INFORMATION PROCESSING IN ELECTRONIC MARKETS: MEASURING SUBJECTIVE INTERPRETATION USING SENTIMENT ANALYSIS

Completed Research Paper

Michael Liebmann
Information Systems Research
University of Freiburg
Platz der Alten Synagoge
79085 Freiburg, Germany
michael.liebmann@is.uni-freiburg.de

Michael Hagenau
Information Systems Research
University of Freiburg
Platz der Alten Synagoge
79085 Freiburg, Germany
michael.hagenau@is.uni-freiburg.de

Dirk Neumann
Information Systems Research
University of Freiburg
Platz der Alten Synagoge
79085 Freiburg, Germany
dirk.neumann@is.uni-freiburg.de

Abstract

Information availability plays an important role in the efficient resource allocation of electronic markets and e-commerce. Most of this information is of qualitative nature containing essential facts that are, however, difficult to decode. Currently, the information processing capabilities of human agents facing such qualitative news is mostly unknown. Accordingly, it is crucial to understand how different decision makers process qualitative information. In this paper we show that sentiment-analysis facilitates research in qualitative information processing. We use a capital market example to demonstrate how investors and analysts perceive novel information. We find that their interpretation is different from one another: investors rapidly translate novel information into transactions, whereas analysts take more time to respond. We further observe that analysts emphasize different parts of information than investors, and are less put-off by complex information. The approach can be applied to other electronic-markets and the e-commerce industry where individuals react upon textual information.

Keywords: Information processing, Sentiment analysis, Text mining

Introduction

The widespread use of Internet technologies has fueled the emergence of electronic markets and e-commerce as fundamental parts of today's economy. Ebay, Alibaba and Google Adword Auctions are prominent examples of those newly established electronic markets. In addition, an increasing number of financial exchanges are instances of electronic markets that have completely replaced floor exchanges, e.g. NASDAQ, XETRA and the Electronic Banking System (EBS). These constitute multiple trillion Dollar transaction volumes that are handled by electronic markets every day. Similarly, the e-commerce industry has gained significant importance with multiple annual double-digit growth over the past decade. Having the importance of electronic markets and e-commerce in mind, it is essential to understand why they perform quite well in their function of efficient resource allocation.

It is well known that market efficiency depends on the availability of information, such as product and pricing information. Both, electronic markets and the e-commerce industry, are information-intensive, thus, facilitate the access to valuable market information: it is evident that decision makers, i.e. consumers, suppliers and intermediaries, benefit from more information to make purchase and sales decisions (Granados et al. 2010). But product and pricing information provided by suppliers and producers do not drive market efficiency alone. Purchasing decisions are increasingly influenced by supplemental product information provided by user and consumer feedback (Dellarocas et al. 2007; Reinstein and Snyder 2005; Chevalier and Mayzlin 2006). Dellarocas et al. (2007), and Chevalier and Mayzlin (2006) show that book and movie sales are related to positive and negative customer ratings, i.e. review quality. Duan et al. (2005), and Liu (2006) find a relation between sales and review volume.

With increasing electronic mobility and facilitated social network communication, the amount of customer feedback and publicly expressed opinions is very likely to further increase. But an increasing amount of qualitative information does not only facilitate purchasing decisions. Hirshleifer et al. (2009) show that processing multiple sources of information at the same time can actually distract decision makers and negatively influence market efficiency. There are three main challenges:

1. *Information inequality*: some information is more relevant for decision makers than other and may receive more attention (Hirshleifer et al. 2009).
2. *Decision maker subjectiveness*: different decision makers may not equivocally process and interpret new information and can disagree on the same content (Varian 1986, Lehavy et al. 2011).
3. *Processing cost*: As most qualitative information is compiled in the shape of unstructured textual data, their processing is more costly as processing quantitative content (Engelberg 2008).

These three challenges affect transparency and, thus, may influence market efficiency. Hence, it is valuable for information systems research to better understand and measure their impact. However, the formal analysis of qualitative textual information is technically challenging and can require extensive research setups. At the same time, incorporating textual data opens up new avenues for research that leads to a number of intriguing research questions: how do different decision makers with varying objectives process the same information? How long does novel information influence different decision makers? What drives decision makers' disagreement?

In this paper we show how text mining and sentiment analysis help to overcome the technical challenges of empirically analyzing qualitative textual data and contribute to answer our overarching questions. As demonstrated by Li (2010a), the increasing availability of large amounts of textual information adds additional dimensions to the examination of market efficiency. In addition, using more powerful techniques in sentiment analysis allows scholars to analyze these untapped pockets of valuable insights. In this paper we shed light into the qualitative information processing by analyzing how different decision makers react to the same information and how they differ in their interpretation. We use a dynamic text mining approach that evaluates the content of text messages and assigns different sentiment values for concurrent interpretations (Liebmann et al. 2011): the *Tonality approach*. The approach quantifies the correlation between the occurrence of particular words in financial news and their observed positive or negative interpretation. As proxies for positive and negative interpretation we use capital market effects, which affords us the ability to quantify such correlations for any kind of asset that reacts to the news. Using Tonality, we can separately calculate values for the tone and informativeness of single text components. Since the Tonality approach is dynamic and can be adjusted for different feedback of the same text, tone and informativeness can be differentiated based on different interpretations. Thus, we can identify which text components are important for one reader but not necessarily for another. As a result, sentiment analysis, and the Tonality approach in particular, help to overcome all three transparency challenges stated above and allow us to better understand the information processing of qualitative information in electronic markets.

To demonstrate the approach we chose data from electronic capital markets and analyze the information processing of corporate announcements. Analyzing capital markets is an appropriate choice for studying the information processing phenomenon, as there exists a rich data set of information already. Research in information processing in capital markets is particularly interesting, as there are different types of decision makers involved (e.g. investors and intermediaries) that actively contribute to the information processing of novel information and, thus, market efficiency. After all, capital market price changes are nothing else than a measurable proxy for the "wisdom of crowds". Besides the advantageous setup of capital markets for this analysis, there is an extensive research body on decision making behavior in finance research available that helps to interpret and benchmark the results for the IS community. We foresee that we can apply the knowledge gained in this analysis, be it the methodology or be it the results, also to other types of electronic markets and to the e-commerce industry where market participants react upon textual information.

In our study, we analyze and compare how different types of decision makers (i.e. investors and financial analysts) interpret and react upon the submission of important corporate announcements. The differentiation between investors and analysts is important for our work as these two groups play an essential role in the discussion of market efficiency (Kothari 2001). At the same time, investors and analysts differ in their objectives and information processing abilities. Thus, we expect a noticeable difference when they interpret and react to the same set of corporate announcements.

As textual information source we use a comprehensive set of corporate announcements and evaluate their content in terms of positive or negative market signals, which we denote as *Tonality*. We also develop a metric that explains how complex the information embedded in these announcements is. This metric facilitates the analysis whether it is *Complexity* that drives disagreement across investors and financial analysts. Our set of corporate announcements includes only stock price relevant facts, such as profit warnings, management changes and takeover considerations. Thus, the data set is well suited to measure and compare information processing of different news recipients, as it contains only low noise in the data.

We find that the same information gets processed quite differently: investors rapidly translate novel information into stock prices on the day of the announcement, whereas financial analysts require significantly more time to update their forecasts to respond to novel information. One might argue this may indicate that analysts fail to fulfill their obligation to serve investors with value-adding information. However, our research shows that the delayed information processing of financial analysts is not obsolete, even if novel information is translated into prices rapidly. It turns out that financial analysts emphasize different parts of novel information than investors. On the one hand, investors are more sensitive towards author suggestions. For them terms such as "*sluggish* corporate performance", "*slump* orders" or "*disappointing* results" provide higher informativeness. Analysts, on the other hand, seem to pay attention to fact-driven communication, expressed by words as "revenues *fell*", "EBIT *decrease*" or "results *below* expectations". As a result, analysts are able to unleash information content into their forecasts even weeks after the original submission, which has been disregarded by swift investor interpretations.

In this context, Jegadeesh and Wu (2011) analyze whether investors under- or overreact to sentiment in annual reports. They find a two week underreaction and conclude that the market does not fully respond to the sentiment of annual reports during the filing period. Our results, on the contrary, document that the market immediately reacts to news sentiment and almost fully translates novel content into stock prices on the announcement day. Markets show some responsiveness to news sentiment during the following trading days, but stop within the first week of the submission. We also find that the relation between stock price development and novel information does not increase during any period after the announcement day. Consequently, we do not find any evidence for a noticeable market underreaction. As shown by You and Zhang (2009), more lengthy disclosures such as annual reports and less novel information result in a measurable market underreaction. Having this fact in mind, we can explain the difference to Jegadeesh and Wu by observing the difference in our data sets: our set of corporate announcements only comprises important and unexpected facts that are simpler to process than lengthy annual reports (data set of Jegadeesh and Wu). Accordingly, it is the lengthy structure of annual reports that may result in a market underreaction.

Analyzing stock price development over a longer time frame we find that the relation between stock prices and news content is highest on the announcement day and gradually deteriorates thereafter, even though it remains significant over several weeks. This demonstrates how fast novel and important information wears-off and newer information interferes. However, since the relation remains significant over several weeks, it shows that the content offered by this data set provides some sustainability and thus is very important for sustainable investor decision making – not only on the day of the announcement. Furthermore, we find that investors can struggle to interpret more complex announcements. We consider an announcement to be more complex the more informative text components it contains. While investors seem to increasingly disagree with increasing Complexity, analysts are less put-off by Complexity.

Empirical research of qualitative textual information processing is still in its infancy and the proposed methods are influenced by a margin of error, so are results. However, by demonstrating a discernible relation between the reaction of decision makers and qualitative content, we believe that sentiment analysis, as a tool in information processing research, will help to better understand which information drives decision makers and contributes to a better understanding in market econometrics.

Related Work

In this section we present related literature in two areas: first, alternative designs of sentiment analysis for corporate disclosures. Second, analyses of information processing and difference in opinion following financial announcements.

Sentiment analysis

In sentiment analysis, texts are associated with a particular value or category in order to answer a certain research question. There is a growing research body on the relationship between tone of financial disclosures and observed stock price reactions (Tetlock et al. 2008; Loughran and McDonald 2011; Li 2010a). Most approaches measure how market participants perceive and react upon important corporate news. They use observed stock price reactions following the announcement, to validate their classification into positive or negative sentiment.

As written text is a form of unstructured information, it first has to be translated into a machine readable representation of the text[1]. This step includes determining the selection of words, which are most relevant for the analysis. Second, an aggregated, quantified metric needs to be calculated for each announcement based on the occurrences of the previously selected words. These two steps have to be performed whenever a researcher conducts sentiment analyses. Thereby, Li (2010) distinguishes *dictionary-based* and *statistical approaches*.

The *dictionary* has a fix set of words and often originates from a certain specific field, e.g. the Harvard-IV General Inquirer (GI), which is a psychosocial dictionary (used by Tetlock et al. 2008). For aggregation into a metric, Tetlock et al. (2008) simply count and add up all occurrences of negative words from the dictionary for each message. While this approach is striking for its simplicity, it suffers from certain drawbacks: few dictionaries exist that are built for the setting of corporate financial statements and thus may not be applicable for a capital market analysis (Loughran and McDonald 2011). Further, the simple dictionary-based approach ignores the context of a sentence. For instance, if a sentence is about cost, then the word '"increase" should be treated as negative: however, it is likely to be a positive word if the sentence is about "revenue". Lastly, some words may change their sentiment depending on the specific context they are used in: while the word "cancer" is a serious disease and labeled negative in the GI dictionary, it may be part of a rather positive corporate announcement of a pharmaceutical company that just received an FDA approval of a new drug or treatment, as shown by Loughran and McDonald (2011).

[1] Creating machine readability contains three sub-steps: Feature extraction (i.e. retrieving words from a text), Feature Selection (i.e. determining which of the retrieved words are relevant), and Feature Representation (i.e. selecting the data format how occurrences of words in a text shall be represented). We follow this approach, but simplify the subsequent description.

To mitigate the shortcomings of predetermined dictionaries, *statistical approaches* have been developed to measure statistical correlations between keywords and observed capital market effects to identify positive or negative content. This task mostly is performed by machine learning approaches (Li 2010b; Hagenau et al. 2011). The result of statistical approaches is a metric that allows the researcher to obtain a quantitative value for each announcement regarding its content (i.e. sentiment). Liebmann et al. (2011) argue that also typical machine learning approaches carry relevant shortcomings that impact a comprehensive sentiment analysis: most machine learning approaches embody a black-box character, which complicates the analysis of single text components. Hence, the authors propose a statistical approach that directly quantifies the tone of single text components in a data set. The observed stock prices are used as proxy for the interpretation of an announcement. The subsequent quantification expresses whether this word is perceived as positive or negative and how relevant it is for the interpretation. Some words are less relevant, whereas other words better help to discriminate between positive and negative interpretations. The authors call their metric *Tonality*. While the sign of Tonality identifies the positive or negative tone of the word, the absolute value reflects its *informativeness*. The higher the absolute Tonality, the more the words discriminates between positive and negative announcements (i.e. higher informativeness).

In this paper, we use the Tonality approach to identify the most informative words for investor and analyst interpretation of corporate news. Since each single word receives a different value for tone and informativeness we are able to analyze on which words investors and analysts emphasize and how they differ when reading the same set of corporate announcements.

Information processing and difference in opinion

In this subsection we elaborate on related literature that provides seminal findings and research setups to analyze the information processing of decision makers in capital markets. Research of information processing in capital markets has been conducted for decades: Ball and Brown (1968) and Bernard and Thomas (1989) and others find stock price drifts following corporate announcements in the same direction as the immediate stock price reaction over several months. They conclude that these drifts can be interpreted as an underreaction of the market to the information provided in the announcements. Ball and Brown (1968) and Bernard and Thomas (1990) did not use sentiment measures as proxies for the announcement content, they simply analyzed submission timing relative to price reactions.

However, using sentiment measures to better model the content of financial disclosures has recently gained attention in research: Tetlock et al. (2008) show that stock price changes correspond to the content (i.e. sentiment) of financial news. They find a discernible relation and provide an empirical research setup using inferential regression analysis. The authors constitute, if they find a relation between sentiment and stock price development, it could not be denied that there is an influential relationship between the two variables. Li (2010), Loughran and McDonald (2011) and Jegadeesh and Wu (2011) confirm their findings in similar analysis setups (but based on different sentiment approaches). Jegadeesh and Wu (2011) demonstrate that market underreactions to novel information in annual reports can be traced using a sentiment approach. The authors analyze whether their reports' sentiment influences stock price returns over several weeks following the publication. They find that stock prices tend to underreact to the information in annual reports over two weeks after the announcement.

Announcement drifts and stock price underreaction are not the only information processing aspects that have been explored: according to the *information processing cost hypothesis*, more complex information in disclosures increases the processing cost for their receivers (Grossman and Stiglitz 1980; Bloomfield 2002). Lehavy et al. (20ll) find that less readable annual reports are associated with greater dispersion, lower accuracy, and greater overall uncertainty in earnings forecasts of financial analysts. But interpretation dispersion is not only a phenomenon among financial analysts. Varian (1986) and Kim and Verrecchia (1991) provide theoretical frameworks to model difference in opinion of investors. Kandel and Pearson (1995) add empirical evidence and a research setup for these considerations. Kandel and Pearson (1995) find that neither investors nor analysts have identical interpretations of public announcements. The authors analyze trading volume in the presence of an announcement and show a noticeable increase in volume that is not explained by strong price changes or other influential factors. They conclude that investors interpreting the same announcement differently may place concurrent trading orders. Considering this, we may find increased trading volume without strong price changes.

A measure for complexity in financial disclosures has been provided by Miller (2010): the author measures announcement complexity simply by taking the length of a disclosure. We argue, that announcement complexity rather depends on the amount of informative text components and not necessarily the pure length. However, we take this variable into consideration when designing our own factors.

Methodology

In this section we elaborate on the sentiment algorithm used to calculate our text metrics *Tonality* and *Complexity*. We define our measures for investor and analyst interpretation and present descriptive statistics on our data set and reason why we chose our particular set of corporate announcements.

Sentiment metric design

As sentiment approach we chose the design of Liebmann et al. (2011) to perform our sentiment analysis. The approach quantifies the statistical correlation between each single word within the textual corpus and the observed stock price reactions on the announcement day. The observed stock price reaction serves as feedback proxy for the interpretation of market participants (e.g. investors). Thus, each announcement can be associated with a positive or negative investor feedback. The approach allows us to calculate a value for each word that reflects its correlation with positive and negative investor feedback (i.e. word-Tonality). The more a particular word occurs in news with positive market feedback while it clusters less in news with negative feedback, the higher is the according correlation with positive news, vice-versa for negative. The approach is based on the Chi² equation with the difference, that positive and negative nuances can be measured independently. We calculate the Tonality of a particular word as follows,

$$Tonality_{Word} = \frac{O_{pos} - E_{pos}}{E_{pos}} - \frac{O_{neg} - E_{neg}}{E_{neg}} \quad . \tag{1}$$

O_{pos} denotes the observed frequency of positive news containing a particular word. E_{pos} denotes the expected frequency that this word occurs in a positive announcement. Expected frequencies can be calculated based on the distribution of positive and negative news, as in the Chi² method. If the occurrence of a particular word is independent of the positive or negative interpretation of the news, it should occur equally often in both. If the observed frequency deviates noticeably from the expected frequency, we cannot per se neglect that this word is independent of a positive or negative interpretation. O_{neg} and E_{neg} are calculated for negative news accordingly. The differences between observed O and expected E are standardized by E so that the words that occur more frequently are not overvalued.

The sign of the word-Tonality value differentiates between positive and negative words, whereas the absolute value reflects how informative the word for the discrimination between positive and negative interpretation is. Words that discriminate stronger between positive and negative announcements are considered to be more informative. A neutral word which does not discriminate at all between positive and negative stock price reactions would receive a word-Tonality of zero.

Applying the Tonality approach, we calculate tone and informativeness for all words in our news corpus. As a results, we obtain 3,227 unique word stems. Based on the distribution of informativeness across all words, we identify the top 600 words as most discriminative between positive and negative stock price signals. We chose to cut-off the word list where the informativeness starts to ascends comparably flat. The more words contain the same or similar informativeness the less they discriminate. As the distribution of the informativeness forms an S-shape, scholars may simply use the saddle points of either sides of the distribution to obtain their cut-offs for particular word lists. Details on the word selection process and how a selection of fewer words reduces information redundancy can be found in Liebmann et al. (2011). To be able to generalize potential findings, we perform the selection process out-of-sample on a different set of corporate announcements (details in data subsection). Based on the resulting word list we calculate measures for announcement sentiment and its complexity.

To obtain a sentiment value that best reflects the likely investor interpretation of a given announcement, all observed word-Tonalities have to be aggregated into a single value (hereafter Tonality). Let TW_i be the word-Tonality of word *i* in our word list *WL* and $f_{i,A}$ be its frequencies in announcement *A*. N_A denotes the total number of words in the announcement:

$$Tonality_A = \frac{1}{N_A} \cdot \sum_{i \in WL} TW_i \cdot f_{i,A} \tag{2}$$

Thus, for each announcement, the word-Tonalities for the occurring words are added up and normalized by the total number of words in the announcement. The normalization is important to avoid long announcements to automatically receive higher Tonality figures. We hypothesize that a positive or negative perception of an announcement does not necessarily require long text. Therefore, we design the Tonality metric as independent from the length of the announcement. A more positive (negative) Tonality value indicates a more positive (negative) investor interpretation.

Based on the assumption that more complex announcements are less likely to be perceived homogeneously across receivers (information processing cost hypothesis, see related work section), we design a Complexity measure. We assume that announcements with more informative words (based on our definition) allow more room for differential interpretation. Since the Tonality word list provides us with the most informative words in an announcement, we simply need to count their number in one message. It could be argued whether the total word count in an announcement is a better measure for complexity, however, since executives and editors tend proliferate news with unrelated and less important facts, we resort to counting only words that provide an informative value-add to the reader. We do not standardize this word count as the absolute amount of incremental information is important for this analysis. We do not expect a linear relationship between announcement Complexity and differential interpretation. Similarly, we presume that doubling the amount of informative words in a news text does not double the interpretation complexity. As an approximation for these assumptions we take the natural logarithm of informative words for each announcement. Considering this, the Complexity measure for announcement *A* is calculated based on the number of informative words $N_{w,A}$ within the announcement:

$$COMPLEXITY_A = \ln\left(N_{w,A}\right) \tag{3}$$

Measuring investor and analyst interpretation

In this subsection we elaborate the different feedback proxies for investor and analyst interpretation. We use the stock price development as proxy to model *investor announcement interpretation*. As important corporate announcements have been found to significantly influence the stock price development on the announcement day, the reaction can be seen as an aggregation of all market participants in one value. In order to reduce influences from overlaying market effects we calculate daily abnormal returns (*AR*) that are inherently firm-specific. We deliberately chose a simple market model [2] (MacKinlay 1997) as benchmark to calculate abnormal returns. Based on these abnormal returns we can measure investor feedback over daily intervals and over longer time frames when aggregating daily abnormal returns in cumulative abnormal returns (*CAR*).

To model a proxy for *analyst interpretation* we use Gleason and Lee (2003) as an orientation[3] and measure how analysts change their forecasts after corporate announcements. As corporate announcements have been found to influence analyst forecasts for earnings per share (EPS), we consider the forecast change as valuable proxy for analyst interpretation. We calculate the percentage change of the consensus following the announcement compared to the trading day immediately preceding (*T-1*). Consequently, the EPS forecast revision magnitude of a given firm on trading day *t* relative to trading day *T-1* is

---

[2] It could be argued that a three factor model is more appropriate to calculate abnormal returns. However, Liebmann et al. (2011) show that there is hardly any difference in choosing abnormal returns or actual raw returns in the announcement analysis. Thus, we consider the error from not choosing a three factor model negligible.

[3] Gleason and Lee (2003) analyze price forecasts of analysts and use closing prices on trading day *T-1* as benchmark.

$$REVISION_t = \frac{EPS_t}{EPS_{T_{-1}}} - 1. \tag{4}$$

As $REVISION_t$ calculates the percentage change on a particular trading day relative to trading day *T-1* it reflects the development over several days. Hence, AR_0 and $REVISION_0$ measure the investor and analyst interpretation on the announcement day and CAR_t and $REVISION_t$ (for $t>0$) are proxies for the development over longer time frames.

Dataset

Based on the efficient market hypothesis, it is most important that information is *novel* and *relevant* to investors when analyzing capital market effects following corporate announcements. Otherwise, it is unlikely to observe a meaningful relation between the announcement and the stock price. Since testing for novelty and stock price relevance are research subjects of their own, we directly select a data set that fulfills these requirements. Thus, we select regulated financial news, that obliges management executives to only publish those new facts that are bound to change the stock price. German regulated Adhoc Announcements (AA) best fulfill these requirements, as they have to be published immediately, are accessible to all investors at the same time and are quality controlled by a public authority prior publishing[4]. Firms have to publish any material facts as AA that are expected to affect the stock price. Ad-hoc announcements typically include facts on unexpected earnings, management changes, M&A transactions, dividends, and other content. German Ad-hoc announcements are found to provide significant stock price and volume reactions (Röder 2000), thus, may provide good potential to make the relation between Tonality and stock price effects more obvious. Furthermore, Ad-hoc announcements are relatively short (between 100 und 500 words) as firms are obliged to only include relevant facts. Thus, lowering noise levels when processing qualitative data.

We take pricing data from XETRA (Exchange Electronic Trading). For the regression analysis, we include a comprehensive set of control variables (e.g. market value, market-to-book ratio, market model regression alpha, share turnover). We exclude penny stocks with a share price below 2 Euro on the announcement date to ensure that findings are not influenced by illiquid stock events. Furthermore, events with extreme stock price effects on the event day were removed (winsorized at the 1% level). We also controlled for missing values and erroneous entries. We required each message to have a minimum of 50 words in total, to separate announcements that only show tables. We impose these filters to limit the influence of outliers and errors.

Table 1 – descriptive statistics

Return interval	# news	Tonality	Market value	Market-to-book	Alpha	Turnover
< -8%	611	-0.0114	2,137	4.81	-0.0015	0.056
-8 to -6%	273	-0.0065	3,731	5.4	-0.0009	0.056
-6 to -4%	448	-0.0048	2,734	3.4	-0.0006	0.083
-4 to –2%	839	-0.0033	4,728	4.03	-0.0005	0.074
-2 to 0%	1225	-0.0021	4,489	3.52	-0.0001	0.064
0 to 2%	1459	0.0012	4,898	3.67	-0.0004	0.069
2 to 4%	1058	0.0015	3,801	3.58	-0.0003	0.065
4 to 6%	662	0.0032	3,199	4.4	-0.0005	0.062
6 to 8%	409	0.0033	2,945	4.36	-0.0007	0.056
> 8%	710	0.0046	1,202	5.14	-0.0015	0.057
Total	**7,694**	**-0.0008**	**3,686**	**4.04**	**-0.0005**	**0.065**

[4] US 8-K filings are similar to German AA, but can be published up to 4 days after news becomes effective.

Table 1 shows descriptive statistics for our news data set relative to the observed stock price effect on the announcement day (clustered in 10 return intervals). A concentration of announcements around zero return can be observed, but still 65% of all events reached abnormal returns larger than +/-2%. Smaller firms (by market value) incur larger stock price returns, in line with Amihud and Mendelson (1988). Even more obvious seems the relation between Tonality and return: Tonality increases strictly monotonic with increasing return, switching from negative to positive at the same interval as returns do.

We analyze the last 12 years of Adhoc announcements published between 1998 and 2010. Our data sample begins with 21,106 announcements. Outlier, erroneous message and penny stock removal yields 12,824 announcements with consistent stock price and textual data. To select our word list and obtain the Tonality values out-of-sample, we split the data set in word list-selection and validation set by 40% to 60%. All subsequent analyses has been performed on the validation set with 7,694 valid Ad-hoc announcements. In the interest of optimal space allocation we omitted a detailed correlation analysis, however all variables are tested for individual and multicollinearity as well for heteroskedasticity.

Results

In this section we empirically investigate the information processing of investors and financial analysts when reading and reacting to important corporate announcements. For this purpose, we use sentiment analysis to measure the relation between investor and analyst behavior and the content in the announcements. The section is structured according to following five research questions that are designed to contemplate the information processing of investors and financial analysts:

1. *Investor timing*: how long do investors require to interpret novel information and translate it into stock prices?
2. *Information sustainability*: over what time frame does novel information influence investor transactions and how does this resonate with traditional explanatory factors?
3. *Analyst timing*: how long do analysts require to interpret novel information and respond to it in their forecasts?
4. *Analyst value-add*: if analysts do require more time than investors to reflect the same information in their forecasts, can we find an informative value-add of their revised forecasts during periods immediately following an important public announcement?
5. *Difference in interpretation*: does announcement complexity drive readers' disagreement and are investors and financial analysts equally influenced by announcement complexity?

Each of these research questions will be examined through a set of regression analyses employing our sentiment measures. For each analysis, we use pooled ordinary least squares regressions (OLS) and robust standard errors clustered by calendar month, to account for seasonal effects caused by accumulated publishing in the last month of each quarter. Tonality serves here as a proxy factor for announcement content as it was designed (out-of-sample) to measure value relevant information. Thus, the use of Tonality allows us to draw inferences on how information effects evolve by analyzing its relation to stock price movements and forecast revisions.

Investor timing

To analyze investor timing, we use a set of regressions - one for each of the eight trading days surrounding the announcement - to investigate how stock price changes relate to announcement content (i.e. Tonality) on a particular day. Thus, the key independent variable here is Tonality. We also incorporate a set of control factors to control for internal effects and autocorrelations. Based on Jegadeesh and Titman (1993) and Chan et al. (1996) we account for momentum induced by cumulative abnormal returns during the six trading weeks prior to the announcement ($CAR_{-29,-4}$) and the market model regression intercept of the previous year (*Alpha*). We also control for market-to-book value (*ln(MTBV)*) and firm size (*ln(MV)*), both factors are taken at the prior year-end. These controls are analogous to the Fama and French (1992) factors. Furthermore, we control for trading volume using the natural logarithm of share turnover (*ln(Turnover)*). The share turnover is calculated as trading volume of the period of interest divided by the total number of shares outstanding. During each regression we control for year-, industry-, and firm effects using a comprehensive set of dummy variables.[5] Furthermore, we tested for heteroscedasticity and multicollinearity[6] to ensure that no abnormality confounds the experiments' results.

Table 2 – Investor Timing

Variable	**AR$_{-3}$**	**AR$_{-2}$**	**AR$_{-1}$**	**AR$_0$**	**AR$_1$**	**AR$_2$**	**AR$_3$**	**AR$_4$**	**AR$_5$**
Announcement content variable									
Tonality	0.0540 (1.21)	0.0469 (1.13)	0.1047 ** (2.28)	1.6107 *** (21.13)	0.1763 *** (4.15)	-0.0054 (-0.13)	0.0622 * (1.78)	0.0669 * (1.75)	0.0232 (0.63)
Control variables									
CAR$_{-29,-4}$	-0.0089 ** (-2.07)	0.0004 (0.09)	0.0116 ** (2.42)	-0.0187 *** (-2.97)	-0.0033 (-0.90)	-0.0041 (-1.25)	0.0024 (0.78)	-0.0048 (-1.27)	0.0047 (1.14)
Alpha	-1.0395 *** (-3.53)	-1.8831 *** (-5.53)	-1.1807 *** (-3.50)	-3.0068 *** (-5.18)	-0.8384 *** (-2.69)	-0.4858 (-1.61)	-0.5289 (-1.56)	-0.8664 *** (-3.76)	-1.0418 *** (-4.13)
ln(MV)	-0.0026 * (-1.84)	-0.0030 ** (-2.05)	-0.0040 ** (-2.55)	-0.0121 *** (-5.22)	-0.0002 (-0.13)	0.0000 (-0.03)	0.0005 (0.45)	0.0004 (0.30)	-0.0004 (-0.35)
ln(MTBV)	-0.0003 (-0.20)	0.0014 (0.87)	0.0013 (0.77)	0.0032 (1.34)	0.0012 (0.81)	-0.0018 (-1.46)	-0.0021 (-1.58)	0.0012 (1.20)	0.0007 (0.70)
ln(Turnover)	0.0004 (0.49)	0.0003 (0.64)	-0.0009 (-0.99)	-0.0010 (-0.94)	-0.0012 * (-1.68)	0.0002 (0.30)	-0.0005 (-0.77)	-0.0005 (-0.76)	0.0001 (0.14)
R^2	**0.061**	**0.066**	**0.070**	**0.180**	**0.077**	**0.084**	**0.073**	**0.065**	**0.073**
Stated: coefficients, t-stat. in parenthesis Observations: 7,694 Clusters: 139 Dummies: industry, firm, year Alpha: *0.1 **0.05 ***0.01									

The results are presented in Table 2: it shows clustered regression coefficients, robust t-statistics and R^2 statistics for each of the regressions. It can be observed that the relation between returns and announcement Tonality peaks on the announcement day and remains significant on day one (for alpha level 0.01). However, a major predictability uplift based on R^2 can only be witnessed on the announcement day, all other days remain relatively flat. It is interesting to observe that the relation between abnormal returns and Tonality starts to evolve already one day before publication. This effect indicates a certain anticipation of the announcement content, even though the R^2 does not increase. Tonality remains strongly significant on trading day one, indicating a slight underreaction on the announcement day, as documented by Jegadeesh and Wu (2011). However, the Tonality t-statistic returns to insignificance on trading day two. Interestingly, it can be noted that the trading days two and three do not appear to follow prior announcement predictability patterns: all control variables forfeit their significance, which seems to gradually return on trading day four and five. After trading day one, Tonality remains largely insignificant (alpha level 0.01).

These findings support the hypothesis that investors rapidly react to novel and relevant information and nearly completely absorb its potential on the announcement day and shortly thereafter. The trading days two and three appear to be announcement resets. As neither of our variables, announcement Tonality and traditional prediction factors, are able to provide an explanation, it allows speculation whether these days are driven by compensating for immediate and short-term over- and underreactions. However, this effect seems to wear-off on trading day four and five.

Information sustainability

Investors and consequently stock prices may react fast, but it does not explain how long novel information influences stock price development over time. To analyze information sustainability, we perform a separate set of regressions. Each regression analyzes the relation between announcement content and stock price development over different periods after the announcement. In contrast to the regression above, we take cumulative abnormal returns (*CAR*) for the following periods: 1, 5, 10, 20, 30 and 40

[5] Dummy variable reasoning: since Tonality reflects value relevant information sentiment, we expect Tonality to fluctuate with economic cycle. Moreover we presume a structural relation between Tonality and industry as well as disclosing firm. As shown by Engelberg (2008), disclosure readability can depend on the characteristics of the underlying industry and/or firm.

[6] All variance inflation factors below 1.50.

trading days (e.g. $CAR_{-1,40}$). As starting point for cumulative returns we use trading day T_{-1} immediately preceding the announcement, thus, $CAR_{-1,0}$ is equivalent to AR_0 in the investor timing analysis above.

The regression results are presented in Table 3: following the Tonality t-statistic, all periods of cumulative returns are significantly impacted by announcement Tonality. However, its impact gradually decreases over the entire period as the robustness of predictability slowly deteriorates. It is interesting to observe that the increased R^2 statistic on the announcement day ($CAR_{-1,0}$) sharply drops in the next period ($CAR_{-1,5}$) and continuously increases over the next 40 trading days. Our interpretation is that during the day of the announcement the stock price movement is primarily driven by the information effect of the announcement, whereas the longer we advance in event time the more typical explanatory factors, expressed as our control variables, influence stock prices. It is thus not astonishing that the development of the control factors support this interpretation: the t-statistics of all control variables (except turnover) monotonically increase from trading day 5 to 40. This is in line with the findings of Jegadeesh and Titman (1993) for momentum factors (*Alpha* and $CAR_{-29,-4}$) and Fama and French (1992) for value and size factors. The authors report structural relations between these factors and medium to long term stock price development over several months.

Table 3 – Information Sustainability

Variable	$CAR_{-1,0}$	$CAR_{-1,5}$	$CAR_{-1,10}$	$CAR_{-1,20}$	$CAR_{-1,30}$	$CAR_{-1,40}$
Announcement content variable						
Tonality	1.6107 *** (21.13)	1.9020 *** (16.76)	1.8843 *** (12.91)	1.8391 *** (10.06)	2.0193 *** (8.88)	1.8670 *** (8.02)
Control variables						
$CAR_{-29,-4}$	-0.0187 *** (-2.97)	-0.0253 ** (-2.02)	-0.0331 ** (-2.16)	-0.0595 *** (-3.07)	-0.0974 *** (-3.52)	-0.1178 *** (-3.90)
Alpha	-3.0068 *** (-5.18)	-6.8180 *** (-6.84)	-12.1564 *** (-8.88)	-24.8995 *** (-12.71)	-37.3318 *** (-12.76)	-48.2875 *** (-15.67)
ln(MV)	-0.0121 *** (-5.22)	-0.0113 *** (-3.41)	-0.0185 *** (-4.22)	-0.0392 *** (-6.74)	-0.0628 *** (-8.25)	-0.0768 *** (-8.88)
ln(MTBV)	0.0032 (1.34)	0.0024 (0.74)	0.0073 (1.59)	0.0162 ** (2.60)	0.0291 *** (4.40)	0.0367 *** (5.09)
ln(Turnover)	-0.0010 (-0.94)	-0.0025 (-1.60)	-0.0024 (1.19)	0.0004 (0.15)	-0.0012 (-0.37)	-0.0055 (-1.52)
R^2	**0.180**	**0.132**	**0.138**	**0.178**	**0.215**	**0.271**
Stated: coefficients, t-stat. in parenthesis	Observations: 7,694	Clusters: 139	Dummies: industry, firm, year		Alpha: *0.1 **0.05 ***0.01	

The gradual deterioration of the relation between announcement content and stock price development underscores the importance of news novelty. It demonstrates that new information starts to degrade in stock prices with increasing age. It also shows that there is no general underreaction through the first two months in event time, as described by Bernard and Thomas (1989) and Jegadeesh and Wu (2011) who found post-announcement drifts after earnings surprises and other important announcement types. Since underreaction is defined as post-announcement abnormal returns of the same sign as announcement date returns, we would not expect a decreasing predictability of Tonality for returns over time.

Analyst timing

Having found that novel information in corporate announcements experiences a fast reflection in stock price movements, but gradually deteriorates over the two following months, we investigate how long it takes, until analysts revise their forecasts for the novel information. In a third clustered OLS regression analysis we assess how the dependency between analysts' EPS forecast revisions and announcement Tonality evolves over event time. We analyze the same event time window as above. Consequently, we successively regress Tonality against different period lengths of forecast revisions[7]: 5, 10, 20, 30 and 40 trading days following the announcement. Again, we include a set of control variables as elaborated above. Additionally, we include analyst coverage (*ACOV*) based on Krishna and Morgan (2004) who

[7] Compare equation (4).

describe an improvement in informational content when an increasing number of experts follows a stock. The regression results are presented in Table 4.

There are two interesting observations: first, the Tonality t-statistics and coefficients monotonically increase over the 40-day period. The relation between announcement content (i.e. Tonality) and forecast revisions reaches statistical significance around 20 trading days following the announcement (for alpha level 0.01), which is noticeably slower than for stock prices. Second, the R^2 statistics similarly increase gradually over the same period.

Table 4 – Analyst Timing

Variable	$Revision_5$	$Revision_{10}$	$Revision_{20}$	$Revision_{30}$	$Revision_{40}$
Announcement content variable					
Tonality	0.0940 (0.28)	0.3091 (0.85)	0.9511 ** (2.05)	1.4985 *** (2.86)	1.8219 *** (3.26)
Control variables					
$CAR_{-29,-4}$	0.1074 ** (2.36)	0.1337 ** (2.44)	0.1951 *** (2.80)	0.2051 ** (2.41)	0.3262 *** (4.14)
Alpha	16.4832 *** (3.05)	18.2961 *** (3.10)	19.9948 *** (2.94)	22.6457 ** (2.57)	32.8376 *** (3.77)
ln(MV)	-0.0373 (-1.54)	-0.0451 (-1.66)	-0.0661 ** (-2.29)	-0.1113 *** (-2.67)	-0.1252 *** (-2.77)
ln(MTBV)	0.0535 ** (2.12)	0.0695 ** (2.48)	0.1085 *** (3.40)	0.1584 *** (3.45)	0.1695 *** (3.39)
ln(Turnover)	-0.0068 (-0.93)	-0.0037 (-0.44)	0.0116 (1.18)	0.0266 * (1.91)	0.0268 * (1.85)
ACOV	0.0020 ** (2.32)	0.0018 * (1.97)	0.0014 (1.30)	0.0036 ** (2.33)	0.0053 *** (3.41)
R^2	**0.174**	**0.173**	**0.219**	**0.250**	**0.283**

Stated: coefficients, t-stat. in parenthesis Observations: 921 Clusters: 136 Dummies: industry, firm, year Alpha: *0.1 **0.05 ***0.01

It should be noted that next to Tonality the variables for firm-size and market-to-book gain significance over time as well. This is very similar to the significance gain of these factors in cumulative returns in the stock price experiment (review Table 3). As discussed, these factors have been identified to add predictability for expected abnormal stock price returns[8] and appear to have a similar impact on forecast revisions. This observation is interesting as we find three model factors simultaneously gaining significance over time. A correlation analysis reveals that there is a positive relation between the market-to-book ratio and Tonality, but no relation between size and Tonality. We find this to be intuitive: high value firms can be seen as more successful, and more successful firms in the past are more likely to report positive news in the future, as long as their prosperity stems from a structural difference to their competition. Not so with the firm-size factor, larger or smaller firms are not more likely to publish important positive or negative news. Thus, we do not find any correlation between these two factors. All found correlations between the regression variables do not bias the analysis[9]. To sum it up, analysts appear to respond to novel information noticeably later than investors.

Analyst value-add

Comparing the experiments for investor and analyst timing shows that investors incorporate novel information into stock prices noticeably faster than analysts into their forecast revisions. This is even more interesting when considering the findings of the information sustainability analysis. The relation between announcement content and stock prices starts to deteriorate during the same time frame as analysts begin to respond to novel information. To illustrate this, we plot the relation of stock price development and forecast revisions against Tonality over time (*c.f.* Figure 1). The graph shows Tonality t-

[8] See Fama and French (1992).

[9] As described above, the test for multicollinearity did not indicate any confounding interactions between regression factors that bias the analysis.

statistics for forecast revisions and cumulative abnormal returns over time (as presented in table 3 and 4). The t-statistic here serves as proxy for explanatory power of stock prices (investor interpretation) and forecast revisions (analyst interpretation).

As described above, the relation between cumulative returns and Tonality slightly deteriorates over time, but clearly remains significant. In contrast, the relation between Tonality and forecast revisions gradually increases over the same period and reaches significance after trading day 20. This suggests that analysts start to respond to novel information while the information effect already fades out of stock prices. However, if the information effect of important public corporate announcements was already digested by investors and reflected in stock prices, what is the value of updated analyst forecasts during that time period? According to Ivkovic and Jegadeesh (2004) who analyze earnings announcements "the market price response to revisions in earnings forecasts and recommendations is the weakest in the period immediately after earnings announcements. These findings indicate that, although analysts' interpretation of the information in quarterly earnings announcements conveys useful information, it is not the dominant source of analysts' value. Indeed, the value of analysts' earnings forecasts and recommendations stems more from their independent collection of information than from their interpretation of public information".

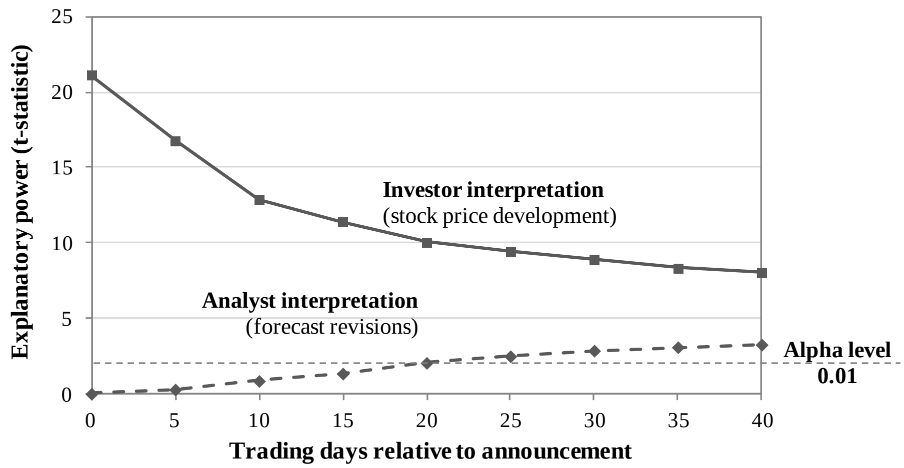

Figure 1. Explanatory power of announcement content for investor and analyst interpretation over time

These observations raise the question whether analysts may provide an informative value-add during periods of important corporate announcements. Employing sentiment analysis allows to dig deeper into this question. If research analysts not just add value from processing private, and thus costly information, but also through their ability to effectively synthesize and interpret public information, we should be able to find an indication through the following experimental setup: if analysts discover additional informative value in announcements that investors (partly) disregard, we still find a significant relation between announcement content and forecast revisions, when controlling for the stock price development after the announcement. Thus, the regression setup requires to include a control factor for the stock price reaction following the announcement[10]. Since the relation between stock prices and Tonality deteriorates over time, as shown above, we test cumulative abnormal returns over different periods in separate regressions.

Table 5 shows the regression results when controlling for different period lengths of stock price return. We find each of the control factors for return significantly related to forecast revisions, indicating that there is some information redundancy between stock price development and forecast revisions. However, Tonality, as expression of the announcement content, remains significant in the presence of each of the

[10] Although Tonality is related to return (see Table 10), tests for multicollinearity did not indicate a bias risk for the regression.

return controls. We interpret this observation as an indicator for additional information translated into forecast revisions, that are not covered by stock prices following the announcement.

Table 5 – Analyst Value-Add

Variable	$Revision_{40}$	$Revision_{40}$	$Revision_{40}$	$Revision_{40}$	$Revision_{40}$
Announcement content variable					
Tonality	1.5037 *** (2.66)	1.5062*** (2.68)	1.6245 *** (2.86)	1.6360 *** (2.93)	1.7070 *** (3.09)
Stock price control variable					
$CAR_{-1,5}$	0.4320 *** (3.43)				
$CAR_{-1,10}$		0.4191*** (3.94)			
$CAR_{-1,20}$			0.2337 ** (2.62)		
$CAR_{-1,30}$				0.2076 *** (2.72)	
$CAR_{-1,40}$					0.1467 ** (2.46)
Control variables					
$CAR_{-29,-4}$	0.3376 *** (4.34)	0.3450*** (4.53)	0.3403 *** (4.38)	0.3446 *** (4.40)	0.3413 *** (4.27)
Alpha	37.1631 *** (4.13)	39.0330*** (4.28)	39.3362 *** (4.04)	40.8133 *** (4.12)	40.7877 *** (3.97)
ln(MV)	-0.1174 ** (-2.61)	-0.1114** (-2.45)	-0.1113 ** (-2.40)	-0.1078 ** (-2.32)	-0.1110 ** (-2.40)
ln(MTBV)	0.1577 *** (3.14)	0.1525*** (3.00)	0.1554 *** (3.01)	0.1558 *** (3.05)	0.1576 *** (3.07)
ln(Turnover)	0.0247 (1.66)	0.0260* (1.77)	0.0266 * (1.82)	0.0262 * (1.79)	0.0280 * (1.92)
ACOV	0.0050 *** (3.12)	0.0050*** (3.10)	0.0052 *** (3.32)	0.0051 *** (3.25)	0.0051 *** (3.29)
R^2	**0.294**	**0.299**	**0.291**	**0.292**	**0.289**
Stated: coefficients, t-stat. in parenthesis Observations: 921 Clusters: 136 Dummies: industry, firm, year Alpha: *0.1 **0.05 ***0.01					

If analysts (at least partly) focus their announcement interpretation on pieces of information that investors generally disregard in the text, a word list comparison between analysts and investors should also show differences. As the Tonality approach identifies the most informative words based on observed effects, the resulting word lists for analysts and investors should deviate: to further analyze this, we specifically design a Tonality word list, as described in section 3, for our forecast revision data sample, hereafter "analyst word list". We want to compare the analyst word list with the list we designed to predict stock price development ("investor word list"). Besides seminal deviations that stem from the different news compositions[11] it is apparent that research analysts focus on different parts of the text than rapidly reacting investors: investors appear to be sensitive to words that rather express the interpretation of the author instead of genuinely stated facts, such as in "*sluggish* corporate performance", "*slump* orders" or "*disappointing* results". As opposed to analysts, who seem to appreciate facts-driven communication, expressed by words as in "revenues *fell*", "EBIT *decrease*" or "results *below* expectations" (Table 6). The investor word list also includes most of these terms, but they receive lower absolute Tonality. The word list comparison does not prove structural differences in the interpretation process of trading investors and analysts, but it provides an indication which words are more relevant for one side or the other.

[11] Since word Tonalities are a fraction divided by the total number of positive and negative announcements, large data sets tend to be less volatile and can contain more words in a final word list.

Table 6 – Investor Word list vs. Analyst Word list – Top Negative Word Stems					
Investor word list*			**Analyst word list***		
No.	**Word**	**Word-Tonality**	**No.**	**Word**	**Word-Tonality**
1	unexpectedli	-1.19	1	fell	-1.14
2	sluggish	-1.18	2	longer	-1.03
3	shortfal	-1.08	3	commit	-1.00
4	warn	-1.07	4	cut	-0.88
5	uncertain	-1.01	5	environ	-0.86
6	slump	-1.00	6	decreas	-0.73
7	downward	-0.99	7	neg	-0.61
8	departur	-0.94	8	below	-0.60
9	disappoint	-0.92	9	restructur	-0.59
10	reluct	-0.88	10	declin	-0.56
*Note: words are reduced to their word stem and thus do not always appear in their original form (e.g. "unexpectedli")					

As a conclusion on these findings, we can state that there are considerable differences between investor and analyst announcement interpretations. They differ in interpretation timing and content focus. In the next subsection we analyze how the measure for announcement complexity helps to understand the differences in interpretation within the groups of investors and analysts.

Difference in opinion

According to the *information processing cost hypothesis*, more complex information in disclosures increases the processing cost for their receivers (Grossman and Stiglitz 1980; Bloomfield 2002). Lehavy, Li, and Merldey (20ll) find that less readable annual reports are associated with greater dispersion in analyst earnings forecasts. Besides analysts, the information processing of investors is also influenced by the readability of corporate disclosures: in our data set, we observe strongly increased trading volumes on the day of the announcement. According to Karpoff (1986), trading volumes are not only driven by strongly changing returns (volume-return relation), but also by difference in opinion of investors. We assume that the probability for investor disagreement during the interpretation of corporate announcements increases with increasing processing cost of the announcements.

Table 7 – Difference in Interpretation		
Variable	**$Volume_0$**	**$Dispersion_{40}$**
Announcement content variables		
Complexity	0.3706 *** (4.05)	0.1326 ** (2.26)
Tonality	-3.9002 (-0.84)	-3.3290 (-0.35)
Control variables		
abs(Return)	6.3360 *** (5.08)	0.3030 (0.44)
Alpha	-55.3420 ** (-2.06)	-30.6190 (-0.82)
ln(MV)	-0.3942 *** (-3.12)	-0.3227 (-1.19)
ln(MTBV)	0.1736 (1.45)	0.3177 (0.98)
ln(Turnover)	-0.0914 (-1.43)	-0.0339 (-0.51)
R^2	**0.459**	**0.189**
Observations / clusters	7,694 / 139	921 / 136
Stated: coefficients, t-stat. in parenthesis Dummies: industry, firm, year Alpha: *0.1 **0.05 ***0.01		

To analyze whether increasing processing cost influences investor and analyst dispersion, we presume that processing cost depends on announcement complexity. Miller (2010) and other authors measure announcement complexity simply by taking the length of a disclosure. We argue that announcement complexity rather depends on the amount of informative text components and not necessarily the pure length. Therefore, we measure announcement Complexity based on the number of informative words (see section 3).

In two separate regressions we analyze whether Complexity drives investor and analyst dispersion: following Karpoff (1986) we measure investor disagreement through abnormal trading volume on the announcement day ($Volume_0$), while controlling for (absolute) returns. To measure analyst dispersion, we take the change in standard deviation of the analyst consensus from one day before to 40 trading days after the announcement ($Dispersion_{40}$). As before, we include a set of relevant control variables for firm-size and performance and additional controls for firm, year, and industry. Table 7 shows the regression results: it can be inferred that both investor and analyst dispersion are significantly influenced by announcement complexity. However, analyst dispersion seems to be less influenced by announcement complexity. The announcement content represented by Tonality does not influence dispersion. As expected, stock price returns are also strongly related to investor dispersion, but it does not dilute the influence of Complexity.

Conclusions

The advent of the Internet has resulted in an increasingly growing body of information. Most of this information is of qualitative nature, which contains essential facts that are, however, difficult to decode. Electronic markets and also e-commerce on the whole seem to perform very well in terms of their information processing capability achieving an efficient resource allocation. Currently, the information processing capabilities of human agents facing qualitative news is mostly unknown. More precisely it is information inequality, decision maker subjectiveness and increased processing cost for qualitative information that impede decision efficiency. Accordingly, it is crucial to understand how different decision makers process qualitative information and how it contributes to their actions.

In this paper we show that sentiment analysis facilitates research in information processing of qualitative information. We use a capital market example to demonstrate how investors and financial analysts perceive and react to novel information. We find that their interpretation and reaction to the same information is quite different from one another: investors react rapidly and translate novel information into stock prices on the day of the announcement, whereas analysts take more time to respond to novel information. But their delayed information processing is not obsolete, even if novel information is translated into prices rapidly. It turns out that financial analysts emphasize different parts of novel information than investors. Additionally, we find that investors increasingly disagree with increasing complexity, whereas analysts are less put-off by more complex information. Another key advantage of the approach is that scholars can use it to analyze the information processing of readers only in the presence of qualitative aspects and information. This is novel in information systems research.

We deliberately chose data from capital markets to demonstrate the benefits of sentiment analysis for information processing research, as capital markets provides a rich data set, different types of decision makers, and a comprehensive research body on information processing to benchmark our results. Capital market prices constitute a proxy for the “wisdom of crowds”. Hence, we believe that we can apply the knowledge gained in this analysis, be it the methodology or be it the results, also to other types of electronic markets and the e-commerce industry where market participants react upon textual information. Presumably, we can apply the Tonality approach also to analyze the effect of textual recommendation on the buying behavior of consumers. We strongly believe that our methodology opens up many intriguing avenues for future research in all e-commerce areas where textual messages or news play a role.

References

Antweiler, W., and Frank, M. 2004. "Is all that talk just noise? The information content of internet stock message boards", *The Journal of Finance* 59, 1259-1294.

Atiase, R. K. 1985. "Predisclosure Information, Firm Capitalization, and Security Price Behavior Around Earnings Announcements", *Journal of Accounting Research* 23, 21- 36.

Ball, R., and Brown, P. 1968. "An Empirical Evaluation of Accounting Income Numbers", Journal of Accounting Research, 6, pp. 159-178.

Bernard, V., and Thomas, J. 1989. "Post-earnings announcement drift: Delayed price response or risk premium?", *Journal of Accounting Research*, Suppl. 27, 1-36.

Bloomfield, R. J. 2002. "The 'incomplete revelation hypothesis' and financial reporting", *Accounting Horizons* 16, 233-243.

Chan, L.K.C., Jegadeesh, N., and Lakonishok, J. 1996. "Momentum strategies", *The Journal of Finance* 51, 1681–1713.

Chevalier, J., Mayzlin, D. 2006. "The effect of word of mouth online: Online book reviews". *Journal of Marketing Research* 43(3), 345-354.

Dellarocas, C., Zhang, M., and Awad, N. F. 2007. "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures", *Journal of Interactive Marketing* Volume 21 (4), Fall 2007, 23-45.

Duan, W., Gu, B., and Whinston, A. 2005. "Do online reviews matter? - an empirical investigation of panel data". *Decision Support Systems* Volume 45(4), 1007–1016.

Engelberg, J. 2008. "Costly information processing: Evidence from earnings announcements", *Working paper available on SSRN*, Northwestern University.

Fama, E. F., 1970. "Efficient capital markets: a review of theory and empirical work", *The Journal of Finance* 25, 383 -417.

Fama, E. F., and French, K.R. 1992. "The cross-section of expected stock returns", *The Journal of Finance* 47, 427–465.

Gleason, C. A., and Lee, C. M. C. 2003 "Analyst forecast revisions and market price discovery", *The Accounting Review*, 78, 193-225.

Granados, N., Gupta A., Kauffman R. J. 2010. „ Information Transparency in Business-to-Consumer Markets: Concepts, Framework and Research Agenda. *Information Systems Research*, 21, 2, June 2010, 207-226.

Grossman, S.J. and Stiglitz, J. E. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review*, 70, pp. 393-408.

Hagenau, M., Liebmann, M., Hedwig, M., and Neumann, Dirk. 2012. „Automated News Reading: Automated news reading: Stock Price Prediction based on Financial News Using Context-Specific Features", *Proceedings of the 45th Hawaii International Conference on System Sciences*, Waleia, Hawaii, January 2012.

Hirshleifer, D., Lim, S. S., and Teoh, S. H. 2009. "Driven to Distraction: Extraneous Events and Underreaction to Earnings News". *The Journal of Finance* 64, 5, 2289–2325 (2009).

Ivkovic, Z., and Jegadeesh, N. 2004. "The Timing and Value of Forecast and Recommendation Revisions", *Journal of Financial Economics* 73, 433-463.

Jegadeesh, N., and Titman, S. 1993. "Returns to buying winners and selling losers: Implications for stock market efficiency", *The Journal of Finance* 48, 65–91.

Jegadeesh, N., and Wu, A. D. 2012. Word Power: A New Approach for Content Analysis (November 1, 2011). *AFA 2012 Chicago Meetings Paper*.

Karpoff, J. 1986. "A Theory of Trading Volume", *Journal of Finance* 41,1069–1087.

Kandel, E., and Pearson, N. D. 1995. "Differential Interpretation of Public Signals and Trade in Speculative Markets", *Journal of Political Economy* 103, 831–872.

Kim, O., and Verrecchia, R. E. 1991. "Trading Volume and Price Reactions to Public Announcements". *Journal of Accounting Research*. Volume 29. pp. 302-321.

Kothari, S.P. 2001. "Capital market research in accounting", *Journal of Accounting and Economics* 31, 105-231.

Krishna, V., and Morgan, J. 2004. "The art of conversation: Eliciting information from experts through multi-stage communication", *Journal of economic theory* 177, 147–179.

Lehavy. R., Li, F., and Merkley, K. 2011. "The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecast", *The Accounting Review* 86(3), 1087–1115

Li, F. 2010a "Textual Analysis of Corporate Disclosures: A Survey of the Literature", *Journal of Accounting Literature* 29, 143-165.

Li, F. 2010b, "The Information Content of Forward-Looking Stements in Corporate Filings - A Naiive Bayesian Machine Learning Approach", *Journal of Accounting Research* 48, 49-102.

Liebmann, M., Hagenau, M., Häussler, M., and Neumann, D. 2011. Effects Behind Words: Quantifying Qualitative Information in Corporate Announcements, *working paper presented at Financial Management Association conference*, Porto, Portugal.

Liu, Y. 2006. "Word-of-mouth for movies: Its dynamics and impact on box office receipts?", *Journal of Marketing*, 70, 74-89.

Loughran, T., and McDonald, B. 2011. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks", *Journal of Finance* 66, 35-65.

MacKinlay, C. 1997. "Event Studies in Economics and Finance", *Journal of Economic Literature,* 35, 13-39.

Miller, B. 2010. "The Effects of Reporting Complexity on Small and Large Investor Trading", *The Accounting Review*.

Reinstein, D., and Snyder, C. 2005. „The influence of expert reviews on consumer demand for experience goods: A case study of movie critics", *Journal of Industrial Economics* 53(1): 27-51.

Röder, K. 2000. "Die Informationswirkung von Ad hoc-Meldungen", *Zeitschrift für Betriebswirtschaft* 70, 567 – 593.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. 2008. "More than words: Quantifying Language to Measure Firms' Fundamentals", *The Journal of Finance* 63, 1437-1467.

Varian, H. R. 1986. "Differences of opinion in financial markets", *Financial Risk: Theory, Evidence, and Implications*, Proceedings of the 11th Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis (Kluver Academic Publishers, Boston MA).

You, H., and Zhang, X. 2009. "Financial reporting complexity and investor underreaction to 10-K information", *Review of Accounting Studies* 14 (4) 559-586.