# Decision Analytics for Initial Public Offerings: How Filing Sentiment Influences Stock Market Returns



Jonas Krinitz and Dirk Neumann

Abstract Companies issuing stocks through an initial public offering (IPO) are obligated to publish relevant information as part of a prospectus. Besides quantitative figures from accounting, this document also contains qualitative information in the form of text. In this chapter, we analyze how sentiment in the prospectus influences future stock returns. In addition, we investigate the impact of pre-IPO sentiment in financial announcements on first-day returns. The results of our empirical analyses using 572 IPOs from US companies suggest a negative link between words linked to uncertainty and future stock market returns for up to 10 trading days. Conversely, we find that uncertainty expressed in pre-IPO announcements is positively linked to first-day stock returns. These insights have implications for research on IPOs by demonstrating that future stock returns are also driven by textual information from the prospectus and assist investors in placing their orders.

#### 1 Introduction

Companies are increasingly turning to stock markets to raise expansion capital to monetize the investments of early private investors. An *initial public offering* (IPO) refers to a type of public offering where shares of stock in a company are sold to the general public, for the first time, on a securities exchange. As a result of regulatory publication requirements, firms must file a prospectus with, e.g., the Securities and Exchange Commission (SEC) in the USA prior to its initial public stock offering (c.f. Kenney and Patton 2013a, Ritter and Welch 2002). Details of the proposed offering are disclosed to potential purchasers in the form of a detailed document known as a *prospectus*. This prospectus contains key figures from financial accounting, as well as information on current and future business

J. Krinitz · D. Neumann (⋈)

models, potential opportunities, and risks threatening the business. The prospectus is used as an instrument to remove or at least alleviate disadvantages stemming from information asymmetries among different investor groups. Obviously, the prospectus plays a major role in the formation of IPO prices, which ultimately determines the amount of capital that accrues from the IPO. Thus, all involved stakeholders in the IPO may need to understand how the prospectus affects the final IPO prices. While the underwriting company has a natural incentive to cast a positive light on their company, investors have a need to decipher the text messages for relevant signals on the future performance of the company.

The theory of market microstructure may fill the bill in explaining the relationship between the prospectus and final IPO price as it refers to the branch of finance that is concerned with the details of how exchange occurs in markets. More specifically, market microstructure theory epitomizes the study of the process and outcomes of exchanging assets under a specific set of rules (O'Hara 1995). One crucial aspect refers to the role of information and how prices reflect them. Interestingly, market microstructure theory is still in its infancy, dealing with information which consists of text messages, such as the prospectus. The theory is, thus, nearly silent in explaining how text messages are processed into prices (c.f. Jegadeesh and Wu 2013, Loughran and McDonald 2011, Tetlock et al. 2008).

This is where *decision analytics*, as a subfield of Information Systems (IS) research, comes into play (Chen et al. 2012) in order to contribute to the understanding of how human agents process information when facing textual news. Besides quantitative information, such as quarterly earnings, the textual content obviously also provides valuable insights. Typically, the sentiment of financial news is utilized to analyze how information is processed (Siering 2013) as investors and analysts perceive novel information differently. Liebmann et al. (2012) come to the conclusion that investors rapidly translate novel information into transactions, whereas analysts wait to respond. Similarly, commodity markets seem to be mostly driven by negative news (Feuerriegel and Neumann 2013, Feuerriegel et al. 2014). Hence, information processing has been subject to many decision analytics research publications since problems span several dimensions, namely, volume, velocity, variety, and veracity (IBM 2013). Thus, these research questions are placed at the intersection between behavioral finance, Information Systems, and decision analytics, which has been receiving great attention lately (Chen et al. 2012).

Altogether, information from an IPO prospectus provides the basis for the decision-making process for potential investors. Thus, it is logical consequence to study the *information processing* of investors facing IPO filings in the context of efficient electronic markets. Though IPO filings consist of a few quantitative facts, a large proportion accounts for qualitative information. While prior research focuses on quantitative facts, such as the number of offered shares, knowledge of the textual components is rare. Hence, this paper examines the statistical relationship between sentiment in IPO filings and the subsequent stock market performance.

As a main contribution to Information Systems research, this chapter sheds some light on information processing along IPO filings. The soft data contained in these filings conveys insights and potentially valuable information that is absent in traditional quantitative numbers. A recent study by Ferris et al. (2013) points out that the extent to which management is confident about the success of its issue and the implications that such beliefs have on IPO pricing are largely ignored in the literature. To close this gap, individual research questions addressing the content of IPO prospectuses, as well as contributions, are as follows.

Confirmatory Analysis: To what extent are the stock market returns of firms driven by the sentiment of their IPO filings?

As a confirmatory analysis of previous research (Ferris et al. 2013, Loughran and McDonald 2013), we investigate how investors react to IPO prospectuses and analyze information processing empirically by exploiting sentiment analysis. We find that sentiment in IPO prospectuses correlates with first-day return. In fact, one standard deviation increase in the sentiment measure is negatively linked to an increase in first-day returns by an economically significant 10.56%.

Research Question 1: How persistent is the influence of filing tone on long-term returns?

According to our results, sentiment is not only positively correlated with returns in the short term, but the polarity of filings also affects the performance in the long term.

Research Question 2: How is the IPO performance affected by pre-IPO news sentiment?

The confirmatory analysis proves that IPO filings have an effect on the subsequent stock market returns. However, the media coverage of firms going public occurs in more depth, since journalists also report on the progress of the IPO, the previous performance of the firm, and the general outlook. Because of these reasons, we investigate the impact of pre-IPO news announcements on first-day returns and, finally, compare the influence of sentiment on IPO filings and news announcement in terms of strength.

The remainder of this chapter is structured as follows. In Sect. 2, we combine previous research on both initial public offerings and approaches for sentiment analysis in order to study literature on their intersection, which deals with the information processing theory of IPO filings. To gauge the sentiment of these IPO filings, we detail our research model in Sect. 3. Using this research model, Sect. 4 analyzes how the sentiment in an IPO prospectus is linked to the corresponding stock market performance and how information processing is influenced by pre-IPO news sentiment. Section 5 concludes the chapter with a summary and an outlook on future research.

#### 2 Related Work

In this section, we present related literature grouped into two categories. First, we revisit literature covering different types of IPO filings. Second, we compare approaches for financial documents to measure their sentiment. Third, we review previous work on information processing of investors when facing IPO filings. All

in all, the following references provide evidence that investigating the effect of sentiment on IPO filings across different time spans, as well as the impact of pre-IPO news sentiment, is both a novel and relevant research question to the Information Systems community.

To a large extent, market efficiency relies upon the availability of information (Fama 1965). Access to market information is promoted with ease in electronic markets, and because of straightforward access, decision-makers (i.e., consumers, suppliers, and intermediaries) can use more detailed information to make purchases and sales more beneficial (e.g., Granados et al. 2010). In fact, it is both native and crucial to IS research regarding how decision-makers process and act upon (qualitative) information in IPO filings. Though information processing has been extensively studied in capital markets, this is not the case when it comes to literature focusing on IPO filings: "An examination of the textual or soft information contained in prospectuses is less common" (Ferris et al. 2013).

### 2.1 IPO Filings

This section introduces the basic filing types required by US firms going public. For US firms, one of the first steps in going public is filing a Form S-1 on the Securities and Exchange Commission's (SEC) Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. Typically, the S-1 offers investors their first detailed glimpse of a firm's business model and financial statements. It is a pre-effective registration statement submitted when a company decides to go public. Hence, an investor can use its content to evaluate potential investment opportunity. If the S-1 is followed by a successful initial public offering, firms file a final prospectus (so-called Form 424), which is published via EDGAR on the day of or a few days following the IPO. Loughran and McDonald (2013) report a median time lag of 88 calendar days between S-1 and Form 424. According to the authors, more calendar time between the S-1 and 424 filings (the final IPO prospectus) is associated with significantly lower first-day returns. Companies with more problematic S-1 filings or facing adverse market conditions are often delayed in issuing stock. Some of the delay in going public could be due to the many days of management needed to properly respond to SEC concerns about the IPO document or a sharp decline in IPO market conditions. However, the overall tone of the S-1 is typically very similar to the tone of the final IPO prospectus. In fact, changes in tone between the S-1 and 424 filings seem to be unrelated to positive revisions in the offer price (Loughran and McDonald 2013).

### 2.2 Sentiment Analysis in the Finance Discipline

Methods that use the textual representation of documents to measure the positivity and negativity of the content are referred to as opinion mining or *sentiment analysis*. In fact, sentiment analysis can be utilized to extract subjective information from text sources, as well as to measure how market participants perceive and react upon financial materials. Here, one uses the observed price reactions following the financial text to validate the accuracy of the sentiment analysis routines. Based upon sentiment measures, one can study the relationship between financial documents and their effect on markets. For example, empirical evidence shows that a discernible relation between news content and its stock market reaction exists (e.g., Antweiler and Frank 2004, Tetlock 2007, among the first).

As sentiment analysis is applied to a broad variety of domains and text sources, research has devised various approaches to measure sentiment. For example, Pang and Lee (2008) provide a comprehensive domain-independent survey. Within finance, recent literature surveys (Minev et al. 2012, Mittermayer and Knolmayer 2006b) compare studies aimed at stock market prediction. For example, dictionarybased approaches are very frequently used in recent financial text mining research (cp. Demers and Vega 2010, Henry 2008, Jegadeesh and Wu 2013, Loughran and McDonald 2011, Tetlock et al. 2008). These methods count the frequency of predefined positive and negative words from a given dictionary—producing results that are straightforward and reliable. Similarly, term-weighting methods also rely upon positive and negative words, but these are classified based on word frequencies in a training set (i.e., the prior probabilities) to assign a weight to each term (Liebmann et al. 2012). Machine learning approaches (e.g., Antweiler and Frank 2004, Li 2010, Mittermayer and Knolmayer 2006a, Schumaker and Chen 2009) offer a broad range of methods but may suffer from overfitting (Sharma and Dey 2012).

In our research framework, we have only a few hundred IPO filings, each linked to a large text basis (i.e., a length of several thousand words) along with a continuous return value. Thus, we have experienced difficulties in achieving robust results with term-weighting as well as with machine learning approaches and focused, instead, on dictionary-based methods. In fact, we have tested all of the above dictionary-based metrics in combination with various dictionaries. In short, we find that dictionaries vary strongly in the strength of the link between sentiment and stock market reaction. Along with previous research on IPOs, the ratio of words labeled as *uncertainty* according to the Loughran and McDonald Financial Sentiment Dictionary (McDonald 2012) outperforms all other approaches.

## 2.3 Information Processing Theory of IPO Filings

This section provides an overview of related research investigating the information processing of IPO filings. First, we review other research studying the information processing of IPO, but without considering the polarity of the prospectus content.

- Arnold et al. (2010) examine the risk factors section of a prospectus. They not
  only count the number of risk factors disclosed in this section but also measure
  the number of words used to explain each of the risk factors. They find that the
  soft information is significantly related to both the initial and subsequent IPO
  returns, but not to ex post measures of investor sentiment.
- Hanley and Hoberg (2012) find that issuers trade off first-day returns and disclosure in amended filings (S-1/A) as a hedge against the risk of future law suits. The two authors show that IPOs having a higher risk of material omissions in their filings are more likely to hedge litigation exposure with higher levels of underpricing. A greater level of disclosure by issuers, as proxied by more meaningful revisions in their S-1/A filings during the bookbuilding period, lowers the probability of being sued by investors. The key assumption of Hanley and Hoberg (2012) is that an issuer with a large pre-IPO price adjustment (but having only revised their filings slightly) is likely to have a material omission in the prospectus as the issuer did not disclose the information underlying the price change.
- To evaluate disclosure style, Loughran and McDonald (2014) create a standardized statistic that aggregates a series of writing components specifically identified by the SEC. The six components are average sentence length, average word length, passive voice, legalese, personal pronouns, and negative/superfluous phrases. In the context of Form 424 filings, the authors find substantial differences in the content of IPO filings. This is a result of the 1998 rule implemented by the SEC requiring firms to use plain English in their prospectus filings.

Closest to our research are the approaches by Ferris et al. (2013), Hanley and Hoberg (2010), and Loughran and McDonald (2013), in terms of both research objectives and methodology. We individually address each of these papers as follows:

- Ferris et al. (2013) examine the effect that conservative or cautionary language (measured using negative tone) in the prospectus might have on IPO performance. As a result, greater conservatism in the prospectus is related to increased underpricing. The authors observe that this relation is stronger for technology than non-technology IPOs. Besides that, they study how conservative language evolves across time. In addition, they find a significant relationship between conservative language from the Loughran-McDonald dictionary and an average industry-adjusted return on assets for the 3 years following the IPO, but give no insights on how returns evolve through time.
- Hanley and Hoberg (2010) examine the information content of IPO initial prospectuses and its effect on pricing. The authors split the information contained in the S-1 into standard and informative components. Interestingly, they find that

IPOs with more informative content in their S-1 have lower offer price revisions and first-day returns. They argue that IPOs with more standard S-1 content would be more likely to solicit information from investors during bookbuilding.

- Liu et al. (2014) explore the effects of pre-IPO media coverage (in the Factiva database) measured by the number of newspaper articles. According to the authors, this news count correlates positively with the stock's long-term value, liquidity, analyst coverage, institutional investor ownership, and first-day returns. Overall, the paper provides evidence for a long-term role for media coverage, consistent with Merton's attention or investor recognition hypothesis.
- Loughran and McDonald (2013) study the relationship between S-1, as well as Form 424 filings, and the IPO first trading day returns. As a result, IPOs with a more uncertain or weak modal tone experience higher first trading-day returns. Moreover, the percentages of uncertain, weak modal, and negative words in the S-1 are much more powerful variables in explaining levels of underpricing than many commonly used IPO control variables, such as venture capital dummy, top-tier underwriter dummy, or trailing annual sales. In fact, the *t*-statistics are higher when using Form 424 instead of S-1. In addition, the tone is positively linked with higher volatility in a 60-day period following the offering, but there is no similar evidence for stock market returns.

Hence, information processing has been the subject of many IS research publications. However, the above research papers concentrate on a partial examination of prospectus sentiment, and, thus, many questions are left unanswered. All listed publications lack (1) an in-depth evaluation of how persistent the influence of filing content on long-term returns is and (2) a comparison between the reception of IPO filings and pre-IPO news sentiment. Consequently, an empirical study analyzing the effects of filing sentiment in the long term and in pre-IPO news seems to be an open research question. Consequently, we pursue a rigorous IS approach: First, we utilize sentiment analysis to extract the polarity of IPO filings. Second, we specify a model to measure the impact of prospectus sentiment empirically.

## 3 Research Methodology: Sentiment Analysis of IPOs

This section introduces our research methodology as depicted in Fig. 1. In a first step, only those IPO prospectuses and news announcements are filtered that fit our research focus. Then, each prospectus and each announcement is subject to *preprocessing* steps which transform the running text into machine-readable tokens. The frequencies of these tokens are aggregated to compute the corresponding *sentiment*.

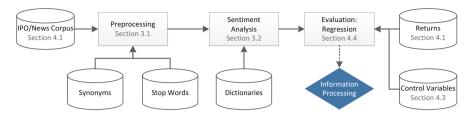


Fig. 1 Research model to study information processing using sentiment analysis of IPO filings

### 3.1 Preprocessing IPO Filings

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

- Cleaning. We use the textual component of IPO filings only, and then, all HTML tags and XBRL syntax are removed. This is consistent with Ferris et al. (2013), as well as Loughran and McDonald (2013).
- **Tokenization.** Each announcement is split into sentences and single words named *tokens* (Grefenstette and Tapanainen 1994).
- Negations. Negations invert the meaning of words and sentences. According to Loughran and McDonald (2011), as well as Tetlock (2007), handling negations is an inevitable prerequisite to measure positive tone. However, in all previous papers, handling negations is neglected. To close this gap, we adopt the following approach (Dadvar et al. 2011): When encountering the word *no*, each of the subsequent three words (i.e., the object) is counted as words from the opposite dictionary. When other negating terms are encountered (*rather*, *hardly*, *couldn't*, *wasn't*, *didn't*, *wouldn't*, *shouldn't*, *weren't*, *don't*, *doesn't*, *haven't*, *hasn't*, *won't*, *hadn't*, *never*), the meaning of all succeeding words is inverted.
- **Stop word removal.** Words without a deeper meaning, such as *the*, *is*, *of*, etc., are named *stop words* and, thus, can be removed. We use a list of 571 stop words (Lewis et al. 2004).
- Synonym merging. Synonyms, though spelled differently, convey the same meaning. Thus, approximately 150 frequent synonyms are grouped and aggregated by their meaning—a method referred to as *pseudoword* generation (Manning and Schütze 1999).
- **Stemming.** Stemming refers to the process of reducing inflected words to their stem (Manning and Schütze 1999). Here, we use the so-called Porter stemming algorithm.

### 3.2 Method for Sentiment Analysis

As shown in a recent empirical evaluation on the robustness of sentiment analysis (Feuerriegel and Neumann 2013), the correlation between sentiment in financial materials and the corresponding stock market reaction varies strongly across different sentiment metrics. Out of the potential dictionaries, the ratio of words labeled as *uncertain*—according to the Loughran and McDonald Financial Sentiment Dictionary (McDonald 2012)—achieves the highest robustness, and, consequently, we rely upon this approach in the following evaluation.

Let us briefly recapitulate this sentiment analysis approach. Let  $W_{\text{total}}(P)$  denote the total number of words in the prospectus P and  $W_{\text{uncertain}}(P)$  denote the number of *uncertainty* words in the prospectus P. Then, the sentiment is defined by

$$Sentiment(P) = \frac{W_{\text{uncertain}}(P)}{W_{\text{total}}(P)} \in [0, 1]. \tag{1}$$

Thus, the sentiment variable *Sentiment(P)* measures the number of *uncertainty* words normalized by the number of total words. We improve the metric, in contrast to Ferris et al. (2013), as well as Loughran and McDonald (2013), by using rules to invert the meaning when encountering negations. Thus, tokens following a negation are not counted as *uncertainty* words. This allows us to further boost the accuracy.

## 4 Empirical Evaluation: Analyzing Sentiment of IPO Filings

Having discussed the steps to compute sentiment values, we apply the sentiment analysis to investigate how stock market returns of initial public offerings are driven by their prospectus. Initially, we choose the IPO filing corpus and gain first insights by descriptive statistics. Then, we specify all control variables, which, ultimately, give the regression design that inspects information processing by linking returns with prospectus sentiment. By varying the time span of returns, we study the persistence of sentiment on stock market performance, and in addition to that, we analyze the reception of news covering initial public offerings.

## 4.1 IPO Filing Corpus, News Corpus, and Stock Market Data

A structured dataset of US initial public offerings is found in the so-called Kenney-Patton IPO database (Kenney and Patton 2013a;b). This database consists of all de novo firms going public on American exchanges and filed with the Securities and Exchange Commission (SEC), where we restrict the corpus to those initial public offerings between the years 2003 and 2010. This accounts for a total of 625 IPOs.

In a next step, we collect the final prospectus (i.e., Form 424) for each of these initial public offerings from the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system.<sup>1</sup> Filings are available online for the 621 out of 625 initial public offerings.

Afterward, we combine the Form 424 filings with their corresponding stock market performance. Here, we use the return after the *n*-th trading day, defined as

$$R(n) = \left(ShareClosingPrice_n - InitialSharePrice\right) / InitialSharePrice. \tag{2}$$

Daily data from stock markets is (mostly) provided by Thomson Reuters Datastream where price data is available for 577 initial public offerings. As in Ferris et al. (2013) and Loughran and McDonald (2013), we require an initial offer price of at least \$5. In order to adjust for sector-specific differences like previous research (Ferris et al. 2013, Loughran and McDonald 2013), we introduce dummy variables for each industry. We use the dummies provided by the Thomson Reuters Business Classification (TRBC),<sup>2</sup> which are classified into 52 industry groups. Because of data availability, we use a total of 572 out of these 577 observations.

Our news corpus originates from the *Thomson Reuters News Archive* for Machine Readable News. We choose Reuters news deliberately because of three reasons: (1) Reuters conveys, in particular, news about stock markets. (2) Reuters news is a third-party content and, thus, gives a certain level of objectivity. (3) In contrast to newspapers, news agencies feature a shorter time lag and lack perturbations by edits. All announcements provided by Reuters arise from January 1, 2003, onward. The announcements come along with stock symbols denoting the firm the content it deals with. Based upon these labels, the news corpus is filtered such that we extract the announcements of firms going public. In addition, we require announcements to be published during the 365 days prior to the IPO. All in all, this set of criteria filters a total of 1085 announcements covering initial public offerings. If no announcement on an IPO is available, we set the sentiment value to 0. If more than one announcement is available, we use the average sentiment value. Furthermore, we incorporate the number of news announcement linked to each IPO into a separate control variable #News. Examples of headlines are given

<sup>&</sup>lt;sup>1</sup>US Securities and Exchange Commission (2013). *EDGAR Company Filings*. Web: http://sec.gov/edgar/searchedgar/companysearch.html, retrieved on June 17, 2013.

<sup>&</sup>lt;sup>2</sup>Thomson Reuters (2014). *Thomson Reuters Business Classification (TRBC)*. Web: http://thomsonreuters.com/business-classification/, retrieved on February 18, 2014.

<sup>&</sup>lt;sup>3</sup>This is achieved by applying a set of filter criteria (Feuerriegel and Neumann 2013): (1) The language must be English. (2) The event type is *Story Take Overwrite* to guarantee that we do not yield an alert but the actual message. (3) Special types of announcements, such as alerts or tabular data, might have limited relevance, and we want to exclude these. Thus, we omit announcements that contain specific words (*advisory*, *chronology*, *feature*, *diary*, *instant view*, *analysts view*, *newsmaker*, *refile*, *rpt*, *schedule*, *table*, *service*, *alert*, *wrapup*, *imbalance*, *update*) in their headline. (4) In order to remove white noise, we require announcements to count at least 50 words.

Date and time	Stock symbol	Headline
Jan 2, 2003, 14:15	SCOR	Cardinal Health completes acquisition of Syncor
Jan 6, 2003, 10:49	CE	Council of Europe plans \$750 mln 2010 bond Monday
Jan 7, 2003, 16:59	ONE	Bank One launches \$1 billion 5-year note sale
Jan 7, 2003, 17:10	PATH	AmeriPath shareholder urges buyout rejection
Jan 7, 2003, 18:38	ONE	New Issue—Bank One sells \$1 bln five-yr notes
Jan 8, 2003, 14:01	MGG	MGM Mirage says it will miss Wall St. profit estimates
Jan 8, 2003, 23:45	MGG	Park Place has not changed Q4 guidance-spokesman
Jan 9, 2003, 10:17	PRO	Provalis expects first annual pre-tax profit

**Table 1** Examples of headlines from news used in news sentiment variable

**Table 2** Descriptive statistics of sentiment, issued prices and returns after n trading days

Variable	Mean	Median	Min	Max	Std. Dev.	Skewness	Kurtosis
$Sentiment_{IPO}$	0.0173	0.0173	0.0125	0.0246	0.0017	0.2094	0.5365
Sentiment <sub>News</sub>	0.0049	0.0000	0.0000	0.0476	0.0074	1.6724	3.1994
#News	2.2945	1.0000	0.0000	108.0000	8.2986	8.8609	90.7178
Issued price	14.2613	14.0000	4.0000	85.0000	6.5506	3.6347	29.3660
R(1)	0.4216	0.0842	-0.8558	35.0000	2.2441	10.5652	134.3806
R(10)	0.4124	0.1037	-0.8452	29.3000	2.0963	9.9248	116.7767
R(21)	0.4304	0.1114	-0.8337	30.8000	2.1816	9.7876	114.0889

in Table 1. From this table, we can see that the news not only reports that a firm is going public but states more insights of the pre-IPO activities.

## 4.2 Descriptive Statistics of Stock Market Returns

In our evaluation, we vary the time span of returns to study the persistence of prospectus sentiment. All descriptive statistics of issued share prices, as well as returns after the n-th trading day, are presented in Table 2. Accordingly, the initial issued prices range from \$4.00 to \$85.00, while the mean initial share price accounts for \$14.2613 with a standard deviation of \$6.5506. Furthermore, the returns after 21 trading days span approximately from -0.8337 to 30.8000, with a nonzero mean of approximately 0.4304. Interestingly, the kurtosis of 134.3806 on the first day of trading is substantially higher than 3, indicating the existence of heavy tails, which are probably caused by price spikes. Both characteristics of nonzero mean and a high kurtosis are consistent with related literature (cf. Jenkinson and Ljungqvist 2001, Ritter and Welch 2002, for recent reviews), where these effects are referred to as underpricing.

#### 4.3 Control Variables

When isolating and extracting the effect of sentiment, we use a wide range of control variables. We control for the development of the stock market and, in addition to that, further control variables which comprise the number of shares, the share overhang, and the underwriter discount. The latter three variables are retrieved from the Kenney-Patton IPO database (Kenney and Patton 2013a;b), which we improved with additional corrections. Similar variables are used by a number of prior papers to explain first-day returns (see, e.g., Ferris et al. 2013, Loughran and McDonald 2013). Individual control variables are as follows:<sup>4</sup>

- *UpRevision* This control variable measures the percentage of the upward revision from the mid-point of the filing range if the offer price is greater than the mid-point; otherwise it is set to zero. As in Ferris et al. (2013) and Loughran and McDonald (2013), including this variable is crucial; when this variable is added last, many others are no longer significant.
- ΔDays<sub>S-1, IPO</sub> To control for possible delays of IPOs, we include a variable measuring the logarithmized calendar days between S-1 and the initial public offering.
- NASDAQ<sub>-15d</sub> We include the general development of a stock market index to
  ensure that the stock market performance is driven by sentiment instead of the
  economic cycle or the general mood of investors. Thus, we choose the return of
  the NASDAQ-100 stock market index of the 15 days prior to the initial public
  offering. Here, values originate from Thomson Reuters Datastream.
- In *Sales* This variable is defined as the logarithm of the number of shares sold to the public in the offering multiplied by the initial share price. Thus, the logarithm of all sales controls the volume of the initial public offering.
- *ShareOverhang* This variable indicates the number of shares retained divided by the number of shares in the initial offering. More precisely, it is defined as

$$ShareOverhang = \frac{S_{\text{outstand}}}{S_{\text{outstand}} + S_{\text{sold}}},$$
 (3)

where  $S_{\text{outstand}}$  is the number of shares outstanding after the offer and  $S_{\text{sold}}$  is the number of shares sold to the public in this offering.

• *UnderwriterDiscount* This variable includes the per share discounts and commissions taken by the underwriters.

By using these control variables, we want to assure that our results measure the impact that comes from the sentiment only and avoid perturbations by other causes, such as changes in fundamental variables or external events. Next, descriptive statistics of all variables are presented in Table 3. In addition to that, we incorporate

<sup>&</sup>lt;sup>4</sup>Although many papers control if the firm is backed by venture capital. However, Loughran and McDonald (2013) find no evidence that this variable has a significant impact, and thus, we omit it.

Variable	Mean	Median	Min	Max	Std. dev.	Skewness	Kurtosis
UpRevision	6.0484	0.0000	0.0000	79.3103	10.2935	2.5742	8.8797
$\Delta Days_{S-1, IPO}$	3.1603	3.0201	0.0000	6.6201	0.8557	-0.0029	5.1859
$NASDAQ_{-15\mathrm{d}}$	0.0065	0.0072	-0.1800	0.0979	0.0373	-0.4665	0.5743
ln Sales	18.3719	18.2869	15.5690	21.5984	0.8806	0.4368	0.9694
ShareOverhang	0.7699	0.7826	0.0004	0.9879	0.0908	-2.7661	16.7556
UnderwriterDiscount	0.9538	0.9100	0.0000	3.8400	0.3752	1.8949	10.6112

**Table 3** Descriptive statistics of control variables

additional dummies into our model to adjust for both seasonal and sector-specific effects.

## 4.4 Regression Design

This section presents the corresponding regression design, i.e.,

$$R(n) = \alpha + \beta_1 Sentiment_{IPO} + \beta_2 UpRevision + \beta_3 \Delta Days_{S-1, IPO} + \beta_4 NASDAQ_{-15d}$$

$$+ \beta_5 \ln Sales + \beta_6 ShareOverhang + \beta_7 UnderwriterDiscount$$

$$+ \sum_{i} \beta_j Dummy_j + \epsilon,$$

$$(4)$$

where n specifies the n-th day of trading,  $\beta_i$  are coefficients to be estimated, and  $\epsilon$  is the error term. Dummies are added for each sector, as well as year, to consider additional external events not covered by the control variables and to handle non-seasonally adjusted time series. The correlation of sentiment values and first-day returns is depicted in Fig. 2. This diagram also features a so-called LOWESS trend line, which is a locally weighted scatterplot smoothing calculated via local regressions. From this LOWESS trend line (smoothing parameter f set to 2/3), we can identify a visible relationship between the sentiment variable and stock market returns. Finally, we give justice to extreme stock price effects and remove outliers at the 0.05% level at both ends.

## 4.5 Results: Linking Sentiment and Stock Market Performance

In this section, we investigate how investors react to IPO prospectuses and analyze the information processing empirically. To succeed in this goal, we exploit sentiment analysis to measure the relationship between investor behavior and the content of IPO filings.

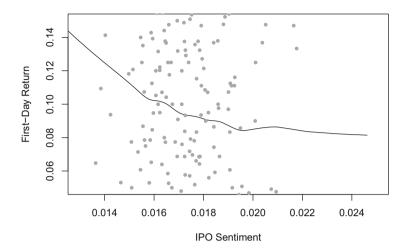


Fig. 2 Scatterplot of first-day returns and sentiment values with the LOWESS trend line

Confirmatory analysis: To what extent are stock market returns of firms driven by the sentiment of their IPO filings?

We use the above regression design from Eq. (4) to measure the impact of sentiment in prospectuses on IPO underpricing. We tested for heteroskedasticity, constant variance, serial correlation, and normally distributed residuals at the 0.01% level to ensure that the results are not confounded. When checking variance inflation factors, we also see no indication of multicollinearity. Independence across IPO filings is given as long as all prospectuses are entirely novel and not based on an interrelated course of events. Under the assumption that first-day returns are jointly multivariate normal, as well as independently and identically distributed through time, the model can be estimated using ordinary least squares (OLS).

Regression results are given in Table 4. According to this table, we observe that prospectus sentiment influences first-day returns significantly. The corresponding t-value accounts for -2.505 with a p-value of 0.013. Furthermore, we find that one standard deviation increase in the sentiment measure is negatively linked to an increase in first-day returns by an economically significant 10.56%. When additionally comparing the coefficients of the fundamental variables and sentiment, we find that the sentiment coefficient (accounting for -61.879) exceeds all other coefficients originating from the fundamental variables greatly. In addition to that, none of the control variables are, except for UpRevision, significant at the 10% level. Altogether, these results provide evidence that the sentiment in IPO prospectuses correlates significantly with first-day returns.

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Sentiment <sub>IPO</sub>	-47.969*	-45.915	-43.319	-41.023	-54.558*	-54.897*	-61.879*
	(-2.046)	(-1.954)	(-1.842)	(-1.740)	(-2.225)	-2.235	-2.505
UpRevision		0.005	0.005	0.006	0.008	0.008	0.009*
		(1.253)	(1.255)	(1.427)	(1.866)	(1.789)	(1.998)
$\Delta Days_{S-1, IPO}$			-0.073	-0.073	-0.083	-0.081	-0.079
			(-1.604)	(-1.612)	(-1.811)	(-1.777)	(-1.729)
$NASDAQ_{-15\mathrm{d}}$				-1.341	-1.445	-1.408	-1.252
				(-1.253)	(-1.351)	(-1.309)	(-1.162)
ln Sales					-0.101	-0.099	-0.081
					(-1.940)	(-1.900)	(-1.434)
Share Overhang						0.151	0.184
						(0.343)	(0.414)
Underwriter							-0.133
Discount							(-0.969)
Intercept	0.601	0.549	0.755	0.726	2.872*	2.733*	2.636*
	(1.102)	(1.006)	(1.347)	(1.295)	(2.317)	(2.094)	(1.972)
AIC	1469.705	1470.004	1469.213	1469.506	1467.41	1469.282	1470.776
BIC	1660.134	1664.761	1668.298	1672.919	1675.151	1681.351	1687.172
Multiple $R^2$	0.081	0.084	0.089	0.091	0.098	0.098	0.099

Table 4 OLS regression linking sentiment and first-day returns

Stated: Coef. and *t*-stat. in parentheses Dummies: Year, sector Obs.: 572 Signif.: \*\*\* 0.001, \*\*0.01, \*0.05

## 4.6 Results: Analyzing the Persistence of Sentiment Impact Over Time

Having discussed the regression design, we proceed to analyze how the sentiment affects n-th trading day returns after the initial public offering.

Research Question 1: How persistent is the influence of filing sentiment on long-term returns?

Thus, we vary the time span of returns across 10 and 21 days of trading, which is equivalent to 2 weeks and 1 month, respectively. The corresponding regression results are presented in Tables 5 and 6. Again, we checked for heteroskedasticity, constant variance, serial correlation, and normally distributed residuals and multicollinearity to ensure that we can estimate the model using OLS. In short, the results can be summarized as a declining influence of sentiment for longer time lags.

In more detail, the coefficients indicating the sentiment reception start at -61.879 for first-day returns but change slightly to -54.302 after the 10th day of trading. Similarly, the *t*-statistic of the sentiment variable drops from -2.505 for first-day returns to -2.266 after the 10th day of trading. Its value shrinks further down to -1.033 when looking at returns after the 21st day of trading. The influence of sentiment on both the first and the 10th day of trading is significant at the 5%

th day of trading
e 10
after the
S
t and return
it and
sentimer
linking
S regression
OLS
w
Table

	(a)	(b)	(c)	(p)	(e)	(f)	(g)	(h)
Sentiment <sub>IPO</sub>	-38.937	-36.711	-33.393	-32.095	-48.438*	-48.578*	-52.406*	-54.302*
	(-1.705)	(-1.605)	(-1.462)	(-1.400)	(-2.035)	(-2.037)	(-2.186)	(-2.266)
UpRevision		0.005	0.005	900.0	*800.0	*800.0	0.010*	*600.0
		(1.394)	(1.400)	(1.491)	(2.048)	(1.999)	(2.195)	(2.158)
$\Delta Days_{S-1, IPO}$			-0.093*	-0.094*	-0.105*	-0.104*	-0.103*	-0.104*
			(-2.11)	(-2.113)	(-2.365)	(-2.344)	(-2.303)	(-2.334)
NASDAQ_15 d				-0.758	-0.883	-0.868	-0.788	-0.761
				(-0.728)	(-0.851)	(-0.831)	(-0.754)	(-0.730)
InSales					-0.122*	-0.121*	-0.102	-0.103
					(-2.413)	(-2.388)	(-1.851)	(-1.867)
ShareOverhang						0.063	0.1111	0.174
						(0.147)	(0.257)	(0.402)
UnderwriterDiscount							-0.126	-0.130
							(-0.947)	(-0.980)
NASDAQ <sub>10d</sub>								2.131
								(1.671)
Intercept	0.442	0.386	0.648	0.632	3.224**	3.166*	2.975*	2.990*
	(0.832)	(0.726)	(1.191)	(1.159)	(2.679)	(2.498)	(2.293)	(2.309)
AIC	1439.971	1439.864	1437.044	1438.466	1434.144	1436.12	1437.497	1436.437
BIC	1630.400	1634.622	1636.129	1641.879	1641.885	1648.189	1653.893	1657.162
Multiple R <sup>2</sup>	0.057	090.0	0.068	0.069	0.080	0.080	0.081	0.086
Stated: Coef. and t-stat. in	n parentheses	I	Dummies: Year, sector	sector	Obs.: 572	2	Signif.: ***0.0	Signif.: *** 0.001, ** 0.01, *0.05
	-						)	

Table 6 OLS regression linking sentiment and returns after the 21st day of trading

	(a)	(b)	(c)	(p)	(e)	(f)	(g)	(h)
Sentiment <sub>IPO</sub>	-14.800	-12.015	-8.333	-7.667	-19.863	-19.996	-23.378	-26.104
	(-0.617)	(-0.501)	(-0.348)	(-0.319)	(-0.792)	(-0.796)	(-0.925)	(-1.033)
UpRevision		0.007	0.007	0.007	*600.0	*800.0	0.010*	0.010*
		(1.615)	(1.624)	(1.660)	(2.028)	(1.982)	(2.108)	(2.134)
$\Delta Days_{S-1, IPO}$			-0.103*	-0.103*	-0.111*	-0.110*	-0.109*	-0.110*
			(-2.220)	(-2.220)	(-2.389)	(-2.369)	(-2.335)	(-2.362)
NASDAQ_15 d				-0.409	-0.505	-0.492	-0.423	-0.248
				(-0.376)	(-0.464)	(-0.449)	(-0.386)	(-0.226)
In Sales					-0.089	-0.089	-0.073	-0.074
					(-1.69)	(-1.67)	(-1.266)	(-1.281)
ShareOverhang						0.056	0.094	0.162
						(0.125)	(0.208)	(0.358)
UnderwriterDiscount							-0.104	-0.117
							(-0.746)	(-0.835)
NASDAQ <sub>21 d</sub>								1.716
								(1.810)
Intercept	0.039	-0.030	0.257	0.249	2.156	2.105	1.956	1.898
	(0.070)	(-0.054)	(0.452)	(0.437)	(1.705)	(1.582)	(1.436)	(1.396)
AIC	1489.858	1489.035	1485.702	1487.548	1486.441	1488.424	1490.481	1488.897
BIC	1680.287	1683.792	1684.787	1690.961	1694.182	1700.493	1706.878	1709.622
Multiple $R^2$	0.052	0.056	0.065	990.0	0.071	0.071	0.071	0.077
Stated: Coef. and t-stat. ir	in parentheses	I	Dummies: Year, sector	sector	Obs.: 572	6	Signif.: *** 0.(	Signif.: *** 0.001, ** 0.01, *0.05

level, while the sentiment coefficient after 21 trading days becomes not significant at common significance levels. Altogether, this supports the hypothesis of a declining influence of sentiment on longer time spans.

### 4.7 Results: Prospectus Sentiment Versus News Sentiment

While the previous sections analyzed the influence of news, the question regarding the relevance of the general news coverage is left unanswered.

Research Question 2: How is the IPO performance affected by pre-IPO news coverage?

We extend the regression of Eq. (4) by an additional variable measuring the sentiment of news announcements related to that initial public offering. Out of all 572 initial public offerings, we find at least one announcement in the news corpus for 207 firms, i.e., 39.6%. We tested for autocorrelation, heteroskedasticity, constant variance, serial correlation, and normally distributed residuals at the 0.01% level to ensure that the results are not confounded. When checking for variance inflation factors, we also see no indication of multicollinearity. Thus, the model can be estimated using ordinary least squares (OLS). The regression results are presented in Table 7. The influence of IPO sentiment remains similar with a coefficient of -66.045 and a corresponding t-value of -2.456. However, the influence of news announcements seems much stronger. Here, the news sentiment accounts for a coefficient of 20.907 with a P-value of 0.00048 and a t-value of 3.518.

This is interesting for two reasons. First, a higher *t*-value indicates a significant link between returns and news, even stronger than the IPO sentiment. Second, while more *uncertainty* words in the prospectus lower the first-day returns, this does not hold for the news announcements. The more *uncertainty* words in the news announcements, the higher the first-day return. While one standard deviation increase in the IPO sentiment measure is linked to a decrease in first-day returns by an economically significant 11.30%, the news sentiment behaves differently. Here, one standard deviation change in news sentiment correlates with an increase in first-day returns by 14.72%. Furthermore, the number of news announcements (given by *#News*) has no impact on IPO prices at any common significance level. Astonishingly, the effect shows a larger magnitude, and at the same time, the more negative the pre-IPO news coverage, the higher the first-day returns.

A possible explanation can be attributed to the investors' psychological constraints in information processing—the processing depends on the attention different information receives. This phenomenon is already addressed in accounting literature (Hirshleifer et al. 2011): Limited investor attention theory for earning announcements suggests that investors neglect information about future earnings in current-period announcements creating a post-earning announcement drift. Subsequently, this mispricing is corrected over time. The same explanation may apply to IPO prospectuses, where mainly the risks of the company are recognized, which may entail the negative coefficient of our IPO sentiment variable. Underestimating

Table 7 OLS regression comparing influence of sentiment in IPO filings and news

	(a)	(p)	(c)	(p)	(e)	(f)	(g)	(h)	(i)
Sentiment <sub>IPO</sub>	-46.617	-44.120	-44.066	-42.432	-39.287	-36.993	-56.361*	-56.439*	-66.045*
	(-1.813)	(-1.732)	(-1.735)	(-1.667)	(-1.540)	(-1.446)	(-2.113)	(-2.113)	(-2.456)
Sentiment <sub>News</sub>		18.654**	20.021***	19.994***	19.488**	19.176**	20.76***	20.727***	20.907***
		(3.167)	(3.385)	(3.381)	(3.294)	(3.238)	(3.503)	(3.489)	(3.518)
#News			-0.010	-0.010	-0.009	-0.009	-0.010	-0.010	-0.010
			(-1.937)	(-1.857)	(-1.777)	(-1.754)	(-1.905)	(-1.905)	(-1.905)
UpRevision				0.005	0.005	0.005	0.008	0.008	0.010*
				(1.065)	(1.086)	(1.232)	(1.825)	(1.781)	(1.969)
$\Delta Days_{\rm S-1, IPO}$					-0.074	-0.074	-0.086	980:0-	-0.084
					(-1.452)	(-1.455)	(-1.69)	(-1.681)	(-1.631)
NASDAQ_15 d						-1.279	-1.372	-1.356	-1.173
						(-1.117)	(-1.204)	(-1.181)	(-1.020)
In Sales							-0.136*	-0.135*	-0.120*
							(-2.430)	(-2.407)	(-1.966)
ShareOverhang								0.053	0.070
								(0.113)	(0.146)
UnderwriterDiscount									-0.130
									(-0.890)
Intercept	0.723	0.566	0.551	0.515	0.709	0.681	3.561**	3.511*	3.507*
	(1.263)	(0.995)	(0.972)	(0.905)	(1.216)	(1.167)	(2.699)	(2.518)	(2.457)
AIC	1385.908	1377.069	1374.979	1375.735	1375.423	1376.05	1371.573	1373.559	1375.057
BIC	1568.072	1563.47	1565.615	1570.608	1574.533	1579.396	1579.155	1585.377	1591.112
Multiple R <sup>2</sup>	0.086	0.105	0.112	0.114	0.118	0.121	0.132	0.132	0.133
Stated: Coef. and t-stat. in parentheses	t. in parenthe	ses	Dummie	Dummies: Year, sector		Obs.: 523		Signif.: *** 0.001, *** 0.01, *0.05	1, **0.01, *0.05

future opportunities may create a post-prospectus announcement drift that was visible in our previous analysis. News announcements prior to the IPO are obviously less suspect to the companies' risks. On the contrary, the positive coefficient suggests that more emphasis is put on the companies' opportunities. Apparently, the degree of investor attention depends on the specific type of information.

#### 5 Conclusion

Before shares of a company are sold to the general public on a security exchange for the first time, regulatory publication requirements force US firms to file an initial public offering prospectus. Studying the information processing of investors facing these filings is an active research question in the context of efficient electronic markets, though knowledge is rare (e.g., Liebmann et al. 2012): "While the accounting numbers in IPO prospectuses are closely studied by investors, analysts, and others involved in the equity issuance process, an examination of the textual or soft information contained in prospectuses is less common" (Ferris et al. 2013). To close this gap, research must harness decision analytics in order to explain the relationship between the textual content of filings and final IPO prices.

In this paper, we address information processing in IPO filings by analyzing how the sentiment in prospectuses influences stock market performances. Using the final filings of 572 US initial public offerings (Form 424) between 2003 and 2010, we can identify soft content as a major driver of stock returns. All in all, we can empirically establish a relationship between sentiment in IPO filings and stock market reaction. One standard deviation increase in the sentiment measure correlates negatively with a change in first-day returns by an economically significant 10.56%. This dependency appears not only on the first day of trading but also for longer phases of up to 10 days of trading. Apparently, "soft information can offer context to financial numbers and share values, provide insight into managerial expectations, and identify important qualifiers or caveats that are absent from purely numerical data" (Ferris et al. 2013). As an explanation, we find supporting evidence that we can expect initial public offerings with substantial uncertain language to have, on average, lower preliminary offer prices. This effect occurs due to the need for bankers to compensate investors for their information production (Loughran and McDonald 2013). According to Ferris et al. (2013), the findings "suggest that when hard information is noisier, textual information is seen as having greater usefulness and is more likely to be factored into IPO pricing." However, a much stronger impact on stock market prices happens not through the IPO prospectus, but comes from the pre-IPO news sentiment. Interestingly, the more uncertainty words that appear in pre-IPO news, the higher the following first-day stock market return. Consequently, this uncertainty puzzle must be resolved. It seems that investors mainly focus on the chances of a company rather than on the risks. This stands in contrast to the prospectus where risk considerations play a major role.

The work presented in this chapter opens several avenues for future research. First, a pooled regression can be utilized to analyze possible differences between *no crisis* and *crisis* periods where investors change their information processing to more cautious behavior. Second, further effort is needed to validate our approach in terms of robustness, and, thus, we plan to extend our analysis further to filings, such as S-1 or, alternatively, transcripts of court decisions. Third, an obvious next step is to bridge the gap between explanatory regression models and a prediction model (Shmueli and Koppius 2011) that can estimate expected stock market returns from IPO filings.

**Acknowledgment** We thank Stefan Feuerriegel for his valuable contributions.

#### References

- Antweiler W, Frank MZ (2004) Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. The Journal of Finance 59(3):1259–1294
- Arnold T, Fishe RPH, North D (2010) The Effects of Ambiguous Information on Initial and Subsequent IPO Returns. Financial Management 39:1497–1519
- Chen H, Chiang RRL, Storey VC (2012) Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly 36(4):1165–1188
- Dadvar M, Hauff C, de Jong F (2011) Scope of Negation Detection in Sentiment Analysis. In: Proceedings of the Dutch-Belgian Information Retrieval Workshop (DIR 2011), pp 16–20
- Demers EA, Vega C (2010) Soft Information in Earnings Announcements: News or Noise? INSEAD Working Paper No. 2010/33/AC. SSRN Electronic Journal
- Fama FE (1965) The Behavior of Stock-Market Prices. The Journal of Business 38(1):34-105
- Ferris SP, Hao Q, Liao MY (2013) The Effect of Issuer Conservatism on IPO Pricing and Performance. Review of Finance 17(3):993–1027
- Feuerriegel S, Neumann D (2013) News or Noise? How News Drives Commodity Prices. In: Proceedings of the International Conference on Information Systems (ICIS 2013), Association for Information Systems
- Feuerriegel S, Lampe MW, Neumann D (2014) News Processing during Speculative Bubbles: Evidence from the Oil Market. In: 47th Hawaii International Conference on System Sciences (HICSS)
- Granados N, Gupta A, Kauffman RJ (2010) Research Commentary–Information Transparency in Business-to-Consumer Markets: Concepts, Framework, and Research Agenda. Information Systems Research 21(2):207–226
- Grefenstette G, Tapanainen P (1994) What is a word, What is a sentence? Problems of Tokenization. Meylan and France, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.5162
- Hanley KW, Hoberg G (2010) The Information Content of IPO Prospectuses. Review of Financial Studies 23(7):2821–2864
- Hanley KW, Hoberg G (2012) Litigation risk, strategic disclosure and the underpricing of initial public offerings. Journal of Financial Economics 103(2):235–254
- Henry E (2008) Are Investors Influenced By How Earnings Press Releases Are Written? Journal of Business Communication 45(4):363–407
- Hirshleifer D, Lim SS, Teoh SH (2011) Limited Investor Attention and Stock Market Misreactions to Accounting Information. Review of Asset Pricing Studies 1(1):35–73
- IBM (2013) The Four V's of Big Data. http://www.ibmbigdatahub.com/infographic/four-vs-big-data

- Jegadeesh N, Wu D (2013) Word Power: A New Approach for Content Analysis. Journal of Financial Economics 110(3):712–729
- Jenkinson T, Ljungqvist A (2001) Going Public: The Theory and Evidence on how Companies Raise Equity Finance, 2nd edn. Oxford University Press, Oxford
- Kenney M, Patton D (2013a) Firm Database of Initial Public Offerings (IPOs) from June 1996 through 2010. Davis and CA
- Kenney M, Patton D (2013b) Guide to the Firm Database of Initial Public Offerings (IPOs) from June 1996 through 2010. Davis and CA
- Lewis DD, Yang Y, Rose TG, Li F (2004) RCV1: A New Benchmark Collection for Text Categorization Research. Journal of Machine Learning Research 5:361–397
- Li F (2010) The Information Content of Forward-Looking Statements in Corporate Filings: A Naïve Bayesian Machine Learning Approach. Journal of Accounting Research 48(5):1049– 1102
- Liebmann M, Hagenau M, Neumann D (2012) Information Processing in Electronic Markets: Measuring Subjective Interpretation Using Sentiment Analysis. In: George JF (ed) Proceedings of the International Conference on Information Systems (ICIS 2012), Association for Information Systems
- Liu LX, Sherman AE, Zhang Y (2014) The Long-Run Role of the Media: Evidence from Initial Public Offerings: (in press). Management Science
- Loughran T, McDonald B (2011) When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance 66(1):35–65
- Loughran T, McDonald B (2013) IPO first-day returns, offer price revisions, volatility, and form S-1 language. Journal of Financial Economics 109(2):307–326
- Loughran T, McDonald B (2014) Regulation and financial disclosure: The impact of plain English. Journal of Regulatory Economics 45(1):94–113
- Manning CD, Schütze H (1999) Foundations of statistical natural language processing. MIT Press, Cambridge and Mass
- McDonald B (2012) Loughran and McDonald Financial Sentiment Dictionary. http://www3.nd.edu/mcdonald/Word\_Lists.html
- Minev M, Schommer C, Grammatikos T (2012) News and stock markets: A survey on abnormal returns and prediction models. Luxembourg, http://publications.uni.lu/record/9947/files/TR. Survey.News.Analytics.pdf
- Mittermayer MA, Knolmayer GF (2006a) NewsCATS: A News Categorization and Trading System. In: Sixth International Conference on Data Mining (ICDM'06), pp 1002–1007
- Mittermayer MA, Knolmayer GF (2006b) Text Mining Systems for Market Response to News: A Survey. Bern and Switzerland, http://pdf.aminer.org/000/246/869/forecasting\_intraday\_stock\_price\_trends\_with\_text\_mining\_techniques.pdf
- O'Hara M (1995) Market Microstructure Theory. Wiley-Blackwell, Oxford and MA
- Pang B, Lee L (2008) Opinion Mining and Sentiment Analysis. FNT in Information Retrieval (Foundations and Trends in Information Retrieval) 2(1–2):1–135
- Ritter JR, Welch I (2002) A Review of IPO Activity, Pricing, and Allocations. The Journal of Finance 57(4):1795–1828
- Schumaker RP, Chen H (2009) Textual analysis of stock market prediction using breaking financial news. ACM Transactions on Information Systems 27(2):1–19
- Sharma A, Dey S (2012) A comparative study of feature selection and machine learning techniques for sentiment analysis. In: Cho Y, Computing ASIGoA (eds) Proceedings of the 2012 Research in Applied Computation Symposium (RACS 2012), ACM, pp 1–7
- Shmueli G, Koppius O (2011) Predictive Analytics in Information Systems Research. MIS Quarterly 35(3):553–572
- Siering M (2013) Investigating the Impact of Media Sentiment and Investor Attention on Financial Markets. In: Aalst W, Mylopoulos J, Rosemann M, Shaw MJ, Szyperski C, Rabhi FA, Gomber P (eds) Enterprise Applications and Services in the Finance Industry, Lecture Notes in Business Information Processing, vol 135, Springer Berlin Heidelberg, Berlin and Heidelberg, pp 3–19

Tetlock PC (2007) Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance 62(3):1139–1168

Tetlock PC, Saar-Tsechansky M, Macskassy S (2008) More than words: Quantifying Language to Measure Firms' Fundamentals. The Journal of Finance 63(3):1437–1467

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

