive news sentiment and news coverage increase first-day returns and thus drive IPO underpricing.

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 et al. 2012; Mittermayer & Knolmayer 2006b; Nassirtoussi et al. 2014) focus on studies aimed at stock market prediction. Financial text mining research prevalently deploys *dictionary-based methods* (cp. Demers & Vega 2010; Henry 2008; Jegadeesh & Wu 2013; Loughran & McDonald 2011; Tetlock et al. 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 & Frank 2004; Li 2010; Mittermayer & Knolmayer 2006a; Schumaker & Chen 2009) represent a variety of methods but may be subject to overfitting (Sharma & Dey 2012). A remedy could 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 et al. (2015b) with the help of Bayesian learning.

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 the textual content of news.

Before analyzing news sentiment, we need to pre-process our news corpus (Feuerriegel & Neumann 2013; Manning & Schütze 1999): first, tokenization splits running text into single words named tokens. Then, we adjust for negations using a rule-based approach to detect negation scopes and invert the meaning accordingly (Dadvar et al. 2011; Pröllochs et al. 2015a). In a next step, we remove so-called stop words, which are words without relevance, such as articles and pronouns (Lewis et al. 2004). Finally, we perform stemming in order to truncate all inflected words to their stem, using the so-called Porter stemming algorithm.

Upon completion of the pre-processing, 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 one such 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 $W_{\mathrm{pos}}(A)$ and negative $W_{\mathrm{neg}}(A)$ words divided by the total count of words $W_{\mathrm{tot}}(A)$ across all announcements A on day t. Let μ denote the mean of the Net-Optimism metric and σ its standard deviation. We can then introduce a standardized sentiment metric $S^*(t)$ formally by

$$S(t) = \frac{\sum_A W_{\text{pos}}(A) - W_{\text{neg}}(A)}{\sum_A W_{\text{tot}}(A)} \in [-1, +1] \qquad \text{and} \qquad S^*(t) = \frac{S(t) - \mu}{\sigma} \in (-\infty, +\infty) \tag{15}$$

in order to facilitate calculations and later comparisons. The standardized sentiment is scaled to a zero mean with a standard deviation of one. The use of the standardized Net-Optimism news sentiment in the following analysis will henceforth be referred to as *news sentiment*.

References

Akhtar, S., Faff, R., Oliver, B., & Subrahmanyam, A. (2011). The Power of Bad: The Negativity Bias in Australian Consumer Sentiment Announcements on Stock Returns. Journal of Banking & Finance, *35*(5), 1239–1249.

Akhtar, S., Faff, R., Oliver, B., & Subrahmanyam, A. (2013). Reprint of: Stock salience and the asymmetric market effect of consumer sentiment news. Journal of Banking & Finance, *37*(11), 4488–4500.

Alfano, S., Feuerriegel, S., & Neumann, D. (2015). Is News Sentiment More Than Just Noise? In 23rd European Conference on Information Systems (ECIS 2015) (Paper 5). Münster, Germany.

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. Econometrica, *21*(4), 503–546.

Angrist, J. D. & Evans, W. N. (1998). Children and Their Parents Labor Supply: Evidence from Exogenous Variation in Family Size. American Economic Review, *88*(3), 450–477.

Angrist, J. D. & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. Journal of Economic Perspectives, *15*(4), 69–85.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On Making Causal Claims: A Review and Recommendations. The Leadership Quarterly, *21*(6), 1086–1120.

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

Bajo, E. & Raimondo, C. (2015). Media Sentiment and the Pricing of IPOs. In 2015 FMA European Conference.

Barberis, N. (2012). *Thirty Years of Prospect Theory in Economics: A Review and Assessment*. Cambridge, MA: National Bureau of Economic Research.

Baumeister, C. & Kilian, L. (2012). Real-Time Forecasts of the Real Price of Oil. Journal of Business & Economic Statistics, *30*(2), 326–336.

Bencivenga, C., D'Ecclesia, R. L., & Triulzi, U. (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., O'Hara, M., & Saar, G. (2009). How Noise Trading Affects Markets: An Experimental Analysis. Review of Financial Studies, *22*(6), 2275–2302.

Bosman, R., Kraussl, R., & Mirgorodskya, E. (2014). The Framing Effect of News on Invesotr Beliefs: An experimental Approach, 1–73.

Brogaard, J., Hendershott, T., & Riordan, R. (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. & Cliff, M. T. (2004). Investor Sentiment and the Near-Term Stock Market. Journal of Empirical Finance, *11*(1), 1–27.

Brown, G. W. & Cliff, M. T. (2005). Investor Sentiment and Asset Valuation. The Journal of Business, *78*(2), 405–440.

Buchinsky, M. (1994). Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression. Econometrica, *62*(2), 405–458.

Buchinsky, M. (1998). Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research. The Journal of Human Resources, *33*(1), 88–126.

Camerer, C., Loewenstein, G., & Rabin, M. (Eds.). (2004). *Advances in Behavioral Economics*. The Roundtable Series in Behavioral Economics. New York: Russell Sage Foundation and Princeton University Press.

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

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

Chernozhukov, V. & Hansen, C. (2004). The Effects of 401(K) Participation on the Wealth Distribution: An Instrumental Quantile Regression Analysis. Review of Economics and Statistics, *86*(3), 735–751.

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

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

Constantinides, G., Harris, M., & Stulz, R. M. (Eds.). (2003). *Financial Markets and Asset Pricing*. Handbook of the Economics of Finance. Amsterdam: Elsevier.

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

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

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

Demirer, R. & Kutan, A. M. (2010). The Behavior of Crude Oil Spot and Futures Prices around OPEC and SPR Announcements: An Event Study Perspective. Energy Economics, *32*(6), 1467–1476.

Edwards, W. (1954). The Theory of Decision Making. Psychological Bulletin, *51*(4), 380–417.

Engelberg, J. E. & Parsons, C. A. (2011). The Causal Impact of Media in Financial Markets. Journal of Finance, *66*(1), 67–97.

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

Feuerriegel, S., Heitzmann, S. F., & Neumann, D. (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. & Neumann, D. (2013). News or Noise? How News Drives Commodity Prices. In Proceedings of the 34th International Conference on Information Systems (ICIS 2013).

Fitzenberger, B., Hujer, R., MaCurdy, T. E., & Schnabel, R. (2002). Testing for Uniform Wage Trends in West-Germany: A Cohort Analysis Using Quantile Regressions for Censored Data. In B. Fitzenberger, R. Koenker, & J. A. F. Machado (Eds.), Economic Applications of Quantile Regression (pp. 41–86). Heidelberg: Physica-Verlag HD.

Fitzenberger, B., Koenker, R., & Machado, J. A. F. (Eds.). (2002). *Economic Applications of Quantile Regression*. Heidelberg: Physica-Verlag HD.

Friedman, M. (1953). *Essays in Positive Economics*. Chicago, IL: University of Chicago Press.

Gerber, A., Hens, T., & Vogt, B. (2002). Rational Investor Sentiment. SSRN Electronic Journal.

Grinblatt, M. & Han, B. (2005). Prospect Theory, Mental Accounting, and Momentum. Journal of Financial Economics, *78*(2), 311–339.

Groth, S. S. & Muntermann, J. (2011). An Intraday Market Risk Management Approach based on Textual Analysis. Decision Support Systems, *50*(4), 680–691.

Harvey, A. C. (1990). *Forecasting, structural time series models, and the Kalman filter*. Cambridge: Cambridge University Press.

Hausman, J. A. (1978). Specification Tests in Econometrics. Econometrica, *46*(6), 1251–1271.

Haven, E., Liu, X., & Shen, L. (2012). De-Noising Option Prices with the Wavelet Method. European Journal of Operational Research, *222*(1), 104–112.

Hendershott, T. & Menkveld, A. J. (2014). Price Pressures. Journal of Financial Economics, *114*(3), 405–423.

Hendershott, T., Menkveld, A. J., Li, S. X., & Seasholes, M. S. (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.

Hong, H., Lim, T., & Stein, J. C. (2000). Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies. Journal of Finance, *55*(1), 265–295.

Jegadeesh, N. & Wu, D. (2013). Word Power: A New Approach for Content Analysis. Journal of Financial Economics, *110*(3), 712–729.

Johnson, L. D. & Sakoulis, G. (2008). Maximizing equity market sector predictability in a Bayesian time-varying parameter model. Computational Statistics & Data Analysis, *52*(6), 3083–3106.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica, *47*(2), 263–292.

Kahneman, D. & Tversky, A. (1984). Choices, Values, and Frames. American Psychologist, *39*(4), 341–350.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering, *82*(1), 35–45.

Kanouse, D. E. (1984). Explaining Negativity Biases in Evaluation and Choice Behavior: Theory and Research. Advances in Consumer Research, *11*(1), 703–708.

Kanouse, D. & Hanson, R. (1972). Negativity in Evaluations. In Attribution (pp. 47–62). Morristown: General Learning Press.

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. & Park, C. (2009). THE IMPACT OF OIL PRICE SHOCKS ON THE U.S. STOCK MARKET. International Economic Review, *50*(4), 1267–1287.

Kilian, L. & Vega, C. (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.

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. New York: Cambridge University Press.

Koenker, R. & Bassett, G. (1978). Regression Quantiles. Econometrica, *46*(1), 33–50.

Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied Linear Regression Models*. New York: McGraw-Hill/Irwin.

Kyle, A. (1985). Continuous Auctions and Insider Trading. Econometrica, *53*(6), 1315–1335.

Lechthaler, F. & Leinert, L. (2012). Moody Oil: What is Driving the Crude Oil Price? Zurich.

Lee, C., Shleifer, A., & Thaler, R. (1991). Investor Sentiment and the Closed-End Fund Puzzle. Journal of Finance, *46*(1), 75–109.

Lewis, D., Yang, Y., Rose, T., & Li, F. (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.

Liebmann, M., Hagenau, M., & Neumann, D. (2012). Information Processing in Electronic Markets: Measuring Subjective Interpretation Using Sentiment Analysis. In Proceedings of the International Conference on Information Systems (ICIS 2012). Association for Information Systems.

Lopes, H. F. & Tsay, R. S. (2011). Particle Filters and Bayesian Inference in Financial Econometrics. Journal of Forecasting, *30*(1), 168–209.

Loughran, T. & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. Journal of Finance, *66*(1), 35–65.

Loughran, T. & McDonald, B. (2013). IPO First-Day Returns, Offer Price Revisions, Volatility, and Form S-1 Language. Journal of Financial Economics, *109*(2), 307–326.

MacGregor, P. (2013). International News Agencies: Global Eyes that Never Blink. In K. Fowler-Watt & S. Allan (Eds.), Journalism (pp. 35–63). Centre for Journalism & Communication Research, Bournemouth University.

MacKinlay, A. C. (1997). Event Studies in Economics and Finance. Journal of Economic Literature, *35*(1), 13–39.

Manning, C. D. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Manoliu, M. & Tompaidis, S. (2002). Energy futures prices: Term structure models with Kalman filter estimation. Applied Mathematical Finance, *9*(1), 21–43.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams Engineering Journal, *5*(4), 1093–1113.

Mendel, B. & Shleifer, A. (2012). Chasing Noise. Journal of Financial Economics, *104*(2), 303–320.

Minev, M., Schommer, C., & Grammatikos, T. (2012). News and Stock Markets: A Survey on Abnormal Returns and Prediction Models. Luxembourg.

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

Mittermayer, M.-A. & Knolmayer, G. F. (2006b). Text Mining Systems for Market Response to News: A Survey. Bern, Switzerland.

Muntermann, J. & Guettler, A. (2007). Intraday Stock Price Effects of Ad Hoc Disclosures: The German Case. Journal of International Financial Markets, Institutions and Money, *17*(1), 1–24.

Narayan, S. & Narayan, P. K. (2016). Are Oil Price News Headlines Statistically and Economically Significant for Investors? Journal of Behavioral Finance (forthcoming).

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

Neumann, A., Siliverstovs, B., & Hirschhausen, C. v. (2006). Convergence of European spot market prices for natural gas? A real-time analysis of market integration using the Kalman Filter. Applied Economics Letters, *13*(11), 727–732.

Owens, J. P. & Steigerwald, D. G. (2006). Noise reduced realized volatility: a kalman filter approach. In Econometric Analysis of Financial and Economic Time Series (Vol. 20, pp. 211–227). Advances in Econometrics. Bingley: Emerald (MCB UP).

Pang, B. & Lee, L. (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.

Peeters, G. (1971). The Positive-Negative Asymmetry: On Cognitive Consistency and Positivity Bias. European Journal of Social Psychology, *1*(4), 455–474.

Peeters, G. & Czapinski, J. (1990). Positive-Negative Asymmetry in Evaluations: The Distinction Between Affective and Informational Negativity Effects. European Review of Social Psychology, *1*(1), 33–60.

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

Pröllochs, N., Feuerriegel, S., & Neumann, D. (2015b). Generating Domain-Specific Dictionaries Using Bayesian Learning. In 23rd European Conference on Information Systems (ECIS 2015). Münster, Germany.

Rozin, P. & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. Personality and Social Psychology Review, *5*(4), 296–320.

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

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

Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating Sentiment in Financial News Articles. Decision Support Systems, *53*(3), 458–464.

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

Sharma, A. & Dey, S. (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) (pp. 1–7). New York, NY: ACM.

Shleifer, A. (2000). *Inefficient Markets: An Introduction to Behavioral Finance*. Clarendon Lectures in Economics. Oxford: Oxford University Press.

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

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

Simon, H. A. (1955). A Behavioral Model of Rational Choice. The Quarterly Journal of Economics, *69*(1), 99.

Simon, H. A. (1962). The Architecture of Complexity. Proceedings of the American Philosophical Society, *106*(6), 467–482.

Staiger, D. & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. Econometrica, *65*(3), 557–586.

Stieglitz, S. & Dang-Xuan, L. (2013). Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. Journal of Management Information Systems, *29*(4), 217–248.

Stock, J. H. & Watson, M. W. (2011). *Introduction to Econometrics* (3rd ed.). Boston, MA: Addison-Wesley.

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., Saar-Tsechansky, M., & Macskassy, S. (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*. Roundtable Series in Behavioral Economics. Princeton, NJ: Princeton University Press.

Tsai, I.-C. (2012). The Relationship Between Stock Price Index and Exchange Rate in Asian Markets: A Quantile Regression Approach. Journal of International Financial Markets, Institutions and Money, *22*(3), 609–621.

Tversky, A. & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. Science, *185*(4157), 1124–1131.

Tversky, A. & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. Science, *211*(4481), 453–458.

Tversky, A. & Kahneman, D. (1986). Rational Choice and the Framing of Decisions. The Journal of Business, *59*(4), 251–278.

Tversky, A. & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty, *5*(4), 297–323.

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press.

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

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

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach* (4th ed.). Mason, OH: South Western, Cengage Learning.

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

Zhang, W. & Semmler, W. (2009). Prospect theory for stock markets: Empirical evidence with time-series data. Journal of Economic Behavior & Organization, *72*(3), 835–849.

Zivot, E. & Andrews, D. W. K. (2002). Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. Journal of Business & Economic Statistics, *20*(1), 25–44.