Summarized Disclosure of Risk Factors

This version: August, 25 2019

Daniel Blaseg
ESADE Business School
Barcelona, Spain
Tel.: +49 157 761 34 355
E-Mail: daniel.blaseg@blaseg.de

Christina E. Bannier
University of Gießen
Gießen, Germany
Tel.: +49 641 99 22 551
E-Mail: christina.bannier@wirtschaft.uni-giessen.de

Abstract

Given the increasing length and complexity of financial market disclosures, high-quality summarization becomes ever more important, especially for retail investors. However, little is known about the informativeness of summaries in disclosure documents such as security prospectuses. In this study, we investigate how firms summarize information in security prospectuses and whether the quality of a summary provides valuable information about a security's future performance. We develop an automatic approach for evaluating the quality of a summary by examining textual disclosures on risk factors in bond prospectuses. Our results suggest that firms with poor performance tend to provide low-quality summaries and engage in impression management. The manipulation includes significant differences in readability, specificity, tonality, use of boilerplate, as well as dissimilarity in content between the summary and the full prospectus. Nevertheless, prospectus summaries are informative for investors as their quality can be shown to be a highly significant predictor of the default of the security.

Keywords: prospectus summary, summary evaluation, textual disclosure, risk factor disclosure

INTRODUCTION

The most valuable currency in financial markets is reliable information. Without it, investors are unable to make informed decisions about where to allocate their capital, which hurts companies' ability to attract it and puts a drag on economic growth. Transparency is an economic engine.

– Michael Bloomberg and Mary Schapiro (Financial Times, May 2014)

Issuers of financial securities nowadays disclose more information than ever (Dyer et al. 2017). This is despite the fact that retail investors are known to be unable to absorb vast quantities of information (Spindler 2011) – a fact that has also been recognized by financial regulatory agencies (e.g., European Securities and Markets Authority 2019; Securities and Exchange Commission 2016). As reduced and streamlined disclosures help retail investors to better analyze a firm's performance and prospects (Saha et al. 2019), regulators around the world increasingly require that issuers of securities release a summary in the security prospectus. While a security prospectus often consists of hundreds of pages and is frequently not read by investors (Beshears et al. 2011), even a summary that conveys only the most essential information, such as the risks and terms of the investment (La Porta et al. 2006), may run up to several pages. Academics and practitioners therefore question whether these short-form disclosure documents function as intended—namely, as an accurate and comprehensive representation of the information provided in the full prospectus (Barth 2015; CFA Institute 2017).

Experimental evidence suggests that summaries of disclosure documents in financial markets are of practical relevance as they have been shown to affect investors' perceptions and decisions (e.g., Cardinaels et al. 2019; Walther 2015). However, little is known about the degree of informativeness of summaries as compared to the original documents: Do issuers of financial products use summaries to provide relevant information to promote market efficiency (Heinle et al. 2018), or as a marketing tool to generate investor interest (Hanley and Hoberg 2010)? Several studies indeed suggest that managers use summaries to engage in impression management, by selectively presenting information that is beneficial to the firm (Cardinaels et al. 2019; Guillamon-Saorin et al. 2012; Huang et al. 2018). Moreover, summaries in prospectuses are often not subject to significant litigation risk, as there is limited liability for the summary (Burn 2016). Therefore, there is potential for biased reporting.

In this study, we try to fill this gap in the literature and investigate the informativeness of prospectus summaries. In particular, we analyze how firms summarize the prospectus information and whether the quality of a summary provides a valuable indication regarding a security's future performance. Examining this issue is essential for several reasons: First, our results provide new insights for regulators to clarify the role of summaries in financial markets. Second, our findings equip investors with the necessary knowledge to evaluate summaries as a basis for their financial decisions. Third, managers and lawyers may learn which attributes determine the perception of the quality of a summary, allowing them to streamline their disclosure practice.

Although the general usefulness of security prospectuses is debatable, recent academic studies find that individual sections do provide valuable information for investors (Hanley and Hoberg 2010). Notably, academics are interested in the disclosure of risk factors. Several studies show that different disclosure characteristics such as quantity, length, specificity, and number of topical keywords, reflect information about firm risk (e.g., Bonsall and Miller 2017; Campbell et al. 2014; Hope et al. 2016; Kravet and Muslu 2013). Though the description of risk factors is an elemental part of a security prospectus both in the United States and the European Union, a separate summary of risk factors must be provided only in the European Union (European Securities and Markets Authority 2012; Securities and Exchange Commission 1998). We therefore focus on the summarized disclosure of risk factors in a European setting in this study.

To investigate the informativeness of summarized risk factors, we use the German small and medium-size (SME) bond market for empirical context. This market offers several benefits: First, the high number of actual defaults in this market segment allows to directly differentiate between low- and high-quality bonds instead of relying on models that solely predict the probability of default. Second, due to the low nominal amount of EUR 1,000, comparably high yields, and well-known issuers' brand names, the main targets of SME bonds are indeed retail investors, which are suspected of being most strongly influenced by prospectus summaries.

To address our research objectives, we employ several consecutive procedures on our empirical dataset. In a first step, we identify and apply different metrics to automatically assess the quality of a summary, where quality is defined as the similarity between the original text and the summary. As the quality of a summary is not only determined by the content but also the way the information is presented (Lloret et al. 2018), we measure it via two standard criteria: (1) the semantic quality (i.e., the divergence between the probability distributions of words in the main text and its summary) and (2) the linguistic quality (i.e., readability, use of boilerplate, specificity, and tonality). To combine these two different aspects into a single measure of summary quality that is to be highly correlated with the "gold standard" of human evaluation, we recruit participants from the Clickworker platform (the German equivalent of Amazon's Mechanical Turk) in a second step to evaluate the quality of the summaries manually. This allows us to assess which aspects of semantic and linguistic quality drive human quality evaluations so that, based on a relative weight analysis, a single summary quality measure can be established. We finally use regression analyses to examine whether this summary quality is associated with various characteristics of the issuing firm. In addition, we study the informativeness of the summaries by investigating the relationship between a summary's quality and the security's subsequent performance.

Our results suggest that firms with poor implicit credit risk frequently provide low-quality summaries and engage in impression management. These firms strategically select, frame, and present textual information in the summary of the security prospectus in a way that is likely to affect positively investors' perceptions and decisions. The manipulations include statistically significant differences in readability, specificity, tonality, use of boilerplate, as well as dissimilarity in content between the summarized and the full risk factor sections in the security prospectus. Nevertheless, discretionary disclosure practice in prospectus summaries can be shown to be informative for investors, as the quality of a summary is a highly significant predictor of the future security performance: A lower summary quality is highly significantly associated with a higher likelihood of a default of the bond. In our analyses, we take great care in establishing that these results remain robust when controlling for the endogenous nature of releasing a low-quality summary.

Our study contributes to the literature on disclosures in financial markets in several ways. To the best of our knowledge, we are the first to analyze the quality of summarized information in security prospectuses empirically, and to relate firms' risk-disclosure practices in summaries to firm performance. Moreover, we contribute to the nascent stream on summarization in accounting and finance (e.g., Cardinaels et al. 2019), by developing a reproducible approach for measuring the quality of a summary.

The remainder of this study is structured as follows: In Chapter 2, we briefly discuss the legal background of summarized disclosure in the European Union. In Chapter 3, we portray the relevant literature and develop the hypotheses. Chapter 4 and Chapter 5 present the empirical study. Finally, Chapter 6 offers a discussion of implications and concludes.

LEGAL BACKGROUND

Prospectus Regulation in the European Union

Mandated disclosure of information is an essential driver for promoting financial market efficiency (Healy and Palepu 2001). Thus, regulators in most countries around the world require that firms offering securities to the public publish a prospectus (La Porta et al. 2006). In the European Union (EU), the Prospectus Directive 2003/71 (Prospectus Directive) serves as the primary legal framework of disclosure rules applicable to firms entering capital markets. By standardizing the requirements for creating, approving, and disseminating a prospectus, it aims at facilitating firms' fund-raising efforts in a uniform European capital market and at ensuring investor protection throughout the EU (Sergakis 2018). The Prospectus Directive and its subordinate legislation, Commission Regulation (EC) No 809/2004 (Prospectus Regulation) define the requirements, minimum content, and layout of the prospectus, preparation details, approval by a competent authority, as well as the filing and publication process when firms offer securities to the public.

The rules were implemented in Germany by the enactment of a new law, the Securities Prospectus Act (Prospectus Act), which came into effect on July 1, 2005. Together with the Prospectus Directive and the Prospectus Regulation, the Prospectus Act provides the legal basis for offering securities to the public and admitting securities to trading in a regulated market within Germany

(Du Vignaux et al. 2006). The Prospectus Directive was amended in December 2010. The Amending Directive 2012/73/EU (Amending Directive) and its subordinate legislation, the Commission Delegated Regulation (EU) No 486/2012 (Amending Regulation), require member states to implement the new legislation in national law by July 1, 2012, and include several changes in the layout and content of the summary, aiming to improve its usefulness for retail investors.[1]

Table 1 provides an overview of the legal framework encompassing the disclosure of risk factors in the summary and the full prospectus before and after the amendment. As the Securities Prospectus Act follows the content and wording of the European prospectus regime very closely, our discussion is based on the European regulatory framework.

Insert Table 1 about here

###1.1

According to Article 25 of the Prospectus Regulation, "each prospectus must include a section which sets forth the risk factors associated with the issuer and the type of security covered by the issue." Usually, the disclosed factors describe risks that are specific to the state, situation, and industry of the issuer, its ability to fulfill its obligations, as well as the security itself, and are material for making an informed investment decision (European Securities and Markets Authority 2012). The specific content, as well as the layout, language, and length of the risk factor section,

[1] Please see Fischer-Appelt (2010) for a discussion of all changes.

is not further defined. The risk factor section in the prospectus, as well as the definition of risk factors, remained unchanged in the Amending Directive and the Amending Regulation.

Legal Background on Summarization of Risk Factors

The Prospectus Directive requires that each prospectus contain a summary of no more than 2,500 words, conveying the essential characteristics of, and risks associated with, the issuer, any guarantor, and the security.[2] To ensure full access to the information, the summary must be brief, and written in a non-technical manner in the language in which the prospectus was initially drawn up. The summary must also include warnings that it is only an introduction to the prospectus, and that "the decision to invest in the securities should be based on consideration of the prospectus as a whole by the investor" (Prospectus Directive, Article 5).

Annex I in the Prospectus Directive prescribes the minimum content requirements of the summary in the form of 27 elements in 11 sections. These elements include details of the offer as well as "key information." While not further defined, key information cover elements such as selected financial data, reasons for the offer, use of proceeds, and risk factors. Thus, issuers are required to summarize the risk factor section. However, the regulation presents no further requirements regarding their specificity and materiality, i.e. issuers can decide how and which risk factors to summarize. Moreover, the regulation does not explicitly define the layout and order of content in the summary and does not allow cross-references to other parts of the prospectus. Most importantly,

[2] Although Rec. 21 in the Prospectus Directive defines a maximum number of 2,500 words, the Federal Financial Supervisory Authority (Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin) regularly approved prospectuses with summaries of more than 5,000 words (Just et al. 2019).

there is no civil liability for misstatements based solely on the summary, unless the summary is misleading, inaccurate, or inconsistent when read with the full prospectus.

The key change in the Amending Directive and the Amending Regulation is a standardization of the format and content of the prospectus summary. The Amending Directive now defines "key information", that must be included in the summary, as "essential and appropriately structured information which is to be provided to investors with a view to enabling them to understand the nature and the risks of the securities that are being offered to them" (Fischer-Appelt 2014). Annex XXII of the Amending Regulation further standardizes the content and layout of the summary in order to enable investors to compare various securities directly. Summaries must now be constructed on a "modular basis according to the Annexes from the Prospectus Regulation on which the prospectus has been based." Although the number of sections in the summary decreased from 11 to 5 in the amendment, the number of elements increased from 27 to 87. However, the required content of the sections and elements differs only marginally, and the different numbers of sections and elements result mainly from changes in the layout requirements. As previously, the summary may not contain cross-references to other parts of the prospectus (Burn 2016). Furthermore, the Amending Regulation now prescribes that the summary must "not exceed 7% of the length of a prospectus or 15 pages, whichever is longer." It also mandates that the summary must be written in "plain language, presenting the information in an easily accessible way."

With regard to liability, the Amending Directive enhances the earlier provision if the summary "does not provide, when reading together with the other parts of the prospectus, key information in order to aid investors when considering whether to invest in such securities." However, the condition that liability arises only "when read together with the other parts of the prospectus" severely limits the scope of this provision.[3] Again, there are no limitations on the length of the risk factor section in the prospectus. The issuer is still responsible for deciding which and how to summarize the risk factors (Fischer-Appelt 2010).

In sum, while the format and layout of content presentation in the summary are more prescriptive under the new prospectus regime, other key aspects remain unchanged. In practice, industry experts expected the content of summaries to not change considerably (Burn 2018; Fischer-Appelt 2010). Indeed, summaries look different and their overall length increased in recent years, but the type of information disclosed did not vary considerably from previous standards (Fischer-Appelt 2012). In the empirical section, we will confirm that the effect of the regulatory change on the content of the summary was quite small. This allows us to consider risk factor summaries both before and after the new regulation in our empirical analysis.

THEORETICAL BACKGROUND

Empirical Literature on Disclosure of Risk Factors

Over the last two decades, regulators around the world have been demanding greater quantity and quality of risk factor disclosures in corporate documents and securities prospectuses (Elshandidy et al. 2018; Hope et al. 2016). Paralleling this development, risk factor reporting has also become an important topic in the academic literature (Schrand and Elliott 1998). While critics argue that managers are likely to publish all possible risks factors to mitigate litigation risk, without considering their materiality, a large body of papers also provides empirical and theoretical evidence that

[3] Although securities litigation due to incorrect or untimely disclosure has increased quickly worldwide over the last few years, no known cases involve the content of summaries (Savitt 2018).

risk factor disclosure is actually informative in different markets (debt and capital markets), corporate documents (10-K forms and annual reports) or regulatory schemes (mandatory, voluntary, and aggregate).

According to Elshandidiy et al. (2018), the literature on risk reporting can be categorized into two main streams: (1) the incentives for and (2) the informativeness of risk factor disclosure. Due to the heavily regulated risk-reporting environment, studies based on U.S. data mainly focus on the informativeness of mandated risk disclosure. Research in other legal environments, such as the European Union, in contrast, also examines the incentives for risk reporting, where in some circumstances (e.g., annual reports) risk disclosure is voluntary (Elshandidy et al. 2015). As our study focuses on the informativeness of summarized risk factor disclosure, we limit the literature discussion to the second stream.

Kravet and Muslu (2013), Filzen (2015), and Campbell et al. (2014) provide evidence that content and tone of risk reporting in 10-K and 10-Q forms, as well as their updates and changes, are informative, and that equity investors accordingly react to such disclosures. In line with these studies, Hope et al. (2016) find that the quality of risk disclosures in terms of more specific risk factors causes stronger market reactions, and propose that risk factor disclosure is not boilerplate per se. These results are, however, not undisputed. Bao and Datta (2014) investigate risk disclosure in 10-K forms via a latent Dirichlet allocation topic model. Focusing on the semantic content of risk disclosure, they find that around two-thirds of risk types are not informative for investors. Moreover, they show that even informative risk factors do not change investors' risk perceptions. Campbell et al. (2018) and Chiu, Kim, et al. (2018) investigate the role of single, specific risk factor disclosures. They observe that the disclosure of tax-related risks provides information about firms' future cash flows, and that the disclosure of customer-related risks contains useful information for suppliers to achieve better investment efficiency.

In the case of debt markets, Chiu, Guan, et al. (2018) provide evidence that the disclosure of risk factors improves investors' assessment of default probability. Not explicitly focusing on the risk factor section but, instead, the full 10-K form, Bonsall and Miller (2017) find that less readable disclosures have adverse effects, and are linked to less favorable ratings, more disagreement between rating agencies, and higher costs of funding.

Beatty et al. (2018) confirm that investors on different capital markets react to unexpected risk factor disclosures. These reactions decline significantly over time, however, especially since the financial crisis, as the disclosures become longer, more redundant, and less specific. Dyer et al. (2017) find that similarity in disclosures increases, while readability and specificity decrease over time. They identify changes in the practices of risk factor disclosures as the main driver of this effect. Brown et al. (2018), in contrast, report that the recent assessment of risk disclosure practices by the Securities and Exchange Commission (SEC) has spillover effects. Even firms that do not received comments letters from the regulator provide more firm-specific disclosures in the following year.

Taken together, the literature on risk reporting generally suggests that the disclosure of risk factors in corporate documents is informative for investors. However, considerable variations in the effect exist, due to different disclosure measures such as content, tone, specificity, readability, similarity, and semantic clues, different time horizons, and dependent variables.

Theoretical Background of Risk Factor Disclosure

While there is no universal theory, earlier studies often refer to agency theory and proprietary cost theory to explain the informativeness of risk disclosures (Elshandidy et al. 2018; Al-Hadi et al. 2016; Abraham and Shrives 2014). Within the agency theory strand, existing research typically assumes that risk disclosure facilitates decision making by minimizing information asymmetries between managers and investors. Unbiased reporting is driven not only by laws mandating the disclosure of risk factors but also by incentives for the firms and managers (Elshandidy et al. 2013) such as reduced cost of capital and increased reputation and compensation (Baginski et al. 2000).

However, there is also an extensive literature that claims that textual disclosure enables managers to further exploit (rather than reduce) information asymmetries between managers and investors through engaging in biased reporting, that is, impression management. Impression management in this respect describes the behavior of managers to strategically select, frame, and present textual information in disclosure documents in a way that is intended to manipulate investors' perceptions and decisions (Aerts 2005; Clatworthy and Jones 2001; Leung et al. 2015; Yuthas et al. 2002). It assumes that investors, at least in the short term, are unable to detect reporting bias, leading to potential capital misallocations, due to non-informative and misleading information (Holthausen 1990). Impression management has been shown to be especially prevalent when firm performance is poor (Bloomfield 2002). It might also be relevant in the case of summarized disclosure of risk factors, as summarization inherently allows for discretion (Leung et al. 2015).

Merkl-Davies and Brennan (2007) categorize impression management strategies into two types, (A) concealment and (B) attribution. The authors identify seven strategies from previous research on the topic: (A1) reading ease manipulation, (A2) rhetorical manipulation, (A3) thematic manipulation, (A4) visual and structural manipulation, (A5) performance comparisons, (A6) selectivity of numerical disclosures, and (B1) attribution of performance. Although several studies document the existence of biased disclosure (e.g., Smith and Taffler 2000), empirical evidence of an association between impression management strategies and firm characteristics, primarily financial performance, is mixed and inconclusive. Different results are derived depending on the type of disclosure documents and performance measures (see Merkl-Davies and Brennan 2007 for a comprehensive review of the literature). Whether these findings are generalizable to the summarized disclosure of risk factors is unclear as none of the studies examining impression management has so far focused explicitly on risk factor disclosures nor on summarized disclosures.

Literature on Summarized Disclosure

To the best of our knowledge, this study is the first to investigate the quality of the summarized disclosure of risk factors in security prospectuses. However, several recent studies examine the role and effects of summarized or simplified disclosure generally. Most closely related to our study, Cardinaels et al. (2019) investigate how algorithm-generated summaries of earnings releases compare to management summaries of the same documents over several dimensions (e.g., bias and trustworthiness), and how summaries affect individual investors' perceptions. The results suggest that algorithm-generated summaries are generally less positively biased, often without missing essential information and informativeness. Furthermore, the authors provide evidence that summaries alter investors' valuation-related perceptions and confidence in their judgments. They conclude that summaries provided by managers are subject to an incremental positive bias compared to the original text, and that summaries affect investors' judgments.

Walther (2015) confirms that summaries have an effect on investor judgment as investors report that short-form disclosures in the form of key investor documents for investment funds are more informative and helpful than the underlying prospectuses. However, the analyses also reveal that many investors are not able to understand the summarized information.

Godwin and Ramsay (2016) present reasons for this lack of understanding. They use an investor survey to investigate the informativeness of short-form disclosure formats as adopted in six jurisdictions. Their results suggest that retail investors evaluate the quality of short-form disclosure documents based on an array of factors, including the informativeness of the document in terms of the critical features and investment risks, the readability, as well as the length and format. Venti (2011) notes that while short-form disclosures are shorter and less complex compared to a full prospectus, they are still not easy to understand as they contain technical jargon and still require basic knowledge of financial reporting.

Burn (2010) discusses the ability of summarized disclosure to improve retail investors' investment decisions. Referring to recent studies from YouGov and IOSCO reports, she states that summaries "can never contain sufficient information to enable investors to take an investment decision," and that short-form disclosures are unlikely to be informative (Burn 2010, p. 167). Similarly, Beshears et al. (2011) show that short-form disclosure documents for mutual funds allow investors to make faster, but not better, decisions in terms of higher returns. Hanley and Hoberg (2010) investigate the informativeness of prospectuses from Initial Public Offerings by decomposing the information into standard and informative components using word content analysis. After separately parsing four sections of the prospectuses, namely the summary, risk factors, use of proceeds, and management's discussion and analysis (MD&A), they find little informative content in the summary, and conclude that this section is mainly used as "a marketing tool to generate interest from investors" (Hanley and Hoberg 2010, p. 2848).

Research Question and Hypotheses

While the earlier literature hence suggests that the disclosure of risk factors is mostly informative for investors, the literature on simplified and summarized disclosure indicates that summaries itself are not informative. The discretionary nature of summaries can be explained by the impression management view rooted in agency theory, supported and facilitated by the fact that the specific summary content is usually unregulated and not subject to high litigation risk. However, we do not expect all firms to actively engage in impression management whenever the opportunity arises. Rather, we suggest that poorly performing firms are more likely to use the summary as a concealment vehicle to hide adverse information and that such discretionary disclosure practice (i.e., a decrease in quality from the full and summarized risk factor section) is informative for investors. This leads to the central hypotheses (stated in the alternative form):

H1: Poorly performing firms are more likely to release a low-quality summary of risk factors, where quality is measured as the similarity between the full and summarized section in the security prospectus.

H2: The quality of a risk factor summary is informative for investors, i.e. it helps to predict the future security performance.

In the next section, we describe the approach for quantifying the summary quality by developing an automatic measurement procedure and explain the empirical design of our study.

EMPIRICAL METHODOLOGY

Sample Construction

Following the global financial crises in 2008, tighter lending conditions (e.g., Basel II and higher equity requirements) restricted access to bank lending for firms around the world. Lending restrictions are especially crucial for SMEs (Altman et al. 2018; Rossi 2016) that tend to be heavily reliant on traditional bank lending even in developed countries such as Germany (Puri et al. 2011). Ownership of German SMEs primarily rests with private investors and families that try to avoid the dilution of ownership and control rights following equity market financing. They were therefore particularly strongly affected by the credit crunch (Audretsch and Elston 1997; Mietzner et al. 2018).

As a consequence of this credit rationing in the aftermath of the financial crisis, several European countries experimented with SME bond markets (European Securities and Markets Authority 2017). While established public capital markets require an issuance volume of at least EUR 100 million for bonds, several exchanges introduced specific segments with a lower minimum issuance volume. This allowed smaller firms to tap a liquid secondary market for public debt by offering titles to retail and professional investors.

In Germany, the market for SME bonds started in 2010 when the Stuttgart Stock Exchange created a segment called bondm. Four other stock exchanges quickly created similar segments: Munich (m:access), Dusseldorf (Mittelstandsmarkt), Frankfurt (Entry Standard), and Hannover/Hamburg (Mittelstandsboerse). The requirements for issuing a bond vary only slightly between these exchanges. Frankfurt and Hannover/Hamburg do not have a formal minimum. Stuttgart and Munich require a minimum issuance volume of EUR 25 million, and Dusseldorf of EUR 10 million. Denominations are usually limited to a maximum of EUR 1,000. Furthermore, only Dusseldorf and Munich require an official bond rating. Frankfurt and Stuttgart waive the rating requirement for public firms. Listing on the Hannover/Hamburg exchange generally requires no rating. All exchanges obligate issuers to mandate an official advisor (often called a coach). The advisor acts as the underwriter and/or issuing bank. It is usually a small unknown bank with only little experience in capital markets (Florstedt 2017).

The SME bond market was conceived favorably by investors at its inception, and the media covered the market extensively. Private investors, in particular, were lured by well-known names and the label “Mittelstand,” which is often linked to high productivity and quality standards and has become an internationally recognized stamp of high quality (Simon 2009). In addition, in the prevailing low-interest environment, investors welcome the opportunity of a new investment class that promises to yield an extra premium.

The German market for SME bonds has grown strongly since its launch in 2010. More than 200 bonds were issued between 2010 and 2018 (Bondguide 2019), and raised a total volume exceeding EUR 11 billion. The segment saw a large number of emissions in the first few years. When the first defaults appeared in 2014, accompanied by extensive press coverage, the emission activity decreased considerably. However, since 2018 the number of emissions has risen again.

As of 2019, more than one-third of all firms issuing an SME bond are unable to meet their obligations at some point or filed for bankruptcy (Bondguide 2019). After the first wave of defaults, Stuttgart and Dusseldorf closed their segments for SME bonds. While the other segments still exist, they were partly restructured. The high number of defaults is not only remarkable per se but also surprising to most market participants. The SME bonds initially had good credit ratings, and participants perceived the issuers as hidden champions (Mietzner et al. 2018). High default rates that deviate from the initial credit risks models indicate not only a vast variation in the quality of the

SMEs issuing bonds but also suggest that the well-established methods for assessing risk failed. Thus, the setting offers an excellent opportunity to examine the relationship between the actual bond defaults and the quality of the disclosed information.

Several researchers have investigated the German SME bond market. Mietzner et al. (2018) show that credit rating agencies regularly understate the credit risk inherent in SME bond offerings, and issue overly favorable ratings. The fact that ratings were too favorable is consistent with the observation that more SME bonds defaulted than would be expected, using historical default probabilities in the respective rating classes. The authors therefore conclude that investors cannot rely on ratings as their primary reference point to assess the investment risk.

Feihle and Lawrenz (2017) examine whether the performance of firms issuing SME bonds differs considerably compared to a matched sample of non-issuing firms. They show that issuers exhibit a substantially lower performance, as shown by a decline in proxies for financial performance, such as cash flows and total profitability, compared to non-issuers. This observation leads them to conclude that most of the unfortunate development in the performance of issuing companies was not predictable from the disclosed financial information before the issuance. The samples of issuing and non-issuing firms are ex-ante indistinguishable, but show different results ex-post.

Utz et al. (2016) empirically investigate the size of the liquidity premium in the spread of SME bonds. The authors find that illiquidity is highly correlated with the yield spread, when controlling for default risk. They predict that defaults in the future will further increase the perceived risk of this investment class, resulting in rising yield spreads, ending up in a vicious circle and market failure. To counter this development, the authors suggest that firms should disclose information more comprehensively and in a more timely manner to decrease yield spreads, thus underpinning the vital role of disclosure in this market.

In sum, current research highlights the critical role of disclosure in this market. Hitherto well-established metrics and proxies such as credit ratings and financial metrics do not seem to allow an adequate assessment of the underlying risk. However, little is known about the role of textual disclosure. The German SME bond market offers a suitable setting to study the role of summarized risk factor disclosure for several reasons: (1) The high number of actual defaults directly differentiates between low- and high-quality firms instead of relying on models that solely predict the default probability. (2) Due to the low nominal amount of EUR 1,000, high yields, and big brand names, mainly retail investors buy SME bonds. As the literature suggests, retail investors are particularly likely to rely on the summary of a prospectus. Thus, the setting allows measuring the role of summarized disclosure in a market where the informativeness of summaries is most relevant. (3) Financial disclosure of large public firms usually makes up only a small portion of the total information surrounding these firms, thus creating noise in the measurement of the level and quality of a firm's disclosure. By investigating the disclosure behavior of new entrants to public capital markets, we can analyze the direct relation between firm disclosure and performance.

Automatic Evaluation of Summaries

Following the tremendous growth in financial disclosures in recent years that tends to overwhelm investors with limited attention spans (Cardinaels et al. 2019; Dyer et al. 2017), regulators around the world have started to require that issuers of financial products release a summary conveying the essential information. In addition to numerous legal aspects and definitions that have to be considered in the case of mandated disclosures, summarizing information is challenging. Radev et al. (2002) define a summary as "a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that."

A summary aims at presenting the critical information of a text in a concise form. If every word and sentence in a text were equally important, summarization would be impossible. Any attempt to condense the text would result in a proportional decrease in informativeness. Fortunately, not every word in a text contains informative content (e.g., stopwords) so that a text can be segmented into more and less informative parts. However, distinguishing between uninformative and informative words is a crucial challenge in summarization (Radev et al. 2002).

Although there are different types of summaries (e.g., indicative, generic, topic-oriented summaries; Sparck-Jones 1999), the summary of a prospectus should be informative. Summarization by humans usually includes the identification of relevant content in the main text (extraction), the reformulation in other, usually fewer, words (abstraction), and the combination and compression of extracted material by squeezing out unimportant parts. Summaries written by humans are typically based on an individual's understanding of the text, thus leaving room for interpretation and impression management (Brandow et al. 1995). Against this backdrop, several researchers develop algorithms that create a summary automatically, by relying on statistical approaches for identifying and extracting informative parts to summarize large amounts of textual information without human efforts.[4]

[4] See Nenkova and McKeown (2011) for a review of the literature on automatized summarization.

However, evaluating the quality of a summary, generated by humans or automatically, is difficult, starting with the question of the dimension to evaluate. Traditional evaluation of summarization involves human judgments of metrics such as readability, conciseness, coherence, grammaticality, and content (Mani 2001). However, human evaluation of summaries requires enormous efforts in terms of time. Thus, in recent years, automatic evaluation of summaries has drawn massive attention in the scientific community. Despite the substantial progress in the assessment of summary quality in previous years, there is no universal approach. In the following, we will therefore identify several well-established metrics from the literature to automatically assess the quality of a summary. The quality of a summary is determined not only by its content but also by the way the information is presented (Lloret et al. 2018). We therefore propose, in line with the literature, to measure the quality of a summary along two standard criteria (Lloret and Palomar 2012): 1) the semantic quality and 2) the linguistic quality.

Measuring the Semantic Quality of Summaries

Previous studies evaluate short-form disclosures by comparing different types of summaries (i.e., a human versus an automatic generated) using text similarity measures, such as cosine similarity (Donaway et al. 2000), *n*-grams overlap, or longest common subsequence (Saggion et al. 2002). Following the successful application of automatic evaluation methods in automatic translation evaluation, such as BLEU (Papineni et al. 2002), which measures the overlap in *n*-grams between translations, Lin (2004) proposed with the ROUGE package several recall measures based on *n*-gram co-occurrence statistics between human- and automatic-generated summaries. However, these summary evaluation approaches rely on the assumption that there is an ideal summary against which they can compare other summaries in order to examine how informative they are. Thus, these methods fail when reference summaries are not available.

Several research projects therefore aim at evaluating summaries without the necessity of a human reference summary. Louis and Nenkova (2009) conducted several experiments to evaluate the semantic quality of a summary by comparing the summary to the original text instead of a human reference. They assume that useful summaries are representative of the input, and anticipate that higher similarity to the input is an indicator of high quality (Louis and Nenkova 2013). Based on Lin et al.'s (2006) work on the evaluation of automatic summaries, Louis and Nenkova (2013) test several input–summary similarity metrics that could be used for a fully automatic evaluation of summaries. The authors provide a freely available Java package, called SIMetrix (Summary Input similarity Metrics), to compute the evaluation metrics efficiently.[5] Their approaches include (1) information-theoretic metrics, such as the Kullback–Leibler (KL) divergence and the Jensen–Shannon (JS) divergence, to measure the divergence between the probability distributions of words in the main text and its summary, (2) vector space similarity metrics using cosine similarity between the term frequency–inverse document frequency (tf–idf) vector representations of the texts, (3) a generative model that compares the different word distributions, (4) topic signatures to quantify the presence of words that are highly descriptive of the input in a summary, and (5) the regression-based combination of all the features mentioned above (Lloret et al. 2018). Similar to Lin et al. (2006), Louis and Nenkova (2013) find that information-theoretic measures are the best for evaluating the semantic quality of summaries (while the JS divergence regularly outperforms the KL divergence). The authors also report that the so-created scores not only replicate scores based

[5] The SIMetrix package is available at http://homepages.inf.ed.ac.uk/alouis/IEval2.html (retrieved May 15, 2019).

on comparison with a human reference summary very accurately but are also highly predictive of scores provided by human raters. The correlations range between 0.77 and 0.93. The JS divergence measures and the combinations of the JS divergence with other metrics, such as cosine similarity, performed best. A disadvantage of their approach is, however, that they use only unigrams. As shown by Barrett et al. (2006), legal texts especially may require the use of *n*-grams (e.g., trigrams or tetragrams).

Similarly, Torres-Moreno et al. (2010) provide further evidence for the effectiveness of automatic summary evaluation based on measuring the divergences between probability distributions. Specifically, they show that the JS divergence can obtain reliable rankings of summaries using a content-based evaluation package called FRESA (Framework for Evaluating Summaries Automatically).[6] The package is coded in Perl, works with different languages (French, Spanish, English, and German), and offers the possibility to calculate the divergences using different kinds of *n*-grams.

More recently, Cabrera-Diego and Torres-Moreno (2018) propose a new method for evaluating the quality of a summary based on the dissimilarity among the summary, its source document, and other summaries of the same source. Their research is based on the idea that one can assess a summary better by increasing the information (in this the case, the number of summaries) available. Similarly to other studies on the topic, the authors find that JS divergence metrics perform the best.

Based on these findings, we measure the semantic quality of a summary using the JS divergence, which is based on the KL divergence, formally defined as:

---

[6] The FRESA package is available at http://dev.termwatch.es/~fresa/FRESA/index.html (retrieved May 15, 2019).

$$Kullback\text{–}Leibler\ Divergence = KLD(P\|Q) = \sum_{\omega} p_{\omega} \log_2(\frac{p_{\omega}}{q_{\omega}})$$

The KL divergence has some particular features. First, it is not symmetric, so that $KLD(P\|Q) \neq KLD(Q\|P)$, resulting in different metrics when measuring the input–summary and the summary–input divergence. Second, the metric is undefined, and gives infinite values when one word exists only in one distribution but not in the other, such as when $p_{\omega} > 0$ but $q_{\omega} = 0$. In contrast, the JS divergence is symmetric (which allows to interpret the JS divergence as a distance), and is always defined. We therefore employ the Jensen–Shannon divergence, which is based on the assumption that the distance between two single distributions cannot be considerably different from the average of the distances from their mean distribution (Louis and Nenkova 2013). Thus, the JS *divergence* is a symmetrized, finite, and smoothed version of the KL divergence (Mehri et al. 2015), and is formally defined as:

$$Jensen\text{–}Shannon\ Divergence = JSD(P\|Q) = \frac{1}{2}[KLD(P\|A) + KLD(Q\|A)]$$

where $A = (P + Q)/2$. The JS divergence is bounded by 1, given the use of the base 2 logarithm $0 \leq JSD(P\|Q) \leq 1$.

Measuring the Linguistic Quality of Summaries

Several aspects, such as the grammaticality, non-redundancy, referential clarity, and tonality, determine the linguistic quality of a summary (Vadlapudi and Katragadda 2010; Yadav and Chatterjee 2016). To assess texts grammatically, they are usually evaluated on how difficult they are to read. A large body of research in accounting and finance (e.g., Loughran and McDonald 2014) documents that the readability of corporate disclosure documents is significantly associated with capital market reactions. Less readable filings are related to higher costs of debt (Bonsall and Miller 2017) and are a sign of poor future performance (Li 2008). Most studies examining readability rely on Fog and Flesch indices as a proxy for readability, which were defined for the English language. However, some measures have been developed for other languages, either by adapting the weights of existing indices or by developing independent measures.

One of the most frequently used readability indices for German is the Wiener Sachtext Formel (WSF), which was developed by Bamberger and Vanecek (1984) to measure the readability of German texts.[7] The WSF is defined as follows:

$$WSF = 0.1935 \times LWS + 0.1672 \times SLW + 0.1297 \times LWC - 0.0327 \times SWS - 0.875$$

where LWS is the average number of long words with three or more syllables, LWC is the average number of long words with six or more characters, SWS the average number of short words with only one syllable, and SLW is the average sentence length in words. Similar to other readability metrics, such as the Flesch Grade Scale, the WSF score is scaled such that it roughly matches school grade levels. Accordingly, for most purposes, the range of the WSF is [4;15]; higher scores indicate more difficult texts. We derive *readability* scores for the summary, and the main document

[7] The use of other readability metrics for German texts, such as the Lix readability index (LIX), deliver similar results.

of each prospectus in the dataset, using the Readability Calculator, a Python package developed by Weiß (2016).[8]

Redundancy describes the unnecessary repetition of facts within a text. Following Lang and Stice-Lawrence (2015), we assume that using ubiquitous words is a sign of redundancy, because a disclosure that reiterates the wording of many other firms is unlikely to be informative. To identify redundancy, we compute a measure of *Boilerplate*, as suggested by Lang and Stice-Lawrence (2015). This measure is defined as the percentage of total words in sentences that contain a tetragram of words that appear in at least 60% of the documents of peer firms, based on two-digit industry codes.

To measure referential clarity, which is usually defined as the ease of identifying what specific phrases in the summary refer to, we follow Hope et al. (2016). We compute the specificity of the summarized disclosure and the full disclosure of risk factors. Specificity is defined as the number of specific entity names, including (1) names of persons, (2) names of locations, (3) names of organizations, (4) quantitative values in percentages, (5) money values in dollars, (6) times, and (7) dates, all scaled by the total number of words in the respective section. Thus, a higher *Specificity* value indicates disclosure of more specific risk factors. To construct the *Specificity* measure, we use the Stanford named entity recognition (NER) tool, which is also available for the German language (Faruqui and Padó 2010).[9]

---

[8] The package is available at https://github.com/zweiss/RC_Readability_Calculator (retrieved May 15, 2019).

[9] The tool is available at https://nlp.stanford.edu/software/CRF-NER.html (retrieved May 15, 2019).

Another critical aspect of assessing the quality of a summary refers to the tonality, which measures the positivity of a text relative to its negativity. As Cardinaels et al. (2019) show, management summaries are usually more positive and less negative in tonality compared to the underlying source document, indicating a positive bias in line with impression management theory. To compute the *tonality* of the summary and the main document, we use Bannier et al.'s (2019) words list, which is a translated and adjusted version of the English-language dictionary compiled by Loughran and McDonald (2011). The Loughran and McDonald (2011) word list has become the standard tool for textual analyses of disclosure documents in finance and accounting (Loughran and McDonald 2016).

To measure positive and negative sentiments in relation to each other, we follow Henry and Leone (2016) and Bannier et al. (2017), and estimate the relative sentiment of the risk sections as follows:

$$Tonality_i = \frac{POS_i - NEG_i}{POS_i + NEG_i}$$

where POS_i is the number of positive words, and NEG_i is the number of negative words. Tonality is scaled between -1, indicating a purely negative tone, and 1, indicating a purely positive tone.

To obtain the degree of change between the summary and the prospectus for each aspect of linguistic quality, we finally measure the relative change between the summarized and full disclosure of risk factors by taking the difference between the natural logarithms (plus 1, to account for negative values) for each metric of linguistic quality x_i (Törnqvist et al. 1985). This difference represents a symmetric and normed indicator of relative change, which is formally defined as:

$$Relative\ Change_i = \ln(1 + x_i^{Full}) - \ln(1 + x_i^{Summary})$$

As the interpretation of the sign of the relative change measure depends on the scale of the underlying variable (e.g., a higher readability change score indicates a decrease in quality, while a higher specificity change score represents an increase in quality), we unify the relative changes in such a manner that negative values always indicate a decrease in quality from the full text to the summary. We therefore multiply the relative change in specificity by -1, which is possible due to the symmetric nature of the relative change measured as the difference in natural logarithms.

Creating a Single Score of Summary Quality

Each of the proposed metrics focusses on a specific aspect of the overall quality of a summary. However, not only might semantic and linguistic quality of a text be independent, e.g. the readability of a summary might be irrelevant from a content perspective, but they might also be of different importance. As a consequence, the evaluation of a summary's quality should be based on the combination of several metrics (Ellouze et al. 2013), which might also be reflected in human evaluation schemes. One of the most representative human evaluation methods is the responsiveness score. It is defined as the degree to which a summary responds to a specific information need, considering the semantic, as well as linguistic, features (Owczarzak et al. 2012). Responsiveness is typically evaluated on a five-point scale, ranging from 1 (very poor) to 5 (very good).

To create a combined measure of a summary's quality, which consists of the five proposed individual metrics discussed above (i.e., divergence, readability, boilerplate, specificity and tonality) and is highly correlated to human evaluation, it is necessary to understand the drivers of human ratings. We therefore recruit participants from Clickworker platform (the German equivalent of Amazon's Mechanical Turk) to evaluate the responsiveness of the summaries. Following similar studies on financial accounting (e.g., Bonner et al. 2014; Cardinaels et al. 2019), we select individuals based on their financial experience and knowledge. Specifically, participants are required to have made at least five bond or stock transactions over the previous two years and to rate their self-reported knowledge about the characteristics and risks of bonds or stocks 6 or higher on a scale of 10. Moreover, the individuals were required to be native German speakers, and to be at least 18 years old.

We provide each person with the summary of the risk factors, the full risk factor section, and a statement that the information must be covered by the summary (which is based on the original wording on the aim of a summary laid out by the Prospective Directive). In line with the legal definition of risk factors, we ask the participants to assign a score that reflects to what extent the summary allows them to assess the risk associated with the security compared to the full version of the risk factors. At least three different persons evaluate each summary, and all firm names are anonymized. The order in which participants are provided the summary or the full risk factor section is randomized. On completion, participants are paid EUR 13.00 per hour via Clickworker and need 32 minutes, on average. A Krippendorf's Alpha of 0.872 indicates an acceptable level of intercoder reliability.

To examine which aspects of linguistic and semantic quality affect the human responsiveness score, we then employ a plain OLS regression model with the human responsiveness scores as the dependent variable and the five automatically retrieved metrics (divergence, readability, boilerplate, specificity and tonality) as independent variables. We obtain the relative importance of the individual metrics by using relative weight analysis (RWA) using RWA-Web provided by Tonidandel and LeBreton (2015).[10] This approach decomposes the total variance in the form of the R-squared of the OLS model into pseudo-orthogonal weights, each reflecting the relative contribution of each independent variable. The relative weights represent the relative effect sizes, and identify which predictors explain non-trivial variance in an outcome, even in the presence of correlated predictors (Kulik et al. 2016; Tonidandel and LeBreton 2011).

These relative weights are used to create a combined metric for the quality of the summary (Z. Lin et al. 2012). To obtain an absolute score that is independent of the scale of the metrics,[11] we create quintiles of the relative change in each metric, multiply them by the respective relative weight, and take the sum of the individual weighted scores.[12] Similar to the human responsiveness score, the summary quality score is scaled between 1 and 5. Higher values indicate higher quality in terms of semantic and linguistic quality.

Measuring the Informativeness of Summary Quality

After having established the summary quality, we use a plain OLS regression to assess which firm- and bond-specific characteristics explain the individual aspects of the summary quality, as well as the overall score. In line with hypothesis H1, we expect that poorly performing firms are more likely to release a low-quality summary. To test the second hypothesis, namely, whether the quality of a summary is informative for investors, we employ a probit regression with the default of the

[10] The tool is available at http://relativeimportance.davidson.edu (retrieved May 1, 2019).

[11] For example, the divergence measure is bounded between 0 and 1, while the other metrics can take negative values.

[12] Alternative scaling approaches, such as Z-score normalization and min-max scaling, yield quantitatively and qualitatively similar results.

bond as the dependent variable. We expect the summary quality to be a statistically significant and negative predictor of this adverse future outcome.

It needs to be recognized, however, that the summary quality is likely to be an endogenous regressor in these analyses. Though we include an extensive array of control variables capturing the firm's quality, as well as industry and time fixed effects that help to alleviate the endogeneity concern, there could still be an omitted variable bias. Managers arguably have private information (i.e., knowledge about the materiality of specific risk factors) that systematically affects firm performance and, therefore, their decision what and how to disclose. Such unobservable factors could bias the probit estimates. The standard textbook solution for such a problem is to employ a two-step instrumental variable estimation approach (Greene 2003; Wooldridge 2002). We use an exogenous variable, called an instrument, in the first estimation stage to obtain the predicted values of the potentially endogenous variable that will be uncorrelated with the error term. We then use these predicted values in the second stage to predict the dependent variable of interest. As the endogenous variable (summary quality) is continuous, and the variable of interest (default) is binary, we employ an IV-probit regression approach. However, the challenging task in such models is to find a valid instrument (Larcker and Rusticus 2010).

A valid instrument in our context is a statistically significant determinant of summary quality without having a direct effect on the bond default. A large body of papers in finance and accounting suggests that litigation risk could be a valid instrument, as it is an essential driver of high-quality disclosure (e.g., Field et al. 2005; Hanley and Hoberg 2012). Managers (and lawyers) who are aware of recent activity of the respective authority should be inclined to avoid the risk of sanctions by releasing a high-quality summary (Dechow et al. 2011; Kim and Skinner 2012). To create a measure of litigation risk, we manually extract the number and types of security-related enforcement actions from the Federal Financial Supervisory Authority (BaFin) annual reports, which are publicly available on the website. An overview of the types of fines proceedings by the BaFin over time is provided in Table A3 in Appendix A. The variable *Fines Proceedings* captures the number of all actions taken by the BaFin during the six months before the issuance of a specific bond. In the next section, show that this variable is indeed a valid instrument.

RESULTS

Descriptive Statistics

We define SME bonds as financial securities that are or were listed in the respective segments on any of the five SME bond exchange segments in Germany and hand-collect the names of the bonds from the webpages of the respective exchanges. The final dataset includes 159 bonds issued between 2010 and 2016, with a total issue volume of more than EUR 7 billion. We retrieve bond and firm characteristics from several databases, such as Dafne and Amadeus. Variable definitions are provided in Table A1.

Textual Characteristics

We initially retrieve the bond prospectuses as PDFs, and convert them into text files using the pdfminer package. We let the SoMaJo Python package[13] (Proisl and Uhrig 2016) split the text of each prospectus into sentences. Next, we follow Li (2008) and Lang and Stice-Lawrence (2015),

[13] The package is available at https://github.com/tsproisl/SoMaJo (retrieved May 15, 2019).

and remove all sentences that do not contain at least 30% alphabetic characters, as well as sentences that have fewer than five alphabetic characters. We then remove stopwords using the Natural Language Toolkit for common German words, Bannier et al.'s (2019) finance-specific list of stopwords,[14] and the names of the firms. Finally, we tokenize and lemmatize all words using the GermaLemma[15] package. Lemmatizing is preferred over stemming, as it not only removes grammatical endings from words but also takes the morphological analysis of the words into account, to combine words with similar meanings.

Insert Table 2 about here

Insert Table 3 about here

Insert Figure 1 about here

Table 2 presents the raw numbers of pages of the relevant parts of the prospectus. On average, an entire prospectus spans about 195 pages, and the summary 14 pages. The risk section consists of

[14] The list is available at http://www.uni-giessen.de/fbz/fb02/forschung/research-clusters/bsfa/textual_analysis (retrieved May 15, 2019).

[15] The package is available at https://github.com/WZBSocialScienceCenter/germalemma (retrieved May 15, 2019).

14 pages, and the summary of the risk section is 3.6 pages, on average, which represents a reduction in terms of pages of roughly 75%.

The number of words before and after removing stopwords is shown in Table 3. Roughly 50% of the words are stopwords. Before removing them, the risk section is roughly 12% of the full prospectus, while the summarized risk section constitutes about one-third of the summary. The reduction from the full risk factor section to the summarized risk factor section is about 80%.

Figure 1 shows the average number of pages and words over time. As one can see, the effect of the regulatory change in mid-2012 on the length of summaries is quite limited. On average, the number of words increases by roughly 3%, and the difference is not statistically significant.

Bond Characteristics

Bond characteristics are reported in Table 4. The average SME bond has a target volume of around EUR 46.6 million,[16] an average spread of 6.6%,[17] more than seven covenants, and a maturity of around five years. Around 40% of the bonds are collateralized, and more than 70% of the bonds have a rating.[18] Most bonds are rated BB or higher, indicating an average probability of default (PoD) of 0.3% to 1.5% over a 12-month time horizon according to the Creditreform Rating Map (2019). Fifteen percent of the bonds have an investment-grade rating (minimum rating of BBB, respectively a PoD of less than 0.1%). Fifty bonds (31.45%) defaulted by March 2019.

[16] On average, the SME bonds raise 82.81% of their target volume, and 46% (73) of the bonds raise the full requested amount.

[17] The average coupon is 7.2%. The yield spread of an SME bond is calculated as the difference between the bond's coupon and the coupon of a benchmark government bond that has a very similar maturity and currency. In this case, I use coupons of 2-, 5-, and 10-year German bonds as the risk-free rate.

[18] These ratings usually come from small, local rating agencies, such as Scope, Euler Hermes, Feri, or Creditreform.

Insert Table 4 about here

Firm Characteristics and Financials

Around 85% of the SME bond issuers are non-public firms, and thus, inexperienced in raising funds from public capital markets. On average, issuing firms have around 765 employees, and are 23 years old at the time of the bond issuance. Based on the Federal Statistical Office's classification of economic activities, the firms issuing SME bonds are active in 10 different industries, mainly power supply construction, and financial services (see Table A2 in Appendix A).

We calculate the implied PoD as the transformed o-score following Ohlson (1980). The result shows that the average implied PoD is roughly 15%. This is more than 10 times as high as the PoD implied by the agency ratings. Furthermore, we see that the average firm has EUR 430,000 of total assets, a liquidity ratio of approximately 90%, and an average leverage ratio of 1.3. The average growth in revenues during the 3-year period before the bond is issued is 1.4%.

Quality of Summaries

As can be seen from Table 5, on average 2.9% of the words in the risk section are boilerplate. Compared to 2.4% of boilerplate words in the summary, the relative change is small but positive. This means that managers provide, on average, higher-quality summaries of risk factors due to less use of boilerplate sentences. This positive result does not hold for specificity. On average, 4.6% of the words in the full sections refer to a specific entity, but this value decreases to 3% in the summary. Managers are hence less specific in the summary, which is usually interpreted as a sign of low-quality disclosure. Though readability is quite good, as indicated by low average scores of 3.27 for the full risk factors and 4.18 for the summarized risk factors, the summaries are more challenging to read as indicated by a negative relative change of -0.186. Unsurprisingly, risk factor disclosures are generally negative in tonality. More importantly, however, the descriptive statistics show that managers use the summary to present information in a more positive light, indicated by a negative relative change in tonality of -0.161. With regard to the semantic quality of the summary, we find that the distance is quite large with an average of 22.8%, indicating low average semantic quality (Mehri et al. 2015).

In sum, the comparison of the summarized risk factors and the full risk factors shows that, on average, the summaries are of lower quality, indicated by lower readability, lower specificity, more positive tonality and high semantic distance. Thus, the negative picture of summaries in the prospectus, as suggested by Hanley and Hoberg (2010), can be confirmed for the risk factor section based on the descriptive statistics.

Insert Table 5 about here

Summary Scores

In order to combine the individual summary quality components into an overall quality score, we run an OLS regression to examine which of the quality metrics drive human responsiveness rating scores of the summaries. The corresponding results are provided in Table 6. They show that each metric is highly statistically significant and positive, indicating that each component is an important aspect when evaluating a summary. Overall, the five metrics jointly explain 60.9% of the total variance, as shown in Column (5).

To obtain relative effect sizes, we employ a relative weight analysis, and report the results in Table 7. The sum of the relative weights equals the R-squared of the OLS regression in Table 6 (0.609). The standardized weights are obtained by dividing the relative weights by the model R-squared and sum up to 1. They quantify the proportional contribution, often referred to as the importance, of each metric to the explained variance.

The standard errors of the relative weights are based on bootstrapping procedures, with 95% bias-corrected confidence intervals using 1,000 replications (Tonidandel and LeBreton 2011). All confidence intervals exclude zero, thus confirming the statistical significance of the semantic and linguistic quality metrics.

Ranking the relative weights demonstrates that semantic quality, measured by the JS divergence, is the most critical driver of human evaluations, contributing 47% of the variance explained. The second most important metric explaining human responsiveness scores is the use of boilerplate sentences in the summary, compared to the full risk factor section, which constitutes 21% of the explained variance. The changes in specificity and readability are both contributing roughly 12%, followed by 7% from the relative change in tonality.

Based on this initial step, we then use the standardized weights to create a single metric of summary quality. We create quintiles of each metric and multiply these values by the respective standardized relative weight. The upper panel in Table 8 presents the summary quality score before and after the weighting. The unweighted scores refer to the average of the quintiles for each metric.

The middle panel of Table 8 reports the descriptive statistics for the human ratings and the scores of the automatic evaluation. On average, human coders evaluate the summary with a score of 2.9. The lowest score for a summary is 1.75, the highest 4.5. The lower panel in Table 8 provides the correlation matrix for the human ratings and the automatic summary scores. The correlations are generally high (with values around 0.80). However, the results suggest that the weighted score performs best in explaining the quality of summaries evaluated by humans.

Insert Table 6 about here

Insert Table 7 about here

Insert Table 8 about here

Determinants of Summaries Quality

To assess whether firm characteristics determine a summary's quality, we employ an OLS regression with each quality metric as well as the overall summary quality score as the dependent variable in turn. The results are given in Table 9 and confirm the conjecture that firm performance is a statistically significant driver of summary quality.

More precisely, firms with high implicit default risks as indicated by the Ohlson o-score disclose statistically significant lower-quality summaries, and make more use of boilerplate sentences. Collateralization significantly explains many of the quality metrics as well as the overall score. The growth in sales over the recent year, the number of covenants, and whether the firm is publicly listed partially explain the quality of a summary. From an economic standpoint, a one standard deviation increase in the o-score decreases the overall quality by 0.163. Given an average of 2.964 and a standard deviation of 0.776 in the summary score, this change is substantial.

All specifications control for industry- and time-fixed effects. To account for inherent problems of heteroskedasticity and multicollinearity in regressions using cross-sectional data, we follow common practice in research on disclosure (e.g., Bushee et al. 2003) and perform several tests to confirm the robustness of our results. First, we compute all standard deviations using White's heteroskedasticity-adjusted standard errors. Second, the analysis of residuals does not indicate problems due to heteroscedasticity or non-normal distributions. Third, the collinearity diagnostics show that all variance inflation factor values range from 1.20 to 1.68, suggesting that multicollinearity is not an issue.

Overall, the results of the regression examining the determinants of summary quality confirm that poorly-performing firms engage in biased reporting; that is, they use the summary for impression management. Managers seem to strategically summarize the risk factors in a manner that is intended to manipulate investors' perceptions and decisions. However, whether firms' use of impression management is also informative for investors, depends on the degree to which the summary bias allows investors to draw inferences about the future market performance of the bonds. We examine this important question in the next section.

Insert Table 9 about here

Informativeness of Summaries

Table 10 presents the results of the probit regression with default of the bond as the dependent variable in all specifications. The reported coefficients are average marginal effects. As indicated by Columns (1) to (5), not all quality metrics predict the default of a bond significantly. Only semantic quality measured by the JS divergence and the relative change in specificity are statistically significant when specified individually. However, the negative signs of all quality metrics indicate that firms that issue high-quality summaries are less likely to default on their bonds.

Joint estimates of the effect of the quality metrics on bond defaults are shown in Column (6). Again, all metrics exhibit the predicted (negative) signals, and are statistically significant. The combined and weighted quality score in Column (7) is highly statistically significant at the 1% level, and economically relevant. A one standard deviation increase in the combined quality score reduces the likelihood of failure by around 15.4%.

Among the control variables Ohlson's o-score and collateralization turn out to be statistically significant predictors of future performance. Interestingly, ratings and spreads do not show significant effects. These findings are in line with existing research on the market segment that emphasize credit rating agencies regularly understate the credit risk inherent in these offerings, and issue overly favorable ratings (e.g., Mietzner et al. 2018).

To evaluate the explanatory power of the probit regression model, we rely on the area under the receiver operating characteristic curve (AURROC), which plots the likelihood of identifying true signals (sensitivity) and false signals (1-specificity) for the entire range of possible cutoff points

(Hosmer and Lemeshow 2012). The AUROCC can be interpreted as the percentage of correctly classified events (sensitivity) versus specificity, expressed as 1 minus the percentage of correctly classified non-events, thus naturally ranging from 0.5 to 1.0. Following Hosmer and Lemeshow (2012), the minimum AUROCC value is 0.7; however, adequate values should range above 0.8. All higher values are excellent. The 0.899 AUROCC value for Column (7) indicates that the model is well suited to differentiate between the two possible outcomes.

Insert Table 10 about here

Robustness of Results

Although all previously discussed results suggest that low-quality firms are more likely to engage in impression management, and that this tactic is informative for investors, unobservable characteristics might bias the results.

Table 11 provides the results of an instrumental variable estimation to account for the endogeneity of the summary quality. Column (1) mirrors the specification of Column (7) in Table 10. Columns (2) and (3) present the results of the IV probit approach. Column (2) suggests that litigation risk, measured by the number of enforcement actions by the BaFin during the six months before the bond emission is a statistically significant predictor of the summary quality. The statistically significant F-statistic confirms the strength of this instrument. The summary quality coefficient in Column (3) exhibits the same direction as the plain probit model at a slightly lower significance level. However, the Wald chi-square test of exogeneity indicates that endogeneity might be not a problem, and that the effect is well explained by the observable variables that are included. Taken together, these results confirm that our main results are robust.

Insert Table 11 about here

CONCLUSION

Discussion and Implications

Automatic evaluations are a challenging, but feasible, way to assess the informativeness of disclosed information from an investor perspective. In this study, we use multiple methods to investigate how to evaluate summarized information in security prospectuses, and examine the informativeness of these documents. Thus, our study responds directly to Barth's (2015, p. 506) call for research on how to "summarize, aggregate, and present information in financial reports to aid investors and other outside providers of capital in their decision making."

In this study, we develop a new measure to automatically evaluate the quality of a prospectus summary, and show that poorly-performing firms use these summaries to engage in impression management. Firms strategically select, frame, and present textual information in disclosure documents in a way that is intended to manipulate investors' perceptions and decisions. The manipulation includes statistically significant differences in readability, specificity, tonality, use of boilerplate, as well as selectivity of content. Automatic evaluations of summaries correlate highly with human evaluations. Further, we show that the historical performance of a firm determines the quality of a summary. Moreover, discretionary disclosure practice in prospectus summaries is informative for investors, as the quality of a summary is a highly statistically significant predictor of the future performance in the form of the default of the bond.

This study contributes to the literature on disclosure in several ways. To the best of our knowledge, this study is the first to empirically analyze the use of impression management strategies in summarized information, and to relate firms' risk-disclosure practices in summaries to future performance. We thus contribute to the nascent stream on summarization in accounting (e.g., Cardinaels et al. 2019), by developing a reproducible approach for measuring the quality of a summary.

Furthermore our study informs policymakers and regulators with regard to the generation and effects of summarization in financial disclosures. The findings suggest that the current regulatory scheme in the European Union leaves room for manipulation, and supports recent calls for higher quality in the summarized presentation of risk factors in prospectuses (CFA Institute 2017).

Moreover, our results provide investors helpful knowledge on how to evaluate textual disclosure. Given the tremendous amount of information in securities markets, summaries will become increasingly important. Investors need to learn how to understand the summarized information.

Limitations and Future Research

Our study is subject to several limitations, which open up avenues for future research. First, the analysis is limited to the evaluation of summaries of risk factors in prospectuses in a European retail debt market setting. Future research could test how different types of investors and stakeholders (e.g., professional investors or analysists) use summarized information, and the consequences. Although risk factors are an important element of a prospectus, future studies could evaluate the role of summarized information based on different outlets (e.g., earnings releases or conference calls). Moreover, it would be valuable to expand the research to other security markets and countries with different regulations.

Second, this research focuses on specific quality metrics for evaluating a summary's quality. However, other elements may affect human assessment. For example, it would be interesting to examine the content of the summary in greater detail, to understand whether firms strategically select certain information or specific risk factors (Campbell et al. 2018; Chiu, Kim, et al. 2018).

Third, new methods, such as machine learning, can improve the summarization of information (Silva et al. 2015). Future work could assess whether such methods could be used to evaluate the quality of a summary. These approaches require reference summaries for training. However, it would be possible to understand further which aspects are the relevant drivers for human assessments.

Fourth, it would be important to expand our knowledge on how summarized information affects investors' perception, and to examine managers' incentives to engage in impression management further. Such insights would help a great deal to implement future regulatory actions.

Finally, the upcoming change in the prospectus law in July 2019 will enable researchers to examine the effectiveness of certain new requirements. There are new summary rules, including a limit on listing only the 15 most material risk factors and a restriction on the length of summaries to a maximum of seven pages. Most interestingly, firms are required to assess the materiality of the risk factors, using a qualitative scale, which examines the informativeness of the disclosed information further (Fagernäs et al. 2019).

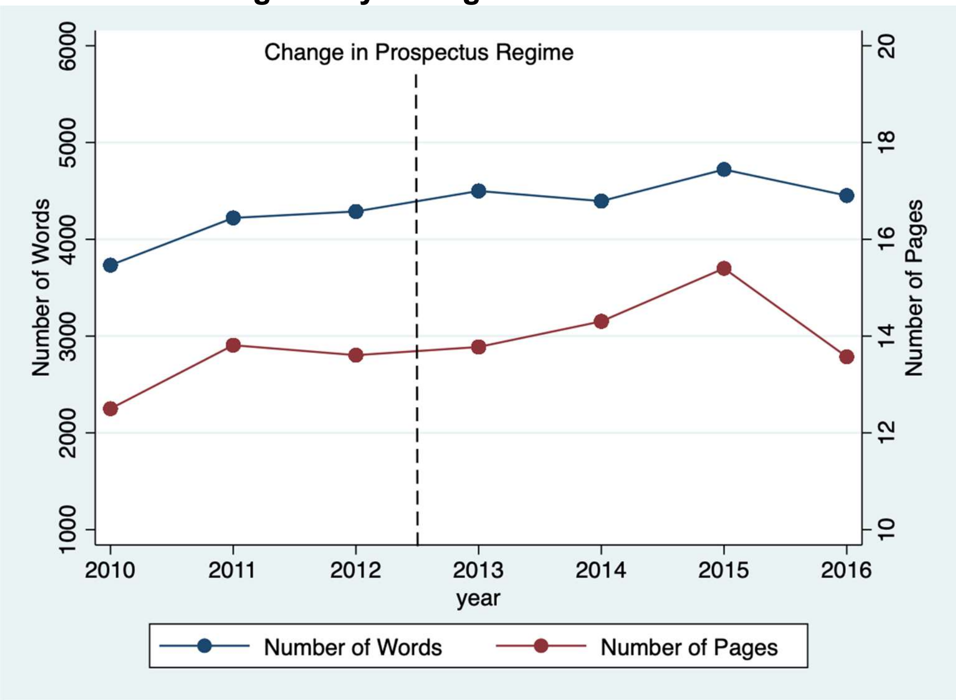
Concluding Remarks

In this study, we investigate the informativeness of summarized risk factor disclosure. In particular, we examine (1) how firms summarize information, by developing an automatic approach for evaluating the quality of a summary, and (2) whether the quality of a summary provides valuable information about a security's future performance.

The results suggest that firms with poor performance often provide low-quality summaries, and engage in impression management. These firms often strategically select, frame, and present textual information in the summary of the security prospectus in a way that is likely to be intended to manipulate investors' perceptions and decisions. The manipulation includes significant differences in readability, specificity, tonality, use of boilerplate, as well as dissimilarity in content between the summary and the full prospectus. Moreover, discretionary disclosure practice in prospectus summaries is informative for investors, as the quality of a summary is a highly statistically significant predictor of future performance in the form of the default of the bond.

FIGURES AND TABLES

Figure 1: Length of Security Prospectuses in the German SME Bond Market Before and After Regulatory Change



Notes: This figure presents the average number of pages and words (after removing stopwords) of the summaries in the prospectuses provided by 159 German SME's issuing a bond between 2010 and 2016 on a regulated stock exchange in Germany over time. The Amending Directive 2012/73/EU was implemented in German law on July 1, 2012. The means before and after the regulatory change are not statistically significant different at a 10%-level.

Table 1: Overview of Legal Framework on Disclosure of Risk Factors in the European Union

Legal Basis	Directive 2003/71/EC Commission Regulation (EC) No 809/2004	Directive 2010/73/EC Commission Regulation (EC) No 486/2012
German Law	Securities Prospectus Law July 1, 2005	Securities Prospectus Law July 1, 2012
Length of Summary	"shall not exceed 2500 words"	"shall not exceed 7 % of the length of a prospectus or 15 pages, whichever is the longer"
Content of Summary	specified (11 sections, 27 elements)	specified (5 sections, 87 elements)
Order of Content in Summary	not specified	specified
Layout of Summary	not specified	specified (Tables)
References in Summary	not allowed	not allowed
Risk Factors in Summary	yes	yes
Risk Factors in Prospectus	yes	yes
Definition of Risk Factors	"material to the securities being offered"	"material to the securities being offered"
Liability for Summary	"only if the summary is misleading, inaccurate or inconsistent when read together with the other parts of the prospectus"	"only if the summary is misleading, inaccurate or incon-sistent when read together with the other parts of the prospectus or it does not provide, when read together with the other parts of the prospectus, key information in order to aid investors when considering whether to invest in such securities"
Summary Exemptions	no exemptions	"non-equity securities having a denomination of at least 100,000 euros"
Language of Summary	"should be written in non-technical language"	"should be drafted in plain language, presenting the information in an easily accessible way"

Table 2: Number of Pages in Security Prospectuses in the German SME Bond Market

	Full Prospectus	Risk Section	Summary	Summarized Risk Section
Mean	195.47	14.15	13.79	3.60
SD	92.76	5.17	4.50	1.57
Median	176.00	14.00	14.00	3.00
Min	59.00	3.00	3.00	1.00
Max	651.00	36.00	34.00	9.00

Notes: This table presents descriptive statistics of the security prospectuses released by 159 German SME's issuing a bond between 2010 and 2016 on a regulated stock exchange in Germany. The number of pages is manually retrieved from the original PDF files.

Table 3: Number of Words in Security Prospectuses in the German SME Bond Market Before and After Removing Stopwords

	Full Prospectus	Risk Section	Summary	Summarized Risk Section
After removing stopwords				
Mean	52,300.00	6,692.33	4,341.13	1,241.09
SD	28,100.00	2,758.67	1,674.36	726.56
Median	47,600.00	6,169.00	4,219.00	1,013.00
Min	13,100.00	876.00	942.00	297.00
Max	204,000.00	18,300.00	12,900.00	4,363.00
After removing stopwords				
Mean	27,764.28	3,253.43	2,159.17	608.44
SD	17,091.00	1,364.70	841.39	360.04
Median	24,676.00	3,013.00	2,085.00	498.00
Min	6,951.00	434.00	464.00	151.00
Max	138,698.00	9,132.00	6,523.00	2,378.00
Share of Stopwords	46.91%	51.39%	50.26%	50.98%

Notes: This table presents descriptive statistics of the security prospectuses released by 159 German SME's issuing a bond between 2010 and 2016 on a regulated stock exchange in Germany. Stopwords are removed by employing the Natural Language Toolkit for common German words as well as by removing finance-specific German stopwords as provided by Bannier et al. (2019), and the names of the firms.

Table 4: Firm and Bond Characteristics

Variable	N	Mean	SD	Median	Min	Max
Employees	159	765.396	578.171	108.000	587.000	2865.000
Firm Age	159	23.604	24.938	1.000	17.000	131.000
Public	159	0.157	0.365	0.000	0.000	1.000
Collateralization	159	0.396	0.491	0.000	0.000	1.000
Covenants	159	7.786	4.929	0.000	8.000	16.000
Spread	159	0.066	0.011	0.033	0.066	0.108
Nominal Volume	159	46.623	49.134	1.000	34.000	300.000
Maturity	159	5.327	1.334	1.000	5.000	10.000
Investment Grade	159	0.151	0.359	0.000	0.000	1.000
No Rating	159	0.283	0.452	0.000	0.000	1.000
O-Score	159	0.151	0.100	-0.007	0.138	0.440
Size	159	5.183	1.506	2.335	5.054	7.626
Liquidity	159	0.899	0.479	0.052	0.879	1.884
Sales Growth	159	0.014	0.044	-0.192	0.012	0.219
Leverage	159	1.311	0.837	0.000	1.250	3.293

Notes: This table presents descriptive statistics on firm and bond characteristics of 159 German SME's issuing a bond between 2010 and 2016 on a regulated stock exchange in Germany. Data were retrieved from the respective prospectus and several databases. Financial metrics and firm characteristics correspond to the year prior to bond issuance. A description of all variables is provided in Table A1 in Appendix A.

Table 5: Descriptive Statistics of Semantic and Linguistic Quality of Security Prospectuses in the German SME Bond Market

	Summarized Risk Factor Section	Full Risk Factor Section	Relative Change
Boilerplate			
Mean	0.024	0.029	0.004
SD	0.013	0.008	0.011
Median	0.020	0.030	0.005
Min	0.000	0.010	-0.020
Max	0.070	0.055	0.035
Specificity			
Mean	0.030	0.046	-0.016
SD	0.019	0.030	0.026
Median	0.030	0.040	-0.010
Min	0.000	0.010	-0.085
Max	0.100	0.130	0.049
Readability			
Mean	4.175	3.271	-0.186
SD	0.774	0.442	0.175
Median	4.280	3.272	-0.201
Min	2.147	2.045	-0.662
Max	7.678	5.001	0.234
Tonality			
Mean	-0.829	-0.865	-0.161
SD	0.096	0.044	0.615
Median	-0.865	-0.870	-0.139
Min	-0.941	-0.971	-2.209
Max	-0.391	-0.730	0.993
Divergence			
Mean	0.228		
SD	0.108		
Median	0.188		
Min	0.097		
Max	0.876		

Notes: This table presents descriptive statistics on the quality of prospectus summaries of 159 German SME's issuing a bond between 2010 and 2016 on a regulated stock exchange in Germany. Relative change is meassured as the difference of the natural logarhitms. Negative values in the relative change indicate a decrease in quality. A description of all variables is provided in Table A1 in Appendix A.

Table 6: OLS Regression Results of Semantic and Linguistic Quality on Human Ratings of Responsiveness

	(1)	(2)	(3)	(4)	(5)	(6)
Relative Change in Boilerplate	20.057 ***					23.823 ***
	(5.038)					(3.486)
Relative Change in Specificity		6.163 ***				8.269 ***
		(2.001)				(1.328)
Relative Change in Readability			1.014 ***			1.029 ***
			(0.308)			(0.211)
Relative Change in Tonality				0.266 ***		0.182 ***
				(0.087)		(0.054)
Divergence					3.266 ***	3.469 ***
					(0.481)	(0.480)
Constant	2.811 ***	3.000 ***	3.089 ***	2.943 ***	2.157 ***	2.359 ***
	(0.054)	(0.062)	(0.079)	(0.056)	(0.116)	(0.121)
Observations	159	159	159	159	159	159
R-Squared	0.111	0.055	0.070	0.060	0.275	0.609

Notes: This table presents the unweighted and weighted score for a summary's quality based on the relative weights analysis from Table 7. The sample consists of 159 firms that issued a bond between 2010 and 2016 on a regulated stock exchange in Germany. The variable definitions are provided in Table A1 in Appendix A. Robust standard errors appear in brackets. Significance levels are as follows: *=10%, **=5%, ***=1%.

Table 7: Relative Weights Analysis of Semantic and Linguistic Quality on Human Ratings of Responsiveness

	Relative Weight	Standardized Relative Weight	Bootstrapped Standard Error	95% Confidence Interval
Relative Change in Boilerplate	0.130	0.213	(0.041)	[0.063 ; 0.223]
Relative Change in Specificity	0.075	0.123	(0.031)	[0.027 ; 0.142]
Relative Change in Readability	0.071	0.116	(0.031)	[0.022 ; 0.140]
Relative Change in Tonality	0.043	0.071	(0.023)	[0.011 ; 0.097]
Divergence	0.291	0.477	(0.046)	[0.205 ; 0.386]
Observations		159		
R-Squared		0.609		

Notes: This table presents the results of the relative weights analysis based on the regression provided in Table 6. The sample consists of 159 firms that issued a bond between 2010 and 2016 on a regulated stock exchange in Germany. The variable definitions are provided in Table A1 in Appendix A.

Table 8: Weighted Score of Summary Quality and Correlations with Human Evaluations of Responsiveness

	Human Rating	Unweighted Score	Weighted Score
Relative Change in Boilerplate		2.918	0.620
Relative Change in Specificity		2.962	0.363
Relative Change in Readability		2.987	0.347
Relative Change in Tonality		2.987	0.212
Divergence		2.975	1.420
Mean	2.900	2.966	2.964
SD	0.673	0.636	0.776
Median	3.000	3.000	3.052
Min	1.750	1.600	1.542
Max	4.500	4.600	4.645
Correlation Matrix			
Human Rating	1.000		
Unweighted Score	0.785	1.000	
Weighted Score	0.838	0.806	1.000
Observations	159	159	159

Notes: This table presents the unweighted and weighted scores for a summary's quality and their correlation with human ratings of responsiveness. Weighted scores are obtained by dividing the relative changes of the content and linguistic quality metrics in quintiles and multiplying them with the relative weights from Table 7. The sample consists of 159 firms that issued a bond between 2010 and 2016 on a regulated stock exchange in Germany. The variable definitions are provided in Table A1 in Appendix A.

Table 9: OLS Regression Results of Firm and Bond Characteristics on Semantic, Linguistic, and Overall Summary Quality

	(1)	(2)	(3)	(4)	(5)	(6)
	Relative Change in Boilerplate	Relative Change in Specificity	Relative Change in Readability	Relative Change in Tonality	Divergence	Summary Quality
Collateralisation	0.001 ***	0.003	0.107 ***	0.277 **	0.014	0.374 **
	(0.000)	(0.005)	(0.034)	(0.129)	(0.023)	(0.154)
Covenants	0.000	0.001	0.000	0.031 ***	0.000	0.002
	(0.000)	(0.000)	(0.003)	(0.009)	(0.002)	(0.014)
Spread	-0.003	-0.172	1.914	3.095	-0.510	-1.525
	(0.008)	(0.253)	(1.627)	(5.343)	(0.805)	(6.301)
Nominalvolume	0.000	0.000	0.000	0.000	0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Maturity	0.000	-0.001	-0.003	-0.028	0.000	0.018
	(0.000)	(0.002)	(0.010)	(0.040)	(0.008)	(0.045)
Investment Grade	0.000	0.005	0.030	0.227 *	0.030	-0.084
	(0.000)	(0.006)	(0.043)	(0.130)	(0.041)	(0.205)
No Rating	0.000	0.000	0.011	0.106	0.010	0.021
	(0.000)	(0.005)	(0.037)	(0.126)	(0.026)	(0.151)
Employees	0.000	0.001	-0.010	-0.175 *	0.006	-0.012
	(0.000)	(0.004)	(0.025)	(0.104)	(0.019)	(0.121)
Firm Age	0.000	0.000	0.001	0.001	0.000	0.001
	(0.000)	(0.000)	(0.001)	(0.002)	(0.000)	(0.003)
Public	0.000	-0.004	0.009	0.095	-0.040 **	-0.028
	(0.000)	(0.006)	(0.037)	(0.112)	(0.018)	(0.187)
O-Score	-0.003 ***	-0.019	-0.097	-0.356	-0.067	-1.625 **
	(0.001)	(0.025)	(0.179)	(0.588)	(0.101)	(0.714)
Size	0.000	0.001	-0.003	-0.040	-0.005	-0.051
	(0.000)	(0.002)	(0.011)	(0.036)	(0.007)	(0.043)
Liquidity	0.000	0.002	0.027	-0.103	-0.015	-0.176
	(0.000)	(0.005)	(0.035)	(0.098)	(0.018)	(0.142)
Sales Growth	0.003 *	-0.097 **	-0.625 **	-1.310	-0.008	-0.809
	(0.002)	(0.046)	(0.290)	(1.425)	(0.125)	(1.046)
Leverage	0.000	0.000	-0.003	-0.057	0.019	-0.011
	(0.000)	(0.003)	(0.020)	(0.064)	(0.015)	(0.087)
Constant	0.001	-0.014	-0.326 **	0.131	0.320 ***	3.681 ***
	(0.001)	(0.029)	(0.160)	(0.553)	(0.100)	(0.743)
Industry FE	yes	yes	yes	yes	yes	yes
Time FE	yes	yes	yes	yes	yes	yes
Observations	159	159	159	159	159	159
R-squared	0.385	0.211	0.218	0.289	0.179	0.254

Notes: This table presents the coefficients from OLS regression, where the dependent variables are the different metrics of content and linguistic quality as well as the overall score respectively. The sample consists of 159 firms that issued a bond between 2010 and 2016 on a regulated stock exchange in Germany. The variable definitions are provided in Table A1 in Appendix A. Robust standard errors appear in brackets. Significance levels are as follows: *=10%, **=5%, ***=1%.

Table 10: Probit Regression Results of Summary Quality, Firm, and Bond Characteristics on Bond Default

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Relative Change in Boilerplate	-21.327					-54.227 *	
	(32.886)					(29.639)	
Relative Change in Specificity		-2.395 **				-2.850 ***	
		(1.066)				(1.012)	
Relative Change in Readability			-0.266			-0.323 *	
			(0.198)			(0.185)	
Relative Change in Tonality				-0.061		-0.078 *	
				(0.053)		(0.044)	
Divergence					-1.256 ***	-1.247 ***	
					(0.319)	(0.296)	
Summary Quality							-0.199 ***
							(0.037)
Collateralisation	-0.229 ***	-0.220 ***	-0.204 ***	-0.229 ***	-0.243 ***	-0.160 **	-0.205 ***
	(0.071)	(0.069)	(0.075)	(0.070)	(0.070)	(0.077)	(0.070)
Covenants	-0.006	-0.003	-0.005	-0.003	-0.006	-0.002	-0.006
	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)
Spread	-1.210	-1.002	-0.026	-0.877	-1.936	-1.244	-2.014
	(3.000)	(2.843)	(2.966)	(2.917)	(2.757)	(2.736)	(2.898)
Nominalvolume	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Maturity	0.021	0.015	0.020	0.020	0.023	0.020	0.029
	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)	(0.020)	(0.020)
Investment Grade	-0.044	-0.010	-0.026	-0.014	-0.065	-0.051	-0.101
	(0.098)	(0.098)	(0.099)	(0.101)	(0.098)	(0.091)	(0.097)
No Rating	-0.018	-0.008	0.004	0.001	-0.033	-0.019	-0.012
	(0.074)	(0.072)	(0.074)	(0.073)	(0.072)	(0.067)	(0.067)
Employees	-0.020	-0.006	-0.018	-0.026	0.023	0.000	0.006
	(0.059)	(0.056)	(0.059)	(0.061)	(0.053)	(0.053)	(0.051)
Firm Age	-0.002	-0.002	-0.002	-0.002	-0.002 *	-0.002 *	-0.002 *
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Public	0.048	0.018	0.045	0.043	-0.036	-0.036	-0.043
	(0.089)	(0.089)	(0.086)	(0.087)	(0.086)	(0.089)	(0.098)
O-Score	1.108 ***	1.113 ***	1.142 ***	1.140 ***	1.078 ***	0.842 ***	0.873 ***
	(0.343)	(0.332)	(0.340)	(0.336)	(0.326)	(0.306)	(0.324)
Size	-0.019	-0.015	-0.023	-0.022	-0.025	-0.026	-0.034
	(0.022)	(0.021)	(0.022)	(0.022)	(0.021)	(0.021)	(0.021)
Liquidity	-0.058	-0.043	-0.034	-0.055	-0.056	-0.046	-0.063
	(0.067)	(0.065)	(0.066)	(0.068)	(0.063)	(0.056)	(0.059)
Sales Growth	-0.521	-0.844	-0.632	-0.691	-0.512	-0.899	-0.642
	(0.747)	(0.736)	(0.757)	(0.754)	(0.675)	(0.664)	(0.670)
Leverage	0.059	0.065	0.057	0.055	0.076 **	0.053	0.053
	(0.038)	(0.038)	(0.038)	(0.038)	(0.034)	(0.034)	(0.033)
Industry FE	yes	yes	yes	yes	yes	yes	yes
Time FE	yes	yes	yes	yes	yes	yes	yes
Observations	159	159	159	159	159	159	159
Pseudo R-squared	0.309	0.324	0.316	0.313	0.375	0.425	0.407
AURROC	0.846	0.862	0.851	0.856	0.89	0.903	0.899

Notes: This table presents the average marginal effects from probit regression, where the dependent variable is the default of a bond. The sample consists of 159 firms that issued a bond between 2010 and 2016 on a regulated stock exchange in Germany. The variable definitions are provided in Table A1 in Appendix A. Robust standard errors appear in brackets. Significance levels are as follows: *=10%, **=5%, ***=1%.

Table 11: IV-Probit Regression Results on Bond Default

	Probit	IV Probit	
		1st Stage	2nd Stage
	(1)	(2)	(3)
Summary Quality	-0.199 ***		-0.534 **
	(0.037)		(0.213)
Fine Proceedings		0.020 ***	
		(0.002)	
Controls	yes	yes	yes
Industry FE	yes	yes	yes
Time FE	yes	yes	yes
Observations	159	159	159
Pseudo R-squared	0.407		
Tests of Weak Instrument:			
Partial F-Statistic		38.891 ***	
Test of No Endogeneity:			
Wald Chi-Squared Test of Exogeneity			1.243 n.s.

Notes: This table presents the average marginal effects from probit and IV-probit regressions. The sample consists of 159 firms that issued a bond between 2010 and 2016 on a regulated stock exchange in Germany. The variable definitions are provided in Table A1 in Appendix A. Robust standard errors appear in brackets. Significance levels are as follows: *=10%, **=5%, ***=1%.

REFERENCES

Abraham, S., & Shrives, P. (2014). Improving the Relevance of Risk Factor Disclosure in Corporate Annual Reports. *British Accounting Review*, *46*(1), 91–107. doi:10.1016/j.bar.2013.10.002

Aerts, W. (2005). Picking up the Pieces: Impression Management in the Retrospective Attributional Framing of Accounting Outcomes. *Accounting, Organizations and Society*, *30*(6), 493–517. doi:10.1016/j.aos.2004.07.001

Altman, E., Esentato, M., & Sabato, G. (2018). Assessing the Credit Worthiness of Italian SMEs and Mini-Bond Issuers. *Global Finance Journal*, (forthcoming), 1–11. doi:10.1016/j.gfj.2018.09.003

Audretsch, D., & Elston, J. (1997). Financing the German Mittelstand. *Small Business Economics*, *9*(2), 97–110. doi:10.1023/A:1007963621438

Baginski, S., Hassell, J., & Hillison, W. (2000). Voluntary Causal Disclosures: Tendencies and Capital Market Reaction. *Review of Quantitative Finance and Accounting*, *15*(4), 371–389. doi:10.1023/A:1012002608615

Bamberger, R., & Vanecek, E. (1984). *Lesen, Verstehen, Lernen, Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Vienna, Austria: Jugend und Volk Verlagsgesellschaft.

Bannier, C., Pauls, T., & Walter, A. (2017). CEO-Speeches and Stock Returns. *Working Paper*, 1–46. doi:10.2139/ssrn.3051151

Bannier, C., Pauls, T., & Walter, A. (2019). Content Analysis of Business Communication: Introducing a German Dictionary. *Journal of Business Economics*, *89*(1), 79–123. doi:10.1007/s11573-018-0914-8

Bao, Y., & Datta, A. (2014). Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures. *Management Science*, *60*(6), 1371–1391. doi:10.1287/mnsc.2014.1930

Barth, M. (2015). Financial Accounting Research, Practice, and Financial Accountability. *Abacus*, *51*(4), 499–510. doi:10.1111/abac.12057

Beatty, A., Cheng, L., & Zhang, H. (2018). Are Risk Factor Disclosures Still Relevant? Evidence from Market Reactions to Risk Factor Disclosures Before and After the Financial Crisis. *Contemporary Accounting Research*, (forthcoming), 1–57. doi:10.1111/1911-3846.12444

Beshears, J., Choi, J., Laibson, D., & Madrian, B. (2011). How Does Simplified Disclosure Affect Individuals' Mutual Fund Choices? In D. Wise (Ed.), *Explorations in the Economics of Aging* (pp. 75–96). London, UK: University of Chicago Press.

Bloomberg, M., & Schapiro, M. (2014, May 24). Give Investors Access to All the Information They Need. *Financial Times*. https://www.ft.com/content/0d9ccea6-db66-11e3-94ad-00144feabdc0

Bloomfield, R. (2002). The Incomplete Revelation Hypothesis and Financial Reporting. *Accounting Horizons*, *16*(3), 233–243. doi:10.2308/acch.2002.16.3.233

Bondguide. (2019). Mittelstandsanleihen. https://www.bondguide.de. Accessed 20 March 2019

Bonner, S., Clor-Proell, S., & Koonce, L. (2014). Mental Accounting and Disaggregation Based on the Sign and Relative Magnitude of Income Statement Items. *Accounting Review*, *89*(6), 2087–2114. doi:10.2308/accr-50838

Bonsall, S., & Miller, B. (2017). The Impact of Narrative Disclosure Readability on Bond Ratings and the Cost of Debt. *Review of Accounting Studies*, *22*(2), 608–643. doi:10.1007/s11142-017-9388-0

Brandow, R., Mitze, K., & Rau, L. (1995). Automatic Condensation of Electronic Publications by Sentence Selection. *Information Processing and Management*, *31*(5), 675–685. doi:10.1016/0306-4573(95)00052-I

Brown, S., Tian, X., & Tucker, J. (2018). The Spillover Effect of SEC Comment Letters on Qualitative Corporate Disclosure: Evidence from the Risk Factor Disclosure. *Contemporary Accounting Research*, *35*(2), 622–656. doi:10.1111/1911-3846.12414

Burn, L. (2010). KISS, but Tell All: Short-Form Disclosure for Retail Investors. *Capital Markets Law Journal*, *5*(2), 141–168. doi:10.1093/cmlj/kmq001

Burn, L. (2016). Capital Markets Union and Regulation of the EU's Capital Markets. *Capital Markets Law Journal*, *11*(3), 352–386. doi:0.1093/cmlj/kmw011

Burn, L. (2018). The Effect of Recent EU Legislation on EU Retail Bond Markets. *Capital Markets Law Journal*, *13*(4), 518–533. doi:10.1093/cmlj/kmy022

Bushee, B. J., Matsumoto, D. A., & Miller, G. S. (2003). Open versus Closed Conference Calls: The Determinants and Effects of Broadening Access to Disclosure. *Journal of Accounting and Economics*, *34*(1–3), 149–180. doi:10.1016/S0165-4101(02)00073-3

Cabrera-Diego, L., & Torres-Moreno, J. (2018). SummTriver: A New Trivergent Model to Evaluate Summaries Automatically without Human References. *Data and Knowledge Engineering*, *113*(1), 184–197. doi:10.1016/j.datak.2017.09.001

Campbell, J., Cecchini, M., Cianci, A., Ehinger, A., & Werner, E. (2018). Tax-Related Mandatory Risk Factor Disclosures, Future Profitability, and Stock Returns. *Review of Accounting Studies*, (forthcoming), 1–39. doi:10.1007/s11142-018-9474-y

Campbell, J., Chen, H., Dhaliwal, D., Lu, H., & Steele, L. (2014). The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings. *Review of Accounting Studies*, *19*(1), 396–455. doi:10.1007/s11142-013-9258-3

Cardinaels, E., Hollander, S., & White, B. (2019). Automatic Summarization of Earnings Releases: Attributes and Effects on Investors' Judgments. *Review of Accounting Studies*, (forthcoming), 1–31. doi:10.1007/s11142-019-9488-0

CFA Institute. (2017). Designing A European Summary Prospectus Using Behavioural Insights. https://www.cfainstitute.org/-/media/documents/article/position-paper/designing-a-european-summary-prospectus.ashx. Accessed 15 May 2019

Chiu, T.-T., Guan, Y., & Kim, J.-B. (2018). The Effect of Risk Factor Disclosures on the Pricing of Credit Default Swaps. *Contemporary Accounting Research*, *35*(4), 2191–2224. doi:10.1111/1911-3846.12362

Chiu, T.-T., Kim, J.-B., & Wang, Z. (2018). Customers' Risk Factor Disclosures and Suppliers' Investment Efficiency. *Contemporary Accounting Research*, (forthcoming), 1–52.

doi:10.1111/1911-3846.12447

Clatworthy, M., & Jones, M. (2001). The Effect of Thematic Structure on the Variability of Annual Report Readability. *Accounting, Auditing & Accountability Journal*, *14*(3), 311–326. doi:10.1108/09513570110399890

Creditreform. (2019). Creditreform RatingMap. https://www.creditreform.de/microsites/creditreform-ratingmap-deutschland.html. Accessed 15 March 2019

Dechow, P., Ge, W., Larson, C., & Sloan, R. (2011). Predicting Material Accounting Misstatements. *Contemporary Accounting Research*, *28*(1), 17–82. doi:10.1111/j.1911-3846.2010.01041.x

Donaway, R., Drummey, K., & Mather, L. (2000). A Comparison of Rankings Produced by Summarization Evaluation Measures. In *Proceedings of the NAACL-ANLP Workshop on Automatic Summarization* (pp. 69–78). Seattle. doi:10.3115/1567564.1567572

Du Vignaux, H., Gouzard, C., Gehringer, A., Byers, D., Cuccia, S., Wagner, H., et al. (2006). The Implementation of the EU Prospectus Directive - A Country-by-Country Analysis. *Capital Markets Law Journal*, *1*(1), 89–112. doi:10.1093/cmlj/kml006

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics*, *64*(2–3), 221–245. doi:10.1016/j.jacceco.2017.07.004

Ellouze, S., Jaoua, M., & Belguith, L. (2013). An Evaluation Summary Method Based on a Combination of Content and Linguistic Metrics. In *Proceedings of the International Conference Recent Advances in Natural Language Processing* (pp. 245–251). Hissar, Bulgaria.

Elshandidy, T., Fraser, I., & Hussainey, K. (2013). Aggregated, Voluntary, and Mandatory Risk Disclosure Incentives: Evidence from UK FTSE All-Share Companies. *International Review of Financial Analysis*, *30*(1), 320–333. doi:10.1016/j.irfa.2013.07.010

Elshandidy, T., Fraser, I., & Hussainey, K. (2015). What Drives Mandatory and Voluntary Risk Reporting Variations across Germany, UK and US? *British Accounting Review*, *47*(4), 376–394. doi:10.1016/j.bar.2014.06.001

Elshandidy, T., Shrives, P., Bamber, M., & Abraham, S. (2018). Risk Reporting: A Review of the Literature and Implications for Future Research. *Journal of Accounting Literature*, *40*(1), 54–82. doi:10.1016/j.acclit.2017.12.001

European Securities and Markets Authority. (2012). *Technical Advice on Possible Delegated Acts Concerning the Prospectus Directive as Amended by the Directive 2010/73/EU*. Paris, France. https://www.esma.europa.eu/system/files_force/library/2015/11/2012-864.pdf

European Securities and Markets Authority. (2017). *Access to Public Capital Markets for SMEs*. Paris, France. https://www.esma.europa.eu/sites/default/files/library/esma22-106-535_smsg_report_on_access_to_public_capital_markets_for_smes.pdf

European Securities and Markets Authority. (2019). *ESMA Guidelines on Risk Factors Under the Prospectus Regulation*. https://www.esma.europa.eu/sites/default/files/library/esma31-62-1217_final_report_on_guidelines_on_risk_factors.pdf

Fagernäs, T., Kanervo, J., Núñez, G., & Alcalá, A. (2019). The Why and How of the New European Union Prospectus Regulation. *Business Law International*, *20*(1), 5–23.

Faruqui, M., & Padó, S. (2010). Training and Evaluating a German Named Entity Recognizer with Semantic Generalization. In *KONVENS* (pp. 129–133). Saarbruecken, Germany.

Feihle, P., & Lawrenz, J. (2017). The Issuance of German SME Bonds and its Impact on Operating Performance. *Schmalenbach Business Review*, *18*(3), 227–259. doi:10.1007/s41464-017-0036-9

Field, L., Lowry, M., & Shu, S. (2005). Does Disclosure Deter or Trigger Litigation? *Journal of Accounting and Economics*, *39*(3), 487–507. doi:10.1016/j.jacceco.2005.04.004

Filzen, J. (2015). The Information Content of Risk Factor Disclosures in Quarterly Reports. *Accounting Horizons*, *29*(4), 887–916. doi:10.2308/acch-51175

Fischer-Appelt, D. (2010). Prospectus Directive Amendments – Discussion of Key Changes. *Law and Financial Markets Review*, *4*(5), 490–498. doi:10.5235/175214410792930014

Fischer-Appelt, D. (2012). The Revised EU Prospectus Regulation: Key Changes to the Contents of Prospectuses. *Law and Financial Markets Review*, *6*(4), 249–257. doi:10.5235/lfmr6.4.249

Fischer-Appelt, D. (2014). Amendments to the EU Prospectus Directive Regime. *Business Law International*, *15*(2), 95–114.

Florstedt, T. (2017). Lehren aus dem Zusammenbruch des Marktes für sog. Mittelstandsanleihen. *Journal of Banking Law and Banking*, *29*(3), 145–158. doi:10.15375/zbb-2017-0304

Godwin, A., & Ramsay, I. (2016). Short-Form Disclosure Documents—An Empirical Survey of Six Jurisdictions. *Capital Markets Law Journal*, *11*(2), 296–316. doi:10.1093/cmlj/kmw002

Greene, W. (2003). *Econometric Analysis* (5th ed.). New York, USA: Prentice Hall.

Guillamon-Saorin, E., Osma, B., & Jones, M. (2012). Opportunistic Disclosure in Press Release Headlines. *Accounting and Business Research*, *42*(2), 143–168. doi:10.1080/00014788.2012.632575

Hanley, K. W., & Hoberg, G. (2010). The Information Content of IPO Prospectuses. *The Review of Financial Studies*, *23*(7), 2821–2864. doi:10.2307/40782968

Hanley, K. W., & Hoberg, G. (2012). Litigation Risk, Strategic Disclosure and the Underpricing of Initial Public Offerings. *Journal of Financial Economics*, *103*(2), 235–254. doi:10.1016/j.jfineco.2011.09.006

Healy, P., & Palepu, K. (2001). Information Asymmetry, Corporate Disclosure, and the Capital Markets: A Review of the Empirical Disclosure Literature. *Journal of Accounting and Economics*, *31*(1), 405–440. doi:10.1016/S0165-4101(01)00018-0

Heinle, M., Smith, K., & Verrecchia, R. (2018). Risk-Factor Disclosure and Asset Prices. *Accounting Review*, *93*(2), 191–208. doi:10.2308/accr-51863

Henry, E., & Leone, J. (2016). Measuring Qualitative Information in Capital Markets Research: Comparison of Alternative Methodologies to Measure Disclosure Tone. *Accounting Review*, *91*(1), 153–178. doi:10.2308/accr-51161

Holthausen, R. (1990). Accounting Method Choice, Opportunistic Behavior, Efficient Contracting, and Information Perspectives. *Journal of Accounting and Economics*, *12*(1–3), 207–218. doi:10.1016/0165-4101(90)90047-8

Hope, O., Hu, D., & Lu, H. (2016). The Benefits of Specific Risk-Factor Disclosures. *Review of Accounting Studies*, *21*(4), 1005–1045. doi:10.1007/s11142-016-9371-1

Hosmer, D., & Lemeshow, S. (2012). *Applied Logistic Regression* (3rd ed.). Hoboken: Wiley.

Huang, X., Nekrasov, A., & Teoh, S. H. (2018). Headline Salience, Managerial Opportunism, and Over- and Underreactions to Earnings. *Accounting Review*, *93*(6), 231–255. doi:10.2308/accr-52010

Just, C., Voß, T., Ritz, C., & Zeising, M. (2019). *WpPG: Wertpapierprospektgesetz und EU-Prospektverordnung* (2nd ed.). Frankfurt, Germany: Beck.

Kim, I., & Skinner, D. (2012). Measuring Securities Litigation Risk. *Journal of Accounting and Economics*, *53*(1–2), 290–310. doi:10.1016/j.jacceco.2011.09.005

Kravet, T., & Muslu, V. (2013). Textual Risk Disclosures and Investors' Risk Perceptions. *Review of Accounting Studies*, *18*(4), 1088–1122. doi:10.1007/s11142-013-9228-9

Kulik, C., Perera, S., & Cregan, C. (2016). Engage Me: The Mature-Age Worker and Stereotype Threat. *Academy of Management Journal*, *59*(6), 2132–2156. doi:10.5465/amj.2015.0564

La Porta, R., Lopez-De-Silanes, F., & Shleifer, A. (2006). What Works in Securities Laws? *Journal of Finance*, *61*(1), 1–32.

Lang, M., & Stice-Lawrence, L. (2015). Textual Analysis and International Financial Reporting: Large Sample Evidence. *Journal of Accounting and Economics*, *60*(2–3), 110–135. doi:10.1016/j.jacceco.2015.09.002

Larcker, D., & Rusticus, T. (2010). On the Use of Instrumental Variables in Accounting Research. *Journal of Accounting and Economics*, *49*(3), 186–205. doi:10.1016/j.jacceco.2009.11.004

Leung, S., Parker, L., & Courtis, J. (2015). Impression Management through Minimal Narrative Disclosure in Annual Reports. *British Accounting Review*, *47*(3), 275–289. doi:10.1016/j.bar.2015.04.002

Li, F. (2008). Annual Report Readability, Current Earnings, and Earnings Persistence. *Journal of Accounting and Economics*, *45*(2–3), 221–247. doi:10.1016/j.jacceco.2008.02.003

Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In *Proceedings of the ACL Text Summarization Workshop* (pp. 74–81). Barcelona, Spain.

Lin, C.-Y., Cao, G., Gao, J., & Nie, J.-Y. (2006). An Information-Theoretic Approach to Automatic Evaluation of Summaries. In *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics* (pp. 463–470). New York, USA. doi:10.3115/1220835.1220894

Lin, Z., Liu, C., Ng, H., & Kan, M.-Y. (2012). Combining Coherence Models and Machine Translation Evaluation Metrics for Summarization Evaluation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* (pp. 1006–1014). Stroudsburg, USA. http://dl.acm.org/citation.cfm?id=2390524.2390659

Lloret, E., & Palomar, M. (2012). Text Summarisation in Progress: A Literature Review. *Artificial Intelligence Review*, *37*(1), 1–41. doi:10.1007/s10462-011-9216-z

Lloret, E., Plaza, L., & Aker, A. (2018). The Challenging Task of Summary Evaluation: An

Overview. *Language Resources and Evaluation*, *52*(1), 101–148. doi:10.1007/s10579-017-9399-2

Loughran, T., & McDonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Distionaries, and 10-Ks. *Journal of Finance*, *66*(1), 35–65. doi:0.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2014). Measuring Readability in Financial Disclosures. *Journal of Finance*, *69*(4), 1643–1671. doi:10.1111/jofi.12162

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, *54*(4), 1187–1230. doi:10.1111/1475-679X.12123

Louis, A., & Nenkova, A. (2009). Automatically Evaluating Content Selection in Summarization without Human Models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 306–314). Singapore. doi:10.3115/1699510.1699550

Louis, A., & Nenkova, A. (2013). Automatically Assessing Machine Summary Content without a Gold Standard. *Computational Linguistics*, *39*(2), 267–300. doi:10.1162/COLI_a_00123

Mani, I. (2001). *Automatic Summarization*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Mehri, A., Jamaati, M., & Mehri, H. (2015). Word Ranking in a Single Document by Jensen-Shannon Divergence. *Physics Letters*, *379*(28–29), 1627–1632. doi:10.1016/j.physleta.2015.04.030

Merkl-Davies, D., & Brennan, N. (2007). Discretionary Disclosure Strategies in Corporate Narratives: Incremental Information or Impression Management? *Journal of Accounting Literature*, *26*(1), 116–196.

Mietzner, M., Proelss, J., & Schweizer, D. (2018). Hidden Champions or Black Sheep? The Role of Underpricing in the German Mini-Bond Market. *Small Business Economics*, *50*(2), 375–395. doi:10.1007/s11187-016-9833-7

Nenkova, A., & McKeown, K. (2011). Automatic Summarization. *Foundations and Trends in Information Retrieval*, *5*(2), 103–233. doi:10.1561/1500000015

Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, *18*(1), 109–131. doi:10.2307/2490395

Owczarzak, K., Conroy, J., Dang, H., & Nenkova, A. (2012). An Assessment of the Accuracy of Automatic Evaluation in Summarization. In *Proceedings of Workshop on Evaluation Metrics and System Comparison for Automatic Summarization* (pp. 1–9). Montréal, Canada. https://www.aclweb.org/anthology/W12-2601

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the Annual Meeting on Association for Computational Linguistics* (pp. 311–318). Philadelphia, USA. doi:10.3115/1073083.1073135

Proisl, T., & Uhrig, P. (2016). SoMaJo: State-of-the-Art Tokenization for German Web and Social Media Texts. In *Proceedings of the Web as Corpus Workshop* (pp. 57–62). Berlin, Germany. doi:http://dx.doi.org/10.18653/v1/W16-2607

Puri, M., Rocholl, J., & Steffen, S. (2011). Global Retail Lending in the Aftermath of the US

Financial Crisis: Distinguishing Between Supply and Demand Effects. *Journal of Financial Economics*, *100*(3), 556–578. doi:10.1016/j.jfineco.2010.12.001

Radev, D., Hovy, E., & McKeown, K. (2002). Introduction to the Special Issue on Summarization. *Computational Linguistics*, *28*(4), 399–408. doi:10.1162/089120102762671927

Rossi, S. (2016). *Access to Bank Credit and SME Financing*. Cham, Switzerland: Springer International Publishing.

Saggion, H., Teufel, S., Radev, D., & Lam, W. (2002). Meta-Evaluation of Summaries in a Cross-Lingual Environment Using Content-Based Metrics. In *Proceedings of COLING* (pp. 849–855). Taipei, Taiwan. doi:10.3115/1072228.1072301

Saha, A., Morris, R., & Kang, H. (2019). Disclosure Overload? An Empirical Analysis of International Financial Reporting Standards Disclosure Requirements. *Abacus*, *55*(1), 205–236. doi:10.1111/abac.12148

Savitt, W. (2018). *The Securities Litigation Review* (4th ed.). London, UK: Law Business Research.

Schrand, C., & Elliott, J. (1998). Risk and Financial Reporting: A Summary of the Discussion at the 1997 AAA/FASB Conference. *Accounting Horizons*, *12*(3), 271–282.

Securities and Exchange Commission. (1998). Plain English Disclosure. https://www.sec.gov/rules/final/33-7497.txt. Accessed 15 May 2019

Securities and Exchange Commission. (2016). Form 10-K Summary. https://www.sec.gov/rules/interim/2016/34-77969.pdf. Accessed 15 May 2019

Sergakis, K. (2018). *The Law of Capital Markets in the EU* (1st ed.). Basingstoke, UK: Palgrave.

Silva, G., Ferreira, R., Lins, R., Cabral, L., Oliveira, H., Simske, S., & Riss, M. (2015). Automatic Text Document Summarization Based on Machine Learning. In *Proceedings of the 2015 ACM Symposium on Document Engineering* (pp. 191–194). New York, USA. doi:10.1145/2682571.2797099

Simon, H. (2009). *Hidden Champions of the Twenty-First Century: The Success Strategies of Unknown World Market Leaders*. Berlin, Germany: Springer.

Smith, M., & Taffler, R. (2000). The Chairman's Statement: A Content Analysis of Discretionary Narrative Disclosures. *Accounting, Auditing & Accountability Journal*, *13*(5), 624–647. doi:10.1108/09513570010353738

Sparck-Jones, K. (1999). Automatic Summarizing: Factors and Directions. In I. Mani & M. Maybury (Eds.), *Advances in Automatic Text Summarisation* (pp. 1–13). Cambridge, USA: MIT Press.

Spindler, G. (2011). Behavioural Finance and Investor Protection Regulations. *Journal of Consumer Policy*, *34*(1), 315–336. doi:10.1007/s10603-011-9165-6

Tonidandel, S., & LeBreton, J. (2011). Relative Importance Analysis: A Useful Supplement to Regression Analysis. *Journal of Business and Psychology*, *26*(1), 1–9. doi:10.1007/s10869-010-9204-3

Tonidandel, S., & LeBreton, J. (2015). RWA Web: A Free, Comprehensive, Web-Based, and User-Friendly Tool for Relative Weight Analyses. *Journal of Business and Psychology*, *30*(2), 207–216. doi:10.1007/s10869-014-9351-z

Törnqvist, L., Vartia, P., & Vartia, Y. (1985). How Should Relative Changes Be Measured? *American Statistician*, *39*(1), 43–46. doi:10.1080/00031305.1985.10479385

Torres-Moreno, J.-M., Saggion, H., Da Cunha, I., SanJuan, E., & Velázquez-Morales, P. (2010). Summary Evaluation with and without References. *Polibits*, *43*(1), 13–19. doi:10.17562/PB-42-2

Utz, S., Weber, M., & Wimmer, M. (2016). German Mittelstand Bonds: Yield Spreads and Liquidity. *Journal of Business Economics*, *86*(1), 103–129. doi:10.1007/s11573-015-0791-3

Vadlapudi, R., & Katragadda, R. (2010). On Automated Evaluation of Readability of Summaries: Capturing Grammaticality, Focus, Structure and Coherence. In *Proceedings of the NAACL HLT 2010 Student Research Workshop* (pp. 7–12). Los Angeles, USA.

Venti, S. (2011). Comment on How Does Simplified Disclosure Affect Individuals' Mutual Fund Choices? In D. Wise (Ed.), *Explorations in the Economics of Aging* (pp. 96–98). London, UK: University of Chicago Press.

Walther, T. (2015). Key Investor Documents and Their Consequences on Investor Behavior. *Journal of Business Economics*, *85*(2), 129–156. doi:10.1007/s11573-014-0724-6

Weiß, Z. (2016). Applying Readability Formulae to the Task of Proficiency Assessment. *Working Paper*, 1–22.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. London, UK: MIT Press.

Yadav, N., & Chatterjee, N. (2016). Text Summarization Using Sentiment Analysis for DUC Data. In *International Conference on Information Technology* (pp. 229–234). Bhubaneswar, India. doi:10.1109/ICIT.2016.12

Yuthas, K., Rogers, R., & Dillard, J. (2002). Communicative Action and Corporate Annual Reports. *Journal of Business Ethics*, *41*(1–2), 141–157. doi:10.1023/A:1021314626311

Appendix A – Variables and Descriptive Sample Statistics

Table A1: Description of Variables

Variable	Description
Boilerplate	Percentage of words that are in boilerplate sentences.
Collateralization	Dummy variable indicating whether the bond is collateralized.
Covenants	Number of covenants provided in the security prospectus.
Default	Dummy variable indicating whether the bond defaulted within the first two years after issuance.
Divergence	Jensen Shannon divergence between the probability distributions of words in the full risk factor section and its summary.
Employees	Number of employees at time of issuance.
Fine Proceedings	Number of all actions taken by the BaFin in the 6 months before the issuance.
Firm Age	Age of firm in years at time of issuance.
Investment Grade	Dummy variable indicating whether the bond has an investment grade rating.
Leverage	Total debt divided by total assets in the year prior the bond issuance.
Liquidity	Current assets divided by short-term liabilities in the year prior issuance.
Maturity	Maturity of bond in years.
No Rating	Dummy variable indicating whether the bond has a rating.
Nominal Volume	Issue volume of bond in million Euro.
O-Score	Ohlson's o-score in the year prior issuance.
Public	Dummy variable indicating whether the bond issuing firm is listed on a stock exchange at the time of issuance.
Readability	Degree of readability easiness based on the Wiener Sachtextformel.
Sales Growth	Average change in revenues of a 3-year period prior issuance.
Size	Natural logarithm of total assets in the year prior the bond issuance.
Specificity	Number of specific words divided by the number of total words.
Spread	Spread of the yield of bond over an equivalent government benchmark bond.
Summary Quality	Weighted quality score of a summary.
Tonality	Number of positive words minus number of negative words divided by total number of words.

Table A2: Industry Statistics

Industry	2010	2011	2012	2013	2014	2015	2016	Total
				Year				
A - Agriculture, forestry and fishing	1	3	1	4	2	0	0	11
C - Manufacturing	1	6	1	2	1	0	1	12
D - Electricity, gas, steam and air conditioning supply	1	7	6	7	5	0	3	29
F - Construction	2	4	5	2	1	1	0	15
G - Wholesale and retail trade	0	0	3	3	5	0	1	12
I - Accommodation and food service activities	1	0	2	1	0	0	1	5
J - Information and communication	0	1	3	7	1	1	1	14
K - Financial and insurance activities	2	2	4	10	3	2	0	23
L - Real estate activities	1	8	7	7	2	1	0	26
N - Administrative and support service activities	1	1	1	6	3	0	0	12
Total	10	32	33	49	23	5	7	159

Table A3: Types of BaFin Fine Proceedings over Time

	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009
Ad Hoc Disclosures	16	22	21	35	25	36	20	23	23	22
Directors Dealings	1	0	1	3	2	2	7	2	3	4
Market Manipulation	6	7	14	17	11	19	7	13	10	6
Notification and Publication Requirements	51	57	152	302	327	342	347	209	196	342
Duties to Provide Information to Securities Holders	0	2	10	12	46	45	78	41	42	21
Short Selling	0	0	7	5	2	3	2	68	18	10
Financial Reporting Requirements	46	78	59	29	63	55	52	44	95	21
Prospectuses	2	0	9	6	6	2	14	6	9	11
Company Takeovers	5	1	0	3	29	15	5	14	16	28
Others	5	2	8	14	2	0	2	0	0	0
Total Number of Fine Proceedings	132	169	281	426	513	519	534	420	412	465