IS HUMAN INFORMATION PROCESSING AFFECTED BY EMOTIONAL CONTENT? UNDERSTANDING THE ROLE OF FACTS AND EMOTIONS IN THE STOCK MARKET

Research-in-Progress

Nicolas Pröllochs
University of Freiburg
Platz der Alten Synagoge
79098 Freiburg, Germany
nicolas.proellochs@is.uni-freiburg.de

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

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

Abstract

The Securities and Exchange Commission (SEC) mandates stock-listed companies in the U. S. to file regulated disclosures that should allow investors to make an informed decision before exercising ownership in stock. We thus hypothesize that investors do not rely solely upon the essential facts but are also impaired by unconscious and idiosyncratic characteristics in their perception. In fact, such affective processing is suggested by behavioral finance and information processing theory, while empirical evidence in large-scale settings remains rare. As a remedy, this paper statistically locates decisive words in financial news that reflect the complete bandwidth of drivers behind investment decisions. According to our results, the decision-making of investors is significantly influenced by emotionally-charged content and non-informative wording.

Keywords: Information processing, decision analytics, behavioral finance, spike and slab regression, text mining, financial news.

Introduction

Various studies in the field of Information Systems have shown that humans do not rely solely upon essential facts when processing given information (Browne and Parsons, 2012). Instead, they can fall victim to affective characteristics in their perception and thus make them vulnerable to information that appeals to superficial emotions or cognition, but which lacks a deeper meaning (Tversky and Kahneman, 1974). Possible effects are of particular interest when studying the information processing in financial markets (Thaler, 2005).

In financial markets, news serves as an important source of information for investors before exercising ownership in stock (Gidófalvi and Elkan, 2001). These compulsory publications must meet specific requirements regarding the content but are not regulated in terms of actual word choice (Carter and Soo, 1999). As a consequence, investors can be distracted by emotionally-charged content and non-informative wording.

In order to better understand human information processing of financial news, one has to study the interplay between individual textual drivers on a word-by-word level (Pröllochs et al., 2015). For this purpose, we utilize Bayesian variable selection to *statistically* locate driving words in corporate disclosures that impact the behavior of investors when making investment decisions on the stock market. Specifically, we draw inferences from SEC-regulated 8-K filings in relationship to stock market returns by measuring the impact of each word on the perception of investors. We then group the extracted words into several semantic categories and separate words that feature an explicit, fact-based statement from words that can be attributed to emotional orientations.

We are able to contribute to the existing research in two respects: first, we utilize an intriguing approach that allows us to study the processing of textual information on a word-by-word level. Instead of replicating human information processing based on predefined word lists, this creates the opportunity to encompass the complete bandwidth of investor perception. To the best of our knowledge, this is the first study that utilizes Bayesian variable selection to explicate the effects of emotionally-charged content on the information processing of financial news. Second, we contribute to Information Systems theory by providing evidence that, in addition to information in the form of essential facts, non-informative wording has a significant effect on human information processing of financial news. This is interesting, since regulated corporate disclosures are intended to serve as a highly objective source of information for investors (Carter and Soo, 1999).

The remainder of this paper is structured as follows. Section 2 introduces related works that also study the information processing of financial news. Following this, Section 3 presents our research methodology, which uses Bayesian variable selection to statistically select stock-market-relevant words. In Section 4, we group the extracted words into predefined semantic categories and assess their individual impact on the stock market. Finally, Section 5 lays out our main findings and provides managerial implications.

Literature Review

In this section, we further discuss the role of financial news as a primary source of information for investors. Based on studies from the related literature, we also detail the role of cognitive biases and show how investors can be distracted by emotionally-charged content and non-informative wording.

Human Information Processing of Financial News

The availability of information forms the basis for all human behavior and decision-making (Vodanovich et al., 2010). This is particularly true in financial markets, where investors continuously seek out new information before making decisions on the stock market (Wilson, 1999). Thus, regulators have made various provisions to ensure that all market participants have access to relevant corporate information. As such, stock-listed firms are obliged to file regulatory disclosures that apply to multiple aspects of business operations. In the U. S., companies are required to file 8-K reports with the Securities and Exchange Commission (SEC) to notify investors of any material events. This covers various circumstances, such as management changes, the departure of directors, bankruptcy and other unspecified events deemed important. All reports are quality-checked by the SEC to ensure that the information meets the requirements (Lee and Lee, 2013). As a result, regulated corporate disclosures serve as a primary source for investors when choosing whether to exercise stock ownership (e. g. Carter and Soo, 1999).

Advances in the field of information retrieval enable researchers in the Information Systems and finance disciplines to analyze how investors perceive this textual information (Nassirtoussi et al., 2014). While different approaches have been proposed (Antweiler and Frank, 2004; Li et al., 2010; Schumaker and Chen, 2009), the tone of financial news is typically measured in prior research using positive and negative word lists. These predefined dictionaries contain opinion words that are typically associated with a distinctly positive or negative fact-based interpretation independent of the corresponding context (Stone, 2002). A frequently utilized method is to measure the tone by calculating the ratio of positive and negative words as normalized by the total number of words in a document (Tetlock et al., 2008). Various studies have shown that the linguistic content of a document is useful in explaining stock market returns. In this context, dictionary-based methods are used to explain stock returns, stock volatility, and firm earnings with regard to the tone of newspapers (e. g. Tetlock, 2007), company press releases (Demers and Vega, 2010; Engelberg, 2008; Henry, 2008), 10-K reports (Feldman et al., 2008; Hanley and Hoberg, 2008; Li, 2008), and social media (Xu and Zhang, 2013). Thus, in our later analysis, we expect words that express a clearly fact-based statement to significantly impact the stock market behavior of investors.

The Role of Cognitive Biases in Financial News

Information processing theory as a model for human thinking and learning builds on the conjecture that humans process any given information, instead of solely responding to external stimuli (Atkinson and Shiffrin, 1968). However, information processing theory also suggests that humans continuously categorize and filter the given information (Schneider and Shiffrin, 1977). This intermediate step becomes particularly relevant when considering the fact that human information behavior is also affected by individual insufficiency regarding information processing skills and skewed by multiple influences that can have cognitive or affective causes (Browne and Parsons, 2012).

Previous research has identified several patterns whereby human decision-making does not conform to pure, rational behavior (Tversky and Kahneman, 1974). Overall, this is justified by the observation that humans can be susceptible to cognitive biases such as overconfidence, overreaction, and other human errors in reasoning and information processing (Friesen and Weller, 2006). As such, a cognitive bias is an anomaly in the perception of every aspect of the environment surrounding the individual (Tversky and Kahneman, 1974). This can be based on a misinterpretation of presented information, fallible memory, or other influences that alter cognitive processes (Phillips et al., 2014). This is particularly relevant for information in the form of financial news, where the nature of language provides arbitrary opportunities to alter the presentation of information (Henry, 2008). This is even reinforced by the fact that financial news must meet specific requirements regarding the content, such as high readability, but is not bound in terms of the actual word choice (Loughran and McDonald, 2014). For instance, Kahneman and Miller (1986) show that different formulations, without distorting or suppressing information, produce opposite responses among subjects. Similarly, the negativity bias (Kanouse, 1984) describes a phenomenon whereby humans put a greater emphasis on negative than positive information. In the context of prospect theory, related works also suggest an asymmetric influence of the tone in news disclosures (Alfano et al., 2015; Hong et al., 2000). Accordingly, DeLong et al. (1990) state that the reaction of investors to semantic content is not necessarily justified by fundamentals, but can be driven by pseudo-signals that some investors believe convey information about future returns.

While individual characteristics of such anomalies can be studied, for instance, in laboratory experiments, human text processing of financial news in non-artificial setups is still a largely unresolved research area (e. g. Thaler, 2005). To overcome this research gap, this study directly locates the information drivers in financial news at the word level. For this purpose, we measure the perception of investors using the real stock market returns of companies. In contrast to experimental studies that typically focus on small observation sets in isolated environments, this allows us to study the effect of wording under non-artificial conditions. As a result, this intriguing approach could give rise to a better understanding of how investors are distracted by emotionally-charged content and non-informative wording. In the next sections, we present our methodology for studying the information processing of financial news on a word-by-word level.

Methodology and Data Sources

This section introduces our research method of studying the human information processing of financial news on a word-by-word level. First, we present the required preprocessing steps and the utilized data sources. Second, we introduce spike and slab regression as a method to extract words that are relevant for investors.

Data Sources

Our dataset consists of Form 8-K filings from the U. S. that inform investors about important corporate events, such as management changes, the departure of directors, bankruptcy, layoffs, and other unspecified events deemed important. This type of documents is a frequent choice in the related literature when it comes to studying information processing of financial news (e. g. Carter and Soo, 1999; Lee and Lee, 2013). For the 8-K filings corpus, we have downloaded all 8-Ks, including amended documents, from the EDGAR website[1] of the years 2004–2013. The complete sample consists of 901,133 filings, which then undergo several filtering steps. First, we select only filings from firms that are listed on the NYSE. Second, in order to gain information about the stock market reaction of investors, we remove filings for which we are not able to match the SEC CIK numbers to Thomson Reuters Datastream. Third, we exclude filings that contain fewer than 200 words (Loughran and McDonald, 2011). These filtering steps result in a final corpus of 76,717 filings.

We measure the impact of the individual documents on stock market returns using an event study methodology (MacKinlay, 1997; Konchitchki and O'Leary, 2011). This is a common way to study the effects of certain events on stock or commodity prices by measuring the impact of a given event without confounding influences. When using such methods, we first have to predict a *normal return* in the absence of such an event. For this purpose, we estimate the normal return according to the so-called *market model*, which assumes a stable linear relation (MacKinlay, 1997) between the market return and the normal return. The market return is modeled using the NYSE Composite Index, along with an event window of 10 trading days prior to the event.[2] Concordant with the related literature, we then calculate the *abnormal return* as the difference between the actual and the normal return on the day of publication (MacKinlay, 1997). The resulting mean change in abnormal returns for all 8-K filings in our sample is 0.7849 %. The abnormal stock market returns are approximately normally distributed with a standard deviation of 4.0969 and median close to zero.

Preprocessing

Before performing the actual spike and slab regression, several operations are involved in a preprocessing phase that transforms the running text into a structured format to allow for further calculations. This includes a myriad of standard routines from the domain of Natural Language Processing. For instance, we use a list of cut-off patterns to extract only the textual components from the documents. We also remove stop words without a deeper meaning and truncate inflected words to their stem (Manning and Schütze, 1999). We then obtain frequencies $f_{t,d}$ of how often term t occurs in the document d. Subsequently, we transform the frequencies $f_{t,d}$ using a common weighting scheme from information retrieval, namely, the term frequency-inverse document frequency or *tf-idf* for short (Salton et al., 1983). The *tf-idf* weighting scheme determines the relative frequency of words in a specific document compared to the inverse proportion of that word over the entire document corpus. For this purpose, the raw frequency $f_{t,d}$ is weighted by the ratio of the total number N of documents divided by the number n_t of documents that contain the term t. This approach is the most common practice in related research, since it allows one to focus on the characteristic terms with a high relevance for a particular document (Aizawa, 2003). For example, words that occur in a smaller share of

1 www.sec.gov.
2 All financial market data covers stocks from the NYSE and is gathered from Thomson Reuters Datastream.

documents are assigned to higher *tf-idf* values than high-frequency contextual stop words, such as the word *"stock"*. Ultimately, the above preprocessing steps result in a so-called *document-term matrix* that serves as the final dataset from which we extract driving words using spike and slab regression.[3]

Extraction of Relevant Words Using Spike and Slab Regression

This paper treats every word from our news corpus as a potential regressor in order to *statistically* extract informative text features from financial news. Therefore, we introduce a linear model which links the occurrences of words to stock market returns and thus can infer a coefficient for each term describing its ex ante effect. This is different from the existing research, which typically relies on *manually-selected* word lists to examine the information processing of investors. In contrast to Jegadeesh and Wu (2013), we do not estimate the linear model using least squares because of multicollinearity issues that are common to this method . Instead, we utilize a Bayesian variable selection method in the form of spike and slab regression, which exhibits several advantages (Varian, 2014). First, this approach overcomes the problem of ex ante selected words, which could potentially lead to the erroneous exclusion of relevant regressors. Second, the regularization property of this method reduces the multicollinearity problem of the classical ordinary least squares technique (Hastie et al., 2009). Third, its variable selection property filters out non-informative noise variables and, therefore, shows how to statistically select words that are relevant to investors.

The *spike and slab* regression extends a standard linear regression model by performing Bayesian-based variable selection (Varian, 2014). Let $\delta \in \{0,1\}^P$ denote which predictors $1,\ldots,P$ to include in the model. Then, the regression is given by

$$y = \mu + X^\delta \beta^\delta + \varepsilon, \qquad \epsilon \sim \mathcal{N}(0,\sigma^2 I), \tag{1}$$

where X^δ and β^δ contain only the non-zero elements of β given by $\delta_j = 1$. Introducing the indicator variable δ_j into the standard linear regression model results in a prior in the form of a mixture of two distributions for each regression coefficient, so-called *spike and slab priors* (Malsiner-Walli and Wagner, 2011). The *spike* component is a distribution with its mass concentrated around zero and the *slab* component is a flat distribution spread over the parameter space. Consequently, variable selection relies on the posterior probability of assigning the corresponding regression effect to the slab component and the selection is thus given by the posterior inclusion probability $p(\delta_j = 1 \mid y)$.

We implement a spike and slab regression for relevant term selection as follows: we treat each 8-K filing of the document-term matrix as an observation, while we use each column, i. e. each word, as an explanatory variables to explain abnormal stock market returns for each filing. We run a Markov-Chain Monte Carlo (MCMC) scheme for $M = 1000$ iterations after a burn-in of 500 draws. Thereafter, we calculate the posterior inclusion probabilities $p(\delta_j = 1)$ and the regression coefficients β_j for each draw. As a result, the magnitude of the posterior inclusion probability $p(\delta_j = 1)$ for each regressor β_j serves as a measure of variable importance.

Preliminary Results

This section introduces our preliminary results. First, we select statistically relevant words in financial news and organize the individual words into exogenously-given semantic categories. Then, we test our hypotheses and provide evidence that investors are distracted by emotionally-charged content.

3 We also tested the raw term frequencies as part of a robustness check. These result into a slightly inferior goodness-of-fit.

Extraction of Relevant Words From Financial News

We use the spike and slab regression method as described in the previous section to extract the statistically relevant terms from the 8-K filings corpus. Accordingly, the output is given by a weighted word list that contains words with positive and negative coefficients. However, ostensibly positive or negative words are not necessarily positive or negative words from a reasonable, fact-based perspective. In order to separate words that convey an explicitly positive or negative statement from words whose interpretation can be attributed to emotional orientation, we proceed by grouping the extracted words according to the semantic categories of the *Harvard IV Psychological Dictionary*[4].

We model the *fact-based category* by classifying the statistically relevant words according to the positive and negative word lists from the Harvard IV Psychological Dictionary (e. g. Tetlock et al., 2008; Xu and Zhang, 2013). The words in this category express a clearly positive or negative self-contained sentiment such as *"improve"* or *"decline."* Although these fact-based words are particularly popular in the related literature for measuring the impact of written texts on stock market behavior (Feldman et al., 2008; Hanley and Hoberg, 2008; Li, 2008), they do not account for the full breadth of potential emotional or cognitive orientations in a document. Thus, we additionally group words that express an emotional orientation, such as *"fascinate"* or *"mood,"* into a semantic *emotional category*. It is worth noting that the dictionary entries in the Harvard IV Psychological Dictionary can be assigned to multiple categories. Thus, we categorize words that are included in both the *fact-based category* and the *emotional category* as belonging to the *emotional facts category*. Examples include words such as *"focus"* or *"apparent."*

Our results show that the perception of investors depends on many terms that do not feature an explicitly positive or negative sentiment according to the Harvard IV Psychological Dictionary. In fact, only a total of 39 out of 114 selected words (34.21%) from the 8-K filings corpus are also included in the fact-based category. This includes 18 words that coincide with a positive stock market response and 21 words that coincide with a negative stock market response. Among these, 69.23 % of the coefficients feature the same polarity as in the Harvard IV Psychological Dictionary (i. e. are included in the positive word list, if the estimated coefficient is positive, or are included in the negative word list, if the estimated coefficient is negative). This provides evidence that words from the fact-based category are not sufficient to encompass the whole bandwidth of textual drivers in financial news. In fact, as reflected in the *emotional* and *emotional facts* categories, 36.84 % of the extracted words from the 8-K filings (16 positive words and 26 negative words) can be attributed to a domain that typically does not contribute to the informative content from a fact-based perspective.

Hypothesis Tests

Next, we control for external effects and assess the individual significance of the semantic word groups on the stock market behavior of investors. For this purpose, we perform an ordinary least squares (OLS) regression using the abnormal log-returns as the dependent variable, while we use an independent variable for each semantic category based on *facts* F, *emotions* E, and *emotional facts* EF. These variables originate from weighting the number of occurrences of each word in a publication according to the estimated coefficients from the spike and slab regression method. In order to control for potential dependencies between the semantic categories, we also include interaction terms for each possible combination between F, E, and EF.

In addition, we include a set of control variables to control for external effects. First, we account for the overall development of the stock market by using the returns of the NYSE Composite Index. Moreover, we take into consideration insider trading and the effects of information leaking before the corporate disclosure. For this purpose, we include the cumulative abnormal returns CAR of the 15 trading days before publication (Jeng et al., 2003). To account for the fact that small firms can react to news disclosures in different ways from

4 Available from http://www.wjh.harvard.edu/~inquirer/homecat.htm.

larger firms, we measure the company size by the market value MV of the company. We also include the market-to-book value $MTBV$ of a given firm to control for both life cycle effects and the future prospects of the company (Fama and French, 1993). Ultimately, we use the market model *Alpha* to account for the usual over- or under-performance of a stock (Jeng et al., 2003). In addition, we integrate dummy variables for months D_{mon}, as well as company sectors D_{sec}. Thus, our regression design is given by

$$\begin{aligned} AR_{\text{log}} = \beta_0 &+ \beta_1\, F + \beta_2\, E + \beta_3\, EF + \beta_4\, F \times E + \beta_5\, F \times EF + \beta_6\, E \times EF \\ &+ \beta_7\, CAR + \beta_8\, MV + \beta_9\, NYSE + \beta_{10}\, Alpha + \beta_{11}\, MTBV \\ &+ \beta_{12,\text{mon}}\, D_{\text{mon}} + \beta_{13,\text{sec}}\, D_{\text{sec}} + \varepsilon, \end{aligned} \tag{2}$$

with coefficients β_i and Gaussian error ε. As a final step, we remove observations for which we are not able to determine the mentioned control variables from the Thomson Reuters Datastream. This filtering leads to a final sample consisting of 43,437 observations. The regression results are provided in Table 1.

Table 1. OLS Regression Linking Word Categories and Abnormal Returns

	(a)	(b)	(c)	(d)	(e)	(f)
F	1.396^{***} (10.116)	1.418^{***} (10.366)	1.419^{***} (10.374)	1.417^{***} (10.364)	1.428^{***} (10.624)	1.427^{***} (10.609)
E	2.122^{***} (5.131)	2.219^{***} (5.415)	2.210^{***} (5.392)	2.210^{***} (5.391)	2.247^{***} (5.575)	2.245^{***} (5.569)
EF	6.933^{***} (4.565)	6.774^{***} (4.501)	6.777^{***} (4.503)	6.776^{***} (4.502)	7.249^{***} (4.898)	7.256^{***} (4.903)
$F \times E$	−1.336 (−1.020)	−1.523 (−1.173)	−1.497 (−1.153)	−1.480 (−1.141)	−1.132 (−0.887)	−1.124 (−0.881)
$F \times EF$	−5.585 (−1.064)	−4.726 (−0.909)	−4.730 (−0.910)	−4.734 (−0.910)	−7.042 (−1.377)	−7.057 (−1.380)
$E \times EF$	31.316 (1.440)	24.457 (1.135)	25.285 (1.173)	25.182 (1.168)	26.374 (1.244)	26.337 (1.242)
CAR		-0.077^{***} (−28.161)	-0.077^{***} (−28.166)	-0.077^{***} (−28.177)	−0.004 (−1.316)	−0.004 (−1.315)
MV			0.000 (1.346)	0.000 (1.354)	0.000 (1.164)	0.000 (1.163)
$NYSE$				−0.020 (−1.243)	−0.024 (−1.524)	−0.024 (−1.523)
$Alpha$					-1.199^{***} (−38.583)	-1.199^{***} (−38.584)
$MTBV$						−0.000 (−0.921)
Intercept	−0.242 (−0.891)	−0.303 (−1.122)	−0.306 (−1.133)	−0.308 (−1.143)	−0.175 (−0.662)	−0.174 (−0.656)
AIC	254,359	253,572	253,573	253,574	252,107	252,108
BIC	255,566	254,788	254,799	254,806	253,348	253,358
Adjusted R^2	0.006	0.024	0.024	0.024	0.056	0.056

Stated: coef. and t-stat. in parenthesis Dummies: month, sector Obs.: 43,437 Signif.: *** 0.001, ** 0.01, * 0.05

As a first finding, we observe that the words from the *fact-based category* have a significant effect on abnormal log-returns. In fact, the coefficient for the *fact-based category* reveals a t-value of 10.609, which is statistically significant at the 0.1 % level. Thus, we confirm our initial expectation that fact-based information significantly impacts the information processing of investors in the stock market. Next, we also note that the t-values for the *emotional category* (5.569) and *emotional facts category* (4.903) are both statistically significant at the 0.1 % level. In order to validate the robustness of this finding, we use a partial F-test to test the null hypothesis that *emotional* and *emotional fact* words jointly have no impact on abnormal log-returns, i. e. $H_0 : \beta_2 = \beta_3 = 0$. We are able to strongly reject H_0 with an F-value of 25.959, which is statistically significant at the 0.1 % level. Thus, we conclude that words from the *emotional* and *emotional facts* categories contribute decisive information in the explanation of the abnormal log-returns. We also note that none of the interaction terms is statistically significant at any common level. Overall, our results disprove the notion that investors rely solely on fact-based information, providing instead strong evidence that investors are significantly impacted by emotionally-charged content and non-informative wording.

Discussion and Managerial Implications

This study contributes to and has implications for the following areas: first, it allows for a deeper understanding of human information processing of financial news. Concordant with information processing (Friesen and Weller, 2006) and Behavioral Finance theories (Jercic and Astor, 2012), we find that the information processing of investors is based not only on a fact-based component but also on an emotional component. Moreover, our results show that emotionally-charged content in financial news significantly affects the decision-making of investors in the stock market. Interestingly, this finding can also be interpreted in terms of the so-called noise trader theory (DeLong et al., 1990). According to this theory, a relevant share of investors can be characterized as being uniformed traders that tend to make investment decisions without the benefit of fundamental data. This type of investor typically overreacts to good and bad news and is vulnerable to noisy signals, such as wording that appeals to superficial emotions or cognition, but which lacks a deeper meaning (Alfano et al., 2015).

Second, this study also has implications for practitioners who are obligated to publish regulated corporate disclosures. Since noise signals in corporate disclosures can have significant effects on investors, companies benefit from a highly self-reflective writing process that gives weight to the affective characteristics of human information processing. Accordingly, those from investor relations departments should consider that, in addition to information in the form of essential facts, the stock market decisions of investors are significantly influenced by news phrasing and word choice. This also supports the experimental study hypothesis of Bosman et al. (2014), which states that manipulating the tone of news without distorting its content influences the stock market reaction of investors. Moreover, the results of this study also reveal that the positive and negative word lists from the Harvard IV Psychological Dictionary do not encompass the complete bandwidth of the information processing of investors. Thus, phrasing corporate disclosures on the basis of these fact-based dictionaries can be misleading in a financial context.

Overall, the results of this study can help both individuals and organizations to automatically detect and respond to critical market developments, to put in place effective warning mechanisms and to develop decision support systems. Moreover, understanding how humans perceive information in the form of written text is not only relevant for the domain of financial markets but for any company that is subject to market dynamics. Since many sources of information (e. g. product reviews, recommendations, forum entries, etc.) consist of a mixture between fact-based information and content that appeals to underlying emotion and subconscious cognition, this Information Systems study also provides ample scope for further research in various domains.

Conclusion and Further Research

Stock-listed companies are required to publish relevant corporate information that has the potential to influence their valuation. Although subject to various regulations, these publications are not regulated in terms of word choice. However, the phrasing of news can influence the perception of investors in the stock market. In fact, investors do not rely solely on essential facts when processing the provided information, but can fall victim to affective characteristics in their perception that make them vulnerable to information that appeals to superficial emotions or cognition. Thus, this paper sheds light on the information processing of investors by extracting qualitative verbal information from financial news.

As its main contribution, this paper utilizes spike and slab regression to study human information processing on a word-by-word level. In contrast to related studies, which typically rely on predefined word lists, our approach allows us to encompass the complete bandwidth of information drivers. By grouping the stock-market-relevant words into semantic categories, we explicate the effects of emotionally-charged content on the behavior of investors. In this context, 36.84 % of the information-driving words from 8-K filings can be assigned to the emotional domain. From our results, we conclude that these words significantly impact the perception of investors in the stock market. Thus, this paper provides evidence that the information processing of investors is based not only on a fact-based component but also on an emotional component.

Overall, this study allows for a deeper understanding of the roles of facts and emotions in the human information processing of financial news. On the road to completing this research in progress, we will expand the study in four directions. First, we will include additional news sources from other financial domains to test the robustness of our findings. Second, we will study the role of emotional content across different news topics. In this regard, one might suspect that investors tend to weight emotional content more heavily for some topics than for others. Third, we will utilize word combinations, i. e. n-grams, to identify groups of emotionally-charged words that frequently occur together. Fourth, it is an intriguing notion to integrate the utilized variable selection method with negation scope detection (Pröllochs et al., 2016) to reduce the possibility of misclassification due to inverted meanings.

References

Aizawa, A. 2003. "An Information-Theoretic Perspective of Tf–Idf Measures," *Information Processing & Management* (39:1), pp. 45–65.

Alfano, S. J., Feuerriegel, S., and Neumann, D. 2015. "Is News Sentiment More than Just Noise?" in *23rd European Conference on Information Systems (ECIS)*, Münster, Germany.

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

Atkinson, R. C., and Shiffrin, R. M. 1968. "Human Memory: A Proposed System and its Control Processes," *Psychology of Learning and Motivation* (2), pp. 89–195.

Bosman, R., Kräussl, R., and Mirgorodskaya, E. 2014. "The Tone Effect of News on Investor Beliefs: An Experimental Approach," *SSRN Electronic Journal* .

Browne, G. J., and Parsons, J. 2012. "More Enduring Questions in Cognitive IS Research," *Journal of the Association for Information Systems* (13:12), pp. 1000–1011.

Carter, M. E., and Soo, B. S. 1999. "The Relevance of Form 8-K Reports," *Journal of Accounting Research* (37:1), pp. 119–132.

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

Demers, E. A., and Vega, C. 2010. "Soft Information in Earnings Announcements: News or Noise?" *SSRN Electronic Journal* .

Engelberg, J. 2008. "Costly Information Processing: Evidence from Earnings Announcements," *SSRN Electronic Journal* .

Fama, E. F., and French, K. R. 1993. "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* (33:1), pp. 3–56.

Feldman, R., Govindaraj, S., Livnat, J., and Segal, B. 2008. "The Incremental Information Content of Tone Change in Management Discussion and Analysis," *SSRN Electronic Journal* .

Friesen, G., and Weller, P. A. 2006. "Quantifying Cognitive Biases in Analyst Earnings Forecasts," *Journal of Financial Markets* (9:4), pp. 333–365.

Gidófalvi, G., and Elkan, C. 2001. "Using News Articles to Predict Stock Price Movements," *Department of Computer Science and Engineering, University of California, San Diego* .

Hanley, K. W., and Hoberg, G. 2008. "Strategic Disclosure and the Pricing of Initial Public Offerings," *US Securities and Exchange Commission Working Paper* .

Hastie, T., Tibshirani, R., and Friedman, J. H. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, New York: Springer, 2nd ed.

Henry, E. 2008. "Are Investors Influenced by How Earnings Press Releases are Written?" *Journal of Business Communication* (45:4), pp. 363–407.

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

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

Jeng, L. A., Metrick, A., and Zeckhauser, R. 2003. "Estimating the Returns to Insider Trading: A Performance-Evaluation Perspective," *Review of Economics and Statistics* (85:2), pp. 453–471.

Jercic, P., and Astor, P. J. 2012. "A Serious Game using Physiological Interfaces for Emotion Regulation Training in the Context of Financial Decision-Making," in *20th European Conference on Information Systems (ECIS)*.

Kahneman, D., and Miller, D. T. 1986. "Norm theory: Comparing Reality to its Alternatives," *Psychological Review* (93:2), p. 136.

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

Konchitchki, Y., and O'Leary, D. E. 2011. "Event Study Methodologies in Information Systems Research," *International Journal of Accounting Information Systems* (12:2), pp. 99–115.

Lee, M. J., and Lee, H. 2013. *Predicting Corporate 8-K Content Using Machine Learning Techniques*, Stanford, CA: Stanford University Press.

Li, F. 2008. "Annual Report Readability, Current Earnings, and Earnings Persistence," *Journal of Accounting and Economics* (45:2-3), pp. 221–247.

Li, X., Shen, J., Gao, X., and Wang, X. 2010. "Exploiting Rich Features for Detecting Hedges and their Scope," in *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pp. 78–83.

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

Loughran, T., and McDonald, B. 2014. "Measuring Readability in Financial Disclosures," *Journal of Finance* (69:4), pp. 1643–1671.

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

Malsiner-Walli, G., and Wagner, H. 2011. "Comparing Spike and Slab Priors for Bayesian Variable Selection," *Austrian Journal of Statistics* (40:4), pp. 241–264.

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

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

Phillips, B., Prybutok, V. R., and Peak, D. A. 2014. "Decision Confidence, Information Usefulness, and Information Seeking Intention in the Presence of Disconfirming Information," *Informing Science: The International Journal of an Emerging Transdiscipline* (17:1), pp. 1–25.

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

Pröllochs, N., Feuerriegel, S., and Neumann, D. 2016. "Negation Scope Detection in Sentiment Analysis: Decision Support for News-Driven Trading," *Decision Support Systems* (88), pp. 67–75.

Salton, G., Fox, E. A., and Wu, H. 1983. "Extended Boolean Information Retrieval," *Communications of the ACM* (26:11), pp. 1022–1036.

Schneider, W., and Shiffrin, R. M. 1977. "Controlled and Automatic Human Information Processing: Detection, Search, and Attention," *Psychological Review* (84:1), pp. 1–66.

Schumaker, R. P., and Chen, H. 2009. "A Quantitative Stock Prediction System Based on Financial News," *Information Processing & Management* (45:5), pp. 571–583.

Stone, P. J. 2002. "General Inquirer Harvard-IV Dictionary," .

Tetlock, P. C. 2007. “Giving Content to Investor Sentiment: The Role of Media in the Stock Market,” *Journal of Finance* (62:3), pp. 1139–1168.
Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. 2008. “More Than Words: Quantifying Language to Measure Firms’ Fundamentals,” *Journal of Finance* (63:3), pp. 1437–1467.
Thaler, R. H. 2005. *Advances in Behavioral Finance*, Roundtable Series in Behavioral Economics, Princeton: Princeton University Press.
Tversky, A., and Kahneman, D. 1974. “Judgment Under Uncertainty: Heuristics and Biases,” *Science* (185:4157), pp. 1124–1131.
Varian, H. R. 2014. “Big Data: New Tricks for Econometrics,” *Journal of Economic Perspectives* (28:2), pp. 3–27.
Vodanovich, S., Sundaram, D., and Myers, M. 2010. “Digital Natives and Ubiquitous Information Systems,” *Information Systems Research* (21:4), pp. 711–723.
Wilson, T. D. 1999. “Models in Information Behaviour Research,” *Journal of Documentation* (55:3), pp. 249–270.
Xu, S. X., and Zhang, X. M. 2013. “Impact of Wikipedia on Market Information Environment: Evidence on Management Disclosure and Investor Reaction,” *MIS Quarterly* (37:4), pp. 1043–1068.