NOISE TRADER BEHAVIOR – A DISAGGREGATED APPROACH TO UNDERSTANDING NEWS RECEPTION IN FINANCIAL MARKETS

Research paper

Hannemann, Felix, University of Freiburg, Germany, felixhannemann1@gmx.de
Pröllochs, Nicolas, University of Freiburg, Germany, nicolas.proellochs@is.uni-freiburg.de
Alfano, Simon, University of Freiburg, Germany, simon.alfano@gmx.de
Neumann, Dirk, University of Freiburg, Germany, dirk.neumann@is.uni-freiburg.de

Abstract

Financial disclosures serve as primary intermediaries between companies and investors. However, investors have different information processing skills and might easily be misled by noisy signals that lack a deeper meaning. In financial markets, this is formalized by the noise trader theory, which groups investors into two categories: (a) informed investors assumed to form rational decisions and (b) noise traders forming beliefs partly based on non-fundamental noise signals and news sentiment. Yet, little is known about how these groups actually interpret textual information in financial statements and how the resulting stock market reaction differs. This work extends previous research by unraveling the role of word choice and semantic orientation in financial disclosures for both investor types. For this purpose, we use Kalman filtering to decompose the stock market reaction following the publication of U.S. regulated Form 8-K filings into a fundamental price component and a noise residual. We then use LASSO regression to identify the statistically relevant words for informed investors and noise traders. According to our results, each investor type assigns significantly different interpretations and degrees of importance to individual words and documents.

Keywords: Noise trader theory, Information processing, Decision-making, Financial markets, News sentiment, Kalman filter
1 Introduction

Information dictates the decision-making of individuals (Vodanovich, Sundaram and Myers, 2010). However, the interpretation of information is often subjective and depends on each individual’s rational capabilities and information processing skills (Wilson, 2000; Schmeling, 2006; Toplak, West and Stanovich, 2014). Consequently, the decision-making and resulting observable actions based on the same information set can vary substantially across different audiences (Riley and Luippold, 2015). Questions of how information impacts human behavior are thus at the heart of Information Systems (IS) research (Browne and Parsons, 2012; Kearney and Liu, 2014).

Financial stock markets offer a particularly interesting and relevant case to study the role of different information processing skills in relation to human decision-making. As an important advantage, financial markets are characterized by a huge amount of relevant information that is publicly available due to regulation (Carter and Soo, 1999; Drake, Thornock and Roulstone, 2012). For instance, in the U.S., the Securities and Exchange Commission (SEC) mandates disclosure at certain crucial points so that investors can make an informed decision before exercising ownership in stocks (Benston, 1973). As such, regulated financial news disclosures, which ensure that all market participants are provided with equal information, serve as a relevant source of information for investors before engaging in financial trading.

Interestingly, existing research already acknowledges that not all investors in financial markets have access to the same information or the ability to process it equally (Black, 1986; De Long et al., 1990a; Shleifer and Summers, 1990). Specifically, according to the noise trader theory, not all investors are “fully rational and their demand for risky assets is affected by their beliefs or sentiments that are not fully justified by fundamental news” (Shleifer and Summers, 1990, p.19). According to the theory, investors are split into two groups based on their rational abilities – informed investors and noise traders (Shleifer and Summers, 1990). While informed investors are assumed to form rational decisions, noise traders are investors who trade based on noisy signals that are not necessarily justified by fundamentals (Black, 1986). However, despite a huge amount of research on noise trading and its intricacies (De Long et al., 1989; Antweiler and Frank, 2004; Schmeling, 2006), little is yet known about how noise traders actually interpret information and how their reaction to financial disclosures differs from that of informed investors. In particular, it is unclear how the textual elements and semantic orientation in financial disclosures affect the decision-making of these distinct investor types.

In order to better understand the reception of language in financial markets, we use a dataset of U.S. regulated Form 8-K filings to investigate how informed investors and noise traders are influenced by specific textual elements and the semantic orientation in financial disclosures. These mandatory disclosures inform investors about important corporate events, such as management changes, the departure of directors, bankruptcy, layoffs, and other unspecified events deemed material to investors. We thereby aim to find language patterns that allow us to draw inferences regarding individual investor behavior. For this purpose, we use Kalman filtering to decompose stock prices into their fundamental component and a noise residual. We then extract the most relevant terms for each investor type using Bayesian learning in the form of the least absolute shrinkage and selection operator (LASSO). Finally, we analyze word relevance and interpretation, semantic dimensions and the relation of fact-based and emotion-induced sentiment to stock market returns. To the best of our knowledge, this is the first study that statistically extracts driving words to separately analyze and compare the reception of financial disclosures by informed investors and noise traders.

As a main contribution, this paper studies how different information processing skills affect the reception of language in financial disclosures. Specifically, our research addresses two main questions: Which textual elements and semantic orientations are relevant for the news reception of informed investors and noise traders? How do differences in individual word relevance and overall document perception relate to the decision making of the distinct investor groups on the stock market? We thereby shed light on the influence of semantic orientation and show that informed investors and noise trad-
ers assign significantly different interpretations and degrees of importance to individual words and documents. While the differing influence of fact-based and emotion-laden sentiment on financial decision-making is clear-cut for informed traders, we find that the effects are interdependent in the case of noise traders. The resulting insights can help researchers in various domains to better understand how individuals read formal written text and how its reception depends on the information processing skills of the audience. In addition, companies and financial professionals can benefit from our findings when composing financial disclosures or attempting to reduce noise trader risk exposure.

The remainder of this paper is structured as follows. Section 2 establishes the theoretical background of our study and presents related research on news reception in financial markets. Section 3 introduces our dataset and presents our research methodology, which uses Kalman filtering and LASSO regularization to statistically extract textual information that is relevant to informed investors and noise traders. Subsequently, Section 4 presents our empirical results. Section 5 then discusses our results and provides managerial implications. Section 6 concludes and provides an outlook on future work.

2 Theoretical Background

2.1 Noise trader theory

Theories from behavioral economics to explain human decision-making are well-established in the IS and Finance literature (Thaler, 2005; Browne and Parsons, 2012). In this context, a particularly interesting concept is the noise trader theory. This theory has been advanced by Black (1986), De Long et al. (1990a) and Shleifer and Summers (1990) and seeks to explain investors' irrationality in trading on financial stock markets. Acknowledging that not all investors possess equal information processing skills, the noise trader theory distinguishes two types of investors: (1) informed investors, who are assumed to trade based on fundamental information and (b) noise traders who form their decisions based on noisy signals or sentiment that they believe to convey relevant information (Black, 1986). Noise trading based on news sentiment can be traced to the human tendency to overweight new information when drawing inferences (Tversky and Kahneman, 1981). In addition, the linguistic style of documents provides a potential means for deception (Fuller et al., 2013).

The changes in stock demand induced by noise traders tend to be correlated and non-random, as judgment biases affecting individual decision-making tend to follow similar patterns (Shleifer and Summers, 1990). Yet, under the noise trader theory, arbitrage (i.e., informed traders benefiting from the fact that noise traders trade without fundamental information) is also limited (Shleifer and Vishny, 1997). The noise traders’ sentiment or beliefs are stochastic and, in part, unpredictable for the informed investors, who are generally risk-averse and bound with a reasonably short time horizon (Shleifer and Summers, 1990). Hence, making a heavy bet against noise trading is risky in the sense that noise traders’ beliefs and their impact on prices might not revert, even in long timeframes (De Long et al., 1990a). Therefore, arbitrage cannot fully eliminate the noise effects on prices through trading. This leads De Long et al. (1990) to conclude that noise itself creates a price-risk. In fact, it may pay for informed investors not to counter, but rather follow noise trader investments directionally. Specifically, informed investors can then clear their arbitrage positions before noise trader beliefs tilt around and prices revert to their fundamental values. This phenomenon is described by the bandwagon effect in an extension to the noise trader theory (De Long et al., 1990b; Shleifer and Summers, 1990). Interestingly, noise trader research suggests that informed investors might earn an average overnight return of 1.16% by mimicking noise trader investments in the case of attention-grabbing stocks (Seasholes and Wu, 2007).

2.2 News reception and news sentiment in financial markets

Advances in the field of information retrieval have enabled researchers in the IS and Finance disciplines to analyze how investors perceive textual information. News sentiment is a common measure to analyze information processing and news reception in financial markets (Alfano, Feuerriegel and Neumann, 2015). In this context, sentiment analysis allows one to operationalize the overall tone of a
document based on the occurrence of opinionated words (Loughran and McDonald, 2016). While several alternative approaches to measuring news sentiment exist (Antweiler and Frank, 2004; Li, 2010; Liu, 2012; Jegadeesh and Wu, 2013), sentiment analysis of financial disclosures typically relies on so-called dictionaries. These dictionaries collect relevant words that are associated with a distinctly positive, negative or emotionally-charged interpretation that is typically independent of the textual context. For the English language, a range of manually compiled dictionaries exists for general (Harvard IV Psychological Dictionary) and finance-oriented purposes (Henry, 2008; Loughran and McDonald, 2011). However, these dictionaries come with several drawbacks, as they are typically manually selected and also implicitly assume the equal importance of all words. Moreover, they are frequently biased due to subjective interpretation and are thus incapable of adequately reflecting the perception of a specific audience. As a remedy, Bayesian variable selection and regularization methods allow for the generation of domain-specific dictionaries that yield superior results in terms of in-sample and out-of-sample explanatory power (Pröllochs, Feuerriegel and Neumann, 2015). Related research also finds that news reception depends on humans’ rational abilities (Wilson, 2000; Loibl and Hira, 2009).

News sentiment is highly relevant for investors in financial markets (Barberis, Shleifer and Vishny, 1998; Baker and Wurgler, 2007), and its role in financial news disclosures is frequently studied in previous work (see Kearney and Liu, 2014 for a comprehensive overview). High levels of journalistic pessimism are found to be linked to lower subsequent stock returns and higher subsequent stock market volatility. Interestingly, this can be explained by the sentiment itself rather than new fundamental information (Tetlock, 2007). Related literature also finds that firms with a positive tone in quarterly earnings press conference calls experience significantly higher stock returns (Price et al., 2012). News sentiment in crude oil news is positively linked to crude oil returns for both informed and uninformed traders (Alfano, Feuerriegel and Neumann, 2015). Subjects expect a higher future return from a given firm when reading an article skewed towards positive language (Bosman, Kräussl and Mirgorodskaya, 2015). In addition, sentiment scores based on financial disclosures can serve as an input to build trading strategies with superior performance (Schumaker et al., 2012).

3 Methodology and Dataset

We use a two-stage approach to study the news reception of informed investors and noise traders. First, we present our approach to decompose stock prices into their fundamental component and a noise residual using the Kalman filter. In concordance with related literature (Alfano, Feuerriegel and Neumann, 2015; Hendershott and Menkveld, 2014; Brown and Cliff, 2004), we then assign the fundamental component of prices to informed investors and the noise residual to the noise trader group. Subsequently, we show how LASSO regression can be used to statistically identify relevant words for the two distinct investor groups. Finally, we introduce our dataset that allows us to examine differences between noise traders and informed investors regarding the reception of textual information.

3.1 Price decomposition using Kalman filtering

In order to measure the stock market behavior of informed investors and noise traders, we first dissect stock prices into their fundamental component and the noise residual. For this purpose, we use the Kalman filter, which is a frequently employed tool to decompose asset prices (Schwartz and Smith, 2000; Brown and Cliff, 2004; Haleh, Moghaddam and Ebrahimijam, 2011; Brogaard, Hendershott and Riordan, 2014).

The Kalman filter is a recursive approach that uses a state-space model to solve linear filtering problems for discrete data series that involve uncertainty. We briefly outline its functionality in the following. In a first step, the Kalman filter uses a ‘predict’ equation to estimate the $t+1$ state of stock prices ($a priori$ estimates) using the price information at time-step $t$. This includes all observations up to time $t = 0, ..., t$ and error covariance estimates. Second, it uses an ‘update’ equation to obtain feedback based on the measured price in $t+1$ (including noise) and to correct its previous $t+1$ price estimate ($a posteriori$ estimates). The estimation of the $t+2$ price is then based on this improved $t+1$ estimate. The Kalman filter thus effectively incorporates all past available information to yield an optimal projection.
of future prices. In addition, it corrects its own estimates based on realized deviations of true prices from estimated prices. More detailed explanations and mathematical specifications are provided in Tussel (2011).

In accordance with the related literature, we assume that informed investors trade based on fundamental information (e.g., historic price information or public news) and that this information is reflected in past stock prices (Thaler, 2005). Consequently, we explain information-based trading with the \textit{a priori} estimates that are used in the Kalman filter to estimate the stock price at time step $t+1$. In the case of noise traders, we assume that the trading behavior is based on noisy signals that these believe to be relevant information. In this context, noise risk is predictable neither on the basis of the \textit{a priori} estimates for prices nor the \textit{a posteriori} updating. The noise residual is thus given by the difference between Kalman estimates for $t+1$ and the observed realizations of the stock price in $t+1$.

Accordingly, we use the Kalman filter to decompose our discrete dataset of stock prices for each company in our sample. As a result, we obtain two disaggregated time-series for the fundamental stock price component and the noise residual. We then assign the fundamental component of prices to the informed investors group and the noise residual to the noise trader group (cf. Alfano, Feuerriegel and Neumann, 2015). It is worth noting that this categorization is also concordant with Brown and Cliff (2004) who separate investors into institutional (informed) investors and amateur (noise) traders using the Kalman filter.

We next aim to statistically identify those words that are relevant for informed investors and noise traders. For this purpose, we rely on a widely accepted regularization approach (Pröllochs, Feuerriegel and Neumann, 2015) that allows us to identify information drivers in textual materials based on a response variable. This approach utilizes the so-called least absolute shrinkage and selection operator (LASSO), which entails several benefits for this task (Tibshirani, 1996; Hastie, Tibshirani and Friedman, 2009; Pröllochs, Feuerriegel and Neumann, 2015). First, this method features a variable selection property that filters out non-informative noise terms and thus allows one to select only decisive variables in a regression model. Furthermore, the LASSO mitigates the issue of multicollinearity as present when estimating via ordinary least squares. Finally, the LASSO through its feature selection character solves the problem of over-fitting by finding a reasonable trade-off between bias and variance, which occurs if the model complexity is too high. This results in parsimonious and interpretable models that are particularly suited to the extraction of the words most relevant to investors.

For each of the models $j = 1,\ldots,P$ regressors, the LASSO imposes an $l_1$-norm penalty term of form $\lambda \sum_{j=1}^{P} |\beta_j|$ as extension to the OLS estimator (Tibshirani, 1996; Hastie, Tibshirani and Friedman, 2009). Here, $\lambda \geq 0$ can be interpreted as a shrinking parameter with increasing $\lambda$ implying higher shrinkage of the model coefficients. Because of the $l_1$-norm penalty term, the LASSO typically produces estimates in which some of the coefficients are set exactly to zero. This gives the LASSO the desired property of variable selection (Tibshirani, 1996; Hastie, Tibshirani and Friedman, 2009). Specifically, we are willing to give up a certain number of regressors to arrive at a parsimonious model and keep only those variables that are relevant. We then choose the shrinkage parameter $\lambda_{\text{BEST}}$ that minimizes the in-sample error using 10-fold cross-validation. Afterwards, we re-fit our model for all observations using $\lambda_{\text{BEST}}$ to determine its coefficients. The magnitude of estimated $\hat{\beta}_{\text{LASSO}}$ coefficients indicates relative importance of variables. Our standard errors stem from the Post-LASSO, and allow for statistical tests that correspond to the use of specific words (Belloni and Chernozhukov, 2013).

We extract words that are statistically relevant for informed investors and noise traders from financial disclosures as follows. We treat each document in our dataset as an observation, while we use each word in the corpus as an explanatory variable to interpret the Kalman filtered price fractions associated to the two investor types. In the case of informed investors, the dependent variable is given by the daily return of the fundamental price component, while the noise residual reflects the stock market reaction of noise traders. This approach results in two distinct lists containing the words that are statistically relevant for the two investor types. The magnitude of the coefficient estimates serves as meas-
3.2 Dataset and descriptive statistics

Our dataset consists of U.S. Form 8-K filings entailing several advantages compared to alternative sources: first, the information in 8-K filings can be regarded as novel, since regulations require firms to file these financial disclosures within four business days of the specified events (Carter and Soo, 1999). Second, all reports are publicly available and quality-checked by the Securities and Exchange Commission to ensure that the content meets formal requirements. Third, Form 8-K filings are a frequently employed dataset in the related literature and their relevance and informativeness on the stock market has been investigated in various publications (Carter and Soo, 1999; Pröllochs, Feuerriegel and Neumann, 2016).

We have gathered all Form 8-K filings, including amendments, spanning the years 2004 to 2013 that are available from the EDGAR website (www.sec.gov/edgar). The collected sample consists of 901,133 filings, which then undergo several filtering steps as follows: first of all, we select only filings from firms that are listed on the New York Stock Exchange (NYSE). In order to be able to gain information about the stock market reactions, we remove filings for which we are not able to match the CIK number to the Thomson Reuters Datastream. Consistent with Loughran and McDonald (2011), we exclude filings that contain fewer than 200 words. To account for extreme stock price effects, we also remove penny stocks with a price below $5 (Zhang, Swanson and Prombutr, 2012). These filtering steps result in a final corpus of 73,986 filings, i.e. 8.21% of the original sample.

Concordant with the related literature, all text data in our corpus undergoes a five-step preprocessing procedure. (1) We clean our corpus by removing irrelevant information such as HTML formatting, and contact addresses. (2) We remove stop words, such as ‘the’, ‘is’, ‘of’ or ‘in’, to omit frequently used words that lack a deeper meaning (Manning and Schütze, 1999). For this purpose, we use a list of 174 stop words (Feinerer, Hornik and Meyer, 2008). (3) We reduce individual words to their stem (cf. Manning and Schütze, 1999) using the Porter stemming algorithm (Porter, 1980). (4) We store frequencies of individual terms per document in a document term matrix that allows for further calculations. In this context, we also remove all sparse terms that occur in less than 5% of all documents (Loughran and McDonald, 2011). (5) We weigh individual terms in accordance with the term frequency-inverse document frequency approach to better reflect the relative importance of words in the corpus (Salton, Fox and Wu, 1983).

Our corpus covers a total of 729 different companies, while the median number of filings per firm is 85.73 (with a standard deviation of 66.73). The total range goes from a minimum of 1 to a maximum of 501 for a single firm. The median length of individual filings amounts to 1001 words. The stock market returns are fairly normally distributed with a mean of 0.0010 and a standard deviation of 0.0420. The mean noise residual of prices is 0.0199 with a standard deviation 0.0587. The mean return on the fundamental price component is 0.0031 with a standard deviation of 0.0023.

4 Empirical Results

4.1 Word reception in financial disclosures

We use the methodology described in the previous section to identify statistically relevant terms for informed investors and noise traders. The output is two weighted word lists that contain words with non-zero positive and negative LASSO coefficients. Table 1 reports the top 15 words with the highest and lowest standardized coefficient for both investor types. In keeping with our pre-processing, the table lists stems instead of complete words. We additionally calculate standard errors via the Post-LASSO (Belloni and Chernozhukov, 2013).

According to Table 1, all of the top 15 positive and negative word stems are strongly statistically significant based on Post-LASSO standard errors. Interestingly, the word lists for both investors uniform-
ly share high, yet imperfect, concordance with the Loughran and McDonald (2011) master dictionary. This highlights their relevance in the finance context (Loughran and McDonald, 2016) and, from a linguistic perspective, suggests that both investor groups react to words from the same financial language family. We observe that the positive word list for informed investor comprises stems concerning obvious positivity (strong, pleas, improv), active management (govern, instruct, split), causal relation (attribut, use, thereunto), and several external factors (oil, secretari, govern). Interestingly, one of the most important stems for informed investors is related to cash flow (flow), which has for long been one of the most relevant information drivers regarding company valuation (e.g., Fisher, 1930). We also see that the informed investors’ most relevant negative words comprise stems related to problems and decline (challeng, declin, difficult), announcements (messag) and, once again, external factors (environ, market). Moreover, we note that informed investors assign high importance to threats to the overall business (folio, enterpris).

### Table 1

<table>
<thead>
<tr>
<th>A. Informed investors relevant words</th>
<th>B. Noise traders relevant words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word stems</td>
<td>LASSO coefficient</td>
</tr>
<tr>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>(top 15 stems)</td>
<td></td>
</tr>
<tr>
<td>attribut</td>
<td>0.6110</td>
</tr>
<tr>
<td>improv</td>
<td>0.6087</td>
</tr>
<tr>
<td>flow</td>
<td>0.0350</td>
</tr>
<tr>
<td>strong</td>
<td>0.0262</td>
</tr>
<tr>
<td>split</td>
<td>0.0242</td>
</tr>
<tr>
<td>increas</td>
<td>0.0242</td>
</tr>
<tr>
<td>unlaw</td>
<td>0.0231</td>
</tr>
<tr>
<td>instruct</td>
<td>0.0219</td>
</tr>
<tr>
<td>secretari</td>
<td>0.0215</td>
</tr>
<tr>
<td>pleas</td>
<td>0.0212</td>
</tr>
<tr>
<td>use</td>
<td>0.0210</td>
</tr>
<tr>
<td>archiv</td>
<td>0.0210</td>
</tr>
<tr>
<td>govern</td>
<td>0.0195</td>
</tr>
<tr>
<td>thereunto</td>
<td>0.0193</td>
</tr>
<tr>
<td>oil</td>
<td>0.0184</td>
</tr>
<tr>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>(top 15 stems)</td>
<td></td>
</tr>
<tr>
<td>folio</td>
<td>-0.0776</td>
</tr>
<tr>
<td>messag</td>
<td>-0.0510</td>
</tr>
<tr>
<td>challeng</td>
<td>-0.0435</td>
</tr>
<tr>
<td>mortgag</td>
<td>-0.0333</td>
</tr>
<tr>
<td>declin</td>
<td>-0.0313</td>
</tr>
<tr>
<td>bill</td>
<td>-0.0306</td>
</tr>
<tr>
<td>environ</td>
<td>-0.0283</td>
</tr>
<tr>
<td>dof</td>
<td>-0.0282</td>
</tr>
<tr>
<td>pattern</td>
<td>-0.0264</td>
</tr>
<tr>
<td>enterpris</td>
<td>-0.0262</td>
</tr>
<tr>
<td>avenu</td>
<td>-0.0257</td>
</tr>
<tr>
<td>billion</td>
<td>-0.0244</td>
</tr>
<tr>
<td>market</td>
<td>-0.0237</td>
</tr>
<tr>
<td>difficult</td>
<td>-0.0227</td>
</tr>
</tbody>
</table>

Total words: 594 (+) and 601 (-)

Significance levels: *** = 1% level, ** = 5% level, * = 10% level

N = 73,986 Form 8-K fillings

Table 1. Table reports the top 15 positive and negative word stems for informed investors (Panel A) and noise traders (Panel B), together with standardized LASSO coefficients. Standard errors are calculated via Post-LASSO (Belloni and Chernozhukov, 2013).
As expected, stems with obvious positive or negative meaning, such as *improv, strong, increas, challeng* or *declin*, can also be found in the respective positive/negative noise trader word list. Aside from these obvious words, however, we observe several more ambiguous stems, such as *civil, rule, mislead*, and *error*. In addition, we find financial year quarter months (*april, octob, august, novemb*) to be relevant to noise traders. The importance of the stem *nonrecur* suggests that noise traders favor one-off effects over continuity. In addition, *advisori* indicates that noise traders assign positive importance to advisory opinion. This suggests a need for reassurance in their decisions. We further note that the negative noise trader word list includes more stems associated with problems or decline as for informed investors (*declin, reduc, challeng, reduc, decreas*). This may indicate a higher degree of pessimism among noise traders, which we will further investigate in Section 4.2.

Next, we compare textual drivers with relevance for both investor types. In total, we find 181 overlapping words in the two lists, which marks 76.7% of the total words relevant to noise traders. Based on these overlapping words informed traders and noise traders frequently interpret words contrarily. We find that 29.4% of all words perceived negatively by informed investors, are interpreted positively by noise traders and vice versa (21.3%). We also conduct a non-parametric Kendall’s Tau test to assess the correlation on ranks between the informed investor and noise trader word lists. We fail to reject the null hypothesis of no correlation between the positive informed and noise trader word lists (Kendall’s tau = 0.1008, p-value = 0.3239). We observe a similar pattern for the negative word lists (Kendall’s tau = 0.1027, p-value = 0.1923). This suggests that the two investor groups are not drawn to entirely different sets of words in general, and yet, the relative importance and interpretation of individual words differs significantly between noise traders and informed investors.

Figure 1 provides further evidence regarding the central finding that informed investors and noise traders react differently to word-level information in financial disclosures. In this figure, we compare the standardized LASSO coefficients across overlapping words of the informed investors and noise trader word lists. Quadrants two and four in Figure 1 show overlapping terms that are interpreted with an opposite polarity. Examples include stems such as *nonrecur* and *advisori* (noise +, informed -) or *consecut* and *answer* (noise -, informed +). Altogether, Figure 1 confirms that noise traders and informed investors share a common set of relevant words, yet they interpret individual words in Form 8-K filings differently, as measured by standardized LASSO coefficient magnitude and sign.

Figure 1. Comparison of standardized LASSO coefficients for 181 overlapping stems across the noise trader and informed investor word lists. The solid line indicates an equal interpretation of words for the two distinct investor types.
4.2 Categorization into semantic dimensions

We now separate words that convey an explicitly positive or negative statement from words whose interpretation can be attributed to emotional orientation. This allows us to investigate semantic dimensions and better understand the drivers of news reception for both investor groups. For this purpose, we count the number of fact-based words and emotion-laden words in the word lists behind Table 1.

In keeping with the related literature, we model the fact-based category using the General Inquirer ‘Positive’ and ‘Negative’ tag categories (Tetlock, Saar-Tsechansky and Macskassy, 2008). The words in this category, e.g., improv or negat, express a clearly positive or negative self-contained sentiment. In addition, we use the eight emotion categories of the NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013) to form a semantic emotional category. These emotion words distinctively express emotional meaning or feelings, such as trust, joy or surprise, that are of great relevance in a financial context e.g., partnership, safe or bankrupt. Finally, we collect those stems that are marked under both categories i.e., with ambiguous meaning in a mixed-meaning category. Corresponding words typically do not express a clear, self-contained sentiment (e.g., opinion or expert).

<table>
<thead>
<tr>
<th>Fact-based words</th>
<th>Emotional words</th>
<th>Mixed-meaning words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td>Positive</td>
</tr>
<tr>
<td><strong>Informed investor word list</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>14.5%</td>
<td>9.5%</td>
</tr>
<tr>
<td>Negative</td>
<td>14.0%</td>
<td>9.0%</td>
</tr>
<tr>
<td>Overall</td>
<td>28.5%</td>
<td>18.5%</td>
</tr>
<tr>
<td><strong>Noise trader word list</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>15.3%</td>
<td>8.1%</td>
</tr>
<tr>
<td>Negative</td>
<td>11.9%</td>
<td>5.1%</td>
</tr>
<tr>
<td>Overall</td>
<td>27.1%</td>
<td>13.1%</td>
</tr>
</tbody>
</table>

Table 2. Relative overlap between informed investor and noise trader word lists according to semantic dimensions based on fact-based words (501 words), emotional words (552 words) and mixed-meaning words (274 words).

Table 2 shows that the relevant words for both investor groups exhibit similar semantic characteristics based on the number of fact-based and emotional words. For both investor groups, the relevant words for information processing in financial disclosures consist of around 28% fact-based words and more than 30% emotion-laden words. This confirms that facts, as well as emotions, are relevant drivers of news reception in financial markets (Lucey and Dowling, 2005). To further investigate the previously identified trend of more negative words being relevant for noise traders, we also split out the fact-word categories ‘positive’ and ‘negative’. Our findings are two-fold. On the one hand, we observe that noise traders assign higher relative importance to stems that express a negative outlook. This might reflect a greater degree of loss aversion (Tversky and Kahneman, 1981) and pessimism. On the other hand, we see that positive words are less relevant for noise traders than for informed investors.

4.3 Document-level reception of financial disclosures

We now analyze how the identified differences regarding the relevance of individual words impact the reception of Form 8-Ks on a document level between noise traders and informed investors. For this purpose, we weigh the number of occurrences of each word in a publication according to the estimated coefficients from the LASSO regression method. This allows us to calculate weighted informed investor and noise trader sentiment scores for each Form 8-K filing in our dataset. According to our results, the Form 8-K sentiment scores differ significantly between informed investors and noise traders. The mean absolute difference in sentiment is 0.2062, which is statistically different from zero at the 1% significance level based on a two-sided t-test (t-value of 322.45). We also note that 17.70% of all documents are assigned to sentiment scores with opposite signs. Altogether, the differences in observed investor document sentiment scores corroborate the notion that noise traders not only assign signifi-
Disaggregated News Reception in Financial Markets

cantly different importance to individual words in financial disclosures, but also that this results in a different reception of entire documents.

Next, we examine the relationship between the document news reception of the two investor types and stock market returns. For this purpose, we perform an ordinary least squares (OLS) regression using abnormal log-returns as the dependent variable, while we use independent variables for weighted document sentiment scores using each combination of the semantic categories and investor types from the previous section. Here, we employ abnormal returns instead of normal stock market returns, as they correct for confounding influences and the general market movement following each firm’s Form 8-K disclosure (cf. MacKinlay, 1997). Our regression design is given by

\[
AR_{ilog} = \beta_0 + \beta_1IF_i + \beta_2IE_i + \beta_3IMM_i + \beta_4IF \times IE_i + \beta_5IF \times IMM_i + \beta_6IE \times IMM_i
\]

\[
+ \beta_7NF_i + \beta_8NE_i + \beta_9NMM_i + \beta_{10}NF \times NE_i + \beta_{11}NF \times NMM_i + \beta_{12}NE \times NMM_i
\]

\[
+ \beta_{13}CAR_i + \beta_{14}NYSE_i + \beta_{15}Alpha_i + \beta_{16}MTBV_i + \beta_{17}MV_i
\]

\[
+ \beta_{18}YearD_{Year} + \beta_{19}SectorD_{Sector} + \epsilon_i
\]

with model coefficients \(\beta_i\) and Gaussian error \(\epsilon_i\). The dependent variable \(AR_{ilog}\) denotes the stock’s daily abnormal log-return following the Form 8-K filing. \(IF_i, IE_i, IMM_i, NF_i, NE_i, NMM_i\) denote the sentiment scores for each filing \(i\) that were derived using the term frequencies in document \(i\), the word overlap with the fact-based \(F\), emotional \(E\) and mixed-meaning categories \(MM\), multiplied by the LASSO coefficients of the generated noise trader \(N\) and informed investor \(I\) word lists. To account for potential relations between the semantic categories, we include interaction terms for all sentiment variables.

In addition, we include a set of control variables to control for external effects. First, we include the cumulative abnormal returns \(\text{CAR}_i\) over a window of 15 trading days before the Form 8-K publication in order to account for potential insider trading and information leakage before the Form 8-K filing day (MacKinlay, 1997). We further account for the overall market development by approximating the returns on the NYSE Composite Index \(\text{NYSE}_i\) and we include the market-model \(\text{Alpha}_i\) to account for the stock’s usual over- or underperformance (Jeng, Metrick and Zeckhauser, 2003). To account for potential relations between firm size and stock price reaction to the Form 8-K disclosure, we include each firm’s market capitalization \(\text{MV}_i\). Taking a cue from Fama and French (1993), we include each firm’s market-to-book ratio \(\text{MTBV}_i\) and account for potential life-cycle effects or future prospects of the firm. We ultimately include dummy variables for each year \(D_{Year}\) and industry sectors \(D_{Sector}\).

The regression results are provided in Table 3 offering three key findings described below. As expected, the stock’s usual over- or underperformance measure \(\text{Alpha}_i\) is highly statistically significant in explaining abnormal log-returns, as it captures general excess return features of an asset over a benchmark portfolio (Jensen, 1968). Concordant with related literature, we observe a relatively low adjusted R-squared of 0.0476 (Tetlock, Saar-Tsechansky and Macskassy, 2008). As a first finding, we observe that both noise trader fact-induced sentiment \(\text{NF}_i\) and emotions-induced sentiment \(\text{NE}_i\) are positive in sign and statistically significant variables by which to explain a firm’s abnormal log-returns following a Form 8-K. Secondly, we also observe that informed investor fact-induced sentiment \(\text{IF}_i\) and emotion-induced sentiment \(\text{IE}_i\) are highly statistically significant and positive in sign. The fact that news sentiment also affects informed investor decision-making points to the bandwagon effect (De Long et al., 1990b), which we will critically review in Section 5.

Lastly, as a highly interesting finding, we observe that the mixed-meaning sentiment for noise traders \(\text{NMM}_i\) is statistically significant (1% level), as opposed to informed investor mixed-meaning sentiment \(\text{IMM}_i\). This postulates that, while informed investors can distinguish between fact and emotion words in terms of meaning, noise trader information processing is more ambiguous. This is confirmed not only by the high statistical significance of the mixed-meaning sentiment for noise traders, but also by the significance on the interaction term \(\text{NF} \times \text{NE}_i\) (at the 5% significance level), suggesting that noise traders’ fact-induced sentiment varies with the emotion-laden content in Form 8-K filings.
Finally, we perform multiple statistical tests to validate the robustness of our results. First, we calculate variance inflation factors (VIF) for all variables in our models. The VIF of all regressors are below the critical threshold of 4. Second, we check our model for possible heteroskedasticity and autocorrelation. We also find our results to be robust with respect to non-log abnormal returns, quadratic model specifications, as well as for subsets of data across different industry sectors and firm size clusters. Finally, we test for robustness across time and split the observation period into three time horizons. We reflect the global financial crisis from the end of 2008 to the end of 2009, a pre-crisis phase and a post-crisis phase. In the case of noise traders, our results confirm all previous findings during all phases. Interestingly, we observe a slightly different pattern for informed investors. In their case, fact-based and emotional sentiment are not significant in the crisis years while they are statistically significant at the 1% level in pre- and post-crisis years. We suggest two potential explanations for this find-
ing: First, during times of severe market crisis, informed investors become more risk-averse and/or cash-constrained. They may thus refrain from making arbitrage bets against sentiment-induced noise trading. Second, noise itself might not be easily identifiable for informed investors in turbulent market times given higher degrees of price volatility, price co-movement and irrational exuberance. We find these results to be robust with respect to various start- and end months of our crisis time horizon and leave further investigation of these observations to future research.

5 Discussion and Managerial Implications

This study contributes to a better understanding of how investors’ decision-making is impacted by the textual information in financial disclosures. Unlike, for example, Kumar and Lee (2006), who study trading behavior in stock portfolios over months and years, we investigate the investors’ immediate stock market reaction to financial information entering the market. In this context, we also address “the need to better understand the process by which individual investors formulate their trading decisions, including an identification of the information sources they use in decision making” (Kumar and Lee, 2006, p. 2485). Consequently, our research is highly relevant for companies who are obligated to publish regulated financial disclosures, as well as financial professionals who use them as an important information source for their investment decisions. In addition, this study not only sheds light on news reception in financial markets but also serves as an interesting starting point for future research that aims at analyzing decision-making in the behavioral sciences and IS.

Overall, our work provides support for the noise trader theory. We find that informed traders and noise traders assign significantly different importance, interpretation and polarity to individual words and entire documents in financial disclosures. Our results further indicate that noise traders process word-level information more ambiguously as compared to informed traders. Specifically, we observe that the influence of fact-based sentiment and emotion-laden sentiment on financial decision-making is clear-cut for informed traders, while the effects are interdependent in the case of noise traders. Consistent with the theory, this results in mandatory financial disclosures being interpreted differently by informed traders and noise traders. While the noise trader theory does not foresee an internalization of sentiment as being decision-relevant for the informed investor group, a further study of the same authors empirically describes the bandwagon effect as a potential explanation for positive feedback trading strategies (De Long et al., 1990b). Informed investors, when they do not expect noise in-/deflated prices to revert back to fundamental values in the short term, might be better off following noise trader investments directionally and then clearing their arbitrage positions before noise trader beliefs tip around. Our findings provide evidence for the bandwagon effect as an extended explanation of the noise trader theory. Interestingly, we also find that informed investors no longer follow news-sentiment-driven decision-making in times of a market crisis.

Our results entail immediate implications for financial professionals, as well as academic researchers from the fields of IS and beyond. Firms that are obligated to file financial disclosures can use our findings to gain insights into the affective characteristics of investors’ bounded-rational decision-making when interpreting textual information. In a next step, managers and investor relations departments can benefit from a self-reflective writing process that avoids noisy signals in their communications, thus helping to prevent stock prices from deviating from expected values. Our results also indicate that noise traders frequently interpret positive words negatively and vice-versa. Hence, management professionals need to be cautious when framing negative content using positive words. In addition, this finding can help informed traders to reduce exposure to noise trader risk or when trying to identify arbitrage positions. Financial regulators may also use our findings as an input to new directives that limit the use of language that may lead to noisy signals and misleading communication in financial disclosures. Finally, a better understanding of how humans with different information processing skills perceive written information is not only relevant for the financial domain, but for any company that is subject to market dynamics. Since many information sources (e.g., product reviews, recommendations, forum entries, etc.) are presented in textual form, this IS study provides ample scope for further research in various domains. Future researchers can use our methodology, based on the Kalman filter.
and LASSO regularization, to effectively exclude manual pre-selection biases and reflect distinct groups of individuals. This provides a promising avenue to study the decision-making processes of groups with different information processing skills.

This study also entails a number of limitations, which can form the basis for further works as follows. First, our approach ignores linguistic devices such as sarcasm and metaphors. Since humans can be contradictory in their statements, negations can also constitute a major challenge. However, an exact understanding of language is currently intractable with computer programs to a large extent. Practical hurdles are the context-dependent nature of language, where the meaning of words relies largely on their position within a sentence, as well as the domain. For instance, “There is continuously growing competitive pressure on our business” features a different semantic orientation than “there is a great opportunity to continuously grow our business”. A second limitation is that, although our analysis extracts the information drivers that statistically reflect the perception of readers, the proposed approach does not allow one to draw a conclusion about a causal connection. Textual analysis, however, can contribute to our ability to understand the impact of information on stock returns, even if news sentiment does not directly cause returns to behave in one way or another (Loughran and McDonald, 2011). Ultimately, our study merely assesses the reception of documents based on the stock market reaction, while deeper insights into the actual reading process remain blurred. To overcome this limitation, it is an intriguing notion to combine our method with complementary approaches from neuroscience and NeuroIS (Dimoka et al., 2012).

6 Conclusion

Questions of how textual information affects human decision-making loom large in the field of IS research and beyond. However, the interpretation of textual information is subjective, and the resulting decision-making depends on individuals’ information processing skills. This is of particular relevance in financial markets, where the noise trader theory categorizes investors into two groups - namely, informed investors assumed to form rational decision and noise traders, assumed to form decisions based on noisy signals or sentiment that they believe to be information. Yet, little is known about how noise traders actually interpret information and how their reaction to financial disclosures differs from that of informed investors.

As its main contribution, this paper uses a dataset of 73,986 U.S. regulated Form 8-K filings to better understand how the decision-making of informed traders and noise traders is influenced by textual information and semantic orientation in financial disclosures. For this purpose, we use Kalman filtering to decompose stock prices of 729 companies, spanning the years 2004 to 2013, into their fundamental component and noise residual. We then use LASSO regression to identify the statistically relevant words for noise traders and informed investors. According to our results, we find support for the noise trader theory. Our results indicate that each investor type assigns significantly different importance and interpretation to individual words and documents. Specifically, we observe that the influence of fact-based sentiment and emotion-laden sentiment on financial decision-making is clear-cut for informed traders, while the effects are interdependent in the case of noise traders.

Overall, this study not only sheds light on news reception in financial markets but also serves as an interesting starting point for future research and practitioners in various domains. Future IS research can build on our work to gain a better understanding of how the reception of language depends on individuals’ information processing skills. Finance researchers can use the findings from this study to further analyze the interplay between informed investors and noise traders. In this context, future studies may focus on differences regarding the reception of optimism and pessimism or might incorporate additional news sources from other financial domains. Ultimately, financial professionals can use our findings to reduce noise trader risk exposure or to find profitable machine-learning-supported trading strategies, while communications professionals can draw upon our insights when composing financial disclosures.
References


Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning Data Mining
Disaggregated News Reception in Financial Markets


36(4), pp. 992–1011.


