2014 47th Hawaii International Conference on System Science

News Processing during Speculative Bubbles: Evidence from the Oil Market

Stefan Feuerriegel
University of Freiburg
Freiburg, Germany
stefan.feuerriegel@is.uni-freiburg.de

Max W. Lampe
University of Freiburg
Freiburg, Germany
lampe142@googlemail.com

Dirk Neumann
University of Freiburg
Freiburg, Germany
dirk.neumann@is.uni-freiburg.de

Abstract

Speculative bubbles are commonly referred to situations where stock prices considerably deviate from their fundamentals until the bubbles bust. Bursting of bubbles such as the dot-com or U. S. housing bubble is very costly, so there is a need for mechanisms to detect them. In this paper, we attempt to predict when bubbles may bust using the sentiment of news announcements. Accordingly, we first try to understand how news reception evolves depending on the market phase (boom or bust). The probability of bubble bursts are calculated on the basis of a Markov-regime switching model. The approach is applied and validated using the oil market which appears to be one of the most important markets in the globalized world. Our methodology can be similarly extended to other markets such as gold or wheat.

1. Introduction

Today's financial markets are getting increasingly vulnerable to speculative bubbles. In the past decades, the U. S. stock market has been troubled by a series of severe price frenzies ranging from the crash of the dot-com bubble in 2000 to the financial crises in 2007. According to Shiller's definition [1], a speculative bubble can be understood as a social epidemic whose contagion is coordinated by the prices. Increasing prices are very quickly disseminated by successful investors to spread news about their success. This in turn attracts more people into the market, which further raises the prices. This is again associated with news releases about their successes, which fuel the bubble growth as a result of the positive feedback loops. Once the bubble bursts, the falling prices kick off the same contagion process in reverse direction as more and more people exit the market creating pessimistic news about the economy [2].

This episodic description already indicates that speculative bubbles are necessarily associated with news and newspapers [1]. It is undisputed that media and news can actually affect asset prices (e. g. [3]–[6]). In speculative bubbles, boom periods are characterized by extreme market exuberances that are followed by bust periods which denote downturns.

Recent financial studies have started to incorporate psychological phenomena into their research designs in order to explain effects on the stock market that seemed irrational or chaotic at the first glance. For example, psychological explanations attribute those speculative exuberances to cognitive biases such as overconfidence. In addition to that, evidence from psychology suggests that investors react differently to news depending on the market state. More precisely, investors tend to react more to news when primed into negative mood states such as bust phases [7]. In this paper, we address the question of whether these psychological biases impact how news is interpreted by investors during boom and bust periods.

In our research design, we adopt the oil market as reference market since the oil market has drawn significant attention in the globalized world. With manufacturing businesses becoming highly dependent on oil, demand for oil has literally exploded, which is reflected by the vast increase in traded oil future volume reaching 138.5 million contracts with each accounting for 1000 barrel in 2007 [8]. In recent years, oil markets observed extreme price peaks. For example, the price of WTI crude oil price started at \$64.59 per barrel on April 3, 2007, then rised to \$145.31 per barrel on July 3, 2008. Apparently, oil prices have increased by 125 % within a period of 15 months. As a consequence of these price developments, the oil market has attracted many speculators and, thus, established the basis for speculative bubbles.

We use the oil market to find out whether investors change their news processing behavior over time exhibiting the animal spirits that Keynes once pro-

978-1-4799-2504-9/14 \$31.00 © 2014 IEEE
DOI 10.1109/HICSS.2014.506

IEEE computer society

posed [9]. More specifically, we hypothesize a structural change in the relationship how news sentiment affects oil prices between *bullish* and *bearish* markets. Although the coefficients measuring absorption are almost equal, we find evidence that the relationship in *bullish* markets is much stronger in terms of significance than during *bearish* periods.

Furthermore, we analyze how magnitude and direction of news sentiment have changed over time. As a result, we identify that markets absorbed news very quickly during the 2008/09 sharp fall in oil prices. This is shown by a high regression coefficient and, thus, gives indication of exaggerated speculation in oil markets. Effectively, this result is consistent with our expectation that investors react more aggressively in bust phases.

Ultimately, we want to provide evidence that news sentiment, in fact, can predict *bullish* and *bearish* regimes. Here, we expect a lagged relationship between news sentiment and the corresponding regime. Outliers in the news sentiment denote those messages which contain news with a diametrically different sentiment than before. Those outliers may indicate a reversal in the market sentiment having the potential to burst the bubble.

The remainder of this paper is structured as follows. In Section 2, we review literature on detecting periodically collapsing bubbles, related IS publications on information processing in commodity markets and compare approaches for sentiment analysis. To gauge the sentiment of oil-related news, we give our research model in Section 3. Finally in Section 4, we compare information processing in *bullish* and *bearish* markets and, afterwards, show a path on how to predict market regimes using news sentiment. Section 5 concludes the paper with a summary and an outlook on future research.

2. Related work

In this section, we present related literature grouped into two categories. First, methods to detect speculative bubbles are revisited. Second, we compare approaches that measure news sentiment. Third, we review previous work on information processing in oil markets. All in all, the following references provide evidence that linking news sentiment with the detection of speculative bubbles is both a novel and relevant research question to the Information Systems community.

2.1. Testing for speculative bubbles

In literature, the concept of speculative bubbles, i. e. the price building process of assets, is examined to a large extent [10]. Opposite to theoretical models, knowledge on detecting bubbles empirically is rare. The most notable directions towards are as follows. One of the first tests developed is the *Cointegration test* [11]. The principle of the test is that a cointegration relationship should exist between prices and fundamentals. Deviations from this relationship should only last for short periods while the relationship between prices and fundamentals should remain throughout the whole time series. Another broadly-used approach is the *Variance Bounds test*. It tests if the variances of prices are justified by the variances of fundamentals [12]. However, both approaches face two crucial drawbacks. First, both tests rely upon a (given) data set of the fundamental values. While fundamentals such as inflation, building permits, etc. can be easily derived in the real estate sector, this turns out to be difficult when it comes to the oil market. Second, a major pitfall of the Cointegration test is the lack of detecting periodically collapsing bubbles [13]. As it is a reasonable assumption that bubbles appearing in oil markets feature periodical characteristics, the Cointegration test shows only limited applicability in the oil domain.

Thus, we rely on so-called *Markov-regime switching methods*. These are not only independent of fundamental variables, but have been proven to work effectively along with testing periodically collapsing bubbles [14]. This approach computes the probability that the market is in either an explosive (*bullish*) or stationary (*bearish*) regime. As a further advantage, it is possible to identify time frames when prices develop speculative bubbles. With this benefit at hand, we decide to integrate the Markov-regime switching method to detect speculative bubbles in the oil domain.

2.2. Methods for sentiment analysis of financial news

Methods that use the textual representation of documents to measure the positivity and negativity of the content are referred to as opinion mining or *sentiment analysis*. In fact, sentiment analysis can be utilized to extract subjective information from text sources as well as to measure how market participants perceive and react upon news. Here, one uses the observed stock price reactions following the news announcement to validate the accuracy of the sentiment analysis routines. Based upon sentiment measures, one can study the relationship between news and their effect on stock markets. On top of that, empirical evidence shows that a discernible relation between news content and its stock market reaction exists [3], [15].

As sentiment analysis is applied to a broad variety of domains and text sources, research has devised various approaches (cf. [16] as a comprehensive domain-independent survey) to measure sentiment. Within finance, recent literature surveys [17], [18] compare studies aiming at stock markets prediction. For example, *dictionary-based approaches* are very frequently (cp. [4], [5], [19], [20]) used in recent financial text mining research. These methods count the frequency of pre-defined positive and negative words from a given dictionary – producing results that are straightforward and reliable. *Machine learning approaches* (e. g. [15], [21]–[23]) offer a broad range of methods, but might suffer from overfitting [24].

In our research framework, we have to process only few data points linked with a large text basis (i. e. hundreds of announcements from a single day) along with a continuous return. Thus, we have experienced difficulties to gain robust results with machine learning and term weighting approaches and focused, instead, on rule-based methods. In fact, we have tested [25] all of the above rule-based metrics in combination with various dictionaries. We find that the Net-Optimism approach [5], [19] along with Henry's Finance-Specific Dictionary [5] outperforms all others.

2.3. Information processing in oil markets

Up to large extent, market efficiency relies upon the availability of information [26]. Access to market information is promoted with ease in electronic markets and, because of the straightforward access, decision makers (i. e. consumers, suppliers and intermediaries) can use more information to make purchases and sales more beneficial (e. g. [27]). In fact, it is both native and crucial to IS research how decision makers process and act upon (qualitative) information in oil markets. While information processing has been extensively studied in capital markets, literature focusing on commodity markets is rare: *"even though information system methodology, i. e. text mining, is common when it comes to financial (stock) market predictions based on analyses of financial (ad-hoc) messages, the literature is remarkably silent about the oil domain"* [28].

According to previous research, text mining approaches can predict the direction of oil price changes [29] and its magnitude [30]. However, these approaches are impractical when it comes to monitoring the effect of news sentiment on the magnitude of commodity price movements. We are aware of only one study [31] that analyzes the long-term effects of news sentiment along with various variables on crude oil prices. Similarly, empirical evidence shows that abnormal returns in commodity markets can be explained, up to a large extent, by news sentiment [25].

3. Research methodology: from news to sentiment

This section introduces our research methodology as depicted in Fig. 1. In a first step, only those news announcements are filtered that fit our research focus. Then, each announcement is subject to *preprocessing* steps (Section 3.1) which transforms the running text into machine-readable tokens. Tokens of all daily announcements are aggregated to compute the corresponding *news sentiment* in Section 3.2. Then, we analyze the influence of news sentiment on abnormal returns (Section 3.3) by performing an *event study*. In order to link sentiment and speculation, we present the Markov-regime switching method (Section 3.4) as an approach to detect periodically collapsing bubbles.

3.1. Preprocessing news announcement

Before performing the actual sentiment analysis, several operations are involved during a preprocessing phase. The individual steps are as follows.

- **Tokenization.** Each announcement is split into sentences and single words named *tokens*.
- **Negations.** Negations invert the meaning of words and sentences. When encountering the word *no*, each of the subsequent three words (i. e. the object) is counted as word from the opposite dictionary. When encountering other negating terms (*rather*, *hardly*, *couldn't*, *wasn't*, *didn't*, *wouldn't*, *shouldn't*, *weren't*, *don't*, *doesn't*, *haven't*, *hasn't*, *won't*, *hadn't*, *never*), the meaning of all succeeding words is inverted [32].
- **Stop word removal.** Words without a deeper meaning such as *the*, *is*, *of*, etc. are named *stop words* and, thus, can be removed. We use a list of 571 stop words [33].
- **Synonym merging.** Synonyms, though are spelled differently, convey the same meaning. Thus, approximately 150 frequent synonyms are grouped and aggregated by their meaning – a method referred to as *pseudoword* generation [34].
- **Stemming.** Stemming refers to the process for reducing inflected words to their stem [34]. Here, we use the so-called Porter stemming algorithm.

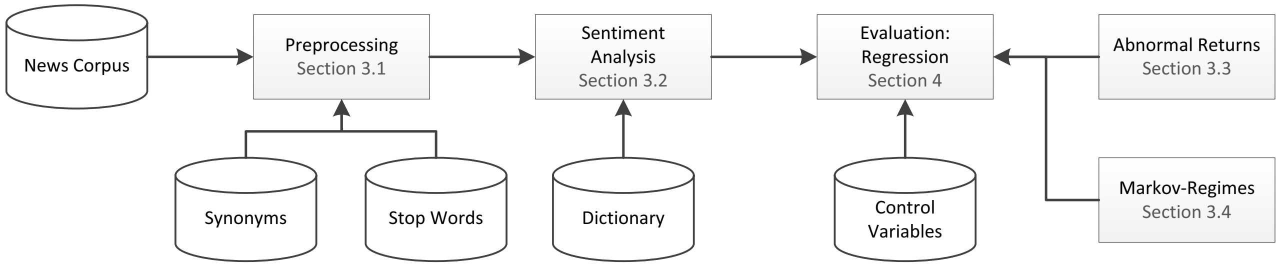

Figure 1. Research model with information processing and event study methodology.

3.2. Method for analyzing news sentiment

As shown in a recent study [25] on the robustness of sentiment analysis, the correlation between news sentiment and abnormal returns in commodity markets varies across different sentiment metrics. A sentiment approach that unfolds a reliable correlation is the Net-Optimism metric [5], [19]. Out of these, Net-Optimism along with Henry's Finance-Specific Dictionary [5] achieves the highest robustness and, consequently, we rely upon this approach in the following evaluation.

Let us briefly recapitulate the Net-Optimism approach. As this measures was originally developed to analyze the sentiment of a single news announcement, we present an extended version that aggregates the sentiment of all announcements from one day into a sentiment value $S_{\mathrm{NO}}(t)$ [7], [25] that represents the daily news stream. Thus, let $W_{\mathrm{tot}}(A)$ denote the total number of words in the announcement A; $W_{\mathrm{neg}}(A)$ denote the number of negative words in the announcement A; and $W_{\mathrm{pos}}(A)$ denote the total number of positive words in the announcement A. Net-Optimism is defined by

$$S_{\mathrm{NO}}(t) = \frac{\sum_A W_{\mathrm{pos}}(A) - W_{\mathrm{neg}}(A)}{\sum_A W_{\mathrm{tot}}(A)}. \tag{1}$$

Thus, Net-Optimism $S_{\mathrm{NO}}(t) \in [-1, +1]$ measures the difference between the count of positive and negative words normalized by the number of total words.

3.3. Event study methodology and abnormal returns

Event studies use financial market data to inspect changes in financial values due to a specific event and measure its impact. Information Systems research exploits event study methodology frequently and turns it into both an effective and widespread approach [35].

For each event of interest, one predicts a *normal* return in the absence of the event and, then, estimates the difference between actual and normal return which is defined as the *abnormal* return [36]. In our research, the event of interest consists of all daily oil-related announcements from the news corpus. We set the period, during which the oil price is examined (i. e. the event window), to the single day of the announcement stream as we are provided with daily financial market data. The normal return is defined as the expected return without conditioning on the event X_τ taking place. The abnormal return is defined by

$$AR(\tau) = R(\tau) - \mathrm{E}(R(\tau) \mid \neg X_\tau) \tag{2}$$

where $AR(\tau)$, $R(\tau)$ and $\mathrm{E}(R(\tau) \mid \neg X_\tau)$ are the abnormal, actual and normal returns in period τ.

As a next step, the normal return is estimated during a time interval named *estimation window*. We calculate the normal return by the so-called *market model*. The market model assumes a stable linear relation between the market return $R_m(t)$ and the normal return, i. e. the return of the market portfolio. More precisely, the market model is

$$R(t) = \alpha + \beta R_m(t) + \varepsilon_t, \tag{3a}$$

$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2 \tag{3b}$$

where $R(t)$ and $R_m(t)$ are returns in period t of oil and on the market portfolio, respectively, and ε_t is the zero mean disturbance term. Here, α, β and σ_ε^2 are the parameters of the market model. These are determined from a regression such as ordinary least squares (OLS) based on the values from the estimation window. Ultimately, the abnormal return in period τ is computed by

$$AR(\tau) := R(\tau) - \alpha - \beta R_m(\tau). \tag{4}$$

3.4. The Markov-regime switching method

When it comes to detecting periodically collapsing bubbles, a frequent approach relies upon so-called Markov-regime switching methods [37], [38]. Here, the market is always in of two *regimes*. While we mathematically denote these states s_t by 0 or 1, a

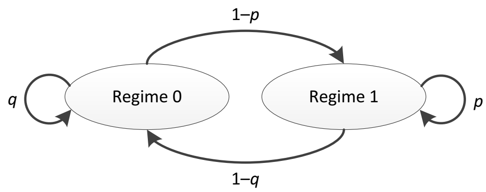

Figure 2. Markov-switching model with two regimes – *bullish* and *bearish* markets – and corresponding transition probabilities.

more obvious interpretation is given by either *bullish* or *bearish* market regimes. As shown in Fig. 2, the market regime can remain constant or change at time $t+1$ where the variables p and q define the probabilities that the model remains in the given regime, whereas $1-p$ and $1-q$ give the so-called *transition* probabilities for switching to the opposite regime. Altogether, we can now specify the probabilities for a transition from s_{t-1} to s_t. Here, the probabilities (given by the probability function Pr) for a state s_t at a time t with a previous states s_{t-1} are as follows

$$\Pr(s_t = 1 \,|\, s_{t-1} = 1) = p, \tag{5a}$$
$$\Pr(s_t = 0 \,|\, s_{t-1} = 1) = 1-p, \tag{5b}$$
$$\Pr(s_t = 0 \,|\, s_{t-1} = 0) = q, \tag{5c}$$
$$\Pr(s_t = 1 \,|\, s_{t-1} = 0) = 1-q. \tag{5d}$$

It is assumed that the parameters governing this approach are time-varying, i. e. changing with the unobserved regime s_t. Then, each of the regimes is modeled by a separate autoregressive process. Given a time series y_t that we want to model, the two-regime Markov switching model is specified by

$$\Delta y_t \tag{6}$$
$$= \begin{cases} \mu_0 + \phi_0 y_{t+1} + \sum_{j=1}^{k} \psi_{0,j} \Delta y_{t-j} + \nu_{0,t}, & \text{if } s_t = 0, \\ \mu_1 + \phi_1 y_{t+1} + \sum_{j=1}^{k} \psi_{1,j} \Delta y_{t-j} + \nu_{1,t}, & \text{if } s_t = 1, \end{cases}$$

depending on the current regime s_t where μ_i, ϕ_i and $\psi_{i,j}$ are real parameters. The sequences $\nu_{0,t}$ and $\nu_{1,t}$ are zero-mean white noise and k is a suitably chosen integer. In order to identify price bubbles [39], we note that the Markov-regime switching approach searches for a regime change in the first differences (i. e. $\Delta y_t = y_t - y_{t-1}$) of a time series. However, the parameters in Eqn. (6) are difficult to estimate because of the piecewise definition and, thus, one combines both cases into a single equation

$$\Delta y_t = \mu_0(1-s_t) + \mu_1 s_t + [\phi_0(1-s_t) + \phi_1 s_t]\, y_{t-1} + \sum_{j=1}^{k} [\psi_{0,j}(1-s_t) + \psi_{1,j} s_t]\, \Delta y_{t-j} + \sigma\varepsilon_t. \tag{7}$$

Here, ε_t is a sequence of independent and identically distributed (i. i. d.) random variables with zero mean and unit variance. Estimating the parameters in Eqn (7) is achieved by the Expectation-Maximization algorithm.

However, in order to interpret s_t as *bullish* and *bearish* regimes, Eqn. (7) must fulfill further conditions. One regime should have an *explosive* characteristics, while the other is stationary. Mathematically speaking, a regime is stationary if $\phi_i < 0$, whereas $\phi_i > 0$ identifies an explosive process. Within the Markov-regime switching framework, the existence of an explosive rational bubble in prices is consistent with $\phi_0 > 0$ or $\phi_1 > 0$. This indicates that one of the states governing the process of interest is characterized by the presence of an explosive regime [39]. Thus, a regime switch can be observed whenever the variable in use changes from ϕ_0 to ϕ_1 or vice versa.

Lastly, fitting Eqn. (7) results in a probability that the model at time t is in an explosive or the stationary regime. Based upon these probabilities, time frames with explosive characteristics are considered as speculative bubbles.

4. Empirical evaluation: analyzing speculation in the oil market

Having discussed the steps to compute the news sentiment, we apply the sentiment analysis to investigate how oil markets are driven by speculative components. First, we specify the news corpus and describe the regression design that inspects information processing in both *bullish* and *bearish* oil markets. Afterwards, we link news sentiment with the Markov-regime switching method to provide evidence that predicting *bullish* and *bearish* regimes is possible.

4.1. News corpus

Our news corpus originates from the *Thomson Reuters News Archive* for Machine Readable News. We choose Reuters news deliberately because of four reasons: (1) Reuters conveys, in particular, news about commodity markets. (2) Reuters news is third-party content and, thus, give a certain level of objectivity. (3) Opposed to newspapers, news agencies feature a shorter time lag and lack from perturbations by edits.

All announcements provided by Reuters arise from the time span January 1, 2003 till May 31, 2012. The announcements come along with additional labels indicating their content. Based upon these labels, the news corpus is filtered such that we extract announcement focusing on the oil market[1]. All in all, this set of criteria filters a total of $339{,}446$ announcements related to crude oil.

4.2. Boom and bust regimes in oil markets

We derive boom and bust phases in oil markets by using the above Markov-regime switching method. The individual steps are as follows. We use monthly nominal WTI crude oil prices from June 1997 till December 2012 as the time series y_t. According to the results from an autocorrelation test, we integrate $k = 2$ autoregressive coefficients. All estimated parameters as well as corresponding standard errors are given in Table 4. Based on these estimates, we compute the probabilities of an explosive regime. The lower plot in Fig. 3 illustrates the smoothed probabilities for the explosive regime. All explosive regimes are highlighted by blue time windows. The upper diagram in Fig. 3 shows how both the real and nominal oil price evolves. Although there is strong increase in price during the 2007–2009 phase, this does not trigger a regime switch because the first differences remain stationary since, in fact, oil markets were even before 2007 very volatile.

4.3. Regression design

In this section, we investigate how investors react to related news announcements and analyze information processing in oil markets empirically. To succeed in this goal, we present the regression design from [25] which links abnormal returns and news sentiment.

Instead of classical commodity prices, we perform an event study to extract the effect of individual events. Let $AR(t)$ define the abnormal return of oil and use a market model. We model the market portfolio

1. This is achieved by applying a set of filter criteria [25]: (1) The language must be English. (2) The event type is *Story Take Overwrite* to guarantee that we not yield an alert but the actual message. (3) Special types of announcements such as alerts or personal opinions might have limited relevance to changes in the oil market and we want to exclude these. Thus, we omit announcements that contain specific words (*advisory*, *chronology*, *corrected*, *feature*, *diary*, *instant view*, *analysts view*, *newsmaker*, *corrected*, *refile*, *rpt*, *schedule*, *table*, *service*, *alert*, *wrapup*, *imbalance*, *update*) in their headline. (4) We use topic code *CRU* to filter announcements that deal with crude oil. (5) We exclude announcements addressing changes in prices to avoid simultaneity. (6) In order to remove white noise, we require announcements to count at least 50 words.

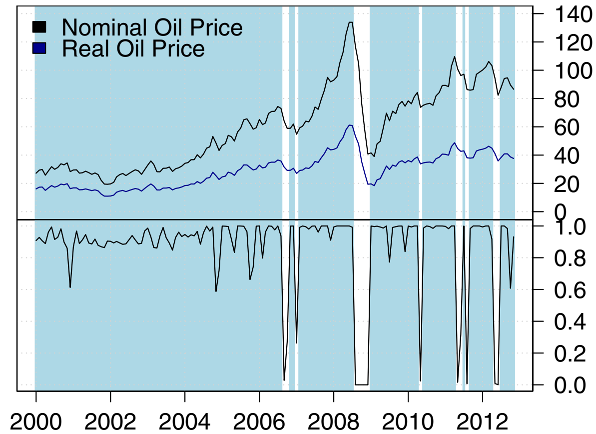

Figure 3. Oil price in \$ per barrel (top) and smoothed probabilities of explosive regime (bottom) from January 1, 2003 till May 31, 2012.

Table 4. Estimated parameters of the Markov-regime switching model.

	Estimate		Standard Error
μ_0	2.1135		1.3001
ϕ_0	-0.186		0.0463
ψ_{01}	0.0361		0.1931
ψ_{02}	0.5767		0.1864
μ_1	-0.6756		0.3101
ϕ_1	0.0528		0.0113
ψ_{11}	-0.0257		0.0734
ψ_{12}	-0.0181		0.0666
p	0.4037	AIC	748.0797
q	0.9088	BIC	831.4315

using a commodity index, namely the Dow Jones-UBS Commodity Index[2] [40], [41] along with an event window of 10 trading days [42] prior to the event. The actual oil price comes from the benchmark oil price in the US (provided by Datastream) which is the West Texas Intermediate (WTI).

The key independent variable to the linear model [25] is the sentiment metric. Instead of a single sentiment $S_{\mathrm{NO}}(t)$, we split this metric into two separate values $S_{\mathrm{Bull}}(t)$ and $S_{\mathrm{Bear}}(t)$ depending on the market regime. These represent the news sentiment if the market is in the corresponding regime and return 0 otherwise. We also incorporate a set of control factors to check for internal (market model α and cumulative

2. The Dow Jones-UBS Commodity Index (formerly Dow Jones-AIG Commodity Index) is a highly liquid and diversified reference for the commodities market consisting of around twenty physical commodities and is used in the model as a proxy for the commodities market performance.

abnormal returns CAR) and external effects[3]. Then, we can specify the regression model with error terms ε_t by

$$AR_{\log}(t) = \beta_0 + \beta_1 S_{\text{Bull}}(t) + \beta_2 S_{\text{Bear}}(t) + \beta_3 \alpha + \beta_4 CAR + \sum_i \gamma_i CV_i(t) + \varepsilon_t. \quad (8)$$

Both the abnormal returns $AR_{\log}(t)$ as well as all control variables $CV_i(t)$ consist of standardized log-returns. In addition to that, we add monthly dummy variables (to consider additional external events not covered by the control variables and to handle non-seasonally-adjusted time series). Finally, we give justice to extreme stock price effects and remove outliers at the $0.05\,\%$ level at both ends.

4.4. Analyzing news reception during speculative bubbles

Having discussed the regression design, we proceed to analyze news reception during speculative bubbles. Here, we use the results from Markov-regime switching approach to identify time frames with *bullish* and *bearish* market behavior.

Research Question 1: Are there differences in how news is absorbed between bullish and bearish markets?

We use the above regression design from Eqn. (8) to measure the impact of news sentiment on oil prices. To cater for two market regimes, we measure news reception in *bullish* and *bearish* market regimes separately. We tested for autocorrelation, heteroskedasticity, constant variance, serial correlation and normally distributed residuals at the $0.01\,\%$ level to ensure that the results are not confounded. When checking Variance Inflation Factors and the condition number of the matrix, we also see no indication of multicollinearity. Independence across announcements is given as long as all announcements are entirely novel and not based on an interrelated course of events. Under the assumption that commodity returns are jointly multivariate normal as well as independently and identically distributed through time, the model can be estimated using Ordinary Least Squares (OLS).

Regression results are given in Table 5. According to this table, we observe that, besides the alpha value from the market model, news sentiment influences abnormal returns significantly. When additionally comparing the coefficients of fundamental variables and news sentiment, we find that news sentiment coefficients (accounting for 1.24 and 1.34) exceed all other coefficients originating from fundamental variables strongly. Further, we notice a high adjusted R^2 value of approximately 0.38 indicating that one of the driving forces on abnormal returns is news sentiment.

When comparing coefficients from *bullish* and *bearish* market regimes, we come up with following findings. First, the coefficients of news sentiment show roughly the same magnitude. Thus, we conclude that we can detect between both market regimes no obvious difference in news receptions. This is contrast to [7] where the magnitude of coefficients during economic expansions and recessions differs more evidently. However, we note that both coefficients are linked with a different level of significance. In fact, a t-value of 25.55 in *bullish* markets is much higher than a t-value of 11.31 in *bearish* markets. Thus, the news reception is more stable and robust in *bullish* markets and, whereas in *bearish* markets, news reception is subject to higher fluctuations. To investigate this characteristic further, we motivate and analyze the following research question.

Research Question 2: Are there differences over time in how news is absorbed by the market?

To investigate news reception over time, we integrate a sentiment variable for each month. Thus, we gain monthly coefficients accounting for news reception. Again, we checked for heteroskedasticity, autocorrelation and multicollinearity to ensure that we can estimate the model using OLS. As a result, we can see the monthly coefficients in Fig. 6 (outliers start at the 1.5-fold of the IQR) that *bullish* regimes are distinguished from *bearish* regimes by outliers. However, a two-sample Wilcoxon test reveals a non-significant difference (P-value of 0.76) between the distributions of monthly coefficients in both regimes.

4.5. Prediction market regimes using news sentiment

As the above Research Questions finds strong evidence that news reception plays an important role in bubble creation, we advance to predict the market regimes based on historic news sentiment.

Research Question 3: Can news sentiment help to predict boom and bust phases in commodity markets?

We estimate the logit regression model

$$M(t) = \alpha + \beta \hat{S}_{\text{NO}}(t - \delta) + \varepsilon_t \quad (9)$$

for predictions (because of that we excluded control variables from time frame t) where the market regime

3. Based upon [25], we restrict our choice of control variables only to the significant values, namely, S&P 500 Index, Wheat Prices and Oil Future Contracts. All of them, show an influence that is significantly smaller than the influence of news sentiment.

Table 5. Pooled regression comparing news reception in bullish and bearish market regimes.

	(1)	(2)	(3)	(4)	(5)	(6)
$S_{\text{Bull}}(t)$ in *Bullish* Markets	1.12^{***}	1.16^{***}	1.25^{***}	1.25^{***}	1.23^{***}	1.24^{***}
	(20.60)	(22.31)	(25.55)	(25.58)	(25.32)	(25.55)
$S_{\text{Bear}}(t)$ in *Bearish* Markets	1.38^{***}	1.35^{***}	1.38^{***}	1.38^{***}	1.41^{***}	1.34^{***}
	(10.43)	(10.61)	(11.58)	(11.61)	(12.07)	(11.31)
Cumulative Abnormal Return CAR		-0.71^{***}	-0.08	-0.08	-0.09	-0.07
		(-14.86)	(-1.55)	(-1.42)	(-1.73)	(-1.34)
Market Model α			-1.11^{***}	-1.11^{***}	-1.09^{***}	-1.12^{***}
			(-19.98)	(-20.07)	(-19.84)	(-20.33)
S&P 500 Index				0.11^{**}	0.10^{**}	0.10^{*}
				(2.80)	(2.60)	(2.51)
Wheat Price					0.24^{***}	0.23^{***}
					(6.23)	(6.04)
Oil Future Contracts						0.12^{*}
						(2.09)
Intercept β_0	1.41^{**}	0.68	0.83^{*}	0.78	0.79	0.90^{*}
	(3.10)	(1.55)	(2.04)	(1.91)	(1.95)	(2.22)
Adjusted R^2	0.1908	0.2574	0.3698	0.3718	0.3811	0.3812
AIC	9818.50	9619.20	9283.72	9277.48	9249.57	9251.94
BIC	10484.72	10291.16	9961.43	9960.93	9938.77	9946.87

Stated: OLS coefficients, t-statistics in parenthesis; Dummies: monthly; Obs.: 4717 Significance: *** 0.001, ** 0.01, * 0.05

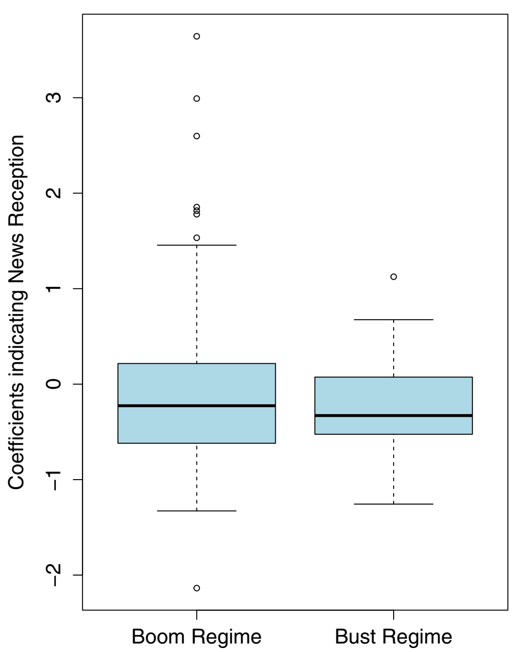

Figure 6. Outliers are more common in bullish markets.

is given by $M(t) \in \{\text{Bull}, \text{Bear}\}$ at time t and ε_t is the error. As market states are provided on a monthly basis, we combine all daily sentiment values to gain a monthly sentiment metric $\hat{S}_{\text{NO}}(t)$. In addition, we hypothesize a time lag δ (in months) which we vary when evaluating the prediction quality across lags. All results are given in Table 7. Based on the given significance levels of β, we find strong evidence that news can be used to predict market regimes. On top of that, a time lag of $\delta = 4$ months achieves, in terms of significance, pseudo-R^2 and AIC, the best quality. Overall, we conclude that one will be able to predict market regimes from historic news sentiment.

5. Conclusion and outlook

Speculative bubbles denote the process when stock prices first inflate way above their fundamental value until they collapse. As recent history has shown, this process can repeat multiple times bringing harm to the overall economy [13]. Bubbles are closely related to speculation since investors act on the basis of their future price expectations. Assuming overconfidence in the market and in the predicted future prices, this group of investors, subsequently, may dominate the market making their predictions – at least for some time – self-fulfilling. Such a self-fulfilling process is limited

Table 7. Logit regression for predicting market regimes across different time lags.

Lag [months]	α	β	Pseudo-R^2	AIC
$\delta = 0$	0.1993***	0.0237	0.021	87.38
	(6.76)	(1.14)		
$\delta = 1$	2.0166***	−0.0354	0.032	86.48
	(6.64)	(−1.34)		
$\delta = 2$	2.2117***	−0.0739*	0.119	80.90
	(6.20)	(−2.47)		
$\delta = 3$	2.3521***	−0.0943**	0.175	77.05
	(5.96)	(−2.92)		
$\delta = 4$	2.7328***	−0.1408***	0.307	67.98
	(5.59)	(−3.64)		
$\delta = 5$	2.22031***	−0.0812**	0.142	78.71
	(6.04)	(−2.66)		

Stated: coef. and z-stat.; Obs. 108 Signif.: *** 0.001, ** 0.01, * 0.05

up to the point until the fundamental values reveal that initial expectations were too optimistic and, afterwards, the bubble bursts [1], [9].

The early detection of price bubbles and, even more important, the moment when they bust is important for modern economies to prevent damage. In this paper, we develop a method that distinguishes any market period into *bullish* and *bearish* phases depending on price movements. In our analysis, we find out that news processing is fundamentally different in *bullish* and *bearish* markets. We use news sentiment as a proxy for the outcome of news processing. Accordingly, news sentiment has a greater impact on the stock market prices, once the economy resides in a bust phase. In addition, we use outlier news announcements – in terms of news sentiment – to find that those extreme messages are more common in *bullish* markets. Intuitively, we expect the news sentiment of messages that were submitted during a boom phase to have the same (positive) news sentiment value. Outliers are different in terms of their sentiment, contradicting the overall market sentiment trend. Lastly, we propose a method how news sentiment can be applied to predict changes in the market regimes.

The work presented in this paper opens several avenues for future research. First, we have looked at news announcements from Reuters only. As all issued announcements are novel, this news corpus seems relevant. Including additional sources such as newspapers, social media or fundamental variables would be of interest. This might be an intriguing way to improve the prediction accuracy of *bullish* and *bearish* market regimes further. Second, further effort is needed to validate our approach in terms of robustness and, thus, we plan to extend our analysis also to other commodities such as gold and wheat.

References

[1] R. J. Shiller, *Irrational Exuberance*, 2nd ed. Princeton University Press, 2005.

[2] G. A. Akerlof and R. J. Shiller, *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton: Princeton University Press, 2009.

[3] P. C. Tetlock, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.

[4] P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More than words: Quantifying Language to Measure Firms' Fundamentals," *The Journal of Finance*, vol. 63, no. 3, pp. 1437–1467, 2008.

[5] E. Henry, "Are Investors Influenced By How Earnings Press Releases Are Written?" *Journal of Business Communication*, vol. 45, no. 4, pp. 363–407, 2008.

[6] T. Loughran and B. McDonald, "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.

[7] D. García, "Sentiment during Recessions," *The Journal of Finance*, vol. 68, no. 3, pp. 1267–1300, 2013.

[8] CBS, "Oil Trading's Powerful "Dark Markets"," 2011. [Online]. Available: http://www.cbsnews.com/2100-18564_162-4188620.html

[9] J. M. Keynes, *The General Theory of Employment, Interest and Money*, London, 1936.

[10] U. Homm and J. Breitung, "Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods," *Journal of Financial Econometrics*, vol. 10, no. 1, pp. 198–231, 2011.

[11] B. T. Diba and H. I. Grossman, "Rational Bubbles in the Price of Gold," 1984.

[12] R. J. Shiller, "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" in *Advances in behavioral finance*. Yale U: New York, 1993, pp. 107–132.

[13] G. W. Evans, "Pitfalls in testing for explosive bubbles in asset prices," *The American Economic Review*, vol. 81, no. 4, pp. 922–930, 1991.

[14] S. van Norden and R. Vigfusson, "Avoiding the Pitfalls: Can Regime-Switching Tests Reliably Detect Bubbles?" *Studies in Nonlinear Dynamics & Econometrics*, vol. 3, no. 1, 1998.

[15] W. Antweiler and M. Z. Frank, "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *The Journal of Finance*, vol. 59, no. 3, pp. 1259–1294, 2004.

[16] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[17] M.-A. Mittermayer and G. F. Knolmayer, "Text Mining Systems for Market Response to News: A Survey," Bern and Switzerland, 2006.

[18] M. Minev, C. Schommer, and T. Grammatikos, "News and stock markets: A survey on abnormal returns and prediction models," Luxembourg, 2012.

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

[20] N. Jegadeesh and A. Di Wu, "Word Power: A New Approach for Content Analysis: AFA 2012 Chicago Meetings Paper," *SSRN Electronic Journal*, 2011.

[21] F. Li, "The Information Content of Forward-Looking Statements in Corporate Filings: A Naïve Bayesian Machine Learning Approach," *Journal of Accounting Research*, vol. 48, no. 5, pp. 1049–1102, 2010.

[22] M.-a. Mittermayer and G. Knolmayer, "NewsCATS: A News Categorization and Trading System," in *Sixth International Conference on Data Mining (ICDM'06)*, 2006, pp. 1002–1007.

[23] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news," *ACM Transactions on Information Systems*, vol. 27, no. 2, pp. 1–19, 2009.

[24] A. Sharma and S. Dey, "A comparative study of feature selection and machine learning techniques for sentiment analysis," in *2012 Research in Applied Computation Symposium (RACS 2012)*, 2012, pp. 1–7.

[25] S. Feuerriegel and D. Neumann, "News or Noise? How News Drives Commodity Prices," in *International Conference on Information Systems (ICIS 2013)*, 2013.

[26] F. E. Fama, "The Behavior of Stock-Market Prices," *The Journal of Business*, vol. 38, no. 1, pp. 34–105, 1965.

[27] N. Granados, A. Gupta, and R. J. Kauffman, "Research Commentary–Information Transparency in Business-to-Consumer Markets: Concepts, Framework, and Research Agenda," *Information Systems Research*, vol. 21, no. 2, pp. 207–226, 2010.

[28] F. Wex, N. Widder, M. Liebmann, and D. Neumann, "Early Warning of Impending Oil Crises Using the Predictive Power of Online News Stories," in *46th Hawaii International Conference on System Sciences (HICSS)*, 2013, pp. 1512–1521.

[29] L. Yu, S. Wang, and K. K. Lai, "A Rough-Set-Refined Text Mining Approach for Crude Oil Market Tendency Forecasting," *International Journal of Knowledge and Systems Sciences*, vol. 2, no. 1, pp. 33–46, 2005.

[30] A. F. Rad, "Assessing the impact of news on oil prices: A text mining approach," Master's Thesis, Luleå University of Technology, Luleå and Sweden, 2009.

[31] F. Lechthaler and L. Leinert, "Moody Oil - What is Driving the Crude Oil Price?" Zürich and Switzerland, 2012.

[32] M. Dadvar, C. Hauff, and F. de Jong, "Scope of Negation Detection in Sentiment Analysis," in *Dutch-Belgian Information Retrieval Workshop (DIR 2011)*, 2011, pp. 16–20.

[33] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li, "RCV1: A New Benchmark Collection for Text Categorization Research," *Journal of Machine Learning Research*, vol. 5, pp. 361–397, 2004.

[34] C. D. Manning and H. Schütze, *Foundations of statistical natural language processing*. Cambridge and Mass: MIT Press, 1999.

[35] Y. Konchitchki and D. E. O'Leary, "Event study methodologies in information systems research," *International Journal of Accounting Information Systems*, vol. 12, no. 2, pp. 99–115, 2011.

[36] A. C. MacKinlay, "Event Studies in Economics and Finance," *Journal of Economic Literature*, vol. 35, no. 1, pp. 13–39, 1997.

[37] S. M. Goldfeld and R. E. Quandt, "A Markov model for switching regressions," *Journal of Econometrics*, vol. 1, no. 1, pp. 3–15, 1973.

[38] J. D. Hamilton, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, vol. 57, no. 2, pp. 357–384, 1989.

[39] S. G. Hall, Z. Psaradakis, and M. Sola, "Detecting periodically collapsing bubbles: a Markov-switching unit root test," *Journal of Applied Econometrics*, vol. 14, no. 2, pp. 143–154, 1999.

[40] R. Demirer and A. M. Kutan, "The behavior of crude oil spot and futures prices around OPEC and SPR announcements: An event study perspective," *Energy Economics*, vol. 32, no. 6, pp. 1467–1476, 2010.

[41] S. Kozicki, E. Santor, and L. Suchanek, "Large Scale Asset Purchases: Impact on Commodity Prices and International Spillover Effects," in *RCEF 2012: Cities, Open Economies, & Public Policy*, 2012.

[42] F. Wirl and A. Kujundzic, "The Impact of OPEC Conference Outcomes on World Oil Prices 1984-2001," *The Energy Journal*, vol. 25, no. 1, pp. 45–62, 2004.