Bournemouth University

The information content of narrative disclosures in financial statements

Dr. Jörg Hering

A thesis submitted in partial fulfilment of the requirements of Bournemouth University for the degree of

Doctor of Philosophy

November 2022

Department of Accounting, Finance and Economics

The Business School, Bournemouth University

Author's declaration

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at Bournemouth University, United Kingdom. I declare that this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that this thesis has not been previously or concurrently submitted, either in whole or in part, for any other qualification at Bournemouth University or other institutions.

Supervisors: Prof. Jens Hölscher & Dr. Phyllis Alexander

Acknowledgements

First and foremost, I would like to thank my supervisors Prof. Jens Hölscher and Dr. Phyllis Alexander for spending valuable time on discussions despite various academic and professional commitments, and for providing major support throughout this dissertation. I would also like to thank participants at doctoral seminars for their useful insights, constructive criticism, and valuable suggestions. Many others have contributed to this work; their contributions are gratefully acknowledged, as they have helped me to build this dissertation.

Finally, and most importantly, I wish to thank my family for their encouragement and selfless support at all times, for which my mere expression of thanks does not suffice. Thank you for your patience and encouragement along the course of this dissertation.

-This dissertation is dedicated to my family-

Introductory statement

As a PhD candidate at the Department of Accounting and Auditing at Friedrich-Alexander University Erlangen-Nuremberg I received a doctorate degree in economics for the thesis entitled *"Textual analysis in accounting and finance"*. That thesis consists of three standalone research studies in the overall field of textual analysis. The first research study develops different information extraction algorithms (*"Annual Report Algorithm"*, *"Items Algorithm"*) enabling users to extract textual information from annual reports on From 10-K filed with the United States Securities and Exchange Commission (SEC) for further analysis. The second research study analysis the market impact of narrative information (overall filing, Form 10-K subsections) and its linguistic tone (textual sentiment) on future stock returns as well as the predictive ability of disclosure tone on other important future firm characteristics (e.g. return on assets, dividend yield, payout ratio). The last research study investigates the stock market impact of (important) corporate events required to be disclosures by public traded companies in current reports on Form 8-K in the United States of America.

As a PhD candidate at the Department of Accounting, Finance and Economics at Bournemouth University I continue my research in the field of textual analysis within the present thesis entitled *"The information content of narrative disclosures in financial statements"*. This thesis again consists of four standalone research studies in the overall field of textual analysis. The first research study investigates the effects of Form 10-K textual sentiment (overall filing, Form 10-K subsections) on the liquidity and risk profiles of U.S. public listed companies. The second research study examines the market reaction and information content of narrative disclosures solely in the notes to the financial statements or *"Footnotes"* (*"Item 8 - Financial Statements and Supplementary Data"*) in the context of managerial behavior (*"Management obfuscation"*). The third research study explores the ability of Form 10-K textual information (type of textual information, measurement of textual information, source of textual information) to differentiate between healthy and unhealthy firms going forward. The last research study analyses the effect of disclosure complexity (not information content or textual sentiment) on various important stock (firm) characteristics (e.g. stock return, trading volume, bid-ask spread, stock return volatility, return on assets, dividend yield, payout ratio) using an alternative measure for disclosure complexity (document file size).

Therefore, each thesis is examining specific and distinct research questions in the field of textual analysis not only providing important empirical evidence for the overall body of textual analysis literature but also for capital market participants and regulatory authorities.

Table of content

Chapter 1 1

Contents 2

1.0 Background and motivation 3

1.1 Capital market efficiency 7

1.2 Research questions 17

1.2.1 Capital market effect of annual report sentiment 17

1.2.2 Management obfuscation in financial statements 19

1.2.3 Explanatory power of disclosure tone on corporate bankruptcy 20

1.2.4 Capital market effect of annual report readability 21

Chapter 2 24

Contents 25

2.0 Introduction 26

2.1 Related literature and hypothesis development 28

2.2 Research design 33

2.2.1 Sample selection and description 33

2.2.2 Regression model, dependent variables, and independent variables 36

2.2.3 Summary statistics and correlations 39

2.3 Empirical results 42

2.4 Robustness tests 48

2.4.1 Sample selection 48

2.4.2 Disclosure length 53

2.4.3 Table content 58

2.5 Sensitivity tests 63

2.5.1 Type of disclosure tone 63

2.5.2 Measurement of disclosure tone 80

2.6 Additional tests 97

2.7 Conclusion 107

Chapter 3 109

Contents 110

3.0 Introduction 111

3.1 Related literature 114

3.2 Hypothesis development 119

3.3 Research design 123

3.3.1 Sample selection and description 123

3.3.2 Regression model, dependent variable, and independent variables 126

3.3.3 Summary statistics and correlations 129

3.4 Empirical results 132

3.5 Robustness tests 138

3.5.1 Sample selection 138

3.5.2 Disclosure length 142

3.5.3 Table content 146

3.6 Sensitivity tests 149

3.6.1 Type of disclosure tone 149

3.6.2 Measurement of disclosure tone 153

3.7 Additional tests 157

3.8 Conclusion 161

Chapter 4 163

Contents 164

4.0 Introduction 165

4.1 Related literature 167

4.2 Hypothesis development 175

4.3 Research design 177

4.3.1 Sample selection and description 177

4.3.2 Model specification 181

4.3.3 Descriptive statistics ... 183
4.4 Empirical results ... 185
4.5 Robustness tests ... 187
4.5.1 Sample selection ... 187
4.5.2 Disclosure length ... 188
4.5.3 Table content ... 189
4.6 Sensitivity tests ... 190
4.6.1 Type of disclosure tone ... 190
4.6.2 Measurement of disclosure tone ... 192
4.7 Additional tests ... 199
4.7.1 Relevance of text sections ... 199
4.7.2 Firm fundamentals ... 204
4.7.3 All available bankrupt and non-bankrupt firm-year observations ... 208
4.7.4 Financial distress ... 211
4.8 Conclusion ... 214
Chapter 5 ... 215
Contents ... 216
5.0 Introduction ... 217
5.1 Related literature ... 220
5.2 Hypothesis development ... 226
5.3 Research design ... 229
5.3.1 Sample selection and description ... 229
5.3.2 Regression model, dependent variables, and independent variables ... 232
5.3.3 Summary statistics and correlations ... 237
5.4 Empirical results ... 241
5.5 Robustness and sensitivity tests ... 249
5.6 Additional tests ... 262
5.7 Conclusion ... 266

Chapter 6 .... **267**

Contents 268

6.0 Summary of main findings 269

6.0.1 Capital market effect of annual report sentiment 269

6.0.2 Management obfuscation in financial statements 270

6.0.3 Explanatory power of disclosure tone on corporate bankruptcy 271

6.0.4 Capital market effect of annual report readability 272

6.1 Limitations and future research avenues 273

References **276**

General appendix 292

Chapter 2: The effect of disclosure tone on market liquidity and market risk 292

Chapter 3: Accounting narrative obfuscation in financial statements 292

Chapter 4: Form 10-K textual analysis and corporate bankruptcy 292

Chapter 5: The effect of annual report readability on financial markets 292

Appendix for Chapter 2 293

Information extraction from financial statements 295

Appendix for Chapter 3 297

Appendix for Chapter 4 299

Appendix for Chapter 5 301

List of tables

Table 1. Sample selection process for liquidity and risk examination ... 34
Table 2. Distribution of Form 10-K filings across industries and fiscal years for liquidity and risk examination ... 35
Table 3. Dependent and independent variables for liquidity and risk examination ... 40
Table 4. Correlations between liquidity, risk, disclosure tone, and fundamental variables ... 41
Table 5. Regression coefficients on negative Form 10-K disclosure tone for liquidity and risk variables including all firm-year observations ... 44
Table 6. Regression coefficients on negative Form 10-K disclosure tone for liquidity and risk variables excluding firm-year observations of the financial industry and and low market capitalization ... 49
Table 7. Regression coefficients on negative Form 10-K disclosure tone for liquidity and risk variables including all firm-year observations using the Loughran and McDonald business dictionary including stop words ... 54
Table 8. Regression coefficients on negative Form 10-K disclosure tone for liquidity and risk variables including table content ... 59
Table 9. Regression coefficients on different Form 10-K disclosure tones for liquidity and risk variables ... 64
Table 10. Regression coefficients on negative Form 10-K disclosure tone changes for liquidity and risk variables ... 82
Table 11. Regression coefficients on negative Form 10-K disclosure tone measurements for liquidity and risk variables ... 88
Table 12. Regression coefficients on negative disclosure tone interaction effects for liquidity and risk variables ... 93
Table 13. Regression coefficients on negative disclosure tone in specific Form 10-K sections for liquidity and risk variables ... 100
Table 14. Sample selection process for stock return examination ... 124
Table 15. Distribution of Form 10-K filings across industries and fiscal years for stock return examination ... 125
Table 16. Dependent and independent variables for stock return examination ... 130
Table 17. Correlations between stock returns, disclosure tone, and firm fundamentals ... 131

Table 18. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market-adjusted model) 135
Table 19. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market model) 136
Table 20. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (4-factor model) 137
Table 21. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market-adjusted model) excluding firm-year observation of the financial industry and low market capitalization 139
Table 22. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market model) excluding firm-year observation of the financial industry and low market capitalization 140
Table 23. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (4-factor model) excluding firm-year observation of the financial industry and low market capitalization 141
Table 24. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (marked-adjusted model) using the Loughran and McDonald business dictionary including stop words 143
Table 25. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (marked model) using the Loughran and McDonald business dictionary including stop words 144
Table 26. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (4-factor model) using the Loughran and McDonald business dictionary including stop words 145
Table 27. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market-adjusted model) including table content 146
Table 28. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market-adjusted model) including table content 147
Table 29. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (4-factor model) including table content 148
Table 30. Regression coefficients on different disclosure tones (changes) in the footnotes sections for abnormal stock return (market-adjusted model) 150

Table 31. Regression coefficients on different disclosure tones (changes) in the footnotes sections for abnormal stock return (market model) 151

Table 32. Regression coefficients on different disclosure tones (changes) in the footnotes sections for abnormal stock return (4-factor model) 152

Table 33. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market-adjusted model) after controlling for net tone levels 154

Table 34. Regression coefficients for the disclosure tone changes in the footnotes sections for abnormal stock return (market model) after controlling for net tone levels 155

Table 35. Regression coefficients for the disclosure tone changes in the footnotes sections for abnormal stock return (4-factor model) after controlling for net tone levels 156

Table 36. Regression coefficients on different disclosure tones in the footnotes sections for return on assets and change in returns on assets 159

Table 37. Sample selection process for corporate bankruptcy examination 179

Table 38. Distribution of Form 10-K filings across industries and fiscal years for corporate bankruptcy examination 180

Table 39. Descriptive statistics for disclosure tones 184

Table 40. Regression coefficients on different Form 10-K disclosure tones (negative, positive, net) for corporate bankruptcy 186

Table 41. Regression coefficients on different Form 10-K disclosure tones for corporate bankruptcy excluding firm-year observations of the financial industry 187

Table 42. Regression coefficients on different Form 10-K disclosure tones for corporate bankruptcy using the Loughran and McDonald business dictionary including stop words 188

Table 43. Regression coefficients on different Form 10-K disclosure tones for corporate bankruptcy including table content 189

Table 44. Regression coefficients on different Form 10-K disclosure tones (uncertainty, litigious, constraining) for corporate bankruptcy 191

Table 45. Regression coefficients on different Form 10-K disclosure tones (changes) for corporate bankruptcy 193

Table 46. Regression coefficients on different Form 10-K disclosure tones (changes) after controlling for tone levels 196
Table 47. Regression coefficients on different disclosure tones in specific Form 10-K sections for corporate bankruptcy 201
Table 48. Regression coefficients on different Form 10-K disclosure tones for corporate bankruptcy after controlling for firm fundamentals........ 206
Table 49. Regression coefficients on different Form 10-K disclosure tones after controlling for firm fundamentals including all available observations (bankrupt and non-bankrupt firm-year observations) 209
Table 50. Regression coefficients on different Form 10-K disclosure tones for future Z-Scores 213
Table 51. Sample selection process for stock characteristics examination......... 230
Table 52. Distribution of Form 10-K filings across industries and fiscal years for stock characteristics examination 231
Table 53. Dependent and independent variables for stock characteristics examination 238
Table 54. Correlations between stock characteristics, annual report readability, and firm fundamentals 239
Table 55. Regression coefficients on Form 10-K readability (gross file size) for abnormal stock return (market-adjusted model, market model, and 4-factor model) 243
Table 56. Regression coefficients on Form 10-K readability (gross file size) for abnormal trading volume (average and median adjusted trading volume) 245
Table 57. Regression coefficients on Form 10-K readability (gross file size) for abnormal bid-ask spread (average and median adjusted bid-ask spread) 246
Table 58. Regression coefficients on Form 10-K readability (gross file size) for abnormal stock return volatility (market-adjusted model, market model, and 4-factor model) 247
Table 59. Regression coefficients on Form 10-K readability (gross file size) for abnormal stock return (market-adjusted model and market model) excluding firm-year observations of the financial industry and low market capitalization 250
Table 60. Regression coefficients on Form 10-K readability (gross file size) for abnormal trading volume (average and median adjusted) excluding firm-year observations of the financial industry and low market capitalization 252

Table 61. Regression coefficients on Form 10-K readability (gross file size) for abnormal bid-ask spread (average and median adjusted) excluding firm-year observations of the financial industry and low market capitalization 253
Table 62. Regression coefficients on Form 10-K readability (gross file size) for abnormal return volatility (market-adjusted model and market model) excluding firm-year observations of the financial industry and low market capitalization 254
Table 63. Regression coefficients on Form 10-K readability (net file size) for abnormal stock return (market-adjusted model, market model, and 4-factor model) 256
Table 64. Regression coefficients on Form 10-K readability (net file size) for abnormal trading volume (average and median adjusted) 258
Table 65. Regression coefficients on Form 10-K readability (net file size) for abnormal bid-ask spread (average and median adjusted) 259
Table 66. Regression coefficients on Form 10-K readability (net file size) for abnormal return volatility (market-adjusted model, market model, and 4-factor model) 260
Table 67. Regression coefficients on Form 10-K readability (gross and net file size) for return on assets 263
Table 68. Regression coefficients on Form 10-K readability (gross and net file size) for dividend yield 264
Table 69. Regression coefficients on Form 10-K readability (gross and net file size) for payout ratio 265

List of figures

Figure 1. Methods for textual analysis. 4

Abstract

The aim of this thesis is to analyse the impact of narrative information in annual reports on financial markets. While the effect of textual corporate information on stock returns is well examined in the literature, other firm characteristics are far less investigated. The current body of literature also neglects the relevance of specific text sections within annual report filings for capital market participants. Both aspects shall be addressed by this thesis. This thesis asks to what extent and in what ways narrative corporate disclosures affect a firm’s stock market characteristics. It does so to better enable academics, company officials, and investors to understand the importance of textual information disclosed in corporate reports. Using regression analysis on corporate textual information and stock characteristics, the thesis shows that narrative disclosures not only affect stock returns but also the liquidity and risk profiles of public listed companies. The thesis also shows that specific text sections within annual reports filed with the United States Securities and Exchange Commission (SEC) are more informative than others. The significance of this thesis is that it informs various corporate stakeholders about the importance of narrative corporate disclosures and how textual information can provide additional value beyond the financial figures disclosed by companies in regulatory filings.

Chapter 1

Introduction

Contents

1.0 Background and motivation 3

1.1 Capital market efficiency 7

1.2 Research questions 17

1.2.1 Capital market effect of annual report sentiment 17

1.2.2 Management obfuscation in financial statements 19

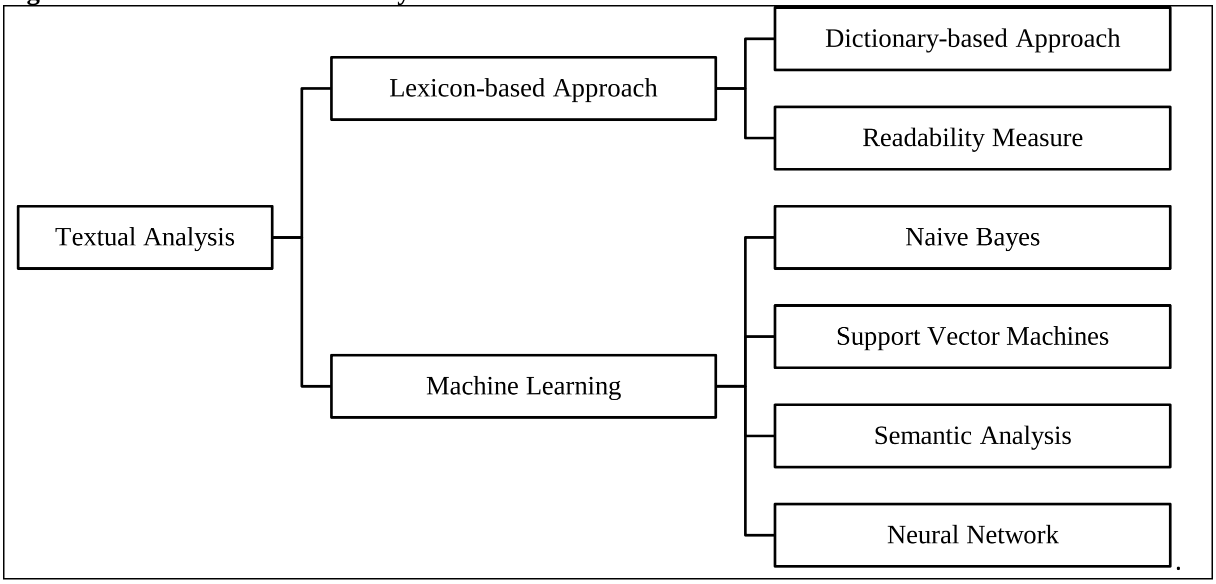
1.2.3 Explanatory power of disclosure tone on corporate bankruptcy 20

1.2.4 Capital market effect of annual report readability 21

1.0 Background and motivation

Because of technological advancements, an increasing amount of corporate information is available for textual analysis, also known as text analytics, computational linguistics, natural language processing, content analysis, and text mining, which has now become a growing field of research in the accounting and finance literature (Bannier et al. 2017, 4; Loughran and McDonald 2016, 1187-1188; Loughran and McDonald 2015, 1; Guo et al. 2016, 153). As a subset of qualitative analysis, textual analysis attempts to extract the information content from text documents (e.g. corporate reports, IPO prospectuses, earnings announcements, management forecasts, analyst presentations, conference calls, press releases, media articles, research notes, regulatory announcements, Internet messages) (Loughran and McDonald 2016, 1188, 1191; Elrod 2009, 4; Guo et al. 2016, 153-154). Despite its limitations or imprecision (e.g. parsing errors, loss of context), the overall aim of textual analysis is to extract the entire information content from a text document (Loughran and McDonald 2016, 1191-1192) and provide (potential) valuable information (Li 2010a, 145; Bannier et al. 2017, 1). In particular, textual analysis aggregates (transforms) the information content of narrative disclosures (words, numbers) into numeric variables for further analysis (Henselmann 2016, 364; Li 2010a, 145; Loughran and McDonald 2016, 1189; 1191). Using different techniques (e.g. lexicon-based approach, machine learning techniques), it transforms plain text into a vector of words to find valuable information (Guo et al. 2016, 154-155; Loughran and McDonald 2016, 1191). Besides manual content analysis, computer programs are used to extract the information content from text documents (computer-based content analysis). Two general methods for conducting computer-based content analysis are the rule-based (lexicon) approach and statistical (machine learning) approach. The rule-based (lexicon) approach distils information content by using a *"dictionary"* or *"word list"* to identify particular attributes of text documents. The statistical (machine learning) approach uses algorithms (e.g. Naïve Bayes, support vector machines, semantic analysis, neural networks) to classify textual information in text documents (Li 2010a, 146-147; Loughran and McDonald 2016, 1198-1200; Bannier et al. 2017, 13). Figure 1 presents an overview of the general methods to conduct textual analysis.

Figure 1. Methods for textual analysis.



Notes: The figure presents an overview of the methods widely used in the accounting and finance literature in the context of textual analysis (Guo et al. 2016, 155).

The actual process of textual analysis is threefold. First, narrative information is retrieved from a specific source of information. After obtaining finance-related firm information, a textual analysis cleans and parses the unstructured text data. Finally, it analyses the given qualitative disclosures in the given text document (Guo et al. 2016, 154-155) and obtains different variables of interest (e.g. readability, similarity, sentiment) to draw inferences (e.g. future stock return, future return on assets). In general, numerus information sources (e.g. regulatory filings, news articles, social media, board postings, Internet content) can be used to obtain these variables of interest (Kearney and Liu 2014, 172-173; Loughran and McDonald 2016, 1188; Loughran and McDonald 2015, 1; Bannier et al. 2017, 2, 4). Besides regulatory filings (non-professional or individual) investors often use various platforms on the internet (e.g. facebook, twitter, reddit) to access and process (corporate) information, especially in recent years. Despite that textual analysis can be applied to various information sources, most literature in capital market research context focuses on annual reports on Form 10-K filed with the United States Securities and Exchange Commission (SEC). This is because this information source is considered the most informative to (professional or institutional) investors (Das 2014, 85), even though other sources of information (e.g. tweets, board postings) have shown to have massive impacts on capital markets (e.g. stock price rallies in meme stocks).

The Securities Exchange Act of 1934 obliges public companies in the United States to file regulatory filings with the SEC (SEC 2013). Among other filings (e.g. quarterly reports on Form 10-Q, current reports on Form 8-K), all public domestic companies are required to file annual reports on Form 10-K with the SEC (Securities Exchange Act 1934, Section 2; Gerdes

2003, 7; SEC 2013). In addition to numerical information (quantitative data), corporate reports contain extensive textual information (qualitative data) (SEC 2009; Henselmann 2016, 362-363; Li 2010a, 143; Bannier et al. 2017, 4). In recent years, investors, researchers, and regulatory authorities have observed a steady increase in the volume of qualitative corporate disclosures, especially in regulatory filings (e.g. annual reports on Form 10-K) (Cazier and Pfeiffer 2016, 1, 9; Loughran and McDonald 2015, 4-5). Given the required accuracy of public corporate disclosures (Securities Exchange Act 1934, Section 13(i)), regulatory filings (potentially) provide capital market participants and researchers with valuable corporate information (if being understood) (Gerdes 2003, 9; Engelberg and Sankaraguruswamy 2007, 3; Chouliaras 2015, 1). For instance, narrative corporate disclosures are an important source of information concerning a company's financial condition (Elrod 2009, 4), and provide a useful context for understanding the financial data provided by a firm's management (e.g. accounting policies, financial risks) (Henselmann 2016, 363). More important, textual disclosures in corporate reports might also contain information (expectations) about a firm's future and financial performance (Chouliaras 2015, 1).

Research provides evidence for the importance and value-relevance of narrative or textual information in regulatory filings beyond financial statements (quantitative data) (Amel-Zadeh and Faasse 2016, 1; Li 2010a, 149; Loughran and McDonald 2016, 1189). While early work in the field of textual analysis considered relevant the readability (and similarity) of narrative corporate information (e.g. Li 2008; Lehavy et al. 2011; Loughran and McDonald 2014; Ditter 2015; Amel-Zadeh and Faasse 2016), recent empirical research documents the implications of the disclosure tone or textual sentiment in regulatory filings on important firm characteristics (e.g. Loughran and McDonald 2011a; Jegadeesh and Wu 2013; Gandhi et al. 2017) (Loughran and McDonald 2016, 1198-1199, 1214-1215). Processing all available quantitative and qualitative data investors (in theory) incorporate all value-relevant information (e.g. sentiment) into a firm's stock price (market value of equity) through trading activities (e.g. stock investments), therefore affecting various stock characteristics. Following the idea that a firm's market value (equity value) is the present value of its expected future cash-flows (discounted future cash-flows) (e.g. Gebhart et al. 2001; Claus and Thomas 2001; Easton 2004) (Zolotoy and Kalev 2012, 2) narrative disclosures in corporate reports can provide important information about a firm's expected future cash-flows (cash-flow news) as well as their discount-rate (discount-rate news) (reasons for unanticipated changes in equity value). In this context note that capital market participants generally act (trade) on a variety of factors including but not limited to firm fundamentals, technical factors, news, and market sentiment (often at the same time).

The explanatory ability of qualitative information in annual reports on future stock returns and corporate earnings is well documented in the literature (e.g. Loughran and McDonald 2011a; Jegadeesh and Wu 2013; Gandhi et al. 2017) most likely due to their increased relevance (financial interest) for a broader range of capital market participants (e.g. stockholders, bondholders, financial institutions) as well as for the overall public (e.g. journalists, labour unions, financial authorities). However, the relevance of narrative disclosures on other important firm (stock) characteristics is far less examined (e.g. stock liquidity, stock volatility). Narrative disclosures potentially influencing the liquidity and risk profile of exchange listed companies specific groups (e.g. fund managers, traders, corporate managers) are affected or at least interested in corporate textual information and their effects on financial markets (e.g. risk-adjusted benchmarking, trading strategies, disclosure choices). In addition, the majority of research studies only examine the relevance of narrative disclosures in the overall (entire) annual report on Form 10-K without even trying to identify the actual or true source of relevant information (specific subsections) within regulatory filings (e.g. *"Item 1A - Risk Factors"*, *"Item 7 - Management's Discussion and Analysis of Financial Condition and Results of Operations"* - *"MD&A"*). Besides using textual measurements (e.g. readability scores) associated with various substantial limitations and drawbacks (e.g. human judgement, computational cost, comparability) empirical results in the overall field of textual analysis are often mixed (e.g. Amel-Zadeh and Faasse 2016; Henselmann and Hering 2017; Seebeck et al. 2018) or even counter-intuitive (e.g. opposite effect of textual information or textual features on firm characteristics). Finally, the implications of different research methodologies and research design choices (e.g. type of textual information, measurement of text length, measurement of textual sentiment) in the field of textual analysis and corresponding inferences are merely examined. To address these gaps, this dissertation consists of four stand-alone research studies motivated by the abovementioned context:

(1) The effect of disclosure tone on market liquidity and market risk;

(2) Accounting narrative obfuscation in financial statements;

(3) Form 10-K textual analysis and corporate bankruptcy; and

(4) The effect of annual report readability on financial markets.

The first study analyses the effect of disclosure tone on a firm's future market liquidity and market risk to complement the question if and how textual information is affecting important stock characteristics other than a company's future stock return. The second study examines whether managers can successfully hide negative corporate information in annual reports on Form 10-K to clarify mixed empirical results in the literature while at the same time trying to find possible reasons for these results in the first place. The third study investigates the association between textual information and future corporate bankruptcies to examine if narrative disclosures can be used to better differentiate healthy and unhealthy companies. The fourth study examines the effect of textual complexity on financial markets using an alternative measurement for annual report readability to determine its validity in the field of textual analysis.

Overall, the four studies concentrate on the information content and value-relevance of the textual information disclosed in regulatory filings required to be made with regulatory authorities in the United States. While the first three studies focus on the implications and effect of textual sentiment in financial markets, the last one concentrates on the effect of textual complexity in regulatory filings on important firm characteristics. This dissertation makes several contributions to the literature on textual analysis in financial economics. The empirical results presented herein are not only useful for academic research, but also of interest to shareholders, bondholders, and other capital market participants, as well as to regulatory authorities.

The remainder of Chapter 1 discusses capital market efficiency and presents a short overview of the four research studies including their associated research questions. Chapters 2 - 5 present the four research studies separately. Chapter 6 concludes the dissertation with a summary of the main findings, limitations, and an outlook for future research avenues.

1.1 Capital market efficiency

Background

The evolution in the field of quantitative finance started in the early 1950s with computerized time series analysis. Since then, several theories on market efficiency have been discussed (Uhr et al. 2014, 1). In fact, economists have long been fascinated by the source of stock market variations (Pesaran 2005, 1). More particularly, to what extent past prices of a stock can be used to meaningful predict the future price of a stock (Fama 1965, 34). However, even after decades of research and thousands of research studies, economists have not yet reached consensus about whether (financial) markets are efficient or not (Lo and MacKinlay 1999, 6). A well-known story in economics tells of a finance professor and student who come across a $100 bill lying on the ground. As the student stops to pick it up, the professor says, *"Don't bother – if it were* *really a $100 bill, it wouldn't be there"* (Malkiel 2003, 60; Lo and MacKinlay 1999, 6). Illustrating what financial economists mean when they say markets are efficient (Malkiel 2003, 60), this humorous example of economic logic also describes the efficient markets hypothesis (EMH), one of the most contested propositions in the social sciences (Lo 2007, 1; Lo and MacKinlay 1999, 6).

The accepted view associated with the EMH is that when information arises (e.g. narrative disclosures in corporate reports), the news spreads quickly and is rapidly incorporated into securities prices (Malkiel 2003, 59), or more generally, that market prices fully reflect all available information (individual investors form expectations rationally, markets aggregate information efficiently, and equilibrium prices incorporate all available information instantaneously) (Lo 2007, 1; 3). Here, *"a market in which prices always fully reflect available information is called efficient"* (Fama 1970, 383). As a result, neither the study of past stock prices to predict future prices (technical analysis) nor the analysis of financial information such as corporate earnings (fundamental analysis) would help investors select undervalued stocks to achieve returns greater than those that could be obtained by holding a randomly selected portfolio of individual stocks (overall stock market index), at least not with comparable risk (Malkiel 2003, 59). Thus, price changes are unpredictable and random (random walk theory) so that even uninformed investors buying a diversified portfolio would obtain a rate of return as generous as that achieved by professional investors (Malkiel 2003, 59; Pesaran 2005, 1).[1] [2]

Versions of market efficiency

The random walk theory of asset prices postulates that in an information efficient market, price changes must be unforecastable and therefore random (Samuelson 1965, 41-42). It says that the future path of the price level of a security (stock) is no more predictable than the path of a series of cumulated random numbers (Fama 1965, 34). *"Why should after-the-fact price changes show any systematic pattern, such as non-bias"* (Samuelson 1965, 41)? *"The series looks like a wandering one, almost as if once a week the Demon of Chance drew a random number from a symmetrical population of fixed dispersion and added it to the current price to determine the next week's price"* (Kendall 1953, 13). In a statistical context the random walk theory implies

1 As Malkiel (2003) noted, *"a blindfolded chimpanzee throwing darts at the Wall Street Journal could select a portfolio that would do as well as the experts"* (Malkiel 2003, 60).

2 Pesaran (2005) explained that the random walk theory of stock prices was preceded by theories that movements in financial markets are related to the business cycle (Pesaran 2005, 1).

that successive price changes are independent, identically distributed random variables (successive price changes are independent and price changes conform to some probability distribution) (Fama 1965, 34-35). This means that knowledge of the sequence of price changes leading up to a certain period is of no help in assessing price changes (probability distribution) during that time period (Fama 1965, 35). Despite empirical evidence countering the random walk hypothesis, no profitable trading strategy could be generated in the presence of transaction costs (Pesaran 2005, 2). *"As far as these tests are concerned, it would seem that any dependence that exists in these series is not strong enough to be used either to increase the expected profits of the trader or to account for the departures from normality that have been observed in the empirical distribution of price changes. That is, as far as these tests are concerned, there is no evidence of important dependence from either an investment or a statistical point of view"* (Fama 1965, 80).

Nevertheless, the random walk model is a statistical statement rather than coherent theory of asset prices (Pesaran 2005, 2; Lo and MacKinlay 1999, 5). Based on the idea that in an informationally efficient market (not to be confused with an allocationally or Pareto-efficient market[3]) price changes cannot be forecasted (random walk theory) (Pesaran 2005, 2; Lo and MacKinlay 1999, 3), Fama (1970) formulated the EMH by distinguishing three forms of market efficiency:

1. The *"weak"* form suggesting that current security prices fully reflect all information about past prices (technical security analysis is useless for investors).[4]

2. The *"semi-strong"* form suggesting that current security prices fully reflect all publicly available information including information about past prices (technical and fundamental security analysis is useless for investors).

[3] Allocationally efficiency is a characteristic of an efficient market and only holds if markets themselves are efficient (e.g. prices fully and instantaneously reflect all available information). A market is allocatively efficient when the marginal rate of return (adjusted for risk) is equal for all producers/borrowers and savers. An allocative efficient market implies that investors provide funds for projects that have the highest net present value and that no "good" investment project remains unfunded (Pareto optimality). Note that (asset) markets can be informationally efficient (e.g. prices fully and instantaneously reflect all available information), without being allocatively efficient (e.g. imperfect competition due to monopolies) (Bauer 2004, 37-40; Koch 2011, 2-3; Gode and Sunder 1997, 603).

[4] As Fama (1991) noted, *"Market efficiency then implies that returns are unpredictable from past returns or other past variables, and the best forecast of a return is its historical mean"* (Fama 1991, 1578).

3. The *“strong”* form suggesting that current security prices fully reflect all public and private information (even inside or monopolistic information is useless for investors).

While the categorization of market efficiency into *“weak”*, *“semi-strong”*, and *“strong”* forms enables pinpointing the level of information at which the hypothesis breaks down, empirical tests are able to provide strong evidence for the EMH in the *“weak”* and *“semi-strong”* forms, but only limited evidence in the *“strong”* form (Fama 1970, 388). *“With few exceptions, the evidence is supportive”* (Fama 1991, 1602).

Besides the different versions of market efficiency, a corresponding market in which current prices reflect all available information requires that:

(i) there are no transactions costs in trading securities,

(ii) all available information is available at no cost to all market participants, and

(iii) all agree on the implications of current information for the current price and distribution of future prices of each security (Fama 1970, 387).

In contrast, *“a frictionless market in which all information is freely available and investors agree on its implications is, of course, not descriptive of markets met in practice”* (Fama 1970, 387). These conditions are sufficient for market efficiency, but not necessary (Fama 1970, 387).[5] As Fama (1970) noted, *“transaction costs, information that is not freely available to all investors, and disagreement among investors about the implications of given information are not necessarily sources of market inefficiency, they are potential sources”* (Fama 1970, 388).

Tests of market efficiency

The theory of efficient markets is concerned with whether prices at any time point fully reflect available information (Fama 1970, 413). Each individual test for capital market efficiency is concerned with the adjustment of security prices to one kind of information generating event (e.g. stock splits, announcements of financial results, security issues, etc.) (Fama 1970, 404).

[5] Note that on complete and competitive markets the receipt of private information cannot create any incentive to trade since *“the mere willingness of the other traders to accept their parts of the bet is evidence to at least one trader that his own part is unfavorable”* (Milgrom and Stokey 1982, 18). Thus, no trade takes place. The idea behind the *“no-trade theorem”* is that the private information will be incorporated into market prices (informational efficiency) before anyone accepts the trade or *“that prices should move more or less automatically with very little trading”* (Bouchaud et al. 2009, 14). This aspect has also been noted by Malkiel (1992): *“Formally, the market is said to be efficient with respect to some information set, if security prices would be unaffected by revealing that information to all participants”* (Malkiel 1992, 127).

The speed of the (stock) price response to new information and its measurement is the central issue for market efficiency (Fama 1991, 1601). However, *"the definitional statement that in an efficient market prices fully reflect available information is so general that it has no empirically testable implications"* (Fama 1970, 384). *"Thus, market efficiency per se is not testable"* (Fama 1991, 1575). According to Lo (2007), *"tests of the EMH are always tests of joint hypotheses. In particular, the phrase 'prices fully reflect all available information' is a statement about two distinct aspects of prices: the information content and the price formation mechanism. Therefore, any test of this proposition must concern the kind of information reflected in prices, and how this information comes to be reflected in prices"* (Lo 2007, 6). The joint hypothesis problem is the main obstacle to inferences about market efficiency (not the ambiguity about information nor trading costs) (Fama 1991, 1575). As stated by Fama (1991), *"it is a disappointing fact that, because of the joint hypothesis problem, precise inferences about the degree of market efficiency are likely to remain impossible"* (Fama 1991, 1576). Nevertheless, event studies (daily data) can attenuate or eliminate the joint hypothesis problem, that market efficiency must be tested jointly with an asset-pricing model (Fama 1991, 1601). The quick (stock) price adjustment following a corporate event or news announcement (firm-specific events) as a typical result in event studies is assumed (evidence) to be consistent with efficiency even though efficiency issues are never entirely resolved (Fama 1991, 1601-1602; 1607)

The *"weak"* form of market efficiency is most tested in the literature (e.g. Samuelson 1965; Mandelbrot 1966; Kendall 1953; Working 1934; Roberts 1959; Osborne 1959; Alexander 1961), and evidence provides strong support for this version of market efficiency (Fama 1970, 414).[6] While economists agree that previous returns cannot be used to make profitable predictions regardless of existing dependencies in historical returns (*"weak"* form of efficiency) (Fama 1970, 399), tests of *"semi-strong"* efficient capital markets are concerned with whether current prices fully reflect all obviously publicly available information (Fama 1970, 404). Furthermore, empirical research (e.g. Fama et. al 1969; Ball and Brown 1968; Waud 1970; Scholes 1969) supports the *"semi-strong"* version of the EMH (and by extension the *"weak"* form of efficiency) (Fama 1970, 404, 408-409; Pesaran 2005, 2). Specifically, the available evidence on the effect of various public announcements on stock returns is consistent with the efficient markets model (Fama 1970, 409; 415). *"In fact, no more than about ten to fifteen percent of the information in the annual earnings announcement has not been anticipated by the month of*

6 Fama (1970) highlighted that most tests of the *"weak"* form of market efficiency stem from the random walk literature (Fama 1970, 388).

the announcement" (Ball and Brown 1968, 175). The overwhelming support for the *"semi-strong"* version of the EMH in the literature comes from the consistency of market efficiency found by empirical studies, not from the quantity thereof (Fama 1970, 409).

In contrast, tests of market efficiency also provide evidence against the *"strong"* version of the EMH (Fama 1970, 409; Fama 1991, 1603). Studies concerned with whether all available information is fully reflected in security prices, including monopolistic information tests for the *"strong"* version of market efficiency (e.g. Niederhoffer and Osborne 1966; Scholes 1969), show that monopolistic access to information can be used to create higher expected trading profits (monopoly profits) (Fama 1970, 409-410; Fama 1991, 1603). However, corporate insiders and stock exchange specialists are the only two groups whose monopolistic access to information has been documented, and there is no evidence for deviations from the strong form of the efficient markets model in other parts of the investment community (Fama 1970, 415-416; Fama 1991, 1603). Being an unrealistic description of reality (Fama 1970, 409), the *"strong"* version of the EMH serves as a benchmark (interpreted in its strictest sense) against which to judge other forms of market efficiencies (deviations) (Fama 1970, 414-415; Pesaran 2005, 2). Consistent with this notion, the *"strong"* form of the EMH is far less tested in the literature than the other forms of capital market efficiency (Fama 1970, 409; 414).

Limitations of market efficiency

Despite being a useful concept, the idea of efficient capital markets is associated with various limitations and critiques. By the start of the twenty-first century, many financial economists and statisticians challenged the dominance of the EMH, contending that stock prices are at least partially predictable when using past stock prices and certain fundamental valuation metrics (Malkiel 2003, 60). In this context, the finance literature investigates numerous anomalies and predictable patterns in stock returns questioning the theory of efficient capital markets (e.g. momentum effect, return reversals, seasonal effects, dividend yield, price-earnings multiple, size effect, price-book ratio).

The original empirical work supporting the notion of randomness in stock prices (random walk theory, EMH) reinforced the view that the stock market has no memory (past stock prices cannot be used to predict future stock prices) (Malkiel 2003, 61). Inconsistent with this notion, the finance literature confirms the short-term momentum effect on financial markets when stock returns are measured over days and weeks (Malkiel 2003, 63). For instance, Lo and MacKinlay (1999) provide evidence that stock prices do not follow random walks. Their empirical results indicate that the stochastic behaviour of weekly returns is inconsistent with the random walk model, especially for stocks with smaller market capitalization (Lo and MacKinlay 1999, 18; 30-32). In particular, they found a significant positive serial correlation for weekly and monthly market index holding-period returns (not individual securities) (Lo and MacKinlay 1999, 18; 27-30). In addition, Lo et al. (2000) found that certain technical patterns based on past stock prices (local extrema of price series) do provide incremental information in forecasting future prices, at least in certain markets (Nasdaq) (Lo et al. 1731-1752). This suggests the possibility that technical analysis can add value to the investment process (not excess trading profits) (Lo et al. 2000, 1753). Furthermore, behaviour finance researchers assume that investors underreact to new information and that news is incorporated over time. Therefore, stock prices show a positive serial correlation over short periods (short-term momentum effect), rejecting the idea of total randomness in stock returns and therefore, the EMH (Malkiel 2003, 61).

In addition to the momentum effect (positive serial correlation), the finance literature also provides evidence for return reversals (negative serial correlation) over longer periods, challenging the EMH (Malkiel 2003, 63). Fama and French (1988) examined the importance of mean reverting price components of long-term stock returns and market efficiency (Fama and French 1988, 247). They found large negative autocorrelations for holding periods beyond a year (3-5 years) due to the mean reversion of returns, suggesting that price variations (return variances) are predictable from past returns (Fama and French 1988, 246-247). Despite being stronger for small firms, the predictability of long-term stock returns also holds true for large companies, indicating that stock prices have both random walk and stationary components (Fama and French 1988, 246-247; 256-259). Investigating the mean reverting behaviour of individual corporate securities, Poterba and Summers (1988) noted a positive autocorrelation in returns over short horizons and negative autocorrelation over longer horizons (Poterba and Summer 1988, 27; 36-38). Confirming previous studies (e.g. Fama and French 1988) on negative autocorrelation in long-term market returns (mean-reversion) (Poterba and Summer 1988, 36), the authors also found (some) long-horizon mean reversion for individual stock prices in relation to the overall market (Poterba and Summer 1988, 43-45). However, based on individual results, they concluded that the random walk hypothesis of stock returns cannot be consistently rejected (Poterba and Summer 1988, 53).[7]

Seasonal and day-of-the-week patterns are other examples of (possible) limited capital market efficiency. Several studies (e.g. Keim 1983; Haugen and Lakonishok 1988; French 1980) show that returns tend to be unusually high during the first two weeks of the year, especially for companies with small market capitalization (January effect), and on the first day of the week (Monday effect). Other research provides evidence for return patterns around the turn of the month (e.g. Lakonishok and Smith 1988) and around holidays (e.g. Ariel 1990) (Malkiel 2003, 64; Fama 1991, 1586-1589). However, these non-random effects (even if dependable) *"do not appear to offer arbitrage opportunities that would enable investors to make excess risk adjusted returns"* (Malkiel 2003, 64). In fact, one explanation for these seasonals might be that investors simply change their trading pattern (seasonals in market microstructure) (Fama 1991, 1587).

Besides capital market anomalies over the years, considerable research has been conducted to determine if future stock returns can be forecasted based on valuation parameters (e.g. price-earnings multiples, dividend yields) and other time series patterns (e.g. short-term interest rates, risk spreads) (Malkiel 2003, 64; 67). Contrary to efficient capital markets and inability to forecast stock returns, studies in the finance literature (e.g. Balvers et al. 1990; Breen et al. 1989; Campbell 1987; Campbell and Shiller 1988; Fama and French 1988; Fama and French 1989; Ferson and Harvey 1993; Kandel and Stambaugh 1996; Pesaran and Timmermann 1994) have predicted future stock returns over different time horizons (days, weeks, months). To this end, they employed dividend yields, price-earnings ratios, interest rates, and macroeconomic variables (Pesaran 2005, 3-4). However, the fact that future returns can be predicted using valuation metrics (dividend yield) is not necessarily inconsistent with capital market efficiency, since it may simply reflect the stock market adjustment to general economic conditions (high dividend yields, high interest rates) (Malkiel 2003, 65). This aspect has been noted by Fama (1991): *"The predictability of stock returns from dividend yields – D/P (or earnings price ratios - E/P) is not in itself evidence for or against market efficiency. In an efficient market, the forecast power of D/P says that prices are high relative to dividends when discount rates and expected returns are low, and vice versa. On the other hand, in a world of irrational bubbles, low D/P signals*

[7] According to Malkiel (2003), the finance literature provides considerable support for return reversals over longer holding periods. However, the evidence for mean reversion differs across research studies and time periods. In addition, return reversals or mean reversion is consistent with the efficient functioning of markets in the context of interest changes and their mean reversal (Malkiel 2003, 63-64).

irrationally high stock prices that will move predictably back toward fundamental values. To judge whether the forecast power of dividend yields is the result of rational variation in expected returns or irrational bubbles, other information must be used. As always, even with such information, the issue is ambiguous" (Fama 1991, 1583).

In addition to valuation metrics, research also documents the predictive ability of firm characteristics on future stock returns (e.g. size effect, price-book ratio) (Malkiel 2003, 67). Keim (1983) examined the empirical relation between (abnormal) stock returns and the market value of firms (Keim 1983, 13). The empirical results showed that the relation between stock returns and the companies' size is negative, even in years when large firms earn larger risk-adjusted returns than small ones (Keim 1983, 13; 16-19). Furthermore, the results show anomalous behaviour in returns related to the size effect, especially in January each year (seasonal effect), which cannot be explained by either risk characteristics or transaction costs (Keim 1983, 14; 20-25). Both results in the context of firm characteristics contradict the traditional capital asset pricing model (CAPM) (Keim 1983, 31) and therefore, the idea of efficient capital markets (EMH). Fama and French (1993) examined common risk factors in the returns on stocks and bonds. In the bond market, maturity and default risk are common risk factors: however, they identified the overall market risk as well as firm size and book-to-market equity (BE/ME) as common risk factors in the stock market (Fama and French 1993, 3). In particular, they constructed portfolios that mimic risk factors related to size and BE/ME, confirming that these factors proxy for common risk factors in stock returns (explain the differences in average returns across stocks) (Fama and French 1993, 4; 21; 24-26). Inconsistent with the idea of efficient capital markets, they concluded that *"the two empirically determined variables, size and book-to market equity, do a good job explaining the cross-section of average returns"* (Fama and French 1993, 4). In contrast, the fact that small companies generate larger returns than bigger ones (size effect) might not indicate capital market inefficiency, but rather a better risk proxy than other measures postulated in the finance literature (beta factor) (Malkiel 2003, 68).

Finally, critics of capital market efficiency and the EMH argue that there are several cases of market history where inefficiency cannot be denied, such as the stock market crash in 1987, Internet bubble in the late 1990s, and cases of irrational stock pricing (Malkiel 2003, 72-76).*"What one can discover from the three levels of market efficiency is that financial markets do not fit in any of these levels"* (Uhr et al. 2014, 1). On the other hand, supporters of the EMH may conclude that *"the evidence is overwhelming that whatever anomalous behavior of stock prices may exist, it does not create a portfolio trading opportunity that enables investors to* *earn extraordinary risk adjusted returns"* (Malkiel 2003, 60). *„In conclusion, only a mixture of inefficiency and efficiency can be the right point of view"* (Uhr et al. 2014, 1).

Market efficiency in practice

"In financial markets the EMH is respected but not worshiped" (Pesaran 2005, 8). It is believed that markets are efficient most of the time, but not all of the time, and that market inefficiencies can arise from important institutional and technological changes. However, it is not possible to foresee when and where such inefficiencies will arise in the future (Pesaran 2005, 8). Moreover, even if anomalies and predictable patterns in stock returns exist, they are neither robust nor dependable through time and could self-destruct in the future (Malkiel 2003, 71). Richard Roll, an academic financial economist and portfolio manager, noted this aspect: *"I have personally tried to invest money, my client's money and my own, in every single anomaly and predictive device that academics have dreamed up. […] I have attempted to exploit the so-called year-end anomalies and a whole variety of strategies supposedly documented by academic research. And I have yet to make a nickel on any of these supposed market inefficiencies. […] If there's nothing investors can exploit in a systematic way, time in and time out, then it's very hard to say that information is not being properly incorporated into stock prices"* (Roll and Shiller 1992, 30-31).

Note though that, *"market traders love volatility as it signals news and change with profit possibilities to exploit"* (Pesaran 2005, 8). The rejection of capital market efficiency (random walk hypothesis, EMH), as documented in the literature, and the (potential) presence of predictable components in the stock market (potentially) enable active investors (e.g. investment firms, portfolio managers, traders) to earn superior returns (Lo and ManKinlay 1999, 8). *"In much the same way that innovations in biotechnology can garner superior returns for venture capitalists, innovations in financial technology can garner equally superior returns for investors"* (e.g. better mathematical models, more accurate identification of investment opportunities, more timely data) (Lo and ManKinlay 1999, 8-10). While exploitable trading opportunities tend to be diversified across markets (bonds, equities, foreign exchange), price differences for different assets in different countries represent the most important trading opportunity in the context of market (in)efficiency (arbitrage trading) (Pesaran 2005, 8). Specifically, profits earned by arbitrageurs and investors in other inefficiencies can be viewed as economic rents for their innovation, creativity, and risk appetite (Lo and ManKinlay 1999, 16).[8] Therefore, *"under the practical version of the efficient markets hypothesis, it is difficult but not impossible to provide investors with consistently superior investment returns"* (Lo and ManKinlay 1999, 9). If utilized, relevant narrative corporate information in regulatory filings might represent such a source of superior returns for investors depending on their effects on financial markets (stock characteristics) and the efficiency of capital markets themselves. Note however, that this thesis and its empirical chapters (event studies) do not directly test the efficiency of capital markets but instead use the general concept of capital market efficiency (incorporation of relevant information) to examine if narrative disclosures in financial statements have important information content or not (market impact).

1.2 Research questions

1.2.1 Capital market effect of annual report sentiment

While the effect of quantitative corporate information (e.g. financial figures, financial ratios) on important firm characteristics is well examined in the literature, the implications of qualitative firm disclosures (e.g. narrative information) are far less analysed. However, because of technological advancements, a growing body of literature in accounting and finance has begun to investigate the relevance of textual corporate disclosures in financial markets. In the textual analysis field most research focuses on the effects and predictive ability of narrative information on future stock returns and future corporate earnings. Consequently, the question of how narrative or textual corporate disclosures affect other important firm characteristics in financial markets arises (e.g. stock liquidity, stock risk).

Research associates textual corporate information and tone thereof (textual sentiment) with stock characteristics and firm fundamentals. For instance, research shows that negative information is predictive for lower stock returns and corporate earnings in the future (e.g. Li 2006; Feldman et al. 2008; Li 2010b; Loughran and McDonald 2011a; Jegadeesh and Wu 2013; Campbell et al. 2014; Chouliaras 2015; Hope et al. 2016; Henry and Leone 2016; Amel-Zadeh and Faasse 2016; Gandhi et al. 2017). This is also true for future dividend payments to a firm's

8 Lo and ManKinlay (1999) highlighted that neither the evidence for nor against the random walk hypothesis is inconsistent with the practical version of the EMH, as market opportunities do not necessarily represent market inefficiencies (Lo and ManKinlay 1999, 16).

shareholders (e.g. Henselmann and Hering 2017). However, the effect of firm news on other important firm (stock) characteristics such as market liquidity and market risk remains unclear.

To address this research gap, this study investigates the effect of accounting narratives and their disclosure tone on a firm's market liquidity and market risk following submission to the SEC. Instead of using one dependent variable for a firm's market liquidity and one dependent variable for its market risk the study uses different proxy variables (abnormal trading volume, abnormal bid-ask spread, abnormal stock return volatility) constructed in different ways (average adjusted, median adjusted, market adjusted model, market model, 4-factor model) to investigate the effect of accounting narratives on stock characteristics in a broader context. In addition, the study examines the effect of different types and measurements of textual sentiment in corporate reports. Instead of using only one specific type and measurement of textual sentiment the study uses multiple types of linguistic tone (negative, uncertainty, litigious, constraining, positive) and tone measurements (sentiment ratio, sentiment ratio change) to better examine how and in which ways stock characteristics are affected by accounting narratives in regulatory filings. Furthermore, the importance of disclosure tone in specific annual report subsections for a firm's market liquidity and market risk is analysed. Instead of analysing the importance of the Form 10-K filing in its entirety or only one specific subsection within the annual report the study looks at all individual subsections within the submission trying to pinpoint the true source of impact on stock characteristics. Finally, the study evaluates different research design choices and their effects on empirical results to better understand their consequences in the field of textual analysis.

The study adds to the literature on textual analysis by providing evidence of an association between disclosure tone in accounting narratives and a firm's future market liquidity and market risk. In addition, the study is able to connect different types and measurements of disclosure tone to proxies for firm liquidity and risk. Furthermore, it demonstrates the effect of disclosure tone in specific Form 10-K subsections on stock characteristics. Finally, the study illustrates that research design choices can affect empirical results, thus contributing further to the overall field of research. Besides academia, the empirical results about the implications of narrative corporate information on the risk and liquidity profiles of public listed companies are also important for investors and regulatory authorities in the context of trading strategies and monitoring activities.

1.2.2 Management obfuscation in financial statements

Annual reports filed with the SEC consist of quantitative and qualitative data. The majority of disclosures in corporate reports are qualitative (textual information) (Henselmann 2016, 362-363; Li 2010a, 143; Bannier et al. 2017, 4). In recent years, it has been noted that the amount of narrative disclosures within corporate reports such as Form 10-K filings has significantly increased (Cazier and Pfeiffer 2016, 2). Given the importance as an information source for various stakeholders (e.g. shareholders, bondholders) (Chouliaras 2015, 1; Elrod 2009, 4; Henselmann 2016, 363), the clarity of narratives in annual reports is vital in understanding and interpreting the information a company discloses (Lo et al. 2017, 2). However, narrative disclosures and especially the way they are presented could also be used by managers to mislead or hide negative information in regulatory filings (Devos and Sarkar 2015, 3).

Based on an ever-increasing volume of corporate disclosures the accounting and finance literature investigates the presence and success of management obfuscation in regulatory filings (e.g. Bloomfield 2002; Li 2008; Devos and Sarkar 2015). While certain subsections of the annual report are processed by capital market participants (e.g. *"Item 7 - Management's Discussion and Analysis of Financial Condition and Results of Operations"* - *"MD&A"*) (Tavcar 1998, 10, 24-25; Humpherys et al. 2011, 585; Czyzewski and Wilkinson 2014, 99-100; Rogers and Grant 1997, 17-18; Amel-Zadeh and Faasse 2016, 7), others are less used by investors when analysing Form 10-K filings (e.g. *"Item 8 - Financial Statements and Supplementary Data"* - *"Footnotes"*) (Sutton et al. 2009, 3; Bedard et al. 2012, 25-26). Research has provided consistent evidence that the *"Footnotes"* are more likely to be used as a strategic deterrence to investors than other parts of the annual report filed with the SEC (Li 2008, 236; Sutton et al. 2009, 3; Bedard et al. 2012, 25-26). Besides being lengthy the notes to the financial statements are more technical than other sections, and thus are the most difficult to read (e.g. complex numerical content, lengthy legal disclosures) (Clatworthy and Jones 2001, 314; Amel-Zadeh and Faasse 2016, 4, 11). As such, they are less used by capital market participants, especially non-professional investors (Sutton et al. 2009, 4-6; Bedard et al. 2012, 25-26; Amel-Zadeh and Faasse 2016, 10-11). This leads to the question of whether managers can successfully hide negative corporate information in the footnotes of annual reports.

Motivated by this question and the mixed empirical evidence (e.g. Li 2008; Miller 2010; Goel and Gangolly 2012; Huan et al. 2014; Devos and Sarkar 2015; Amel-Zadeh and Fasse 2016; Henselmann and Hering 2017; Seebeck et al. 2018), this study examines the presence and success of accounting narrative obfuscation in annual reports filed with the SEC. The study analyses if company officials are able to hide negative information in the footnotes sections of annual reports. It also investigates if managers use positive statements in corporate reports to frame negative company news. Instead of focusing solely on the success of management obfuscation (market impact of negative information in the footnotes sections) the study also tries to examine the management behavior itself (tone management or framing behavior) to shed a more broader light on the overall concept of management obfuscation in regulatory filings (market reaction and original management behavior). In addition to investigating whether capital market participants react to different types of corporate information in the notes to the financial statements, the study also explores the predictive ability of the footnotes sections in annual reports regarding a firm's future accounting profitability. In doing so the study deeply questions the possibility of managers being able to successfully hide adverse information in the footnotes sections and the potential negative outcome of such a behavior. Finally, the study evaluates the impact of different research design choices on empirical results to better understand their consequences in the field of textual analysis.

The study contributes to the literature on qualitative information in several ways. It provides empirical evidence confirming that managers are unable to hide negative information in the footnotes sections of annual reports regardless of a firm's accounting profitability. It also shows that company officials do not frame negative news in the footnotes sections of annual reports. Furthermore, the study reveals that narrative disclosures in the notes to the financial statements are not predictive of future corporate earnings. Finally, the study illustrates that research design choices can have a substantial effect on empirical results and corresponding inferences. Besides academia, the empirical results are also important for managers and investors as well as for regulatory authorities. While company officials may use the results to reduce the impact of negative corporate information on financial markets, investors may use the findings to change their analysis of regulatory filings. The results could also be used to improve the information requirements of public traded companies and capital market efficiency by financial regulators.

1.2.3 Explanatory power of disclosure tone on corporate bankruptcy

Corporate bankruptcy is associated with tremendous economic losses. Given these losses for stockholders, bondholders, and other stakeholders, research has attempted to identify signals for corporate failure. Over decades, researchers in accounting and finance developed models using financial ratios and market variables to predict future firm bankruptcies (e.g. Altman Z-Score). Despite these efforts, few significant improvements have been made (Lopatta et al.

2017, 316), raising the question of how to improve or complement the quantitative bankruptcy prediction models used today.

Research shows that at least certain financial variables can be used to predict future firm bankruptcy (Altman 1968, 594). Other than quantitative models to forecast future firm failure, research is limited on whether textual information (e.g. accounting narratives) and its characteristics (e.g. linguistic tone) can be used for bankruptcy prediction. Very few studies in the field of accounting and finance and in the field of textual analysis provide evidence that textual corporate information can be used to better predict the future demise of companies (e.g. Henselmann and Scherr 2013; Mayew et al. 2015; Lopatta et al. 2017).

To expand this research field, this study investigates the predictive ability of textual information in annual reports on future firm bankruptcies. Instead of focusing on only one type of textual information, the study examines the predictive power of various types (negative, positive, overall, uncertainty, litigious, constraining) and measurements (sentiment ratio, sentiment ratio change) of narrative disclosures regarding future firm failures in much more detail. The study also investigates the importance of specific Form 10-K subsections or disclosure outlets and their information content in the context of corporate bankruptcy to identify the location of the most valuable textual information in regulatory filings. Finally, the study analyses the validity of textual information in quantitative bankruptcy prediction models and effect of various research design choices on empirical results better understand their consequences in the field of textual analysis.

The study adds to the literature in various ways. It shows that textual information in annual reports on Form 10-K filed with the SEC is associated with future firm bankruptcies. In addition, it connects different types of disclosure tone to future firm failures. Furthermore, the study provides evidence that the disclosure tone of narrative information in specific Form 10-K subsections is predictive for future corporate bankruptcies. The empirical results are also valuable for various groups including investors, practitioners, and regulatory authorities. Investors and practitioners can use the results to better evaluate financial investments and to improve their daily professional activities. Among other activities, regulatory authorities can use the results to improve their economic forecasts and risk monitoring activities.

1.2.4 Capital market effect of annual report readability

Over the years, different streams of research (e.g. readability, similarity, sentiment) focused on the implications of the textual characteristics disclosed in corporate reports on important firm characteristics. Besides developing new techniques to analyse the information content of text documents, such as measuring their tone sentiment, researchers have, especially in recent years, also tried to address the concerns and limitations of the established research techniques applied in the field of textual analysis (e.g. readability analysis). While the effect of textual complexity on various firm characteristics is well examined in the literature (e.g. Li 2008; Loughran and McDonald 2009), traditional readability measures are also associated with various significant drawbacks (e.g. human judgement, computational costs, comparability). Subsequently, a question regarding alternative readability measurements and implications thereof is raised.

Research shows that annual report readability affects various firm characteristics. For instance, high Form 10-K textual complexity is negatively associated (predictive) with a firm's stock return, corporate earnings, liquidity, and risk following SEC submission (e.g. Li 2008; Loughran and McDonald 2009; Miller 2010; Lehavy et al. 2011; Lawrence 2013; Loughran and McDonald 2014). Furthermore, few studies (e.g. Loughran and McDonald 2014) use alternative readability measures (e.g. document file size) to confirm the effects of textual complexity on firm characteristics (risk) (Loughran and McDonald 2014, 1660). However, the effect of textual complexity using alternative readability measures on various other company variables remains unanalysed in the literature.

To address this research gap, this study investigates the effect of annual report readability on important firm characteristics using document file size as an alternative measure for textual complexity. Instead of examining one single stock characteristic it examines a wide range of important stock features of exchange listed companies to better understand the implications of textual complexity on capital markets. In particular, the study examines the effect of textual complexity on a firm's stock return, market liquidity, and market risk after an annual report has been submitted. In addition, the study investigates whether disclosure complexity is predictive of a firm's future accounting profitability. Furthermore, the study analyses the association between Form 10-K readability and future dividend payments to investors again using an alternative readability measure. It does so to understand not only the implications for stock characteristics over a short time period around the SEC submission but also to explore the importance and the implications of textual complexity on capital markets over a longer time period after the annual report filing. Finally, it evaluates the effect of different research design choices on empirical results to better understand their consequences in the field of textual analysis.

The study contributes to the literature on textual analysis by providing evidence associating the document file size of an annual report as an alternative measurement of its readability with a firm's market performance, market liquidity, and market risk. In addition, the study shows that the complexity of a Form 10-K filing is predictive of lower corporate earnings in the future. It also indicates an association between annual report readability and future dividend payments to investors. Finally, the study demonstrates that research design choices can have an effect on empirical results in the field of textual analysis. Besides academia, the empirical results are important for investors and regulatory authorities. Knowledge about the effects of textual complexity on a variety firm characteristics may help investors and financial regulators to better understand the implications of textual complexity on financial markets in the context of investments and risk monitoring activities.

Chapter 2

The effect of disclosure tone on market liquidity and market risk

Contents

2.0 Introduction 26

2.1 Related literature and hypothesis development 28

2.2 Research design 33

2.2.1 Sample selection and description 33

2.2.2 Regression model, dependent variables, and independent variables 36

2.2.3 Summary statistics and correlations 39

2.3 Empirical results 42

2.4 Robustness tests 48

2.4.1 Sample selection 48

2.4.2 Disclosure length 53

2.4.3 Table content 58

2.5 Sensitivity tests 63

2.5.1 Type of disclosure tone 63

2.5.2 Measurement of disclosure tone 80

2.6 Additional tests 97

2.7 Conclusion 107

“*[…] disclosures tend to reduce the degree of information asymmetry between insiders and outsiders of a firm and also among various groups of current and potential investors of a firm”* (Kothari et al. 2009, 1645).

2.0 Introduction

Domestic companies in the US are obligated to file annual reports on Form 10-K with the United States Securities and Exchange Commission (SEC). According to the SEC a Form 10-K filing provides *“a comprehensive overview of the company's business and financial condition and includes audited financial statements”* (SEC 2009). Annual reports are highly relevant for capital market participants, as company officials use Form 10-K filings to inform investors about their expectations of a firm's future (Chouliaras 2015, 1). Besides quantitative information, annual reports contain a large amount of qualitative or textual information (narrative disclosures) (Li 2010a, 143; Bannier et al. 2017, 4; Lo et al. 2017, 2; Lehavy et al. 2011, 1089). Both quantitative and qualitative disclosures are important sources of information for shareholders, bondholders, and other capital market participants. In addition to the information content of quantitative disclosures (financial figures), qualitative disclosures (narrative information) can provide a useful context for understanding the financial data companies report in annual reports on Form 10-K. Furthermore, narrative disclosures might also reveal non-public corporate information to shareholders and other investors (Li 2010a, 143-144; Bannier et al. 2017, 5).

Textual analysis, as part of the literature on qualitative information (Loughran and McDonald 2011a, 37; Loughran and McDonald 2016, 1188), tries to identify important and value-relevant information about a firm's future in text documents (Grant and Conlon 2006, 119). In recent years, a stream of research has emerged to measure disclosure tone (textual sentiment) in regulatory filings (e.g. annual reports), media articles (e.g. newspapers), and Internet messages (e.g. board postings) (Kearney and Liu 2014, 173-174). Based on annual reports on Form 10-K filed with the SEC, Loughran and McDonald (2011a) developed a domain-specific business dictionary containing different word lists to measure the tone of textual information in financial disclosures (e.g. *“negative”*, *“positive”*, *“uncertainty”*, *“litigious”*) (Loughran and McDonald 2011a, 35-37; Loughran and McDonald 2015, 2). Transformed into quantitative measures, empirical studies connect the disclosure tone (linguistic style) of narrative information to a firm's future performance (e.g. future stock return, future return on assets) (Loughran and McDonald 2011a, 37).

Previous literature investigates the effect of disclosure tone in regulatory filings on a firm's future stock market performance and future accounting profitability (e.g. Li 2006; Feldman et al. 2008; Li 2010a; Loughran and McDonald 2011a; Loughran and McDonald 2011c; Jegadeesh and Wu 2013; Campbell et al. 2014; Chouliaras 2015; Hope et al. 2016; Henry and Leone 2016; Amel-Zadeh and Faasse 2016; Henselmann and Hering 2017). However, less is known about other important firm characteristics relevant for capital market participants (e.g. trading volume, bid-ask spread, stock return volatility). Thus, using textual sentiment analysis, this study examines the effect of the disclosure tone in accounting narratives on a firm's market liquidity and market risk in the days around the SEC submission. It estimates a regression model using an ordinary least squares regression with a firm's market liquidity and market risk as the dependent variables and several proxies for firm characteristics as independent variables. The regression results indicate that the disclosure tone in a firm's Form 10-K filing is associated with its market liquidity and market risk.

The study contributes to the literature in several ways. First, it provides empirical evidence that narrative disclosures in Form 10-K filings affect a firm's market liquidity and market risk. Second, the study connects different types and measurements of disclosure tone to a firm's market liquidity and market risk. Third, it demonstrates an association between narrative disclosures in specific Form 10-K subsections and stock characteristics, even after controlling for the information content of the entire annual report. Finally, the study illustrates the effect of different research design choices on empirical results in the field of textual analysis. Besides academia, the results of this study are also important for investors and regulatory authorities. The results help investors better understand narrative corporate information and their effects on financial markets. In addition, the results may contribute to improving stock trading strategies based on textual information. Furthermore, this study elucidates the relevance of the corporate disclosures required by regulatory authorities. Finally, it encourages the use of corporate reports as an information source for financial regulators in the context of their risk monitoring activities.

The remainder of Chapter 2 proceeds as follows. The next section presents the related literature and develops the hypotheses. Section 2.2 illustrates the sample selection process and research methodology. Section 2.3 provides the empirical results. Sections 2.4 and 2.5 report the robustness and sensitivity tests. The additional tests are presented in Section 2.6, and Section 2.7 concludes the Chapter.

2.1 Related literature and hypothesis development

Given the increasing amount of corporate information and computational power available for research (Loughran and McDonald 2016, 1226), a growing body of accounting and finance literature examines the association between textual features (e.g. disclosure tone) and firm characteristics (Ditter 2015, 92).

Li (2006) examined the association between the disclosure tone in annual reports and a firm's future stock market performance. Analysing the linguistic tone of Form 10-K filings in their entirety (submission including exhibits), Li found that the *"risk sentiment"* in annual reports is negatively related to stock returns following the SEC submission (Li 2006, 12-14, 26). By counting words related to risk and uncertainty (*"risk sentiment"*), the author further identified an association between the disclosure tone in regulatory filings and a firm's future corporate earnings (Li 2006, 4, 7, 12-14, 26). Feldman et al. (2008) evaluated the information content of non-financial data in annual and quarterly reports filed with the SEC (Form 10-K; Form 10-Q). Measuring disclosure tone by counting positive and negative words in the *"MD&A"* section ("*Management's Discussion and Analysis of Financial Condition and Results of Operations*") (Feldman et al. 2008, 3, 22, 40), the authors associated changes in textual sentiment with a firm's stock market performance over a short period following the SEC submission (Feldman et al. 2008, 26, 34). Li (2010) further examined the value-relevance of the *"MD&A"* section, focusing on the *"Forward-Looking Statements"* (*"FLS"*) disclosed in annual and quarterly reports. Based on a learning algorithm (Li 2010b, 1051, 1060-1061, 1094), he noted a positive relationship between the disclosure tone of the *"FLS"* and a firm's future corporate earnings (Li 2010b, 1050, 1073-1076, 1094). Consistent with the idea that that the *"MD&A"* section is intended to assess an enterprise's liquidity the author also found an association between the disclosure tone (positive sentiment) and a firm's corporate liquidity (operating cash flows divided by liabilities) in the year following the SEC submission (Li 2010b, 1050, 1073-1076, 1094). Developing an alternative word weighting scheme based on the market impact (*"word power"*), Jegadeesh and Wu (2013) significantly related the disclosure tone in annual reports to stock returns around the filing dates (Jegadeesh and Wu 2013, 712-713, 721-724, 728). Consistent with naïve expectations, the authors found that negative disclosures are associated with lower subsequent stock returns. On the other hand, positive accounting narratives cause positive market reactions (Jegadeesh and Wu 2013, 713, 721-723, 728). Furthermore, the authors provided evidence for an association between the disclosure tone in annual reports on Form 10-K and long-term capital market effects (Jegadeesh and Wu 2013, 713, 726, 728-729). Chouliaras

(2015) examined the effect of the disclosure tone in Form 10-K filings on future stock returns in the context of a companies' previous stock market performance (Chouliaras 2015, 2-3). He observed that changes in disclosure tone are related to subsequent stock returns (Chouliaras 2015, 10, 29). The author showed that the effect (impact) of the tone in a regulatory filing depends on a firm's stock market performance prior to the release of the annual report. Specifically, he provided evidence that firms with good stock market performance prior to the SEC submissions are negatively affected by more pessimistic disclosures while others are not (Chouliaras 2015, 13-14, 32). Amel-Zadeh and Faasse (2016) examined the value-relevance of the *"MD&A"* and *"Footnotes"* sections in Form 10-K filings (Amel-Zadeh and Faasse 2016, 2). The authors confirmed that the disclosure tone in the *"MD&A"* section is related to a firm's long-term stock return after an annual report was filed with the SEC (Amel-Zadeh and Faasse 2016, 24, 41). In addition, they demonstrated that tone differences between the *"MD&A"* and *"Footnotes"* sections (*"sentiment spread"*) are positively associated with future stock returns and corporate earnings (Amel-Zadeh and Faasse 2016, 25-27, 41-42). Based on their results, they concluded that Form 10-K filings and specific subsections contain valuable information for capital market participants (Amel-Zadeh and Faasse 2016, 30). Gandhi et al. (2017) used the disclosure tone in Form 10-K filings to measure financial distress (Gandhi et al. 2017, 3), finding that more pessimistic annual reports are predictive for market delistings, lower dividends, and lower corporate earnings (Gandhi et al. 2017, 4-5, 10, 14-20, 31-34). Given their results, they concluded that the disclosure tone in regulatory filings can be used to identify firms in financial distress (Gandhi et al. 2017, 21-22).

While the effect of textual sentiment in regulatory filings on a firm's future stock market performance and accounting profitability is well documented in the literature, empirical studies examining the association between linguistic tone and a firm's market liquidity and market risk are limited.

Loughran and McDonald (2011a) were among the first to examine the effect of the disclosure tone in Form 10-K filings on a firm's market liquidity and market risk. They showed that the tone of narrative information in annual reports is positively related to a companies' trading volume and volatility (Loughran and McDonald 2011a, 56-57). In another study, Loughran and McDonald (2011c) further confirmed that certain worrisome phrases are positively associated with a firm's stock return volatility, analyst forecast dispersion, and litigation risk (Loughran and McDonald 2011c, 91, 94-95). Kravet and Muslu (2013) also examined the effect of the disclosure tone (related to risk) in Form 10-K filings (Kravet and Muslu 2013, 1094), noting that changes in risk-related sentences within annual reports are associated with a higher stock return volatility around the filing dates (Kravet and Muslu 2013, 1105). In addition, their regression results revealed that the amount of risk-related information is positively associated with a firm's trading volume in the immediate days following the SEC submission (Kravet and Muslu 2013, 1106-1108). Campbell et al. (2014) analysed the value-relevance of the disclosure tone in a specific subsection of the annual report on Form 10-K (*"Item 1A - Risk Factors"*). The analysis indicated that the level of disclosures related to risk is associated with important firm characteristics. Among other things, the authors found a negative relationship between the amount of risk disclosures in *"Item 1A"* and a firm's bid-ask spread following the SEC submission (Campbell et al. 2014, 429-434). Finally, Hope et al. (2016) measured the amount of specific risk disclosures (*"Specificity"*) in *"Item 1A - Risk Factors"* of the annual report on Form 10-K. They observed that a high level of specific information related risk-factors (e.g. names of persons, locations, quantitative values) is associated with a firm's abnormal trading volume in the immediate days following the SEC submission (Hope et al. 2016, 1024-1025).

In general, company officials have strong incentives to avoid the disclosure of unfavourable information in regulatory filings (e.g. promotions, compensation) (Kothari et al. 2009, 1643-1644, 1647). However, the failing to disclose negative information (e.g. material risks) may expose managers to legal liabilities (e.g. lawsuits, claims payments) (Campbell et al. 2014, 402; Skinner 1994, 43-44). Given a firm's tendency to disclose negative information only if deemed necessary, this study assumes that negative corporate disclosures are credible in nature and therefore informative for capital market participants. Based on the "*efficient markets hypothesis*" (EMH) and the idea that narrative disclosures in regulatory filings such as annual reports on Form 10-K represent incremental information for investors, the study expects that a higher level of this information is associated with a higher level of market liquidity (higher trading volume, lower bid-ask spread). Despite that negative corporate disclosures might reduce the liquidity of a security by causing greater information asymmetries between informed and uninformed investors (Campbell et al. 2014, 404), previous research positively associated a higher level of negative corporate disclosures with a firm's market liquidity (e.g. Loughran and McDonald 2011a; Kravet and Muslu 2013; Campbell et al. 2014; Hope et al. 2016).

Based the idea that corporate management disclose meaningful and relevant information to capital market participants one would expect that other types of textual information (e.g. positive information) would have an effect on stock characteristics such as a firm's market liquidity.

For instance, one would expect that a higher amount of positive information is positively affecting the market liquidity of a company since positive statements (in principle) represent an incentive to invest into a firm. Inconsistent with this notion, prior empirical research (e.g. Loughran and McDonald 2011a) was not able to connect positive textual sentiment to a firm's market liquidity in the expected manner (even negative effect on liquidity). One explanation for this result might be that company officials have incentives to disclose meaningless positive information to present themselves in the best way possible (Rogers et al. 2011, 2158) or even to frame negative information in financial statements (management obfuscation) (Czerney et al. 2017, 19; Amel-Zadeh and Fasse 2016, 11; Loughran and McDonald 2011a, 38; Loughran and McDonald 2016, 1217). In addition, it might be also true that despite representing credible positive corporate information market participants simply neglect positive statements when assessing Form 10-K filings knowing about the general incentives of company officials to disclose this kind of information or assign a lower credibility to it. Thus, the following hypotheses are proposed:

Hypothesis 1a. There is a positive relationship between negative Form 10-K textual sentiment (disclosure tone) and a firm's trading volume (higher market liquidity).

Hypothesis 1b. There is a negative relationship between negative Form 10-K textual sentiment (disclosure tone) and a firm's bid-ask spread (higher market liquidity).

Theoretical research (e.g. Diamond and Verrecchia 1991; Verrecchia 2001; Healy and Palepu 2001; Easley et al. 2002; Easely and O'Hara 2004; Bushman and Smith 2001; Core 2001; Lambert et al. 2007) suggests that corporate disclosures lower a firm's cost of capital in several ways (e.g. reduced information asymmetries between a firm's investors, improved estimation of a firm's return on assets, improved assessment of a firm's market beta) and therefore, the (market) risk associated with the company. Consistently, a higher level or amount of incremental narrative information should be associated with lower market risks (e.g. stock return volatility) (Campbell et al. 2014, 403).

However, negative or unfavourable disclosures increase a firm's cash flow risk (Kothari et al. 2009, 1647). Based on the idea that a higher level of pessimistic information is associated with a higher volatility of future corporate earnings (increased cash flow risk), investors are less able to forecast a firm's future cash flows. Thus, the risk associated with a specific firm and therefore its market risk increases (Campbell et al. 2014, 403; Ball and Kothari 1989, 53).

Consistently, previous research has associated the disclosure tone in Form 10-K filings with firms' market risk (e.g. Loughran and McDonald 2011a; Loughran and McDonald 2011c; Kravet and Muslu 2013; Campbell et al. 2014). Despite that a higher level of negative corporate disclosures can also be negatively (not positively) associated with firms' risk (higher transparency by firm management), this study expects that negative information in regulatory filings increases investors' perception of risk associated with a firm and its market risk.

Again, if company officials in fact disclosure meaningful and relevant information to capital market participants one would expect (in principle) that other types of narrative disclosures would have an effect on stock characteristics such as a firm's market risk. For instance, positive corporate disclosures conveying (unexpected) incremental or value-relevant information regarding a firm's performance should be positively associated with a firm's market risk (irrespectively of the sign of the news). Despite this notion, only a few research studies (e.g. Loughran and McDonald 2011a) found a positive relationship between positive Form 10-K textual sentiment and a firm's market risk. In addition, one might also argue that positive information decrease the risk in a firm's expected future cash flows and therefore its risk profile (increased certainty with which investors can forecast future cash flows) (Kothari et al. 2009, 1647; Campbell et al. 2014, 403). Given the very limited empirical evidence concerning other types of textual sentiment in the context of a firm's market risk the following hypothesis is proposed:

Hypothesis 2. There is a positive relationship between negative Form 10-K textual sentiment (disclosure tone) and a firm's volatility (higher market risk).

2.2 Research design

2.2.1 Sample selection and description

All annual reports filed with the SEC between 1993 and 2016 were downloaded (Form 10-K; Form 10-K405) using the SEC EDGAR database (*"Form Index Files"*). Note that the time period was chosen due to data availability reasons on various information platforms (e.g. SEC EDGAR database, commercial data providers, public research repositories). An overview of the annual report submissions filed with the SEC is presented in the appendix.[9] For each annual report filing, the text version of the submission was retrieved from the EDGAR database (*"Complete Submission Text File"*). The text version contains all information disclosed within a particular SEC submission (e.g. core Form 10-K document, exhibits) (Bodnaruk et al. 2015, 643; Loughran and McDonald 2011a). The study applied an information extraction algorithm (regular expressions) to extract all narrative disclosures made within a submission. A detailed overview of the parsing procedure applied in this study is presented in the appendix. In addition to the entire Form 10-K submission (core document including exhibits), all individual subsections contained in the Form 10-K filings were extracted (non-relevant documents were deleted). Based on the extracted text sections (narrative information), word counts were obtained to quantify the disclosure tone (textual sentiment) in financial statements (Loughran and McDonald business dictionary).

For the analysis, several databases were used to match stock market data and accounting variables. For each annual report submitted to the SEC, the Central Index Key (CIK) from the document name war extracted. The CIK number was extracted to obtain the International Securities Identification Number (ISIN) and accounting variables for each financial statement (filer) from Standard & Poor's Capital IQ database. Based on the ISIN, stock market variables were retrieved for each company around its annual report filing from the Thomson Reuters DataStream platform. To include an annual report on Form 10-K in the data sample, accounting data (return on assets) and stock market data (market capitalization, price-to-book ratio, share turnover, stock return volatility, market risk) had to be available in addition to an exchange listing on the New York Stock Exchange (NYSE), National Association of Securities Dealers Automated Quotations (NASDAQ) or American Stock Exchange (NYSE MKT). Table 1 provides an overview of the sample selection process and number of annual reports included in the analysis.

9 Please note that the Form 10-K sample is identical to the one used in Henselmann and Hering (2017).

The majority of annual reports in the data sample were filed by companies in the *"Banking"* industry (12.70%), followed by *"Business Services"* (10.57%), *"Trading"* (6.34%), and *"Electronic Equipment"* (6.02%). In general, the distribution of financial statements among certain industries in the data sample is unequal (Fama-French 48 industry classification). While the distribution of firm-year observations across industries is unequal, the distribution of annual reports on Form 10-K over time is more balanced. The distribution of annual report submissions by certain industries and fiscal years is presented in Table 2.

Table 1. Sample selection process for liquidity and risk examination.

	Firm-year observations	**Observations removed**
All observations (Form 10-K & Form 10-K405)	189,998	
Less		
Missing 10-K report (database error)	189,997	1
Removing duplicates CIK and fiscal-year	184,089	5,908
Removing late filers (>100 days after accounting period)	163,838	20,251
Missing SIC (from filing)	162,759	1,079
Merge with Capital IQ data		
Missing ISIN	118,822	43,937
Missing return on assets	112,675	6,147
Missing NYSE, AMEX or NASDAQ exchange listing	81,749	30,926
Merge with DataStream data		
Missing market capitalization	78,754	2,995
Missing Price-to-Book and ratio < 0	75,205	3,549
Missing share turnover	73,088	2,117
Missing volatility	66,996	6,092
Missing market risk	63,493	3,503
Other data filters		
Number of words in Form 10-K (including exhibits) >= 2,000	63,311	182
Observations in final sample		
Firm-year sample	63,311	
Number of unique firms	7,221	-
Average number of years per firm	8.77	-

Notes: The table shows the sample selection process and number of Form 10-K filings included in the analysis.

Table 2. Distribution of Form 10-K filings across industries and fiscal years for liquidity and risk examination.

Panel A: Distribution of Form 10-K filings across industries

Fama-French industry	Filings (Num.)	Filings (%)	Filings (∑%)	Fama-French industry	Filings (Num.)	Filings (%)	Filings (∑%)
Business Services	7,767	12.38	12.38	Apparel	796	1.27	86.95
Trading	4,348	6.93	19.31	Banking	726	1.16	88.11
Electronic Equipment	4,185	6.67	25.98	Electrical Equipment	703	1.12	89.23
Pharmaceutical Products	4,028	6.42	32.40	Personal Services	706	1.12	90.35
Retail	3,210	5.11	37.51	Entertainment	676	1.08	91.43
Insurance	2,656	4.23	41.74	Real Estate	618	0.98	92.41
Petroleum and Natural Gas	2,478	3.95	45.69	Business Supplies	605	0.96	93.37
Utilities	2,453	3.91	49.60	Printing and Publishing	561	0.89	94.26
Medical Equipment	2,344	3.74	53.34	Almost Nothing	531	0.85	95.11
Computers	2,335	3.72	57.06	Rubber and Plastic Products	443	0.71	95.82
Machinery	2,201	3.51	60.57	Recreation	402	0.64	96.46
Wholesale	2,102	3.35	63.92	Aircraft	307	0.49	96.95
Measuring and Control Equipment	1,574	2.51	66.43	Non-Metallic and Industrial Metal Minin	269	0.43	97.38
Transportation	1,572	2.50	68.93	Beer & Liquor	248	0.40	97.78
Communication	1,527	2.43	71.36	Fabricated Products	193	0.31	98.09
Chemicals	1,280	2.04	73.40	Shipping Containers	195	0.31	98.40
Restaraunts, Hotels, Motels	1,208	1.92	75.32	Agriculture	168	0.27	98.67
Healthcare	1,197	1.91	77.23	Textiles	168	0.27	98.94
Food Products	1,017	1.62	78.85	Shipbuilding, Railroad Equipment	156	0.25	99.19
Construction Materials	919	1.46	80.31	Precious Metals	146	0.23	99.42
Construction	874	1.39	81.70	Defense	130	0.21	99.63
Automobiles and Trucks	872	1.39	83.09	Coal	101	0.16	99.79
Consumer Goods	817	1.30	84.39	Candy & Soda	71	0.11	99.90
Steel Works Etc	812	1.29	85.68	Tobacco Products	62	0.10	100.00

Panel B: Distribution of Form 10-K filings across fiscal years

Fiscal year	Filings (Num.)	Filings (%)	Filings (∑%)	Fiscal year	Filings (Num.)	Filings (%)	Filings (∑%)
1993	255	0.41	0.41	2005	3,088	4.92	48.75
1994	409	0.65	1.06	2006	3,105	4.95	53.69
1995	1,075	1.71	2.77	2007	3,099	4.94	58.63
1996	1,891	3.01	5.78	2008	3,045	4.85	63.48
1997	2,555	4.07	9.86	2009	3,124	4.98	68.46
1998	2,830	4.51	14.36	2010	3,118	4.97	73.43
1999	3,068	4.89	19.25	2011	3,068	4.89	78.32
2000	3,080	4.91	24.16	2012	3,109	4.95	83.27
2001	3,111	4.96	29.12	2013	3,243	5.17	88.44
2002	3,093	4.93	34.05	2014	3,263	5.20	93.64
2003	3,054	4.87	38.91	2015	3,231	5.15	98.79
2004	3,083	4.91	43.83	2016	760	1.21	100.00

Notes: The table illustrates the number of annual reports in the data sample across Fama-French industries (Panel A) and across fiscal years (Panel B).

2.2.2 Regression model, dependent variables, and independent variables

Regression model

To examine the effect of the disclosure tone in annual reports on a firm's market liquidity and market risk, the study estimates the following ordinary least squares regression (OLS)[10] :

$$\begin{aligned} Y_{i;t} = \beta_0 + \beta_1 * Tone_{Neg;Full\ Report;i;t} + \beta_2 * Ln\ MVE_{i;t} + \beta_3 * PtB_{i;t} \\ +\beta_4 * ROA_{i;t} + \beta_5 * Turnover_{i;t} + \beta_6 * Volatility_{i;t} + \beta_7 * Market\ Risk_{i;t} \\ + \sum Year\ Fixed\ Effects + \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \quad (1)$$

Where: the dependent variable ($Y_{i;t}$) is a firm's market liquidity (abnormal trading volume, abnormal bid-ask spread) and market risk (abnormal stock return volatility) around its Form 10-K filing. $Tone_{Neg;Full\ Report;i;t}$ is the disclosure tone (textual sentiment) in the entire annual report on Form 10-K (submission including exhibits) as measured by the number of negative words in the SEC filing. The control variables are described below. Year Fixed Effects and Industry Fixed Effects represent the included fixed effects.

Dependent variables

To test whether there is an association between Form 10-K textual sentiment and a firm's market liquidity and market risk, several market variables are obtained (dependent variables).[11]

The first measure of a firm's market liquidity in the days around its Form 10-K filing is the abnormal trading volume (ATV):

$$ATV_{i;[-1,+3];Ave} = \sum_{t-1}^{t+3} \left(Share\ Turnover_t - \emptyset\ Share\ Turnober_{[\mathrm{t-156,t-6}]} \right) \quad (2)$$

Where: $ATV_{i;[-1,+3];Ave}$ is the abnormal trading volume of a certain stock (i) calculated as the cumulative daily share turnover (proportion of traded shares) during the event period ([t-1, t+3]) minus the average share turnover during the non-event period ([t-156, t-6]). Note that the abnormal trading volume is also calculated by adjusting for the median share turnover during the

[10] Note that linear regression models are most commonly used in the accounting and finance literature when examining the effect of textual information on firm characteristics (Kearney and Liu 2014, 177). Further note that the study used a linear regression model over various machine learning models (e.g. neural networks) to better examine the relationship between input (independent) and output (dependent) variables.

[11] Note that the study focused on short-term capital market effects following the Form 10-K submission to avoid biases due to other effects not being captured by the model (e.g. other corporate information not being disclosed in the annual report, other corporate events, market trends).

non-event period ($ATV_{i;[-1,+3];Med}$) (Ditter 2015, 122, 146, 148; Atiase and Bamber 1994, 313-314; Bamber et al. 1997, 585, 588-589). Further note that the starting date of the event period was chosen since prior research (e.g. Morse 1981; Bamber 1987) suggests that the bulk of the trading volume reaction occurs on days -1 and 0 (Atiase and Bamber 1994, 313) and is well established in the literature (e.g. Kravet and Muslu 2013; Blankespoor et al. 2014; Ditter 2015).

The second measure of a firm's market liquidity during the event window starting one day before the initial filing date of the annual report is the abnormal bid-ask spread (ASpread):

$$ASpread_{i;[-1,+3],Ave} = \sum_{t-1}^{t+3} \left(Spread_t - \emptyset\, Spread_{[\mathrm{t-156,t-6}]} \right) \tag{3}$$

Where: $ASpread_{\,i;[-1,+3];Ave}$ is the abnormal bid-ask spread of a certain stock (i) calculated as the cumulative daily bid-ask spread during the event period ([t-1, t+3]) less the average bid-ask spread during the non-event period ([t-156, t-6]). Note that the abnormal bid-ask spread is also calculated by adjusting for the median bid-ask spread during the non-event period ($ASpread_{\,i;[-1,+3];Med}$) (Ditter 2015, 122, 146, 148; Bushee et al. 2010, 6-7; Blankespoor et al. 2014, 1475-1476, 1478).

To assess the association between the disclosure tone in annual reports and a firm's market risk, the study determined a company's abnormal stock return volatility (ARV) during a five-day event window around the time of the Form 10-K submission:

$$ARV_{i;[-1,+3];MAM} = \sum_{t-1}^{t+3} |(Return_{i;t} - Return_{S\&P\ 500;t})| \tag{4}$$

Where: $ARV_{i;[-1,+3];MAM}$ is the abnormal stock return volatility of a company (i) calculated as the sum of the absolute daily abnormal stock returns during the event period ([t-1, t+3]) (Ditter 2015, 122, 147; Bailey et al. 2003, 2493-2494). The abnormal stock returns are computed using the S&P 500 stock market index as a benchmark portfolio (marked-adjusted model - MAM).

In addition, a market model (MM) was employed to calculate a firm's abnormal stock return volatility (ARV) during the event period ([t-1, t+3]):

$$ARV_{i;[-1,+3];MM} = \sum_{t-1}^{t+3} |Return_{i;t} - (\hat{\alpha}_i + \hat{\beta}_i * Return_{S\&P\ 500;t})| \tag{5}$$

Where: $\hat{\alpha}_i$ (intercept) and $\hat{\beta}_i$ (slope) are estimated using an ordinary least squares (OLS) regression of the daily returns of the company (i) on the daily returns of the market index (S&P 500) during the non-event period ([t-250, t-21]).

Finally, a 4-factor model (4 Factor) was implemented to obtain the abnormal stock return volatility (ARV) of a company during the event period ([t-1, t+3]):

$$ARV_{i;[-1,+3];4\,Factor} = \sum_{t-1}^{t+3} |Return_{i;t} - Return_{i;t;4\,Factor}| \tag{6}$$

Where: $Return_{i;t;4\,Factor}$ is the expected return of a company (i) on day (t) based on a 4-factor model. Using the data provided by Kenneth R. French daily stock returns were regressed on the excess market returns (MKT-RF), a size factor (small-minus-big), a value factor (high-minus-low), and a momentum factor (up-minus-down) over an estimation window ([t-250, t-21]) relative to the Form 10-K filing dates to estimate the corresponding factor loadings for the model (Amel-Zadeh and Faasse 2016, 16-17).

Note that the study followed prior research studies in the field of textual analysis (e.g. Amel-Zadeh and Faasse 2016) in defining the non-event period to avoid biases caused by abnormal trading behaviour before the release date of the annual report. Further note that dividends and other capital gains being paid (total return) were accounted for.

Main variable of interest

The study investigates the effect of the disclosure tone (textual sentiment) in financial statements on a firm's market liquidity and market risk. To quantify the linguistic tone in regulatory filings, it used the Loughran and McDonald business dictionary (domain-specific) and its word lists (*"negative"*, *"positive"*, *"uncertainty"*, *"litigious"*, *"constraining"*).[12] Following prior work (e.g., Henry and Leone 2016), an equal weighting scheme was applied to quantify the disclosure tone in the annual reports.[13]

[12] Note that the study used a domain-specific dictionary (Loughran and McDonald's business dictionary) over a general-purpose dictionary (e.g. Harvard dictionary) to effectively gauge narrative corporate information in a capital market setting (Henry and Leone 2016, 157; Loughran and McDonald 2016, 1203-1204). Furthermore, the dictionary approach to measuring disclosure tone is considered objective and replicable (Loughran and McDonald 2016, 1200).

[13] Note that the study used an equal weighting scheme over a term-frequency inverse document frequency (tf.idf) weighting scheme as the former is considered objective and replicable (Tetlock et al. 2008, 1440; Henry and Leone 2016, 158).

The measure for the disclosure tone in financial statements is the number of negative words scaled by the text length of a particular Form 10-K report:

$$Tone_{Neg;Full\ Report;i;t} = \frac{Negative_{Full\ Report;i;t}}{Total_{Full\ Report;i;t}} \quad (7)$$

Where: $Tone_{Neg;Full\ Report;i;t}$ is the negative disclosure tone (textual sentiment) in the entire annual report on Form 10-K (submission including exhibits), $Negative_{Full\ Report;i;t}$ is the number of negative words in the Form 10-K filing, and $Total_{Full\ Report;i;t}$ is the number of all words appearing in the SEC submission.

Note that the Loughran and McDonald business dictionary (excluding stop words) was used to determine the text length of a certain Form 10-K filing. Furthermore, this process of quantifying the disclosure tone of text documents is objective and replicable (Rogers et al. 2011, 2162; Loughran and McDonald 2016, 1200).

Control variables

To control for common firm fundamentals in each regression, the natural logarithm of the market value of equity and price-to-book ratio as well as a firm's return on assets were included as control variables. Following previous literature (e.g. Ditter 2015; Blankespoor et al. 2014; You and Zhang 2009), the average daily share turnover and stock return volatility in the prior fiscal period were further included as control variables in addition to the firm's present market risk (Ditter 2015, 123). In addition, all regressions employed year and industry fixed effects. Standard errors were clustered by firm. Finally, all included variables were winsorized at the 1% level. A detailed overview of all variables is presented in the appendix.

2.2.3 Summary statistics and correlations

This section presents the summary statistics for the market, disclosure, and accounting variables and the correlations between the key variables of interest in the study. The mean (median) abnormal trading volume (ATV) of a company during the event period has a positive value of 0.36 (0.05) percent. During the event period the abnormal bid-ask spread (ASpread) of a company in the data sample is -0.09 (-0.07) percent. The average abnormal stock return volatility (ARV) in the days around the Form 10-K filing is 10.06 (7.41) percent. Panel A of Table 3 provides the summary statistics for the dependent variables used as proxies for a firm's market liquidity and market risk around the filing date of the SEC submission.

The mean (median) ratio of negative information in an annual report is 3.06 (3.04) percent. The mean market value of a firm included in the analysis is $3,410M ($519M). The average price-to-book ratio is 3.05 (1.89). The return on assets has a mean (median) value of 2.96 (3.86) percent. The mean (median) daily share turnover of the companies included in the data sample is 0.69 (0.48) percent. The average (median) volatility of a stock examined in this study is 3.00 (2.53) percent. The mean (median) market risk of a firm in the data sample is 0.91 (0.86). Panel B of Table 3 presents the summary statistics for the market, disclosure, and accounting variables used as independent variables in the analysis (Panel B).

Form 10-K textual sentiment generally relates to liquidity and risk variables in the expected manner. A higher amount of negative annual report information is correlated with a lower abnormal bid-ask spread and a higher abnormal stock return volatility. Conversely, a more negative annual report is correlated with a lower abnormal trading volume. Textual sentiment in annual reports on Form 10-K also generally relates to other variables in the expected manner. A more negative Form 10-K filing is correlated with a lower price-to-book ratio, lower return on assets, higher share turnover, higher volatility, and higher market risk. Conversely, a more pessimistic disclosure tone in annual reports is correlated with a higher market value of equity. Table 4 provides the correlations for the disclosure tone in annual reports on Form 10-K and firm fundamentals.

Table 3. Dependent and independent variables for liquidity and risk examination.

Panel A: Dependent variables

Variable	Num.	Mean	Median	St. Dev.
$\mathrm{ATV}_{[-1,3];\mathrm{Ave}}$	59,968	0.0036	0.0005	0.0267
$\mathrm{ATV}_{[-1,3];\mathrm{Med}}$	59,968	0.0107	0.0027	0.0279
$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	33,047	-0.0009	-0.0007	0.0549
$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$	33,047	0.0094	0.0002	0.0559
$\mathrm{ARV}_{[-1,3];\mathrm{MAM}}$	57,739	0.1006	0.0741	0.0851
$\mathrm{ARV}_{[-1,3];\mathrm{MM}}$	57,739	0.0980	0.0709	0.0853
$\mathrm{ARV}_{[-1,3];4\ \mathrm{Factor}}$	57,739	0.0987	0.0717	0.0849

Panel B: Independent variables

Variable	Num.	Mean	Median	St. Dev.
$\mathrm{Tone}_{\mathrm{Neg;Full\ Report}}$	63,311	0.0306	0.0304	0.0086
Ln MVE	63,311	20.1406	20.0675	1.9114
PtB	63,311	3.0509	1.8900	4.0555
ROA	63,311	0.0296	0.0386	0.0985
Turnover	63,311	0.0069	0.0048	0.0068
Volatility	63,311	0.0300	0.0253	0.0174
Market risk	63,311	0.9053	0.8641	0.5530

Notes: The table provides an overview of stock characteristics (market variables) around the Form 10-K submission and common firm fundamentals included in the analysis as control variables.

Table 4. Correlations between liquidity, risk, disclosure tone, and fundamental variables.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1: $\mathrm{ATV}_{[-1,3];\mathrm{Ave}}$	1.00	0.95***	-0.01**	-0.03***	0.33***	0.33***	0.34***	0.02***	0.04***	0.03***	0.05***	0.06***	-0.02***	0.06***
2: $\mathrm{ATV}_{[-1,3];\mathrm{Med}}$	0.92***	1.00	-0.01**	-0.04***	0.36***	0.37***	0.37***	0.05***	0.05***	0.05***	0.02***	0.26***	0.07***	0.12***
3: $\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	-0.02***	-0.01**	1.00	0.85***	0.07***	0.06***	0.06***	0.00	0.00	0.01	-0.01	0.02***	-0.02***	0.03***
4: $\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$	-0.04***	-0.04***	0.75***	1.00	0.18***	0.17***	0.18***	-0.03***	-0.19***	-0.01**	-0.07***	-0.09***	0.12***	-0.11***
5: $\mathrm{ARV}_{[-1,3];\mathrm{MAM}}$	0.27***	0.31***	0.05***	0.17***	1.00	0.98***	0.96***	0.06***	-0.41***	-0.02***	-0.23***	0.04***	0.59***	0.02***
6: $\mathrm{ARV}_{[-1,3];\mathrm{MM}}$	0.27***	0.32***	0.05***	0.16***	0.96***	1.00	0.98***	0.07***	-0.40***	-0.02***	-0.23***	0.06***	0.59***	0.05***
7: $\mathrm{ARV}_{[-1,3];\mathrm{4\ Factor}}$	0.27***	0.32***	0.05***	0.16***	0.93***	0.96***	1.00	0.06***	-0.41***	-0.01**	-0.24***	0.05***	0.60***	0.03***
8: $\mathrm{Tone}_{\mathrm{Neg;Full\ Report}}$	-0.01**	0.02***	-0.02***	-0.01**	0.07***	0.08***	0.07***	1.00	0.00	-0.03***	-0.20***	0.12***	0.11***	0.09***
9: Ln MVE	0.06***	0.08***	0.06***	-0.11***	-0.45***	-0.42***	-0.43***	-0.03***	1.00	0.15***	0.31***	0.33***	-0.48***	0.28***
10: PtB	0.03***	0.07***	0.03***	-0.06***	-0.13***	-0.11***	-0.11***	-0.11***	0.36***	1.00	-0.04***	0.06***	-0.01**	0.04***
11: ROA	0.07***	0.07***	0.04***	-0.04***	-0.21***	-0.19***	-0.19***	-0.26***	0.39***	0.33***	1.00	0.07***	-0.35***	0.01**
12: Turnover	-0.03***	0.16***	0.12***	-0.02***	-0.01	0.03***	0.02***	0.11***	0.50***	0.21***	0.22***	1.00	0.19***	0.41***
13: Volatility	-0.07***	0.01	-0.04***	0.04***	0.61***	0.62***	0.63***	0.13***	-0.50***	-0.14***	-0.26***	0.14***	1.00	0.22***
14: Market Risk	0.03***	0.09***	0.06***	-0.04***	0.04***	0.09***	0.06***	0.09***	0.28***	0.11***	0.05***	0.47***	0.27***	1.00

Notes: The table provides the correlation results for Form 10-K textual sentiment (disclosure tone) and firm characteristics. Pearson correlation coefficients are presented in the upper triangle. Spearman correlation coefficients are presented in the lower triangle.

2.3 Empirical results

Consistent with *Hypothesis 1a* the regression coefficient on negative disclosure tone for a firm's market liquidity as measured by its abnormal trading volume is significantly positive. A higher level on negative textual sentiment in Form 10-K is positively associated with a firm's trading activity around the SEC submission. Despite being negative in nature, capital market participants seem to interpret pessimistic or negative corporate disclosures as a sign of transparency by company officials (credible information), resulting in higher liquidity in a firms' shares (higher abnormal trading volume). This result is consistent with previous literature documenting a positive relationship between negative Form 10-K textual sentiment and a firm's market liquidity (e.g. Loughran and McDonald 2011a). Therefore, the first hypothesis is confirmed.

Inconsistent with *Hypothesis 1b,* the regression coefficient on negative disclosure tone for a firm's market liquidity as measured by its abnormal bid-ask spread is insignificant (average based abnormal bid-ask spread). A higher amount of negative corporate information in annual reports does not seem to reduce information asymmetries between investors, as reflected in a higher market liquidity (lower abnormal bid-ask spread). This result is inconsistent with previous empirical works in the field of textual analysis, which document a negative (positive) relationship between negative disclosure tone and information asymmetries among investors (market liquidity). Thus, the second hypothesis cannot be confirmed.

Consistent with *Hypothesis 2*, the regression coefficient on negative disclosure tone for a firm's market risk as measured by its abnormal stock return volatility is significantly positive. Despite being associated with higher market liquidity, a higher portion of negative textual information in Form 10-K filings seems to increase investors' risk perception of a specific company. This result is consistent with previous empirical research showing the same association between pessimistic disclosure tone and a firm's market risk (e.g. Loughran and McDonald 2011a). Based on this result, the final hypothesis of the study is confirmed. Table 5 presents the estimated OLS coefficients on negative Form 10-K disclosure tone.

When interpreting the results presented in this study note that certain variables not included in the study might explain the empirical findings (omitted variable bias). Besides other variables associated with the main variable of interest and the dependent variables (e.g. credit rating changes) the corporate governance of a company might influence the disclosure behaviour of a firm in regulatory filings while at the same time affecting various stock characteristics (despite standardized filing, disclosure, and auditing requirements). Please further note the differences in explanatory power for each dependent variable (stock characteristic) which might be explained by their variability itself and various market environments and market forces over time.

Overall, these results confirm that the disclosure tone in annual reports on Form 10-K affects a firm's market liquidity and market risk in the days around the SEC submission. More important, these results reveal that regulatory filings in general and disclosure tone in particular contain relevant textual information for shareholders and other capital market participants.

The empirical results of this study also support the notion that the disclosure tone of narrative information in regulatory filings can affect investors' trading activities on financial markets. Given the narrative information in a firm's annual report, a lower stock trading volume might lead to investors' inability to appropriately position themselves in a particular stock. The empirical results further show that regulatory authorities require relevant information from companies through corporate reporting. Based on the results of this study, financial regulators may also consider textual information besides quantitative information when monitoring the risk of specific companies or the overall stock market. More broadly, the results also fit into the general idea of the efficient markets hypothesis (EMH) in a sense that capital market participants do seem to (quickly) incorporate not only quantitative (financial) but also qualitative (textual) information about a company into its stock market characteristics through their (trading) activities and behavior on financial markets.

Table 5. Regression coefficients on negative Form 10-K disclosure tone for liquidity and risk variables including all firm-year observations.

Panel A: Abnormal trading volume (ATV)

Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**0.0526*** **	**0.0753*** **		**0.191*** **	**0.0773*** **
			(2.943)	**(4.156)**		**(9.539)**	**(4.248)**
Ln MVE		3.79e-06		-1.09e-05	-0.000719***		-0.000734***
		(0.0377)		(-0.109)	(-7.077)		(-7.241)
PtB		0.000188***		0.000194***	0.000238***		0.000244***
		(4.754)		(4.911)	(5.839)		(5.994)
ROA		0.00768***		0.00858***	0.00554***		0.00647***
		(4.135)		(4.587)	(2.941)		(3.408)
Turnover		-0.0573		-0.0675	0.978***		0.968***
		(-1.122)		(-1.317)	(19.22)		(18.94)
Volatility		-0.0419***		-0.0450***	0.0392***		0.0360***
		(-3.676)		(-3.934)	(3.348)		(3.074)
Market Risk		0.00200***		0.00197***	0.000176		0.000143
		(5.905)		(5.810)	(0.513)		(0.415)
Constant		0.00203	0.00236	0.000674	0.0172***	0.00231	0.0158***
		(0.588)	(0.851)	(0.195)	(5.192)	(0.718)	(4.776)
N		59,968	59,968	59,968	59,968	59,968	59,968
R^2		0.0243	0.0216	0.0247	0.0815	0.0344	0.0818
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$
$\mathbf{Tone_{Neg;Full\ Report}}$	-		**0.0612**	**0.0482**		**-0.251*** **	**-0.183*** **
			(1.196)	**(0.912)**		**(-4.166)**	**(-3.235)**
Ln MVE		-0.00167***		-0.00167***	-0.00327***		-0.00325***
		(-6.324)		(-6.352)	(-11.48)		(-11.39)
PtB		-7.47e-06		-4.01e-06	4.43e-05		3.12e-05
		(-0.0750)		(-0.0403)	(0.428)		(0.301)
ROA		-0.0112**		-0.0108**	-0.00398		-0.00575
		(-2.142)		(-2.038)	(-0.728)		(-1.039)
Turnover		0.125**		0.118**	-0.349***		-0.325***
		(2.560)		(2.435)	(-6.939)		(-6.441)
Volatility		-0.319***		-0.320***	0.317***		0.323***
		(-5.232)		(-5.235)	(5.438)		(5.518)
Market Risk		0.00665***		0.00662***	-0.00859***		-0.00848***
		(6.729)		(6.738)	(-8.698)		(-8.635)
Constant		0.0372***	0.00252	0.0365***	0.0646***	-0.00289	0.0671***
		(5.048)	(0.494)	(4.931)	(8.008)	(-0.498)	(8.198)
N		33,047	33,047	33,047	33,047	33,047	33,047
R^2		0.0177	0.0137	0.0177	0.0614	0.0219	0.0618
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel C: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**0.931*****	**0.288*****		**0.998*****	**0.310*****
			(14.56)	**(6.404)**		**(15.51)**	**(6.782)**
Ln MVE		-0.00949***		-0.00955***	-0.00871***		-0.00878***
		(-36.06)		(-36.35)	(-32.55)		(-32.84)
PtB		0.000607***		0.000626***	0.000632***		0.000653***
		(6.211)		(6.393)	(6.420)		(6.612)
ROA		-0.0516***		-0.0479***	-0.0508***		-0.0468***
		(-9.983)		(-9.243)	(-9.794)		(-9.004)
Turnover		0.450***		0.409***	0.460***		0.416***
		(6.664)		(6.054)	(6.813)		(6.160)
Volatility		1.802***		1.789***	1.902***		1.889***
		(42.39)		(42.09)	(43.93)		(43.62)
Market Risk		0.00532***		0.00522***	0.00761***		0.00751***
		(6.094)		(5.986)	(8.709)		(8.596)
Constant		0.226***	0.0469***	0.221***	0.209***	0.0472***	0.203***
		(28.28)	(4.936)	(27.34)	(25.82)	(4.995)	(24.94)
N		57,739	57,739	57,739	57,739	57,739	57,739
R^2		0.395	0.197	0.396	0.389	0.186	0.390
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel D: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**0.964******	**0.300******
			(14.83)	**(6.598)**
Ln MVE		-0.00900***		-0.00906***
		(-33.22)		(-33.52)
PtB		0.000471***		0.000491***
		(4.990)		(5.194)
ROA		-0.0481***		-0.0442***
		(-9.473)		(-8.708)
Turnover		0.372***		0.330***
		(5.576)		(4.942)
Volatility		1.967***		1.954***
		(45.42)		(45.09)
Market Risk		0.00413***		0.00403***
		(4.746)		(4.634)
Constant		0.222***	0.0539***	0.217***
		(26.66)	(5.435)	(25.82)
N		57,739	57,739	57,739
R^2		0.398	0.185	0.398
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for negative textual sentiment in Form 10-K for a firm's market liquidity and market risk. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 3 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

2.4 Robustness tests

2.4.1 Sample selection

To test whether the empirical results presented of the study are robust, a series of corresponding tests were conducted. As a first robustness test, the study applied additional and more restrictive data filters to the data sample and overall sample selection process. From the selected sample, firm-year observations of the financial industry were excluded (SIC codes 6021-6799). Substantial accounting differences between financial and non-financial entities as well as differences in their regulatory framework might affect the empirical results (Ditter and Scherr 2015, 60). In addition, the study removed firms with a low market capitalization (10 percent quantile of market capitalization) from the data set. By eliminating low-priced firms, the study addresses concerns that the regression results are influenced by bid-ask bounces around the filing date of an annual report while investigating the effect of the disclosure tone on a firm's market liquidity and market risk (Loughran and McDonald 2011a, 40; Loughran and McDonald 2015, 3). Applying these additional and more restrictive data filters to the data sample, the analysis was rerun.

The regression results confirm the findings. The regression coefficient on negative disclosure tone for a firm's abnormal trading volume is significantly positive (higher market liquidity). Using this restricted data sample to conduct the analysis, negative Form 10-K disclosure tone is again not significantly related to a firm's abnormal bid-ask spread (average based abnormal bid-ask spread) during the event period. As before, the amount of negative information in financial statements is significantly positively related a firm's abnormal stock return volatility around the SEC filing (higher market risk). Table 6 presents the estimated OLS coefficients on negative Form 10-K disclosure tone.

Table 6. Regression coefficients on negative Form 10-K disclosure tone for liquidity and risk variables excluding firm-year observations of the financial industry and and low market capitalization.

Panel A: Abnormal trading volume (ATV)							
Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**0.107*****	**0.123*****		**0.260*****	**0.117*****
			(4.347)	**(4.898)**		**(9.399)**	**(4.640)**
Ln MVE		-0.000167		-0.000194	-0.000950***		-0.000976***
		(-1.167)		(-1.360)	(-6.597)		(-6.800)
PtB		0.000235***		0.000241***	0.000282***		0.000287***
		(4.808)		(4.920)	(5.657)		(5.761)
ROA		0.00783***		0.00923***	0.00662***		0.00795***
		(3.204)		(3.742)	(2.664)		(3.169)
Turnover		-0.0227		-0.0377	0.961***		0.947***
		(-0.380)		(-0.627)	(16.24)		(15.91)
Volatility		-0.0323*		-0.0389**	0.0875***		0.0812***
		(-1.676)		(-2.013)	(4.413)		(4.086)
Market Risk		0.00236***		0.00234***	-0.000102		-0.000127
		(4.988)		(4.933)	(-0.214)		(-0.266)
Constant		0.00503	0.00122	0.00283	0.0211***	0.000301	0.0190***
		(1.199)	(0.403)	(0.678)	(5.173)	(0.0850)	(4.666)
N		41,258	41,258	41,258	41,258	41,258	41,258
R^2		0.0255	0.0239	0.0263	0.0741	0.0280	0.0748
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$
$\mathbf{Tone_{Neg;Full\ Report}}$	**-**		**0.00754**	**0.0134**		**-0.0867*****	**-0.0697****
			(0.285)	**(0.481)**		**(-2.804)**	**(-2.323)**
Ln MVE		-0.000784***		-0.000787***	-0.00159***		-0.00158***
		(-5.052)		(-5.066)	(-9.656)		(-9.609)
PtB		-1.81e-06		-1.29e-06	2.19e-05		1.92e-05
		(-0.0405)		(-0.0289)	(0.461)		(0.404)
ROA		-0.00160		-0.00150	0.000277		-0.000272
		(-0.625)		(-0.580)	(0.102)		(-0.0998)
Turnover		0.0851***		0.0836***	-0.139***		-0.131***
		(3.434)		(3.362)	(-5.467)		(-5.160)
Volatility		-0.179***		-0.180***	0.0944***		0.0986***
		(-5.642)		(-5.621)	(2.951)		(3.054)
Market Risk		0.00330***		0.00329***	-0.00360***		-0.00358***
		(5.823)		(5.829)	(-6.253)		(-6.238)
Constant		0.105***	0.0873***	0.104***	0.143***	0.117***	0.145***
		(22.79)	(25.69)	(22.49)	(26.57)	(27.23)	(26.46)
N		22,828	22,828	22,828	22,828	22,828	22,828
R^2		0.0238	0.0191	0.0238	0.0469	0.0208	0.0472
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel C: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**1.049*****	**0.320*****		**1.093*****	**0.339*****
			(15.29)	**(6.141)**		**(15.90)**	**(6.470)**
Ln MVE		-0.00879***		-0.00886***	-0.00845***		-0.00852***
		(-29.71)		(-30.02)	(-28.29)		(-28.66)
PtB		0.000561***		0.000571***	0.000602***		0.000612***
		(5.703)		(5.816)	(6.075)		(6.196)
ROA		-0.0431***		-0.0392***	-0.0425***		-0.0384***
		(-7.804)		(-7.043)	(-7.722)		(-6.927)
Turnover		0.479***		0.439***	0.520***		0.477***
		(7.097)		(6.502)	(7.719)		(7.087)
Volatility		1.787***		1.768***	1.854***		1.834***
		(35.44)		(35.01)	(36.14)		(35.74)
Market Risk		0.00725***		0.00721***	0.00773***		0.00769***
		(7.319)		(7.287)	(7.789)		(7.757)
Constant		0.208***	0.0390***	0.202***	0.201***	0.0392***	0.195***
		(25.01)	(4.660)	(24.05)	(23.77)	(4.711)	(22.87)
N		39,855	39,855	39,855	39,855	39,855	39,855
R^2		0.360	0.177	0.361	0.358	0.170	0.359
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel D: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**1.062*** **	**0.343*** **
			(15.70)	**(6.693)**
Ln MVE		-0.00849***		-0.00856***
		(-28.53)		(-28.92)
PtB		0.000426***		0.000436***
		(4.542)		(4.677)
ROA		-0.0387***		-0.0346***
		(-7.298)		(-6.493)
Turnover		0.446***		0.403***
		(6.749)		(6.108)
Volatility		1.881***		1.861***
		(37.28)		(36.84)
Market Risk		0.00445***		0.00441***
		(4.542)		(4.507)
Constant		0.208***	0.0444***	0.202***
		(24.26)	(5.234)	(23.37)
N		39,855	39,855	39,855
R^2		0.360	0.168	0.361
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for negative Form 10-K textual sentiment for a firm's market liquidity and market risk. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words).

2.4.2 Disclosure length

The results could be influenced by the way the overall text length of financial statements was measured in this study (Loughran and McDonald business dictionary excluding stop words). Loughran and McDonald (2011a) and Loughran and McDonald (2014) originally counted the total number of words (more than two characters) appearing in an annual report (including stop words) to determine the entire text length of a specific Form 10-K filing (Loughran and McDonald 2011b, 2; Loughran and McDonald 2014, 1669). To examine whether the results are sensitive to the document length of an annual report, in this study, a modified word count was obtained for the overall text length of a Form 10-K filing (Loughran and McDonald dictionary including stop words) for each submission in the data sample, and the analysis was run again.

The regression results confirm the findings. Based on the Loughran and McDonald business dictionary (including stop words), there is a positive association between the negative Form 10-K disclosure tone and a firm's abnormal trading volume around the filing date (higher market liquidity). The regression results again do not show a significant relationship between the negative textual sentiment in annual reports and a firm's abnormal bid-ask spread (average based abnormal bid-ask spread) in the days a Form 10-K filing is made with the SEC. The regression coefficient on negative tone for a firm's abnormal stock return volatility is again significantly positive (higher market risk). Table 7 presents the estimated OLS coefficients on negative disclosure tone for a firm's market liquidity and market risk (Loughran and McDonald dictionary including stop words).

Table 7. Regression coefficients on negative Form 10-K disclosure tone for liquidity and risk variables including all firm-year observations using the Loughran and McDonald business dictionary including stop words.

Panel A: Abnormal trading volume (ATV)

Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**0.0988*** **	**0.141*** **		**0.363*** **	**0.144*** **
			(2.805)	**(3.947)**		**(9.200)**	**(4.011)**
Ln MVE		3.79e-06		-1.07e-05	-0.000719***		-0.000734***
		(0.0377)		(-0.107)	(-7.077)		(-7.235)
PtB		0.000188***		0.000194***	0.000238***		0.000244***
		(4.754)		(4.916)	(5.839)		(5.998)
ROA		0.00768***		0.00854***	0.00554***		0.00642***
		(4.135)		(4.563)	(2.941)		(3.381)
Turnover		-0.0573		-0.0671	0.978***		0.968***
		(-1.122)		(-1.309)	(19.22)		(18.95)
Volatility		-0.0419***		-0.0447***	0.0392***		0.0363***
		(-3.676)		(-3.913)	(3.348)		(3.093)
Market Risk		0.00200***		0.00197***	0.000176		0.000140
		(5.905)		(5.803)	(0.513)		(0.408)
Constant		0.00203	0.00243	0.000768	0.0172***	0.00251	0.0159***
		(0.588)	(0.872)	(0.222)	(5.192)	(0.776)	(4.802)
N		59,968	59,968	59,968	59,968	59,968	59,968
R^2		0.0243	0.0216	0.0247	0.0815	0.0343	0.0818
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$
$\mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	**-**		**0.134**	**0.105**		**-0.505*** **	**-0.331*** **
			(1.361)	**(1.039)**		**(-4.375)**	**(-3.073)**
Ln MVE		-0.00167***		-0.00168***	-0.00327***		-0.00324***
		(-6.324)		(-6.360)	(-11.48)		(-11.37)
PtB		-7.47e-06		-3.12e-06	4.43e-05		3.06e-05
		(-0.0750)		(-0.0314)	(0.428)		(0.295)
ROA		-0.0112**		-0.0107**	-0.00398		-0.00562
		(-2.142)		(-2.029)	(-0.728)		(-1.016)
Turnover		0.125**		0.117**	-0.349***		-0.327***
		(2.560)		(2.424)	(-6.939)		(-6.486)
Volatility		-0.319***		-0.320***	0.317***		0.322***
		(-5.232)		(-5.237)	(5.438)		(5.501)
Market Risk		0.00665***		0.00662***	-0.00859***		-0.00848***
		(6.729)		(6.734)	(-8.698)		(-8.635)
Constant		0.0372***	0.00239	0.0365***	0.0646***	-0.00276	0.0668***
		(5.048)	(0.469)	(4.932)	(8.008)	(-0.477)	(8.176)
N		33,047	33,047	33,047	33,047	33,047	33,047
R^2		0.0177	0.0137	0.0177	0.0614	0.0220	0.0618
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel C: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**1.750*** **	**0.548*** **		**1.876*** **	**0.585*** **
			(13.94)	**(6.242)**		**(14.86)**	**(6.574)**
Ln MVE		-0.00949***		-0.00955***	-0.00871***		-0.00878***
		(-36.06)		(-36.35)	(-32.55)		(-32.84)
PtB		0.000607***		0.000628***	0.000632***		0.000655***
		(6.211)		(6.407)	(6.420)		(6.626)
ROA		-0.0516***		-0.0480***	-0.0508***		-0.0470***
		(-9.983)		(-9.264)	(-9.794)		(-9.032)
Turnover		0.450***		0.410***	0.460***		0.417***
		(6.664)		(6.068)	(6.813)		(6.178)
Volatility		1.802***		1.790***	1.902***		1.890***
		(42.39)		(42.10)	(43.93)		(43.63)
Market Risk		0.00532***		0.00521***	0.00761***		0.00749***
		(6.094)		(5.971)	(8.709)		(8.581)
Constant		0.226***	0.0480***	0.221***	0.209***	0.0485***	0.203***
		(28.28)	(5.031)	(27.37)	(25.82)	(5.098)	(24.97)
N		57,739	57,739	57,739	57,739	57,739	57,739
R^2		0.395	0.196	0.396	0.389	0.186	0.390
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel D: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**1.807*** (14.18)**	**0.565*** (6.392)**
Ln MVE		-0.00900*** (-33.22)		-0.00906*** (-33.53)
PtB		0.000471*** (4.990)		0.000493*** (5.208)
ROA		-0.0481*** (-9.473)		-0.0443*** (-8.736)
Turnover		0.372*** (5.576)		0.331*** (4.961)
Volatility		1.967*** (45.42)		1.954*** (45.10)
Market Risk		0.00413*** (4.746)		0.00402*** (4.619)
Constant		0.222*** (26.66)	0.0551*** (5.531)	0.217*** (25.85)
N		57,739	57,739	57,739
R^2		0.398	0.184	0.398
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for negative Form 10-K textual sentiment for a firm's market liquidity and market risk. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (including stop words).

2.4.3 Table content

The results of this study could be further influenced by the fact that HTML tables and their corresponding content are excluded from the analysis. EDGAR filers might use HTML table tags (<Table>) to structure and disclose certain textual information in annual reports on Form 10-K filed with the SEC. Thus, the exclusion of tables in financial statements and their corresponding content could affect the results. To address concerns that tables contain relevant narrative information when examining the effect of disclosure tone on firm characteristics, the study includes tables and their corresponding content and the analysis was rerun.

The regression results confirm the findings. When tables and their corresponding content are considered, the regression coefficients on the negative disclosure tone for the annual report on Form 10-K are unchanged (abnormal trading volume, abnormal bid-ask spread, abnormal stock return volatility). Table 8 presents the estimated OLS coefficients on negative Form 10-K disclosure tone. Based on these results, it is concluded that all findings seem unaffected by the robustness tests (sample selection, disclosure length, table content).

Table 8. Regression coefficients on negative Form 10-K disclosure tone for liquidity and risk variables including table content.

Panel A: Abnormal trading volume (ATV)

Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**0.0468****	**0.0775*****		**0.196*****	**0.0781*****
			(2.551)	**(4.153)**		**(9.575)**	**(4.168)**
Ln MVE		4.99e-06		-2.07e-07	-0.000717***		-0.000722***
		(0.0497)		(-0.00207)	(-7.062)		(-7.122)
PtB		0.000191***		0.000197***	0.000240***		0.000246***
		(4.815)		(4.964)	(5.896)		(6.041)
ROA		0.00774***		0.00880***	0.00560***		0.00667***
		(4.166)		(4.684)	(2.970)		(3.498)
Turnover		-0.0551		-0.0652	0.979***		0.969***
		(-1.080)		(-1.274)	(19.28)		(19.01)
Volatility		-0.0411***		-0.0444***	0.0400***		0.0366***
		(-3.599)		(-3.882)	(3.416)		(3.117)
Market Risk		0.00196***		0.00192***	0.000142		9.92e-05
		(5.780)		(5.658)	(0.412)		(0.288)
Constant		0.00211	0.00263	0.000550	0.0173***	0.00244	0.0157***
		(0.611)	(0.944)	(0.159)	(5.219)	(0.754)	(4.736)
N		60,113	60,113	60,113	60,113	60,113	60,113
R^2		0.0243	0.0215	0.0246	0.0816	0.0344	0.0820
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$
$\mathbf{Tone_{Neg;Full\ Report}}$	**-**		**0.0758**	**0.0573**		**-0.196*** **	**-0.182*** **
			(1.413)	**(1.009)**		**(-3.120)**	**(-3.003)**
Ln MVE		-0.00167***		-0.00168***	-0.00328***		-0.00327***
		(-6.349)		(-6.357)	(-11.49)		(-11.46)
PtB		-7.29e-06		-3.01e-06	4.40e-05		3.04e-05
		(-0.0732)		(-0.0303)	(0.425)		(0.294)
ROA		-0.0113**		-0.0107**	-0.00404		-0.00610
		(-2.157)		(-2.005)	(-0.739)		(-1.094)
Turnover		0.125**		0.117**	-0.349***		-0.326***
		(2.570)		(2.426)	(-6.948)		(-6.472)
Volatility		-0.320***		-0.322***	0.317***		0.322***
		(-5.258)		(-5.261)	(5.425)		(5.496)
Market Risk		0.00668***		0.00664***	-0.00857***		-0.00844***
		(6.761)		(6.771)	(-8.693)		(-8.620)
Constant		0.0373***	0.00241	0.0365***	0.0646***	-0.00425	0.0672***
		(5.067)	(0.475)	(4.929)	(8.014)	(-0.739)	(8.226)
N		33,071	33,071	33,071	33,071	33,071	33,071
R^2		0.0178	0.0138	0.0178	0.0614	0.0216	0.0618
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel C: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**1.238***** **(18.74)**	**0.341***** **(7.223)**		**1.312***** **(19.74)**	**0.369***** **(7.709)**
Ln MVE		-0.00948*** (-36.06)		-0.00951*** (-36.25)	-0.00870*** (-32.52)		-0.00873*** (-32.73)
PtB		0.000611*** (6.265)		0.000633*** (6.455)	0.000637*** (6.478)		0.000660*** (6.679)
ROA		-0.0516*** (-9.990)		-0.0467*** (-8.986)	-0.0508*** (-9.795)		-0.0454*** (-8.720)
Turnover		0.450*** (6.668)		0.404*** (5.976)	0.460*** (6.821)		0.410*** (6.076)
Volatility		1.803*** (42.46)		1.787*** (42.06)	1.903*** (44.00)		1.886*** (43.59)
Market Risk		0.00531*** (6.093)		0.00516*** (5.929)	0.00760*** (8.716)		0.00744*** (8.543)
Constant		0.226*** (28.28)	0.0406*** (4.282)	0.219*** (27.09)	0.208*** (25.81)	0.0409*** (4.338)	0.201*** (24.67)
N		57,862	57,862	57,862	57,862	57,862	57,862
R^2		0.395	0.201	0.396	0.389	0.191	0.390
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel D: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{Tone_{Neg;Full\ Report}}$	+		**1.281*** **	**0.358*** **
			(19.09)	**(7.506)**
Ln MVE		-0.00898***		-0.00901***
		(-33.19)		(-33.40)
PtB		0.000474***		0.000497***
		(5.032)		(5.250)
ROA		-0.0482***		-0.0429***
		(-9.494)		(-8.450)
Turnover		0.372***		0.323***
		(5.570)		(4.842)
Volatility		1.967***		1.950***
		(45.49)		(45.05)
Market Risk		0.00411***		0.00396***
		(4.734)		(4.560)
Constant		0.222***	0.0474***	0.214***
		(26.64)	(4.789)	(25.55)
N		57,862	57,862	57,862
R^2		0.398	0.189	0.399
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for negative Form 10-K textual sentiment for a firm's market liquidity and market risk. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words).

2.5 Sensitivity tests

2.5.1 Type of disclosure tone

Apart from the negative disclosure tone, capital market participants might react to other types of textual sentiment (e.g. *"uncertainty"*, *"litigious"*, *"constraining"*, *"positive"*). Prior literature demonstrated that other types of linguistic tone are connected to firm characteristics after an annual report has been filed with the SEC (Loughran and McDonald 2011a, 55-56). To analyse the effect of other types of disclosure tone in financial statements on a firm's market liquidity and market risk, the study employed the following regression model using an OLS regression:

$$\begin{aligned} Y_{i;t} = \beta_0 + \beta_1 * Tone_{Type;Full\ Report;i;t} + \beta_2 * Ln\ MVE_{i;t} + \beta_3 * PtB_{i;t} \\ +\beta_4 * ROA_{i;t} + \beta_5 * Turnover_{i;t} + \beta_6 * Volatility_{i;t} + \beta_7 * Market\ Risk_{i;t} \\ + \sum Year\ Fixed\ Effects + \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \quad (8)$$

Where: the dependent variable ($Y_{i;t}$) is a firm's market liquidity (abnormal trading volume, abnormal bid-ask spread) and market risk (abnormal stock return volatility) around its Form 10-K filing. $Tone_{Type;Full\ Report;i;t}$ is the disclosure tone (textual sentiment) in the entire annual report on Form 10-K (submission including exhibits) as measured by the number of uncertain, litigious, constraining, or positive words in the SEC filing. The control variables are described in the appendix. Year Fixed Effects and Industry Fixed Effects represent the included fixed effects.

Consistent with previous studies in the field of textual analysis, other types of disclosure tone embedded in annual reports on Form 10-K are associated with a firm's market liquidity and market risk in the expected manner. More litigious corporate information is significantly positively associated with a firm's abnormal trading volume (higher market liquidity) and abnormal stock return volatility (higher market risk). In addition, more constraining textual information is also significantly positively related to a firm's abnormal trading volume (higher market liquidity) and abnormal stock return volatility (higher market risk). These results are in line with previous empirical research showing the same association (e.g. Loughran and McDonald 2011b; Kravet and Muslu 2013). Other types of textual information in Form 10-K filings cannot be connected to a firm's stock market characteristics in the expected manner (at least when examining average based dependent variables). Table 9 presents the estimated OLS coefficients on different types of disclosure tone in Form 10-K filings.

Table 9. Regression coefficients on different Form 10-K disclosure tones for liquidity and risk variables.

Panel A: Abnormal trading volume (ATV)

Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{Tone_{Unc;Full\ Report}}$	+		**0.0132**	**0.0237**		**0.136*** **	**0.0336**
			(0.448)	**(0.796)**		**(4.166)**	**(1.120)**
Ln MVE		3.79e-06		5.35e-06	-0.000719***		-0.000717***
		(0.0377)		(0.0533)	(-7.077)		(-7.059)
PtB		0.000188***		0.000189***	0.000238***		0.000239***
		(4.754)		(4.771)	(5.839)		(5.864)
ROA		0.00768***		0.00778***	0.00554***		0.00568***
		(4.135)		(4.164)	(2.941)		(2.999)
Turnover		-0.0573		-0.0584	0.978***		0.977***
		(-1.122)		(-1.143)	(19.22)		(19.18)
Volatility		-0.0419***		-0.0419***	0.0392***		0.0392***
		(-3.676)		(-3.677)	(3.348)		(3.344)
Market Risk		0.00200***		0.00199***	0.000176		0.000152
		(5.905)		(5.851)	(0.513)		(0.440)
Constant		0.00203	0.00334	0.00164	0.0172***	0.00449	0.0166***
		(0.588)	(1.174)	(0.469)	(5.192)	(1.338)	(4.960)
N		59,968	59,968	59,968	59,968	59,968	59,968
R^2		0.0243	0.0214	0.0243	0.0815	0.0324	0.0815
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$
$\mathbf{Tone_{Unc;Full\ Report}}$	**-**		**0.140****	**0.0921**		**-0.119***	**-0.121***
			(2.273)	**(1.497)**		**(-1.684)**	**(-1.852)**
Ln MVE		-0.00167***		-0.00165***	-0.00327***		-0.00329***
		(-6.324)		(-6.255)	(-11.48)		(-11.53)
PtB		-7.47e-06		-4.66e-06	4.43e-05		4.06e-05
		(-0.0750)		(-0.0468)	(0.428)		(0.393)
ROA		-0.0112**		-0.0109**	-0.00398		-0.00441
		(-2.142)		(-2.080)	(-0.728)		(-0.805)
Turnover		0.125**		0.121**	-0.349***		-0.345***
		(2.560)		(2.502)	(-6.939)		(-6.868)
Volatility		-0.319***		-0.317***	0.317***		0.316***
		(-5.232)		(-5.220)	(5.438)		(5.415)
Market Risk		0.00665***		0.00659***	-0.00859***		-0.00850***
		(6.729)		(6.705)	(-8.698)		(-8.669)
Constant		0.0372***	0.000743	0.0350***	0.0646***	-0.00495	0.0675***
		(5.048)	(0.143)	(4.655)	(8.008)	(-0.857)	(8.157)
N		33,047	33,047	33,047	33,047	33,047	33,047
R^2		0.0177	0.0138	0.0177	0.0614	0.0212	0.0615
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel C: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\mathbf{Tone_{Unc;Full\ Report}}$	+		**0.628*** **	**0.0316**		**0.666*** **	**0.0264**
			(6.257)	**(0.462)**		**(6.596)**	**(0.383)**
Ln MVE		-0.00949***		-0.00949***	-0.00871***		-0.00871***
		(-36.06)		(-36.06)	(-32.55)		(-32.53)
PtB		0.000607***		0.000607***	0.000632***		0.000633***
		(6.211)		(6.222)	(6.420)		(6.429)
ROA		-0.0516***		-0.0515***	-0.0508***		-0.0507***
		(-9.983)		(-9.922)	(-9.794)		(-9.740)
Turnover		0.450***		0.449***	0.460***		0.459***
		(6.664)		(6.641)	(6.813)		(6.794)
Volatility		1.802***		1.802***	1.902***		1.902***
		(42.39)		(42.39)	(43.93)		(43.92)
Market Risk		0.00532***		0.00530***	0.00761***		0.00759***
		(6.094)		(6.062)	(8.709)		(8.681)
Constant		0.226***	0.0584***	0.226***	0.209***	0.0597***	0.208***
		(28.28)	(6.215)	(28.01)	(25.82)	(6.332)	(25.54)
N		57,739	57,739	57,739	57,739	57,739	57,739
R^2		0.395	0.191	0.395	0.389	0.180	0.389
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel D: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{Tone_{Unc;Full\ Report}}$	+		**0.618***** **(6.060)**	**0.0128** **(0.185)**
Ln MVE		-0.00900*** (-33.22)		-0.00900*** (-33.21)
PtB		0.000471*** (4.990)		0.000471*** (4.993)
ROA		-0.0481*** (-9.473)		-0.0480*** (-9.432)
Turnover		0.372*** (5.576)		0.372*** (5.567)
Volatility		1.967*** (45.42)		1.967*** (45.42)
Market Risk		0.00413*** (4.746)		0.00412*** (4.733)
Constant		0.222*** (26.66)	0.0663*** (6.737)	0.222*** (26.41)
N		57,739	57,739	57,739
R^2		0.398	0.179	0.398
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Panel E: Abnormal trading volume (ATV)							
Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{Tone_{Lit;Full\ Report}}$	+		**0.0272*** **	**0.0251*** **		**0.0383*** **	**0.0258*** **
			(3.790)	**(3.489)**		**(4.847)**	**(3.528)**
Ln MVE		3.79e-06		-2.90e-05	-0.000719***		-0.000753***
		(0.0377)		(-0.289)	(-7.077)		(-7.382)
PtB		0.000188***		0.000190***	0.000238***		0.000239***
		(4.754)		(4.804)	(5.839)		(5.890)
ROA		0.00768***		0.00779***	0.00554***		0.00566***
		(4.135)		(4.195)	(2.941)		(3.003)
Turnover		-0.0573		-0.0588	0.978***		0.977***
		(-1.122)		(-1.152)	(19.22)		(19.20)
Volatility		-0.0419***		-0.0428***	0.0392***		0.0383***
		(-3.676)		(-3.753)	(3.348)		(3.272)
Market Risk		0.00200***		0.00203***	0.000176		0.000203
		(5.905)		(5.984)	(0.513)		(0.590)
Constant		0.00203	0.00242	0.00166	0.0172***	0.00502	0.0168***
		(0.588)	(0.880)	(0.485)	(5.192)	(1.536)	(5.128)
N		59,968	59,968	59,968	59,968	59,968	59,968
R^2		0.0243	0.0216	0.0245	0.0815	0.0323	0.0817
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel F: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$
$\mathbf{Tone_{Lit;Full\ Report}}$	-		**0.0167**	**0.0263**		**-0.107*****	**0.00960**
			(0.751)	**(1.185)**		**(-4.576)**	**(0.436)**
Ln MVE		-0.00167***		-0.00170***	-0.00327***		-0.00329***
		(-6.324)		(-6.427)	(-11.48)		(-11.48)
PtB		-7.47e-06		-6.78e-06	4.43e-05		4.46e-05
		(-0.0750)		(-0.0681)	(0.428)		(0.430)
ROA		-0.0112**		-0.0112**	-0.00398		-0.00396
		(-2.142)		(-2.130)	(-0.728)		(-0.724)
Turnover		0.125**		0.123**	-0.349***		-0.350***
		(2.560)		(2.517)	(-6.939)		(-6.942)
Volatility		-0.319***		-0.320***	0.317***		0.317***
		(-5.232)		(-5.257)	(5.438)		(5.428)
Market Risk		0.00665***		0.00666***	-0.00859***		-0.00858***
		(6.729)		(6.738)	(-8.698)		(-8.693)
Constant		0.0372***	0.00333	0.0374***	0.0646***	-0.00558	0.0647***
		(5.048)	(0.659)	(5.093)	(8.008)	(-0.984)	(8.023)
N		33,047	33,047	33,047	33,047	33,047	33,047
R^2		0.0177	0.0137	0.0177	0.0614	0.0217	0.0614
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel G: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\mathbf{Tone_{Lit;Full\ Report}}$	+		**-0.0711*** **	**0.0845*** **		**-0.0445* **	**0.0999*** **
			(-2.818)	**(4.377)**		**(-1.733)**	**(5.054)**
Ln MVE		-0.00949***		-0.00960***	-0.00871***		-0.00884***
		(-36.06)		(-36.33)	(-32.55)		(-32.90)
PtB		0.000607***		0.000612***	0.000632***		0.000638***
		(6.211)		(6.273)	(6.420)		(6.493)
ROA		-0.0516***		-0.0513***	-0.0508***		-0.0504***
		(-9.983)		(-9.917)	(-9.794)		(-9.716)
Turnover		0.450***		0.445***	0.460***		0.454***
		(6.664)		(6.598)	(6.813)		(6.734)
Volatility		1.802***		1.799***	1.902***		1.898***
		(42.39)		(42.35)	(43.93)		(43.89)
Market Risk		0.00532***		0.00541***	0.00761***		0.00772***
		(6.094)		(6.200)	(8.709)		(8.838)
Constant		0.226***	0.0711***	0.225***	0.209***	0.0718***	0.207***
		(28.28)	(7.547)	(28.16)	(25.82)	(7.604)	(25.70)
N		57,739	57,739	57,739	57,739	57,739	57,739
R^2		0.395	0.190	0.396	0.389	0.179	0.390
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel H: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{Tone_{Lit;Full\ Report}}$	+		**-0.0533****	**0.102*****
			(-2.069)	**(5.168)**
Ln MVE		-0.00900***		-0.00913***
		(-33.22)		(-33.56)
PtB		0.000471***		0.000477***
		(4.990)		(5.067)
ROA		-0.0481***		-0.0476***
		(-9.473)		(-9.392)
Turnover		0.372***		0.366***
		(5.576)		(5.494)
Volatility		1.967***		1.963***
		(45.42)		(45.38)
Market Risk		0.00413***		0.00424***
		(4.746)		(4.876)
Constant		0.222***	0.0780***	0.220***
		(26.66)	(7.896)	(26.54)
N		57,739	57,739	57,739
R^2		0.398	0.178	0.398
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Panel I: Abnormal trading volume (ATV)							
Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{Tone_{Con;Full\ Report}}$	+		**0.129*** **	**0.163*** **		**0.261*** **	**0.146*** **
			(3.572)	**(4.474)**		**(6.340)**	**(3.945)**
Ln MVE		3.79e-06		2.32e-05	-0.000719***		-0.000701***
		(0.0377)		(0.231)	(-7.077)		(-6.892)
PtB		0.000188***		0.000196***	0.000238***		0.000245***
		(4.754)		(4.956)	(5.839)		(6.017)
ROA		0.00768***		0.00807***	0.00554***		0.00589***
		(4.135)		(4.348)	(2.941)		(3.132)
Turnover		-0.0573		-0.0632	0.978***		0.973***
		(-1.122)		(-1.238)	(19.22)		(19.11)
Volatility		-0.0419***		-0.0417***	0.0392***		0.0394***
		(-3.676)		(-3.656)	(3.348)		(3.368)
Market Risk		0.00200***		0.00195***	0.000176		0.000132
		(5.905)		(5.766)	(0.513)		(0.385)
Constant		0.00203	0.00222	-3.49e-05	0.0172***	0.00393	0.0153***
		(0.588)	(0.806)	(-0.0101)	(5.192)	(1.204)	(4.636)
N		59,968	59,968	59,968	59,968	59,968	59,968
R^2		0.0243	0.0216	0.0247	0.0815	0.0328	0.0817
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel J: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$
$\mathbf{Tone_{Con;Full\ Report}}$	-		**-0.0447**	**-0.132**		**-0.278****	**-0.326*****
			(-0.380)	**(-1.090)**		**(-2.106)**	**(-2.614)**
Ln MVE		-0.00167***		-0.00168***	-0.00327***		-0.00331***
		(-6.324)		(-6.344)	(-11.48)		(-11.58)
PtB		-7.47e-06		-1.37e-05	4.43e-05		2.88e-05
		(-0.0750)		(-0.138)	(0.428)		(0.279)
ROA		-0.0112**		-0.0115**	-0.00398		-0.00466
		(-2.142)		(-2.187)	(-0.728)		(-0.848)
Turnover		0.125**		0.129***	-0.349***		-0.338***
		(2.560)		(2.650)	(-6.939)		(-6.700)
Volatility		-0.319***		-0.319***	0.317***		0.317***
		(-5.232)		(-5.236)	(5.438)		(5.438)
Market Risk		0.00665***		0.00672***	-0.00859***		-0.00842***
		(6.729)		(6.818)	(-8.698)		(-8.576)
Constant		0.0372***	0.00398	0.0385***	0.0646***	-0.00510	0.0679***
		(5.048)	(0.770)	(5.091)	(8.008)	(-0.890)	(8.234)
N		33,047	33,047	33,047	33,047	33,047	33,047
R^2		0.0177	0.0136	0.0177	0.0614	0.0213	0.0616
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel K: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\mathbf{Tone_{Con;Full\ Report}}$	+		**1.278***** **(9.584)**	**0.251***** **(2.688)**		**1.330***** **(9.856)**	**0.281***** **(2.946)**
Ln MVE		-0.00949*** (-36.06)		-0.00946*** (-35.86)	-0.00871*** (-32.55)		-0.00868*** (-32.34)
PtB		0.000607*** (6.211)		0.000617*** (6.316)	0.000632*** (6.420)		0.000644*** (6.534)
ROA		-0.0516*** (-9.983)		-0.0510*** (-9.852)	-0.0508*** (-9.794)		-0.0501*** (-9.653)
Turnover		0.450*** (6.664)		0.441*** (6.517)	0.460*** (6.813)		0.449*** (6.648)
Volatility		1.802*** (42.39)		1.802*** (42.39)	1.902*** (43.93)		1.902*** (43.93)
Market Risk		0.00532*** (6.094)		0.00525*** (6.013)	0.00761*** (8.709)		0.00754*** (8.620)
Constant		0.226*** (28.28)	0.0548*** (5.758)	0.223*** (27.53)	0.209*** (25.82)	0.0561*** (5.911)	0.205*** (25.09)
N		57,739	57,739	57,739	57,739	57,739	57,739
R^2		0.395	0.192	0.395	0.389	0.181	0.389
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel K: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{Tone_{Con;Full\ Report}}$	+		**1.285*****	**0.240****
			(9.509)	**(2.547)**
Ln MVE		-0.00900***		-0.00897***
		(-33.22)		(-33.05)
PtB		0.000471***		0.000481***
		(4.990)		(5.094)
ROA		-0.0481***		-0.0475***
		(-9.473)		(-9.350)
Turnover		0.372***		0.363***
		(5.576)		(5.433)
Volatility		1.967***		1.967***
		(45.42)		(45.42)
Market Risk		0.00413***		0.00407***
		(4.746)		(4.670)
Constant		0.222***	0.0624***	0.219***
		(26.66)	(6.290)	(26.02)
N		57,739	57,739	57,739
R^2		0.398	0.180	0.398
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Panel L: Abnormal trading volume (ATV)							
Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{Tone_{Pos;Full\ Report}}$	+		**-0.0503**	**-0.0533**		**0.0195**	**-0.0402**
			(-1.221)	**(-1.276)**		**(0.418)**	**(-0.950)**
Ln MVE		3.79e-06		1.40e-05	-0.000719***		-0.000711***
		(0.0377)		(0.139)	(-7.077)		(-6.982)
PtB		0.000188***		0.000188***	0.000238***		0.000238***
		(4.754)		(4.759)	(5.839)		(5.841)
ROA		0.00768***		0.00745***	0.00554***		0.00537***
		(4.135)		(3.995)	(2.941)		(2.841)
Turnover		-0.0573		-0.0573	0.978***		0.978***
		(-1.122)		(-1.123)	(19.22)		(19.22)
Volatility		-0.0419***		-0.0417***	0.0392***		0.0393***
		(-3.676)		(-3.661)	(3.348)		(3.361)
Market Risk		0.00200***		0.00201***	0.000176		0.000185
		(5.905)		(5.941)	(0.513)		(0.537)
Constant		0.00203	0.00414	0.00247	0.0172***	0.00637*	0.0175***
		(0.588)	(1.464)	(0.710)	(5.192)	(1.890)	(5.265)
N		59,968	59,968	59,968	59,968	59,968	59,968
R^2		0.0243	0.0214	0.0243	0.0815	0.0318	0.0815
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel M: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$
$\mathbf{Tone_{Pos;Full\ Report}}$	-		**-0.173**	**-0.192**		**-0.609*****	**-0.392*****
			(-1.423)	**(-1.528)**		**(-4.159)**	**(-2.863)**
Ln MVE		-0.00167***		-0.00163***	-0.00327***		-0.00321***
		(-6.324)		(-6.190)	(-11.48)		(-11.27)
PtB		-7.47e-06		-6.02e-06	4.43e-05		4.73e-05
		(-0.0750)		(-0.0603)	(0.428)		(0.456)
ROA		-0.0112**		-0.0120**	-0.00398		-0.00562
		(-2.142)		(-2.258)	(-0.728)		(-1.012)
Turnover		0.125**		0.124**	-0.349***		-0.351***
		(2.560)		(2.541)	(-6.939)		(-6.963)
Volatility		-0.319***		-0.318***	0.317***		0.319***
		(-5.232)		(-5.220)	(5.438)		(5.460)
Market Risk		0.00665***		0.00667***	-0.00859***		-0.00854***
		(6.729)		(6.752)	(-8.698)		(-8.665)
Constant		0.0372***	0.00554	0.0386***	0.0646***	-0.000577	0.0676***
		(5.048)	(1.053)	(5.179)	(8.008)	(-0.0966)	(8.244)
N		33,047	33,047	33,047	33,047	33,047	33,047
R^2		0.0177	0.0137	0.0177	0.0614	0.0219	0.0617
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel N: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\mathbf{Tone_{Pos;Full\ Report}}$	-		**-0.142**	**-0.120**		**-0.0373**	**-0.0970**
			(-0.851)	**(-1.092)**		**(-0.223)**	**(-0.874)**
Ln MVE		-0.00949***		-0.00947***	-0.00871***		-0.00869***
		(-36.06)		(-35.79)	(-32.55)		(-32.34)
PtB		0.000607***		0.000608***	0.000632***		0.000633***
		(6.211)		(6.221)	(6.420)		(6.428)
ROA		-0.0516***		-0.0522***	-0.0508***		-0.0513***
		(-9.983)		(-10.03)	(-9.794)		(-9.827)
Turnover		0.450***		0.450***	0.460***		0.460***
		(6.664)		(6.661)	(6.813)		(6.810)
Volatility		1.802***		1.802***	1.902***		1.903***
		(42.39)		(42.39)	(43.93)		(43.92)
Market Risk		0.00532***		0.00534***	0.00761***		0.00763***
		(6.094)		(6.121)	(8.709)		(8.730)
Constant		0.226***	0.0698***	0.227***	0.209***	0.0704***	0.209***
		(28.28)	(7.338)	(28.27)	(25.82)	(7.399)	(25.80)
N		57,739	57,739	57,739	57,739	57,739	57,739
R^2		0.395	0.190	0.395	0.389	0.179	0.389
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel O: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{Tone_{Pos;Full\ Report}}$	-		**-0.142**	**-0.125**
			(-0.848)	**(-1.132)**
Ln MVE		-0.00900***		-0.00897***
		(-33.22)		(-33.01)
PtB		0.000471***		0.000472***
		(4.990)		(5.002)
ROA		-0.0481***		-0.0486***
		(-9.473)		(-9.529)
Turnover		0.372***		0.372***
		(5.576)		(5.573)
Volatility		1.967***		1.967***
		(45.42)		(45.41)
Market Risk		0.00413***		0.00416***
		(4.746)		(4.772)
Constant		0.222***	0.0775***	0.223***
		(26.66)	(7.786)	(26.64)
N		57,739	57,739	57,739
R^2		0.398	0.178	0.398
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for different Form 10-K disclosure tones for a firm's market liquidity and market risk. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 3 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

2.5.2 Measurement of disclosure tone

The accounting and finance literature uses various methods to quantify the disclosure tone in text documents. For instance, numerous studies used the ratio of certain words (e.g. *"negative"*, *"positive"*) in relation to the entire text length to measure the disclosure tone of narrative information (tone level) (e.g., Loughran and McDonald 2011a; Gandhi et al. 2017). Other empirical works examined the effect of tone changes on future firm characteristics (tone differences) (e.g. Chouliaras 2015; Henry and Leone 2016). While a market response to tone levels indicates that investors react to differences in disclosure tone across companies, a market impact to a change in tone suggests that capital market participants rather react to changes in the linguistic tone in a firm's annual reports over time (Loughran and McDonald 2016, 1219).

To examine the effect of disclosure tone in financial statements on a firm's market liquidity and market risk based on different tone measurements, the study estimates the following OLS regression:

$$\begin{aligned} Y_{i;t} = \beta_0 + \beta_1 * \Delta\, Tone_{Neg;Full\ Report;i;t} + \beta_2 * Ln\ MVE_{i;t} + \beta_3 * PtB_{i;t} \\ +\beta_4 * ROA_{i;t} + \beta_5 * Turnover_{i;t} + \beta_6 * Volatility_{i;t} + \beta_7 * Market\ Risk_{i;t} \\ + \sum Year\ Fixed\ Effects + \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \tag{9}$$

Where: the dependent variable ($Y_{i;t}$) is a firm's market liquidity (abnormal trading volume, abnormal bid-ask spread) and market risk (abnormal stock return volatility) around its Form 10-K filing. $\Delta\, Tone_{Neg;Full\ Report;i;t}$ is the change in disclosure tone in a firm's annual report filing as measured by the difference in disclosure tone (negative textual sentiment) between the current and previous Form 10-K submission. The control variables are described in the appendix. Year Fixed Effects and Industry Fixed Effects represent the included fixed effects.

Interestingly, while the level of negative disclosure tone in annual reports across companies is relevant for capital market participants, the change in tone in a firm's Form 10-K filing is not associated with proxies for its market liquidity and market risk. The regression coefficient for negative disclosure tone for a firm's abnormal trading volume (ATV) around the filing date is insignificantly negative. In addition, the regression coefficient on negative disclosure tone for a firm's abnormal bid-ask spread is also insignificant. Furthermore, the regression coefficient on tone change for a firm's abnormal stock return volatility (ARV) around the filing date is insignificantly positive. Overall, the results suggest that a firm's stock market characteristics are not affected by changes in the disclosure tone over time. Table 10 presents the estimated OLS coefficients on negative Form 10-K disclosure tone change.

Table 10. Regression coefficients on negative Form 10-K disclosure tone changes for liquidity and risk variables.

Panel A: Abnormal trading volume (ATV)

Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\Delta\ \mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	**?**		**-0.0218**	**-0.0167**		**0.0109**	**-0.00831**
			(-1.095)	**(-0.841)**		**(0.545)**	**(-0.415)**
Ln MVE		-0.000114		-0.000114	-0.000817***		-0.000817***
		(-1.007)		(-1.001)	(-7.079)		(-7.076)
PtB		0.000132***		0.000131***	0.000185***		0.000185***
		(2.902)		(2.897)	(4.071)		(4.069)
ROA		0.00846***		0.00844***	0.00654***		0.00653***
		(3.810)		(3.802)	(2.909)		(2.905)
Turnover		-0.0167		-0.0165	0.976***		0.976***
		(-0.299)		(-0.296)	(17.44)		(17.44)
Volatility		-0.0386***		-0.0384***	0.0521***		0.0522***
		(-2.734)		(-2.718)	(3.580)		(3.589)
Market Risk		0.00216***		0.00216***	0.000286		0.000285
		(5.723)		(5.718)	(0.745)		(0.742)
Constant		0.00200	0.00141	0.00198	0.0176***	0.00508	0.0176***
		(0.560)	(0.522)	(0.554)	(5.191)	(1.565)	(5.188)
N		48,449	48,449	48,449	48,449	48,449	48,449
R^2		0.0244	0.0221	0.0244	0.0823	0.0322	0.0823
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$\text{ASpread}_{[-1,3];\text{Ave}}$	$\text{ASpread}_{[-1,3];\text{Ave}}$	$\text{ASpread}_{[-1,3];\text{Ave}}$	$\text{ASpread}_{[-1,3];\text{Med}}$	$\text{ASpread}_{[-1,3];\text{Med}}$	$\text{ASpread}_{[-1,3];\text{Med}}$
$\Delta\ \mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	**?**		**-0.0185**	**0.0106**		**0.0234**	**-0.0359**
			(-0.307)	**(0.175)**		**(0.390)**	**(-0.602)**
Ln MVE		-0.00155***		-0.00155***	-0.00299***		-0.00298***
		(-5.959)		(-5.959)	(-10.62)		(-10.61)
PtB		6.15e-05		6.16e-05	9.22e-05		9.18e-05
		(0.657)		(0.658)	(0.942)		(0.938)
ROA		-0.0152***		-0.0152***	-0.00724		-0.00726
		(-2.988)		(-2.986)	(-1.342)		(-1.344)
Turnover		0.0912**		0.0910**	-0.342***		-0.342***
		(2.019)		(2.017)	(-7.254)		(-7.252)
Volatility		-0.335***		-0.335***	0.281***		0.281***
		(-5.518)		(-5.513)	(4.792)		(4.797)
Market Risk		0.00678***		0.00678***	-0.00774***		-0.00775***
		(6.809)		(6.807)	(-7.712)		(-7.716)
Constant		0.0263***	-0.00612	0.0263***	0.0462***	-0.0196***	0.0462***
		(3.707)	(-1.290)	(3.707)	(5.898)	(-3.852)	(5.898)
N		29,239	29,239	29,239	29,239	29,239	29,239
R^2		0.0206	0.0158	0.0206	0.0608	0.0228	0.0608
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel C: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
Δ Tone$_{Neg;Full\ Report}$	**?**		**0.236*** **	**0.0349**		**0.229*** **	**0.0214**
			(4.287)	**(0.673)**		**(4.118)**	**(0.410)**
Ln MVE		-0.00927***		-0.00928***	-0.00858***		-0.00858***
		(-33.02)		(-33.02)	(-30.18)		(-30.18)
PtB		0.000528***		0.000528***	0.000547***		0.000547***
		(5.090)		(5.095)	(5.221)		(5.224)
ROA		-0.0522***		-0.0521***	-0.0517***		-0.0517***
		(-8.787)		(-8.783)	(-8.673)		(-8.671)
Turnover		0.401***		0.401***	0.402***		0.402***
		(5.494)		(5.489)	(5.526)		(5.524)
Volatility		1.847***		1.847***	1.931***		1.931***
		(37.12)		(37.11)	(38.17)		(38.15)
Market Risk		0.00443***		0.00444***	0.00664***		0.00664***
		(4.580)		(4.586)	(6.881)		(6.884)
Constant		0.217***	0.0583***	0.217***	0.200***	0.0583***	0.200***
		(25.77)	(6.156)	(25.77)	(23.70)	(6.219)	(23.70)
N		46,882	46,882	46,882	46,882	46,882	46,882
R^2		0.392	0.185	0.392	0.385	0.174	0.385
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel D: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\boldsymbol{\Delta\ \mathbf{Tone}_{\mathbf{Neg;Full\ Report}}}$	**?**		**0.244*** **	**0.0341**
			(4.407)	**(0.660)**
Ln MVE		-0.00867***		-0.00867***
		(-30.13)		(-30.13)
PtB		0.000434***		0.000435***
		(4.296)		(4.301)
ROA		-0.0500***		-0.0499***
		(-8.598)		(-8.593)
Turnover		0.330***		0.330***
		(4.548)		(4.544)
Volatility		2.001***		2.001***
		(39.52)		(39.49)
Market Risk		0.00292***		0.00292***
		(3.012)		(3.017)
Constant		0.203***	0.0580***	0.203***
		(23.56)	(6.023)	(23.56)
N		46,882	46,882	46,882
R^2		0.392	0.173	0.392
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for negative Form 10-K textual sentiment (change) for a firm's market liquidity and market risk. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 3 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

Despite this result, changes in disclosure tone over time can have a significant effect on stock characteristics (Miwa 2020, 2). Based on the idea that tone levels alone do not reflect the importance of new corporate disclosures in annual reports, market participants might interpret changes in textual sentiment as incremental information beyond financial ratios as well as the tone levels themselves (Miwa 2020, 5). Thus, a firm's market liquidity and market risk could be influenced by time-series changes in textual sentiment given the specific tone level of the Form 10-K filing and disclosed financial figures.

To examine the effect of disclosure tone changes in annual reports on a firm's market liquidity and market risk in the context of tone levels, the following regression model was estimated using OLS regression:

$$\begin{aligned} Y_{i;t} = \beta_0 &+ \beta_1 * \Delta\, Tone_{Neg;Full\ Report;i;t} + \beta_2 * Tone_{Neg;Full\ Report;i;t} \\ &+ \beta_3 * Ln\ MVE_{I;t} + \beta_4 * PtB_{i;t} + \beta_5 * ROA_{i;t} + \beta_6 * Turnover_{i;t} \\ &+ \beta_7 * Volatility_{i;t} + \beta_8 * Market\ Risk_{i;t} + \sum Year\ Fixed\ Effects \\ &+ \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \tag{10}$$

Where: the dependent variable ($Y_{i;t}$) is a firm's market liquidity (abnormal trading volume, abnormal bid-ask spread) and market risk (abnormal stock return volatility) around its Form 10-K filing. $\Delta\, Tone_{Neg;Full\ Report;i;t}$ is the change in disclosure tone in a firm's annual report filing as measured by the difference in disclosure tone between the current and previous Form 10-K submission (time series). $Tone_{Neg;Full\ Report;i;t}$ is the disclosure tone (textual sentiment) in the entire annual report on Form 10-K (submission including exhibits) as measured by the ratio of negative words in the SEC filing (cross section). The control variables are described in the appendix. Year Fixed Effects and Industry Fixed Effects represent the included fixed effects.

The regression coefficient on $\Delta\, Tone_{Neg;Full\ Report;i;t}$ (tone change) for a firm's abnormal trading volume is significantly negative. While a higher level of negative Form 10-K textual sentiment ($Tone_{Neg;Full\ Report}$) is positively associated with abnormal trading volumes a (positive) time-series change in negative disclosure tone (more pessimistic corporate disclosures) is negatively associated with a firm's abnormal trading volume. In general, this result is inconsistent with the notion that market participants regard a higher level of negative corporate information in Form 10-K filings as a sign of transparency and as associated with higher trading volumes.

While more negative information in comparison to other firms might be considered a sign of transparency, a higher level of negative information over time seems to be interpreted as a negative sign reducing investors' trading activities in a certain stock. Other proxies for a firm's market liquidity (abnormal bid-ask spread) are not affected by time-series changes in Form 10-K textual sentiment when controlling for tone levels.

The regression coefficient on $\Delta\ Tone_{Neg;Full\ Report}$ for a firm's abnormal stock return volatility is significantly negative. Again, this result is not consistent with naïve expectations. While a higher level of negative information in the annual report ($Tone_{Neg;Full\ Report}$) is associated with a higher market risk (cross section), a higher level of pessimistic disclosures over time reduces a firm's abnormal stock return volatility (time series). Table 11 presents the estimated OLS coefficients on negative Form 10-K disclosure tone changes.

Table 11. Regression coefficients on negative Form 10-K disclosure tone measurements for liquidity and risk variables.

Panel A: Abnormal trading volume (ATV)

Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\boldsymbol{\Delta\ Tone_{Neg;Full\ Report}}$	**?**		**-0.0620*****	**-0.0649*****		**-0.103*****	**-0.0555****
			(-2.717)	**(-2.833)**		**(-4.377)**	**(-2.395)**
$\boldsymbol{Tone_{Neg;Full\ Report}}$	+		**0.0836*****	**0.103*****		**0.237*****	**0.100*****
			(3.559)	**(4.297)**		**(9.009)**	**(4.168)**
Ln MVE		-0.000114		-0.000132	-0.000817***		-0.000835***
		(-1.007)		(-1.166)	(-7.079)		(-7.248)
PtB		0.000132***		0.000140***	0.000185***		0.000194***
		(2.902)		(3.094)	(4.071)		(4.260)
ROA		0.00846***		0.00964***	0.00654***		0.00770***
		(3.810)		(4.302)	(2.909)		(3.399)
Turnover		-0.0167		-0.0299	0.976***		0.963***
		(-0.299)		(-0.532)	(17.44)		(17.13)
Volatility		-0.0386***		-0.0423***	0.0521***		0.0484***
		(-2.734)		(-2.980)	(3.580)		(3.318)
Market Risk		0.00216***		0.00211***	0.000286		0.000243
		(5.723)		(5.607)	(0.745)		(0.632)
Constant		0.00200	-0.000512	1.08e-05	0.0176***	-0.000369	0.0157***
		(0.560)	(-0.189)	(0.00300)	(5.191)	(-0.116)	(4.574)
N		48,449	48,449	48,449	48,449	48,449	48,449
R^2		0.0244	0.0225	0.0250	0.0823	0.0354	0.0828
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$	$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$
$\Delta\ \mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	**?**		**-0.0437**	**-0.00577**		**0.124***	**0.0311**
			(-0.674)	**(-0.0896)**		**(1.864)**	**(0.478)**
$\mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	**-**		**0.0540**	**0.0362**		**-0.215*****	**-0.148****
			(1.070)	**(0.694)**		**(-3.548)**	**(-2.570)**
Ln MVE		-0.00155***		-0.00155***	-0.00299***		-0.00297***
		(-5.959)		(-5.980)	(-10.62)		(-10.55)
PtB		6.15e-05		6.45e-05	9.22e-05		8.01e-05
		(0.657)		(0.689)	(0.942)		(0.817)
ROA		-0.0152***		-0.0148***	-0.00724		-0.00872
		(-2.988)		(-2.889)	(-1.342)		(-1.592)
Turnover		0.0912**		0.0863*	-0.342***		-0.323***
		(2.019)		(1.896)	(-7.254)		(-6.746)
Volatility		-0.335***		-0.336***	0.281***		0.285***
		(-5.518)		(-5.513)	(4.792)		(4.848)
Market Risk		0.00678***		0.00676***	-0.00774***		-0.00766***
		(6.809)		(6.823)	(-7.712)		(-7.668)
Constant		0.0263***	-0.00714	0.0257***	0.0462***	-0.0156***	0.0485***
		(3.707)	(-1.499)	(3.598)	(5.898)	(-2.897)	(6.089)
N		29,239	29,239	29,239	29,239	29,239	29,239
R^2		0.0206	0.0158	0.0206	0.0608	0.0234	0.0611
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel C: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\Delta\ \mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	**?**		**-0.229*****	**-0.0960***		**-0.272*****	**-0.124****
			(-3.457)	**(-1.672)**		**(-4.060)**	**(-2.145)**
$\mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	+		**0.967*****	**0.278*****		**1.041*****	**0.309*****
			(12.26)	**(5.066)**		**(13.13)**	**(5.565)**
Ln MVE		-0.00927***		-0.00933***	-0.00858***		-0.00864***
		(-33.02)		(-33.28)	(-30.18)		(-30.46)
PtB		0.000528***		0.000549***	0.000547***		0.000570***
		(5.090)		(5.275)	(5.221)		(5.422)
ROA		-0.0522***		-0.0486***	-0.0517***		-0.0478***
		(-8.787)		(-8.125)	(-8.673)		(-7.955)
Turnover		0.401***		0.363***	0.402***		0.360***
		(5.494)		(4.941)	(5.526)		(4.917)
Volatility		1.847***		1.836***	1.931***		1.918***
		(37.12)		(36.90)	(38.17)		(37.93)
Market Risk		0.00443***		0.00434***	0.00664***		0.00654***
		(4.580)		(4.487)	(6.881)		(6.775)
Constant		0.217***	0.0360***	0.212***	0.200***	0.0343***	0.194***
		(25.77)	(3.671)	(24.72)	(23.70)	(3.550)	(22.63)
N		46,882	46,882	46,882	46,882	46,882	46,882
R^2		0.392	0.191	0.392	0.385	0.181	0.386
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel D: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\Delta\ \mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	**?**		**-0.243***** **(-3.632)**	**-0.110*** **(-1.911)**
$\mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	+		**1.012***** **(12.64)**	**0.305***** **(5.507)**
Ln MVE		-0.00867*** (-30.13)		-0.00872*** (-30.42)
PtB		0.000434*** (4.296)		0.000458*** (4.513)
ROA		-0.0500*** (-8.598)		-0.0461*** (-7.878)
Turnover		0.330*** (4.548)		0.289*** (3.955)
Volatility		2.001*** (39.52)		1.988*** (39.26)
Market Risk		0.00292*** (3.012)		0.00282*** (2.907)
Constant		0.203*** (23.56)	0.0346*** (3.497)	0.197*** (22.55)
N		46,882	46,882	46,882
R^2		0.392	0.180	0.393
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for negative Form 10-K textual sentiment (change) and negative disclosure tone in the annual report for a firm's market liquidity and market risk. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 3 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

To further examine the effect of tone changes on a firm's market liquidity and market risk the present study examined the interaction effect between tone changes and tone levels. Investors might react differently to disclosure tone changes depending on the tone level of a firm's annual report. A more negative disclosure tone in comparison to the previous Form 10-K filing (tone change) might have a bigger effect when the annual report contains less negative information compared to other firms (tone level), suggesting a negative development or outlook for the firm. Thus, the study estimates the following regression model using OLS regression:

$$\begin{aligned} Y_{i;t} = \beta_0 &+ \beta_1 * \Delta\, Tone_{Neg;Full\ Report;i;t} + \beta_2 * Tone_{Neg;Full\ Report;i;t} \\ &+ \beta_3 * \Delta\, Tone_{Neg;Full\ Report;i;t} * Tone_{Neg;Full\ Report;i;t} + \beta_4 * Ln\ MVE_{I;t} \\ &+ \beta_5 * PtB_{i;t} + \beta_6 * ROA_{i;t} + \beta_7 * Turnover_{i;t} + \beta_8 * Volatility_{i;t} \\ &+ \beta_9 * Market\ Risk_{i;t} + \sum Year\ Fixed\ Effects \\ &+ \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \tag{11}$$

Where: the dependent variable ($Y_{i;t}$) is a firm's market liquidity (abnormal trading volume, abnormal bid-ask spread) and market risk (abnormal stock return volatility) around its Form 10-K filing. $\Delta\, Tone_{Neg;Full\ Report;i;t} * Tone_{Neg;Full\ Report;i;t}$ is the interaction effect between the disclosure tone change and disclosure tone level of an annual report on Form 10-K. The control variables are described in the appendix. Year Fixed Effects and Industry Fixed Effects represent the included fixed effects.

Interestingly, the interaction effect between the change in disclosure tone and level of disclosure tone is not significantly associated with proxies for a firm's market liquidity (abnormal trading volume, abnormal bid-ask spread). Capital market participants seem to pay less attention to more pessimistic disclosures in a firm's annual report over time, regardless of how negative the current report is compared to other companies. On the contrary, a positive change in negative annual report sentiment (more negative Form 10-K information) is positively associated with a firm's market risk (abnormal stock return volatility) given a certain amount of negativity in the current report compared to other firms. Table 12 presents the estimated OLS coefficients on the interaction effects between tone changes and tone levels.

Table 12. Regression coefficients on negative disclosure tone interaction effects for liquidity and risk variables.

Panel A: Abnormal trading volume (ATV)

Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\Delta\ Tone_{Neg;Full\ Report}$	?		0.0273	0.0186		-0.0437	0.00328
			(0.415)	(0.283)		(-0.641)	(0.0493)
$Tone_{Neg;Full\ Report}$	+		0.0867***	0.105***		0.239***	0.102***
			(3.663)	(4.386)		(9.021)	(4.216)
$\boldsymbol{\Delta\ Tone * Tone}$	**?**		**-2.875**	**-2.692**		**-1.905**	**-1.894**
			(-1.344)	**(-1.259)**		**(-0.861)**	**(-0.878)**
Ln MVE		-0.000114		-0.000134	-0.000817***		-0.000837***
		(-1.007)		(-1.185)	(-7.079)		(-7.260)
PtB		0.000132***		0.000140***	0.000185***		0.000194***
		(2.902)		(3.093)	(4.071)		(4.260)
ROA		0.00846***		0.00967***	0.00654***		0.00773***
		(3.810)		(4.317)	(2.909)		(3.410)
Turnover		-0.0167		-0.0298	0.976***		0.963***
		(-0.299)		(-0.530)	(17.44)		(17.13)
Volatility		-0.0386***		-0.0422***	0.0521***		0.0485***
		(-2.734)		(-2.974)	(3.580)		(3.322)
Market Risk		0.00216***		0.00211***	0.000286		0.000239
		(5.723)		(5.596)	(0.745)		(0.623)
Constant		0.00200	-0.000526	4.08e-05	0.0176***	-0.000378	0.0157***
		(0.560)	(-0.194)	(0.0113)	(5.191)	(-0.119)	(4.581)
N		48,449	48,449	48,449	48,449	48,449	48,449
R^2		0.0244	0.0226	0.0250	0.0823	0.0354	0.0828
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$
$\Delta\ Tone_{Neg;Full\ Report}$	?		-0.461*	-0.448*		-0.253	-0.191
			(-1.798)	(-1.746)		(-0.978)	(-0.748)
$Tone_{Neg;Full\ Report}$	-		0.0457	0.0273		-0.223***	-0.152***
			(0.904)	(0.524)		(-3.669)	(-2.646)
$\boldsymbol{\Delta\ Tone * Tone}$	**?**		**12.16***	**12.88***		**10.98**	**6.469**
			(1.714)	**(1.813)**		**(1.506)**	**(0.904)**
Ln MVE		-0.00155***		-0.00155***	-0.00299***		-0.00297***
		(-5.959)		(-5.969)	(-10.62)		(-10.55)
PtB		6.15e-05		6.46e-05	9.22e-05		8.02e-05
		(0.657)		(0.691)	(0.942)		(0.817)
ROA		-0.0152***		-0.0149***	-0.00724		-0.00874
		(-2.988)		(-2.896)	(-1.342)		(-1.595)
Turnover		0.0912**		0.0861*	-0.342***		-0.323***
		(2.019)		(1.892)	(-7.254)		(-6.748)
Volatility		-0.335***		-0.336***	0.281***		0.285***
		(-5.518)		(-5.518)	(4.792)		(4.845)
Market Risk		0.00678***		0.00677***	-0.00774***		-0.00766***
		(6.809)		(6.833)	(-7.712)		(-7.664)
Constant		0.0263***	-0.00684	0.0259***	0.0462***	-0.0153***	0.0486***
		(3.707)	(-1.430)	(3.623)	(5.898)	(-2.839)	(6.096)
N		29,239	29,239	29,239	29,239	29,239	29,239
R^2		0.0206	0.0159	0.0207	0.0608	0.0235	0.0610
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel C: Abnormal return volatility (ARV)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\Delta\ Tone_{Neg;Full\ Report}$	?		-0.787***	-0.336*		-0.903***	-0.463**
			(-3.980)	(-1.874)		(-4.442)	(-2.514)
$Tone_{Neg;Full\ Report}$	+		0.947***	0.269***		1.019***	0.297***
			(11.96)	(4.896)		(12.80)	(5.334)
$\boldsymbol{\Delta\ Tone * Tone}$	**?**		**17.95*** **	**7.715**		**20.32*** **	**10.91***
			(2.836)	**(1.358)**		**(3.141)**	**(1.879)**
Ln MVE		-0.00927***		-0.00932***	-0.00858***		-0.00863***
		(-33.02)		(-33.26)	(-30.18)		(-30.44)
PtB		0.000528***		0.000549***	0.000547***		0.000571***
		(5.090)		(5.274)	(5.221)		(5.419)
ROA		-0.0522***		-0.0488***	-0.0517***		-0.0480***
		(-8.787)		(-8.143)	(-8.673)		(-7.981)
Turnover		0.401***		0.363***	0.402***		0.359***
		(5.494)		(4.940)	(5.526)		(4.915)
Volatility		1.847***		1.835***	1.931***		1.918***
		(37.12)		(36.90)	(38.17)		(37.94)
Market Risk		0.00443***		0.00435***	0.00664***		0.00655***
		(4.580)		(4.499)	(6.881)		(6.793)
Constant		0.217***	0.0361***	0.212***	0.200***	0.0344***	0.194***
		(25.77)	(3.672)	(24.68)	(23.70)	(3.553)	(22.58)
N		46,882	46,882	46,882	46,882	46,882	46,882
R^2		0.392	0.191	0.392	0.385	0.182	0.386
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel D: Abnormal return volatility (ARV)				
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\Delta\ Tone_{Neg;Full\ Report}$	?		-0.940***	-0.484***
			(-4.598)	(-2.636)
$Tone_{Neg;Full\ Report}$	+		0.988***	0.292***
			(12.29)	(5.253)
$\boldsymbol{\Delta\ Tone * Tone}$	**?**		**22.43*** **	**12.06** **
			(3.464)	**(2.093)**
Ln MVE		-0.00867***		-0.00871***
		(-30.13)		(-30.38)
PtB		0.000434***		0.000458***
		(4.296)		(4.511)
ROA		-0.0500***		-0.0463***
		(-8.598)		(-7.907)
Turnover		0.330***		0.288***
		(4.548)		(3.952)
Volatility		2.001***		1.988***
		(39.52)		(39.27)
Market Risk		0.00292***		0.00283***
		(3.012)		(2.926)
Constant		0.203***	0.0348***	0.197***
		(23.56)	(3.501)	(22.50)
N		46,882	46,882	46,882
R^2		0.392	0.181	0.393
Fixed Effects		Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for the interaction effects between disclosure tone changes and disclosure tone levels. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 6 (increased due to the high correlation caused by the interaction effect). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

2.6 Additional tests

Apart from the entire annual report, many studies examined the information content and value-relevance of specific text sections within the overall filing (e.g. Loughran and McDonald 2011a; Campbell et al. 2014; Hope et al. 2016; Amel-Zadeh and Faasse 2016; Henselmann and Hering 2017). Annual reports on Form 10-K contain more than 20 text sections (*"Items"*), each assigned to certain aspects of the company (e.g. *"Item 1 - Business"*, *"Item 1A - Risk Factors"*, *"Item 3 - Legal Proceedings"*, *"Item 5 - Shareholder Matters"*) (SEC 2009). However, most of narrative information within a Form 10-K filing is disclosed in a few text sections (e.g. *"Item 7 - Management's Discussion and Analysis of Financial Condition and Results of Operations"*, *"Item 8 - Financial Statements and Supplementary Data"*) **(**Hering 2016, 48-49; Henselmann and Hering 2017, 99; Cazier and Pfeiffer, 2016, 8; Li 2008, 227; Amel-Zadeh and Faasse 2016, 4, 12). For instance, *"Item 7"* the *"MD&A"* section, provides important forward-looking information and contextual disclosures (Tavcar 1998, 10, 24-25; Humpherys et al. 2011, 585; Czyzewski and Wilkinson 2014, 99-100; Rogers and Grant 1997, 17-18; Amel-Zadeh and Faasse 2016, 3, 7; SEC 2003; SEC 2003b). *"Item 8"* or the *"Footnotes"*, contains important narrative disclosures regarding a firm's financial performance and future risks (Leder 2003, 17; Heidari and Felden 2015, 1; Amel-Zadeh and Faasse 2016, 25; Czerney et al. 2017, 8-9). Consistently, research shows that certain items in annual reports are more important to investors than others (Sutton et al. 2009, 5; Bedard et al. 2012, 25-26).

To examine whether the disclosure tone of specific text sections rather than the tone of the overall filing affects a firm's market liquidity and market risk in the days around the SEC submission, the following regression model was estimated using OLS:

$$\begin{aligned} Y_{i;t} = \beta_0 &+ \beta_1 * Tone_{Neg;Full\ Report;i;t} + \beta_2 * Tone_{Neg;Section\ X;i;t} \\ &+ \beta_3 * Ln\ MVE_{i;t} + \beta_4 * PtB_{i;t} + \beta_5 * ROA_{i;t} + \beta_6 * Turnover_{i;t} \\ &+ \beta_7 * Volatility_{i;t} + \beta_8 * Market\ Risk_{i;t} \\ &+ \sum Year\ Fixed\ Effects + \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \tag{12}$$

Where: the dependent variable ($Y_{i;t}$) is a firm's market liquidity (abnormal trading volume, abnormal bid-ask spread) and market risk (abnormal stock return volatility) around its Form 10-K filing. $Tone_{Neg;Full\ Report;i;t}$ is the negative disclosure tone of the entire annual report (submission including exhibits) and $Tone_{Neg;Section\ X;i;t}$ is the pessimistic tone of a specific text section within the Form 10-K filing. The control variables are described in the appendix. Year Fixed Effects and Industry Fixed Effects represent the included fixed effects. Note that the study also includes the textual sentiment of the entire filing (Form 10-K submission including exhibits) as a control variable to proxy for textual characteristics in the overall SEC submission as requested in the literature (Loughran and McDonald 2016, 1218).

In general, the regression results reveal that the disclosure tone in the Form 10-K filing in its entirety (submission including exhibits) is more important to capital market participants than the linguistic style of particular annual report subsections. However, the linguistic style in several prominent subsections of the annual report is associated with a firm's market liquidity and market risk in the days around the SEC submission.

Considering a firm's market liquidity, the negative disclosure tone in *"Item 1 - Business"* is significantly positively associated with ATV in the days around the SEC submission (higher market liquidity). Interestingly, the negative disclosure tone in *"Item 8 - Financial Statements and Supplementary Data"* is negatively associated with a firm's trading activity around the SEC submission (lower market liquidity) despite the entire report showing the expected positive sign (higher market liquidity). Capital market participants seem to interpret a higher level of negative disclosures in the entire report as a sign of transparency by company officials, resulting in a higher ATV (higher market liquidity), while pessimistic textual information in the context of financial figures (*"Footnotes"*) increases the expected risk associated with a specific firm (lower market liquidity). This result is not in line with previous empirical findings showing an significant association between negative (risk) disclosures in specific Form 10-K subsections and a firm's abnormal trading volume in the immediate days following the SEC submission (e.g. Hope et al. 2016).

The abnormal bid-ask spread (ASpread) of a firm as another proxy for a company's market liquidity seems unaffected by the negative disclosure tone of specific annual report subsections. Despite showing the expected negative sign, the regression coefficients are insignificant (e.g. *"Item 1"*, *"Item 7"*, *"Item 8"*). This result is also not in line with previous empirical findings showing a significant negative association between negative (risk) disclosures in specific Form 10-K subsections and a firm's abnormal bid-ask spread following the SEC filing (e.g. Campbell et al. 2014). In addition, the regression coefficient for the disclosure tone in the entire report (submission including exhibits) included in each regression is also insignificant (average based abnormal bid-ask spread).

Consistent with naïve expectations, a more negative disclosure tone in certain subsections of the annual report is associated with a higher abnormal stock return volatility (higher market risk). For instance, the regression coefficient on negative textual sentiment for *"Item 1"* is significantly positive. In this section, companies are required to provide information about a firm's business and its principal products including the status of those products (SEC Regulation S-K 2018, §229.101 Item 101a; SEC Regulation S-K 2018, §229.101 Item 101c (1i); SEC Regulation S-K 2018, §229.101 Item 101c (1ii)). In addition, there is a positive association between the negative textual sentiment in *"Item 5"* and a firm's abnormal stock return volatility (ARV). This section provides a firm's investors with important information about current and future dividend payments, and with detailed insights into share repurchases during the fiscal period (SEC Regulation S-K 2018, §229.201 Item 201c (1)). Furthermore, the regression coefficient on negative disclosure tone for *"Item 7"* is significantly positive. According to the SEC, this subsection of the annual report provides a narrative explanation of a company's financial statements (SEC 2003; SEC 2003b) and important forward-looking information (e.g. corporate earnings, cash flows). However, the regression results also reveal that the disclosure tone in specific subsections is not more strongly related to a company's market risk than the textual sentiment in the overall report. Table 13 presents the estimated OLS coefficients on the disclosure tone for specific Form 10-K subsections.

Table 13. Regression coefficients on negative disclosure tone in specific Form 10-K sections for liquidity and risk variables.

Panel A: Abnormal trading volume (ATV)

Variable	Expected sign	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$
$Tone_{Neg;Full\ Report}$	+	0.0387*	0.122***	0.0652***	0.0519***	0.0612***	0.0629***
		(1.915)	(2.998)	(3.354)	(2.670)	(3.091)	(3.235)
$\mathbf{Tone_{Neg;Item\ 1}}$	+	**0.0350****					
		(2.563)					
$\mathbf{Tone_{Neg;Item\ 1A}}$	+		**-0.0132**				
			(-1.096)				
$\mathbf{Tone_{Neg;Item\ 3}}$	+			**-0.00322**			
				(-1.369)			
$\mathbf{Tone_{Neg;Item\ 5}}$	+				**0.0136**		
					(0.890)		
$\mathbf{Tone_{Neg;Item\ 7}}$	+					**-0.00304**	
						(-0.309)	
$\mathbf{Tone_{Neg;Item\ 8}}$	+						**-0.0230***
							(-1.879)
Ln MVE		4.31e-05	-0.000565***	2.95e-05	7.00e-05	8.18e-05	-3.55e-06
		(0.396)	(-2.772)	(0.274)	(0.659)	(0.752)	(-0.0328)
PtB		0.000167***	0.000125*	0.000162***	0.000162***	0.000148***	0.000171***
		(3.756)	(1.762)	(3.790)	(3.763)	(3.423)	(3.981)
ROA		0.00607***	0.00805**	0.00627***	0.00720***	0.00683***	0.00661***
		(2.786)	(2.261)	(2.934)	(3.321)	(3.145)	(3.104)
Turnover		-0.0816	0.174*	-0.0903	-0.0948	-0.0954	-0.0728
		(-1.344)	(1.884)	(-1.532)	(-1.556)	(-1.592)	(-1.220)
Volatility		-0.0382***	-0.0745***	-0.0369***	-0.0366***	-0.0313**	-0.0390***
		(-3.077)	(-2.627)	(-2.980)	(-2.976)	(-2.528)	(-3.143)
Market Risk		0.00179***	0.00374***	0.00205***	0.00177***	0.00181***	0.00188***
		(4.601)	(6.314)	(5.243)	(4.375)	(4.581)	(4.772)
Constant		-9.10e-05	0.00739	-0.000326	-0.00129	-0.00150	0.00200
		(-0.0284)	(1.054)	(-0.103)	(-0.399)	(-0.464)	(0.558)
N		43,109	20,354	43,828	40,526	42,351	43,568
R^2		0.0233	0.0248	0.0230	0.0223	0.0231	0.0236
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel B: Abnormal trading volume (ATV)							
Variable	Expected sign	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$Tone_{Neg;Full\ Report}$	+	0.0358*	0.125***	0.0627***	0.0510***	0.0648***	0.0640***
		(1.778)	(3.070)	(3.225)	(2.617)	(3.251)	(3.277)
$\mathbf{Tone_{Neg;Item\ 1}}$	+	**0.0404***** **(2.968)**					
$\mathbf{Tone_{Neg;Item\ 1A}}$	+		**-0.0142** **(-1.167)**				
$\mathbf{Tone_{Neg;Item\ 3}}$	+			**-0.00228** **(-0.951)**			
$\mathbf{Tone_{Neg;Item\ 5}}$	+				**0.0154** **(0.997)**		
$\mathbf{Tone_{Neg;Item\ 7}}$	+					**-0.0113** **(-1.130)**	
$\mathbf{Tone_{Neg;Item\ 8}}$	+						**-0.0275**** **(-2.229)**
Ln MVE		-0.000692***	-0.00111***	-0.000714***	-0.000676***	-0.000668***	-0.000750***
		(-6.373)	(-5.545)	(-6.661)	(-6.320)	(-6.103)	(-6.937)
PtB		0.000218***	0.000168**	0.000208***	0.000207***	0.000197***	0.000219***
		(4.687)	(2.324)	(4.653)	(4.608)	(4.338)	(4.846)
ROA		0.00382*	0.00552	0.00409*	0.00453**	0.00434**	0.00444**
		(1.740)	(1.539)	(1.898)	(2.093)	(1.992)	(2.069)
Turnover		0.990***	1.144***	0.981***	0.987***	0.987***	0.995***
		(16.89)	(12.88)	(17.13)	(16.65)	(16.71)	(17.11)
Volatility		0.0350***	0.0431	0.0357***	0.0316**	0.0406***	0.0338***
		(2.727)	(1.521)	(2.805)	(2.499)	(3.172)	(2.646)
Market Risk		-8.41e-05	0.00198***	0.000219	-6.21e-05	-4.18e-05	8.29e-05
		(-0.213)	(3.352)	(0.555)	(-0.152)	(-0.104)	(0.208)
Constant		0.0152***	0.0176**	0.0151***	0.0145***	0.0143***	0.0177***
		(4.823)	(2.496)	(4.823)	(4.648)	(4.534)	(5.180)
N		43,109	20,354	43,828	40,526	42,351	43,568
R^2		0.0828	0.0959	0.0825	0.0817	0.0819	0.0837
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel C: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$
$Tone_{Neg;Full\ Report}$	-	-0.0108	0.0281	0.0306	-0.0186	-0.0236	0.0237
		(-0.148)	(0.417)	(0.439)	(-0.259)	(-0.343)	(0.343)
$\mathbf{Tone_{Neg;Item\ 1}}$	**-**	**-0.0244**					
		(-0.648)					
$\mathbf{Tone_{Neg;Item\ 1A}}$	**-**		**0.0166**				
			(0.711)				
$\mathbf{Tone_{Neg;Item\ 3}}$	**-**			**0.00282**			
				(0.406)			
$\mathbf{Tone_{Neg;Item\ 5}}$	**-**				**-0.000832**		
					(-0.0142)		
$\mathbf{Tone_{Neg;Item\ 7}}$	**-**					**-0.0528**	
						(-1.476)	
$\mathbf{Tone_{Neg;Item\ 8}}$	**-**						**-0.00336**
							(-0.117)
Ln MVE		-0.00169***	-0.00168***	-0.00181***	-0.00204***	-0.00162***	-0.00193***
		(-4.681)	(-4.842)	(-5.022)	(-5.446)	(-4.478)	(-5.515)
PtB		7.14e-05	2.45e-05	9.46e-05	0.000113	4.24e-05	9.36e-05
		(0.493)	(0.171)	(0.647)	(0.674)	(0.290)	(0.621)
ROA		-0.0120*	-0.0111	-0.00929	-0.00870	-0.00696	-0.00733
		(-1.656)	(-1.597)	(-1.265)	(-1.128)	(-0.949)	(-1.020)
Turnover		0.110*	0.0861	0.0998	0.133*	0.0721	0.0881
		(1.652)	(1.317)	(1.468)	(1.872)	(1.079)	(1.344)
Volatility		-0.305***	-0.300***	-0.296***	-0.338***	-0.252***	-0.320***
		(-3.598)	(-3.700)	(-3.475)	(-3.849)	(-2.987)	(-3.844)
Market Risk		0.00751***	0.00730***	0.00717***	0.00664***	0.00664***	0.00772***
		(5.804)	(5.647)	(5.448)	(4.852)	(5.015)	(5.791)
Constant		0.0375***	0.0561***	0.0356***	0.0425***	0.0359***	0.0397***
		(3.883)	(4.187)	(3.777)	(4.392)	(3.701)	(4.379)
N		18,968	19,551	18,981	16,574	18,406	19,301
R^2		0.0210	0.0160	0.0208	0.0199	0.0203	0.0223
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel D: Abnormal bid-ask spread (ASpread)							
Variable	Expected sign	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$
$Tone_{Neg;Full Report}$	-	-0.235***	-0.171**	-0.211***	-0.271***	-0.275***	-0.207***
		(-3.021)	(-2.366)	(-2.747)	(-3.428)	(-3.663)	(-2.791)
$\mathbf{Tone_{Neg;Item\ 1}}$	-	**-0.0492**					
		(-1.238)					
$\mathbf{Tone_{Neg;Item\ 1A}}$	-		**-0.0225**				
			(-0.888)				
$\mathbf{Tone_{Neg;Item\ 3}}$	-			**0.00147**			
				(0.200)			
$\mathbf{Tone_{Neg;Item\ 5}}$	-				**0.0412**		
					(0.733)		
$\mathbf{Tone_{Neg;Item\ 7}}$	-					**-0.0134**	
						(-0.380)	
$\mathbf{Tone_{Neg;Item\ 8}}$	-						**0.0184**
							(0.610)
Ln MVE		-0.00334***	-0.00348***	-0.00349***	-0.00370***	-0.00329***	-0.00357***
		(-8.596)	(-9.147)	(-8.820)	(-8.807)	(-8.281)	(-9.340)
PtB		0.000102	6.97e-05	9.75e-05	0.000153	6.38e-05	0.000129
		(0.662)	(0.459)	(0.632)	(0.864)	(0.412)	(0.813)
ROA		-0.00672	-0.00482	-0.00279	-0.00316	-0.00198	-0.000298
		(-0.879)	(-0.661)	(-0.365)	(-0.392)	(-0.258)	(-0.0394)
Turnover		-0.343***	-0.346***	-0.339***	-0.299***	-0.350***	-0.348***
		(-4.972)	(-5.031)	(-4.767)	(-4.017)	(-5.035)	(-5.104)
Volatility		0.318***	0.300***	0.329***	0.277***	0.343***	0.311***
		(3.992)	(3.857)	(4.024)	(3.328)	(4.222)	(3.956)
Market Risk		-0.00730***	-0.00709***	-0.00768***	-0.00811***	-0.00783***	-0.00695***
		(-5.602)	(-5.439)	(-5.693)	(-5.827)	(-5.827)	(-5.152)
Constant		0.0692***	0.111***	0.0679***	0.0742***	0.0678***	0.0698***
		(6.541)	(5.121)	(6.489)	(6.787)	(6.440)	(6.984)
N		18,968	19,551	18,981	16,574	18,406	19,301
R^2		0.0616	0.0584	0.0627	0.0619	0.0619	0.0607
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel E: Abnormal stock return volatility (ARV – marked adjusted model)							
Variable	Expected sign	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$
$Tone_{Neg;Full\ Report}$	+	0.236***	0.437***	0.323***	0.306***	0.280***	0.311***
		(4.433)	(5.262)	(6.199)	(5.809)	(5.225)	(6.022)
$\mathbf{Tone_{Neg;Item\ 1}}$	+	**0.162*** **					
		(4.675)					
$\mathbf{Tone_{Neg;Item\ 1A}}$	+		**-0.0572** **				
			(-2.252)				
$\mathbf{Tone_{Neg;Item\ 3}}$	+			**-0.00764**			
				(-1.231)			
$\mathbf{Tone_{Neg;Item\ 5}}$	+				**0.114** **		
					(2.548)		
$\mathbf{Tone_{Neg;Item\ 7}}$	+					**0.0560** **	
						(2.028)	
$\mathbf{Tone_{Neg;Item\ 8}}$	+						**-0.0215**
							(-0.754)
Ln MVE		-0.00989***	-0.00854***	-0.00998***	-0.00998***	-0.00994***	-0.0101***
		(-33.22)	(-19.12)	(-33.77)	(-32.67)	(-33.02)	(-34.16)
PtB		0.000808***	4.00e-05	0.000826***	0.000875***	0.000833***	0.000830***
		(6.759)	(0.265)	(7.065)	(7.074)	(7.103)	(6.898)
ROA		-0.0623***	-0.0271***	-0.0615***	-0.0642***	-0.0594***	-0.0638***
		(-10.08)	(-3.013)	(-9.925)	(-10.10)	(-9.471)	(-10.33)
Turnover		0.661***	0.0191	0.636***	0.713***	0.643***	0.633***
		(7.820)	(0.165)	(7.635)	(8.054)	(7.550)	(7.640)
Volatility		1.625***	2.279***	1.621***	1.553***	1.609***	1.598***
		(34.05)	(25.88)	(34.14)	(32.29)	(33.66)	(33.93)
Market Risk		0.00663***	0.00295**	0.00720***	0.00761***	0.00740***	0.00769***
		(6.367)	(1.983)	(6.938)	(7.030)	(6.982)	(7.370)
Constant		0.234***	0.209***	0.235***	0.235***	0.234***	0.236***
		(24.00)	(8.642)	(24.26)	(23.75)	(23.92)	(24.11)
N		41,409	19,818	42,114	38,910	40,653	41,895
R^2		0.393	0.405	0.391	0.392	0.392	0.392
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel F: Abnormal stock return volatility (ARV – market model)							
Variable	Expected sign	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$Tone_{Neg;Full\ Report}$	+	0.272***	0.457***	0.351***	0.344***	0.309***	0.339***
		(5.058)	(5.550)	(6.651)	(6.448)	(5.679)	(6.484)
$\mathbf{Tone_{Neg;Item\ 1}}$	+	**0.146***					
		(4.167)					
$\mathbf{Tone_{Neg;Item\ 1A}}$	+		**-0.0579**				
			(-2.288)				
$\mathbf{Tone_{Neg;Item\ 3}}$	+			**-0.00715**			
				(-1.141)			
$\mathbf{Tone_{Neg;Item\ 5}}$	+				**0.136***		
					(2.990)		
$\mathbf{Tone_{Neg;Item\ 7}}$	+					**0.0731***	
						(2.616)	
$\mathbf{Tone_{Neg;Item\ 8}}$	+						**-0.0247**
							(-0.853)
Ln MVE		-0.00907***	-0.00806***	-0.00916***	-0.00915***	-0.00912***	-0.00928***
		(-29.87)	(-18.12)	(-30.33)	(-29.32)	(-29.59)	(-30.72)
PtB		0.000852***	2.78e-05	0.000865***	0.000919***	0.000889***	0.000868***
		(7.062)	(0.182)	(7.326)	(7.354)	(7.479)	(7.134)
ROA		-0.0605***	-0.0248***	-0.0595***	-0.0622***	-0.0573***	-0.0622***
		(-9.754)	(-2.763)	(-9.569)	(-9.738)	(-9.097)	(-10.02)
Turnover		0.669***	0.0285	0.648***	0.727***	0.653***	0.649***
		(7.889)	(0.250)	(7.749)	(8.174)	(7.650)	(7.851)
Volatility		1.733***	2.337***	1.729***	1.663***	1.714***	1.702***
		(35.41)	(26.65)	(35.60)	(33.68)	(34.94)	(35.19)
Market Risk		0.00847***	0.00648***	0.00904***	0.00928***	0.00915***	0.00952***
		(8.094)	(4.354)	(8.653)	(8.530)	(8.581)	(9.069)
Constant		0.214***	0.194***	0.214***	0.215***	0.214***	0.216***
		(21.50)	(8.383)	(21.74)	(21.25)	(21.32)	(21.75)
N		41,409	19,818	42,114	38,910	40,653	41,895
R^2		0.389	0.402	0.386	0.387	0.387	0.387
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Panel G: Abnormal stock return volatility (ARV – market model)							
Variable	Expected sign	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$Tone_{Neg;Full\ Report}$	+	0.288***	0.441***	0.344***	0.349***	0.318***	0.343***
		(5.361)	(5.384)	(6.518)	(6.546)	(5.844)	(6.573)
$\mathbf{Tone_{Neg;Item\ 1}}$	+	**0.113*** **					
		(3.242)					
$\mathbf{Tone_{Neg;Item\ 1A}}$	+		**-0.0595** **				
			(-2.376)				
$\mathbf{Tone_{Neg;Item\ 3}}$	+			**-0.00285**			
				(-0.455)			
$\mathbf{Tone_{Neg;Item\ 5}}$	+				**0.136*** **		
					(2.993)		
$\mathbf{Tone_{Neg;Item\ 7}}$	+					**0.0739*** **	
						(2.672)	
$\mathbf{Tone_{Neg;Item\ 8}}$	+						**-0.0259**
							(-0.896)
Ln MVE		-0.00954***	-0.00771***	-0.00963***	-0.00973***	-0.00963***	-0.00976***
		(-31.00)	(-17.42)	(-31.45)	(-30.76)	(-30.86)	(-31.85)
PtB		0.000647***	7.20e-05	0.000670***	0.000705***	0.000686***	0.000678***
		(5.645)	(0.475)	(5.976)	(5.943)	(6.068)	(5.869)
ROA		-0.0560***	-0.0243***	-0.0545***	-0.0565***	-0.0522***	-0.0574***
		(-9.176)	(-2.733)	(-8.879)	(-8.980)	(-8.407)	(-9.392)
Turnover		0.576***	-0.0439	0.551***	0.631***	0.561***	0.548***
		(6.884)	(-0.390)	(6.722)	(7.235)	(6.674)	(6.739)
Volatility		1.792***	2.459***	1.785***	1.715***	1.767***	1.763***
		(36.45)	(28.49)	(36.69)	(34.65)	(35.90)	(36.32)
Market Risk		0.00529***	0.00150	0.00597***	0.00634***	0.00609***	0.00637***
		(5.114)	(1.017)	(5.753)	(5.871)	(5.759)	(6.114)
Constant		0.231***	0.192***	0.231***	0.234***	0.232***	0.233***
		(22.41)	(8.305)	(22.63)	(22.52)	(22.51)	(22.51)
N		41,409	19,818	42,114	38,910	40,653	41,895
R^2		0.397	0.407	0.394	0.395	0.395	0.395
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for negative textual sentiment for specific Form 10-K text sections. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). Note that model diagnostics for each corresponding baseline regression are discussed in Section 2.3 Empirical results.

2.7 Conclusion

This study examines the effect of the disclosure tone in annual reports on important firm characteristics. Based on a large sample of Form 10-K filings filed with the SEC, the analysis reveals that the linguistic tone of accounting narratives is associated with a firm's market liquidity and market risk. In particular, the study provides empirical evidence that the negative disclosure tone in annual reports increases a firm's market liquidity in the days a Form 10-K filing is made (higher abnormal trading volume). Furthermore, the study shows that pessimistic corporate disclosures in annual reports on Form 10-K increase a firm's market risk (higher abnormal stock return volatility). Consistent with previous research and based on the EMH, which argues that both quantitative and qualitative information should be incorporated on financial markets (Amel-Zadeh and Faasse 2016, 1), this study concludes that regulatory filings, and in particular, disclosure tone, contain valuable information for capital market participants. The study and its results further underscore the importance of narrative information for investors and regulatory authorities in the context of trading activities and disclosure requirements.

This study is subject to a number of limitations. First, the applied information extraction algorithm to utilize narrative information disclosed in regulatory filings might contain errors (e.g. parsing errors). This concern was mitigated by using an information extraction procedure tested for accuracy. Second, regulatory changes and corresponding changes in the information content of individual subsections over time might affect the regression results (e.g. *"Item 1A - Risk Factors"*). This concern is mitigated by the fact that several findings of this study are not based on individual subsections of the annual report, but on the Form 10-K filing in its entirety. Third, other regulatory filings or disclosures as well as particular corporate events not subject to Form 10-K filing requirements might affect the empirical results of this study. These concerns were mitigated by analysing the effect of disclosure tone on a firm's market liquidity and market risk over a short period around the SEC submission.

Future research avenues could analyse the effect of the disclosure tone in Form 10-K filings on other important firm fundamentals not addressed in this study (e.g. revenue growth, cost of capital). Furthermore, future research should explore the predictive ability of the disclosure tone in annual reports on corporate bankruptcy. Finally, the effect of the disclosure tone in other regulatory filings made with the SEC on a firm's market liquidity and market risk could be investigated (e.g. quarterly reports on Form 10-Q, current reports on Form 8-K). In conclusion, as Loughran and McDonald (2016) noted, *"with increasing computational power and an explosion of digital text available for research, there is much yet to be done"* (Loughran and McDonald 2016, 1226).

Chapter 3

Accounting narrative obfuscation in financial statements

Contents

3.0 Introduction 111

3.1 Related literature 114

3.2 Hypothesis development 119

3.3 Research design 123

3.3.1 Sample selection and description 123

3.3.2 Regression model, dependent variable, and independent variables 126

3.3.3 Summary statistics and correlations 129

3.4 Empirical results 132

3.5 Robustness tests 138

3.5.1 Sample selection 138

3.5.2 Disclosure length 142

3.5.3 Table content 146

3.6 Sensitivity tests 149

3.6.1 Type of disclosure tone 149

3.6.2 Measurement of disclosure tone 153

3.7 Additional tests 157

3.8 Conclusion 161

"While there are many different techniques for reading a 10-K, depending on an individual's investment style, many professional investors start reading these filings from the back to the front, reading the footnotes even before they read the financials. By skimming the footnotes for red flags, such as aggressive accounting policies […] many pros are able to make a quick decision on whether it pays to invest additional time on research" (Leder 2003, 17).

3.0 Introduction

Domestic companies in the United States are obligated to file annual reports on Form 10-K with the United States Securities and Exchange Commission (SEC). Besides quantitative information, annual reports contain a large amount of qualitative or textual information (narrative disclosures) (Li 2010a, 143; Bannier et al. 2017, 4; Lo et al. 2017, 2; Lehavy et al. 2011, 1089). In particular, annual reports on Form 10-K filed with the SEC consist of more than 20 sections (*"Items"*), each assigned to certain aspects of a company (e.g. *"Item 1 - Business"*; *"Item 1A - Risk Factors"*; *"Item 3 - Legal Proceedings"*) (SEC 2009). The majority of narrative disclosures within a Form 10-K filing are concentrated in certain parts of the annual report. On average, a few items within the Form 10-K filing (*"Item 1 - Business"*, *"Item 1A - Risk Factors"*, *"Item 7 - Management's Discussion and Analysis of Financial Condition and Results of Operations"*, *"Item 8 - Financial Statements and Supplementary Data"*) represent about 70% of the overall narrative information reported by firms in this filing form type (Hering 2016, 48-49; Henselmann and Hering 2017, 99; Cazier and Pfeiffer, 2016, 8; Li 2008, 227; Amel-Zadeh and Faasse 2016, 4, 12). According to the SEC, a Form 10-K filing provides *"a comprehensive overview of the company's business and financial condition and includes audited financial statements"* (SEC 2009). Annual reports are highly relevant for capital market participants, because company officials use Form 10-K filings to inform investors about their expectations of a firm's future (Chouliaras 2015, 1). Textual information can provide a useful and valuable context for understanding the financial data disclosed by a firm in an annual report. Accounting narratives might also reveal non-public corporate information to capital market participants (Li 2010a, 143-144; Bannier et al. 2017, 5).

Research shows that certain text sections in financial statements are more important to investors than others (Sutton et al. 2009, 5; Bedard et al. 2012, 25-26). Providing important forward-looking information and contextual disclosures, *"Item 7"*, the *"MD&A"* section, is the most read section in annual reports on Form 10-K (Tavcar 1998, 10, 24-25; Humpherys et al. 2011, 585; Czyzewski and Wilkinson 2014, 99-100; Rogers and Grant 1997, 17-18; Amel-Zadeh and Faasse 2016, 3, 7; SEC 2003a). However, this section is not directly subject to an auditing process (Czerney et al. 2017, 29; Humpherys et al. 2011, 585; Bedard et al. 2012, 19). *"Item 8"* the notes to the financial statements, or the *"Footnotes"*, contains important narrative disclosures regarding a firm's financial performance and future risks (Leder 2003, 17; Heidari and Felden 2015, 1; Amel-Zadeh and Faasse 2016, 25; Czerney et al. 2017, 8-9; Lee 2012, 1159). *"Notes, or footnote disclosures, are an integral part of the financial statements, providing additional information on balances and transactions that help users better assess a company's financial position and performance"* (Czerney et al. 2017, 8). More specifically, the footnotes state information about significant accounting policies, income taxes, pension plans, stock options (SEC 2007), acquisitions and divestitures, corporate debts, and commitments and contingencies. Despite being audited and therefore the most reliable section it is less used by capital market participants when analysing annual reports on Form 10-K (Sutton et al. 2009, 3; Bedard et al. 2012, 25-26; Czerney et al. 2017, 2). Footnotes are more technical than other sections, and thus are the most difficult to read (Clatworthy and Jones 2001, 314; Amel-Zadeh and Faasse 2016, 4, 11). As such, they are less used by capital market participants, especially non-professional investors (Sutton et al. 2009, 4-6; Bedard et al. 2012, 25-26; Amel-Zadeh and Faasse 2016, 10-11). This aspect has been noted by Amel-Zadeh and Faasse (2016): *"Investors muted short-term reaction to the footnotes could be because there is no marginal information value compared to the MD&A, or because the footnotes are more difficult or costly to process"* (Amel-Zadeh and Faasse 2016, 28).

Representing most Form 10-K disclosures, the clarity of narratives is vital understanding and interpreting the information contained in annual reports (Lo et al. 2017, 2). However, narrative disclosures, and especially the way they are presented, could also be used by corporate management to mislead investors (Devos and Sarkar 2015, 3). Having strong incentives to avoid the disclosure of negative information in regulatory filings (Kothari et al. 2009, 1643-1644, 1647; Bloomfield 2002, 238; Amel-Zadeh and Fasse 2016, 10) company officials might obfuscate negative corporate information to prevent and/or delay its incorporation into stock prices (*"management obfuscation hypothesis"*) (Bloomfield 2002, 238; Li 2008, 224; Devos and Sarkar 2015, 5). In this context, managers might strategically hide (shift) negative corporate information in the more complex and difficult to read footnotes section, hoping that capital market participants rather concentrate on other less complex sections of the annual report when analysing a Form 10-K filing (e.g. *"MD&A"*). Consequently, this leads to the question of whether company officials are (really) able to hide negative corporate information in the footnotes sections of annual reports, despite their important information content and relevance for investors.

While the academic literature investigates the information content and usefulness of accounting reports and whether various characteristics of financial statements are proxies of attempts to obfuscate information, less is known about the notes to the financial statements (*"Footnotes"*) (Devos and Sarkar 2015, 1). To address this research gap this study uses textual analysis to investigate whether managers can hide negative news in the footnotes and whether optimistic statements are used to frame pessimistic corporate information. Based on a sample of more than 30,000 footnotes sections in annual reports filed with the SEC, the study examines the presence and success of accounting narrative obfuscation in financial statements. It estimates a regression model using ordinary least squares (OLS) with a firm's short-term stock return as the dependent variable and several proxies for firm characteristics as independent variables. The empirical results show that management obfuscation in the notes to the financial statements is either *"not present"* (obfuscation by word choice), or *"has failed"* (obfuscation by disclosure outlet) even when considering risk adjusted stock returns and firms with low corporate profitability.

The study contributes to the literature in several ways. First, it shows that companies cannot hide negative information in financial statements, regardless of their profitability. Second, it indicates that even if corporate profitability is low, company officials do not attempt to mislead investors. Third, the study demonstrates that the notes to the financial statements, despite their relevance to stock characteristics, are not predictive of future corporate earnings. Finally, the study illustrates the effect of different research design choices on empirical results in the field of textual analysis. Besides academia, the results of this study are also important for managers and investors, and for regulatory authorities. Having discretion over the content of annual reports and their subsections, the study may help managers reduce the impact of negative information on financial markets. Observing a limited market reaction to important accounting narratives, investors may use the empirical results to adjust the analysis of regulatory filings and their information content. Financial regulators could use the study to improve the information requirements of public traded companies and overall market efficiency.

The remainder of Chapter 3 proceeds as follows. The next section presents the related literature. Section 3.2 develops the hypotheses. Section 3.3 describes the sample selection process and research methodology. Section 3.4 presents the empirical results. Sections 3.5 and 3.6 present the robustness and sensitivity tests. The additional tests are presented in Section 3.7 and finally, Section 3.8 concludes the Chapter.

3.1 Related literature

Li (2008) first demonstrated a correlation between textual characteristics and firm performance for a large sample of annual reports on Form 10-K filed with the SEC (Li 2008, 222). He found that the annual report of firms with higher corporate earnings were easier to read (lower Fog Index) (Li 2008, 234-235). However, a detailed analysis revealed that managers do not appear to obfuscate lower earnings by making annual reports more complex (Li 2008, 237). Analysing the writing style of the *"MD&A"* section, he found no evidence that firms with poor results strategically hide negative information (Li 2008, 240-241). On the contrary, company officials may structure annual reports opportunistically (lengthy disclosures) to delay the incorporation of adverse information into stock prices (Li 2008, 243-245).

As such, Li (2008) was unable to confirm a firm's ability to soften or delay the incorporation of negative information into stock prices by producing lengthy disclosures. Similarly, Bloomfield (2008) concluded that the idea of obfuscation in financial statements is *"not particularly robust"* (Bloomfield 2008, 249). Discussing alternative explanations for the correlation between the disclosure length or textual complexity of annual reports and firm performance (e.g. ontology, management by exception, litigation), Bloomfield (2008) detailed what could drive the textual characteristics of Form 10-K (length, readability) other than managers' intent to distract investors (Bloomfield 2008, 249-250).

Miller (2010) examined the effect of Form 10-K reporting complexity (length, readability) on investors trading behaviour (trading volume) (Miller 2010, 2108). He found an association between the complexity of annual reports (length) and trading activities. In particular, longer Form 10-K filings are associated with lower levels of aggregate trading volume (Miller 2010, 2108, 2119, 2123-2125, 2138). Interestingly, less readable annual reports (higher Fog Index) are not significantly related to lower abnormal trading volumes in the immediate days following the SEC submission, especially considering large investors (Miller 2010, 2123-2124, 2127-2129). Contrary to the idea of accounting narrative obfuscation by company officials, he concluded that large investors seem to prefer more data in financial statements written in a more technical style (less readable) (Miller 2010, 2136).

In addition, Rogers et al. (2011) examined the relation between disclosure tone and shareholder litigation (Rogers et al. 2011, 2156). Analysing the language of quoted portions (earnings announcements) used in class action lawsuits, they determined that plaintiffs focus on optimistic language in their class action complaints (Rogers et al. 2011, 2157, 2169, 2179). Finding a strong link between disclosure tone and shareholder litigation, the results confirm that sued firms disclose substantially more optimistic statements in their earnings announcements than non-sued firms (Rogers et al. 2011, 2157, 2171-2172, 2179-2177, 2179). The results further indicated that optimism and insider selling jointly affect litigation risk and that firms can mitigate their exposure to litigation by ensuring that optimistic statements are not contradicted by insider selling (Rogers et al. 2011, 2157, 2174, 2176, 2179). The findings suggest that monitoring and adjusting disclosure tone are means of reducing a firm's litigation risk (Rogers et al. 2011, 2157-2158). The authors concluded that *"managers can reduce litigation risk by dampening the tone of their earnings announcements either by decreasing their use of positive language or by tempering their optimism with statements that are less favorable"* (Rogers et al. 2011, 2179).

Goel and Gangolly (2012) investigated the linguistic features of narrative disclosures in fraudulent annual reports to detect corporate fraud (Goel and Gangolly 2012, 75, 78, 87). They found no difference in the count of complex sentential structures in fraudulent and non-fraudulent annual reports (Goel and Gangolly 2012, 84). Despite disclosing more negative and positive words, the relative distribution of word categories in Form 10-K filings reveals that both fraudulent and non-fraudulent annual reports share the same percentage of negative and positive words (Goel and Gangolly 2012, 84). While non-fraudulent annual reports disclose more active voice sentences, fraudulent financial statements state more passive voice sentences (Goel and Gangolly 2012, 85). Consistent with the idea that in cases of fraud, company officials deliberately use uncertainty to make accounting reports unclear and ambiguous, the authors observed that fraudulent Form 10-K filings exhibit higher levels of uncertainty than non-fraudulent submissions (Goel and Gangolly 2012, 85). Based on their results, they concluded that qualitative information in annual reports on Form 10-K and their linguistic features are useful detecting corporate fraud (Goel and Gangolly 2012, 87).

Huang et al. (2014) studied whether company officials engage in *"tone management"* for strategic purposes and how capital market participants react to this strategic behaviour (Huang et al. 2014, 1083-1084). According to them, the tone in voluntary earnings press releases (earnings announcements) is more positive for firms that are small, profitable, and growing (Huang et al. 2014, 1091). They also noted that an abnormal positive tone predicts negative future earnings and cash flows (Huang et al. 2014, 1083, 1085-1086, 1096-1097, 1111). Their results further confirm that an abnormal positive tone is associated with a positive stock market reaction in the immediate days following the release, reversing in the subsequent quarters (*"return reversal"*)

(Huang et al. 2014, 1083, 1086, 1088, 1105-1107, 1111). Finding that constrained firms are more likely to employ tone management, the authors concluded that managers use tone strategically to mislead investors about future firm fundamentals (Huang et al. 2014, 1083, 1086, 1088, 1111).

Devos and Sarkar (2015) explored the association between the number of footnotes in annual reports and a firm's reporting quality (auditor reputation) and its earnings persistence (Devos and Sarkar 2015, 2). They identified a negative relationship between a firm's reporting quality (Big 4 auditors) and the number of footnotes in its Form 10-K filing (Devos and Sarkar 2015, 12, 24). Lending support for managers trying to hide adverse information by adding more notes to financial statements (*"management obfuscation hypothesis"*) (Devos and Sarkar 2015, 9), they demonstrated a negative relationship between the number of footnotes and a firm's earnings persistence (earnings one and two years ahead) (Devos and Sarkar 2015, 12-13, 25). However, the results showed that firms using reputed auditors (Big 4 auditors) and disclosing more footnotes have higher earnings persistence (Devos and Sarkar 2015, 12-13; 26). Despite the alternative explanation that losses and short-term profits are simply more difficult to explain (see Bloomfield 2008) (Devos and Sarkar 2015, 13), the authors concluded that firms (non-Big 4 auditors) can obfuscate their annual reports by disclosing more footnotes in them (Devos and Sarkar 2015, 15).

Amel-Zadeh and Fasse (2016) examined the information content of the *"MD&A"* and *"Footnotes"* sections of annual reports (Amel-Zadeh and Fasse 2016, 2). They discovered that *"MD&A"* disclosures are more informative to capital market participants than the information stated in the notes to the financial statements (Amel-Zadeh and Fasse 2016, 4, 20-21, 28, 38). Consistent with naïve expectations, their results showed that (net) positive disclosure tone in the *"MD&A"* section is predictive of higher future stock returns. Contrary to naïve expectations, they found a negative association between (net) positive tone in the footnotes sections and future stock returns, indicating that company officials might *"obscure negative information contained in the quantitative details of the footnotes with a positive tone in their text"* (Amel-Zadeh and Fasse 2016, 5, 24-26, 41, 42). Furthermore, they find suggestive evidence of managers obfuscating adverse information through placement and tone (*"impression management"*), and a positive association between sentiment and earnings management (Amel-Zadeh and Fasse 2016, 5, 25-28, 41, 43).

In later research, Czerney et al. (2017) studied the use of narrative disclosure tone in the footnotes sections of annual reports using XBRL data (eXtensible Business Reporting Language) (Czerney et al. 2017, 1). Contrary to the idea of accounting narrative obfuscation, they showed that the overall or net linguistic tone (positive minus negative sentiment) in the footnotes section is positively correlated with a firm's prior stock return. This suggests that the tone of footnote disclosures reflects a company's underlying economic situation. Consistent with naïve expectations, they also demonstrated a positive association between corporate earnings and disclosure tone in this section of the Form 10-K filing (Czerney et al. 2017, 18, 40-41). The regression results further indicated that negative information (prior stock market performance) is disclosed more promptly in the footnotes section and that Big 4 audit firms constrain management's use of optimistic language when disclosing good news to reduce a firm's (and auditor's) litigation exposure (Czerney et al. 2017, 18-19, 40-41).

Henselmann and Hering (2017) explored the relevance of linguistic tone in various Form 10-K subsections on financial markets (Henselmann and Hering 2017, 77). They were, however, unable to show any significant connection between the overall (net) textual sentiment (ratio of negative minus positive words) in the footnotes section of an annual report and a firm's subsequent stock market performance (Henselmann and Hering 2017, 111). However, contrary to the idea of managers being able to hide negative information in the notes to the financial statements, the regression results confirmed a significant association between the change in overall (net) textual sentiment in the footnotes and a firm's stock market performance around the SEC submission (Henselmann and Hering 2017, 122-123, 125). In addition, the results support the notion that managers do not frame negative information in the footnotes with meaningless positive statements (Henselmann and Hering 2017, 123, 125). Not using risk-adjusted stock returns or differentiating between the accounting profitability across firms (different possible obfuscation behaviour and outcomes), the authors concluded that management obfuscation in financial statement footnotes is either *"not present"* or *"has failed"* (Henselmann and Hering 2017, 123).[14]

[14] Note that the present study is closely related to Henselmann and Hering (2017), as it uses the same data sample, the same parsing procedures, and the same measurement of textual sentiment. However, the present study also uses risk-adjusted stock returns to examine the market reaction to narrative information in the footnotes sections of annual reports. In addition, the present study interacts disclosure tone changes and corporate earnings based on the idea that obfuscation behaviour and its outcome might differ depending on a company's accounting profitability. Furthermore, this study examines the market reaction to different types of disclosure tone and the predictive ability of disclosure tone changes in the notes to the financial statements on a firm's future accounting profitability.

Seebeck et al. (2018) investigated firms' strategic use of XBRL extensions in the footnotes sections of annual reports to hide negative corporate information in regulatory filings. Countering to the idea of management obfuscation, the authors found that the tone (sentiment) of XBRL tag names do in fact represent a firm's economic situation (Seebeck et al. 2018, 107-108, 129). Consistent with the idea that negative corporate information is simply harder to explain (ontology), the results indicated a positive association between the number of tags used and their negative sentiment (see Bloomfield 2008) (Seebeck et al. 2018, 108, 129). Failing to provide evidence of managers using XBRL extensions to hide negative news in the footnotes sections of annual reports, the authors demonstrated a negative (not positive) association between the number of XBRL extensions (obfuscation behaviour) and their negative linguistic tone (Seebeck et al. 2018, 108-109, 130). Ignoring the possibility that bad news is simply harder to communicate, the authors used the sentiment spread between XBRL tag types (extension tags, standard tags) as an obfuscation measure. Finding a positive association between negative earnings (losses) and sentiment spread, they claimed that managers strategically use XBRL extensions to hide adverse corporate information in the footnotes of annual reports (Seebeck et al. 2018, 109-110, 131). Furthermore, the authors were unable significantly associate earnings management and the strategic tagging of unfavourable information in this subsection of the Form 10-K filing (Seebeck et al. 2018, 111-112). Finally, the authors explored the predictive ability of tagged information in the notes to the financial statements. They found a significant negative association between negative corporate information (XBRL extensions) and future corporate earnings (not management obfuscation), underscoring the importance of this disclosure outlet for investors (Seebeck et al. 2018, 110-111, 132). Despite their ambiguous results, they concluded that their findings *"provide strong evidence for management obfuscation and can be the basis for more future research in this field"* (Seebeck et al. 2018, 113).

Overall, empirical results regarding whether managers can hide negative information in the notes to the financial statements are mixed, especially in the context of certain firm features (e.g. corporate earnings). In addition, prior research fails to include important control variables when examining the information content of individual Form 10-K text sections (e.g. information content of other text sections). Furthermore, prior research does not consider the effects of different information types (e.g. *"uncertainty"*, *"litigious"*, *"constraining"*) and different tone measurements (e.g. tone level, tone change) on their empirical results. Finally, various research design choices in the field of textual analysis can have significant effects on research results, limiting their comparability (e.g. construction of dependent variables, sample selection process, construction of text scores, inclusion of table content). This study attempts to fill this research gap while also addressing the various limitations of previous studies.

By examining the stock market reaction to the footnotes sections of annual reports this study investigates whether investors do process even the most complex disclosure outlet within a Form 10-K filing (*"efficient markets hypothesis"*) or if company officials are really able to hide negative news in this specific subsection of the annual report (*"management obfuscation hypothesis"*). If capital market participants process the notes to the financial statements and a corresponding market reaction to the information content is found the idea of managers being able to successfully *"burry"* negative corporate information by using the footnotes section as a disclosure outlet is refuted. The same inference holds true for companies with low corporate profitability where investors are assumed to neglect narrative (soft) information in favor of quantitative (hard) information. If a company with low earnings endure the same (negative or positive) market reaction to (negative or positive) information in the footnotes as a company with high earnings the idea that former and its management can obfuscate adverse news by using the footnotes section is also refuted. Empirical evidence for a market reaction to the (complex) information content of the notes to the financial statements is significant since the entire field of research associated with management obfuscation is largely based on the idea that market participants can be fooled by complex disclosures hoping to prevent the incorporation of negative information into a firm's stock price.

3.2 Hypothesis development

The *"management obfuscation hypothesis"* argues that company officials obfuscate negative (adverse) corporate information to prevent and/or delay its incorporation into stock prices (Bloomfield 2002, 238; Li 2008, 224; Devos and Sarkar 2015, 5). Managers might strategically hide (shift) negative corporate information in the footnotes section (discretion over content), hoping that capital market participants concentrate on other sections of the annual report when analysing a Form 10-K filing (e.g. *"MD&A"*, *"Risk Factors"*). Note that this behaviour potentially applies to all companies regardless of their accounting profitability. Even if being read by investors, complex (non-transparent) footnote disclosures might increase the costs of extracting negative information, which will therefore be less reflected in market prices (*"incomplete revelation hypothesis"*) (Grossman and Stiglitz 1980, 404-405; Bloomfield 2002, 234-235; Li 2008, 224; Devos and Sarkar 2015, 5; Miller 2010, 2108, 2138; Amel-Zadeh and Fasse 2016, 2). *"As a result, it matters not only what is disclosed, but also where and in combination with* *what else"* (Amel-Zadeh and Fasse 2016, 2). In addition, research in psychology (e.g. Kahneman and Tversky, 1973; Nisbett and Ross, 1980) showed that people tend to underweight abstract and statistical information (e.g. *"Footnotes"*) (Amel-Zadeh and Fasse 2016, 10). In fact, research confirms that the footnotes are more likely to be used as a strategic deterrence to investors than other annual report subsections such as the *"MD&A"* (Li 2008, 236). Bloomfield (2002) noted this type of managerial behaviour: *"Managers seek to boost stock prices by hiding bad news in footnotes, and regulators work hard to defeat such efforts, even though the efficient markets hypothesis asserts that information is reflected in prices no matter how obscure its presentation"* (management obfuscation by disclosure outlet) (Bloomfield 2002, 233). This leads to the assumption that companies, regardless of their accounting profitability (high and low profitable companies), might be able to successfully hide negative corporate information in the footnotes since investors do not process the information disclosed in the footnotes. If true, a higher amount of negative information in the footnotes would not be associated with a negative market impact (regardless of accounting profitability).

Research further supports the idea that the obfuscation behaviour and outcome thereof might differ depending on a company's accounting profitability. For instance, companies with a lower profitability tend to have even longer and more complex footnote disclosures than profitable firms (Li, 2008, 234-236) potentially negatively affecting their incorporation into stock prices because investors may elect not to process the information at all (Miller 2010, 2108). Faced with terrible corporate earnings, investors may concentrate on *"hard"* quantitative information alone (balance sheet, income statement) rather than on lengthy and complex *"soft"* footnote disclosures (accounting narratives) that enable managers to hide negative textual information in the notes to the financial statement (under reaction to adverse information). This leads to the assumption that (at least) companies with a low accounting profitability might be able to successfully hide negative corporate information in the footnotes since investors do not (or less) process the information disclosed by companies with low corporate earnings. If true, a higher amount of negative information in the footnotes and its negative market impact would be less profound for companies with a low accounting profitability.

To empirically test whether managers are able to use the footnotes sections of annual reports to successfully obfuscate negative corporate information, the following hypotheses are proposed:

Hypothesis 3a. Companies are able to obfuscate negative information in financial statements by choosing the footnotes sections in annual reports as a disclosure outlet to display pessimistic corporate information since investors do not process the (negative) information content in this section regardless of a firm's accounting profitability.

Hypothesis 3b. Low profitable companies are able to obfuscate negative information in financial statements by choosing the footnotes sections in annual reports as a disclosure outlet to display pessimistic corporate information since investors do not process the (negative) information content in this section when corporate profitability is low.

In general, a firm's management has strong incentives to avoid the disclosure of negative information in regulatory filings such as annual reports (e.g. promotions, outside employment opportunities) (Kothari et al. 2009, 1643-1644, 1647; Bloomfield 2002, 238; Amel-Zadeh and Fasse 2016, 10). However, company officials do have incentives to disclose positive information (accounting narratives) in financial statements (e.g. management compensation, capital costs) (Twedt and Rees 2011, 21; Kothari et al. 2009, 1641-1643; Campbell et al. 2014, 401). *"Managers frequently use discretion to describe their firms' results in a favourable light"* (Rogers et al. 2011, 2158). Besides choosing a particular disclosure outlet in the annual report on Form 10-K (*"Footnotes"*) to display pessimistic information, managers might also use positive statements to frame negative corporate disclosures (*"impression management"*) (Czerney et al. 2017, 19; Amel-Zadeh and Fasse 2016, 11). Note that this behaviour potentially applies to all companies regardless of their accounting profitability. Loughran and McDonald (2016) highlighted this behaviour: *"Rarely does management negate a negative word to make a positive statement. Positive words, on the other hand, in addition to their positive usage, are just as frequently used to frame a negative statement"* (Loughran and McDonald 2016, 1217). Hoping to mitigate the effect of negative corporate disclosures on financial markets, positive statements might be used to frame or obfuscate negative (adverse) information in financial statements (management obfuscation by word choice) (Czerney et al. 2017, 7; Huang et al. 2014, 1090). This leads to the assumption that companies, regardless of their accounting profitability (high and low profitable companies), might use positive statements in the footnotes section to frame negative corporate information. If true, a higher amount of positive information in the footnotes would not be associated with a positive (less negative) market impact (regardless of accounting profitability).

Again, research supports the notion that the extent to which managers engage in *"impression management"* depends on a firm's accounting profitability. For instance, companies that just meet or beat past earnings and analysts' forecasts are more likely to disclose information with an abnormal positive tone (orthogonal to the underlying fundamentals) (Huang et al. 2014, 1098). Forced to disclose negative information (explanation of low current earnings), managers of low profitable companies might be more willing to frame or *"manage"* the tone of their firm's accounting narratives. In addition, company officials responsible for low corporate earnings might report textual information diverging from the underlying accounting reality (*"managerial bias"*) (Amel-Zadeh and Fasse 2016, 25). This leads to the assumption that (at least) companies with a low accounting profitability might use meaningless positive information (more often) to frame negative information in the footnotes than companies with a high accounting profitability (high corporate earnings). If true, a higher amount of positive information in the footnotes would not be associated with a positive (less negative) market impact for companies with a low accounting profitability.

However, shareholder litigation is an important mechanism to limit company officials' opportunistic disclosure behaviour (Rogers et al. 2011, 2157; Gandhi et al. 2017, 4). According to the Securities Exchange Act of 1934, *"it shall be unlawful for any person, directly or indirectly, by the use of any means [...] to make any untrue statement of a material fact or to omit to state a material fact necessary in order to make the statements made [...] not misleading"* (SEC 1934, §240.10b-5(b)).

To empirically test whether managers use positive statements in the footnotes sections of annual reports to frame negative corporate information, the following hypotheses are proposed:

Hypothesis 4a. Companies use meaningless positive statements in the footnotes sections of annual reports to frame pessimistic corporate disclosures represented by an inverse market reaction to positive information (negative not positive market impact to a higher amount of positive statements).

Hypothesis 4b. Low profitable companies use meaningless positive statements in the footnotes sections of annual reports to frame pessimistic corporate disclosures represented by an inverse market reaction to positive information (negative not positive market impact to a higher amount of positive statements).

3.3 Research design

3.3.1 Sample selection and description

Using the EDGAR database (*"Form Index Files"*), all annual reports (Form 10-K; Form 10-K405) filed with the SEC between 1993 and 2016 were downloaded. Note that the time period was chosen due to data availability reasons on various information platforms (e.g. SEC EDGAR database, commercial data providers, public research repositories). The number of annual reports filed with the SEC is shown in the appendix.[15] For each submission, the text version (*"Complete Submission Text File"*) was retrieved from the EDGAR database. The text version contains all information disclosed within a particular SEC submission (e.g. core Form 10-K document, exhibits) (Bodnaruk et al. 2015, 643; Loughran and McDonald 2011a, 1). An information extraction algorithm (regular expressions) was applied to extract all narrative disclosures within a submission.[16] In addition to the footnotes section of the annual report, the core Form 10-K document (including relevant exhibits) embedded in each submission was also extracted.[17] Based on the extraction results, word counts were obtained (Loughran and McDonald business dictionary) to quantify the disclosure tone in the financial statements (footnotes sections, entire submissions including exhibits).

For the analysis, several databases were used to match stock market data and accounting variables. For each annual report, the Central Index Key (CIK) number was extracted from the document name on the SEC server. The CIK number was used to obtain the International Securities Identification Number (ISIN) and accounting variables for each financial statement (filer) from Standard & Poor's Capital IQ database. Based on the ISIN, stock returns were retrieved for each company around its annual report filing date as well as for the Standard & Poor's 500 stock index (S&P 500) from the Thomson Reuters DataStream platform. To include a firm-year observation in the data sample, accounting data (return on assets, change in return on assets, leverage ratio) and stock market data (market capitalization, price-to-book ratio) had to be available in addition to an exchange listing on the New York Stock Exchange (NYSE), National Association of Securities Dealers Automated Quotations (NASDAQ), or American

---

15 Please note that the Form 10-K sample is identical to the one used in Henselmann and Hering (2017).

16 Note that the information extraction algorithm applied in this study is based on the *"Annual Report Algorithm"* and *"Items Algorithm"* designed by Hering (2016). Furthermore, the parsing procedures are identical to those used in Henselmann and Hering (2017). For a detailed overview of the parsing rules applied in this study, please refer to the appendix.

17 Note that the study also includes the disclosure tone (textual sentiment) of the entire filing (submission including exhibits) as a control variable to proxy for the linguistic tone in the overall submission.

Stock Exchange (NYSE MKT). Table 14 provides an overview of the data filters and overall sample selection process.

The majority of annual reports in the data sample were filed by companies in the *“Business Services”* industry (12.38%), followed by *“Trading”* (6.93%), *“Electronic Equipment”* (6.67%), and *“Pharmaceutical Products”* (6.42%). In general, the distribution of financial statements among certain industries in the data sample is unequal (Fama-French 48 industry classification). While the distribution of firm-year observations across industries is unequal, the distribution of annual reports on Form 10-K over time is more balanced. On average, more than 3,000 annual reports from each year were included in the analysis. Table 15 presents the distribution of annual submissions by certain industries and fiscal years.

Table 14. Sample selection process for stock return examination.

	Firm-year observations	**Observations removed**
All observations (Form 10-K & Form 10-K405)	189,998	
Less		
Missing 10-K report (database error)	189,997	1
Removing duplicates CIK and fiscal-year	184,089	5,908
Removing late filers (>100 days after accounting period)	163,838	20,251
Missing SIC (from filing)	162,759	1,079
Merge with Capital IQ data		
Missing ISIN	118,822	43,937
Missing return on assets	112,675	6,147
Missing change in return on assets	107,018	5,657
Missing leverage	94,319	12,699
Missing NYSE, AMEX or NASDAQ exchange listing	69,546	24,773
Merge with DataStream data		
Missing market capitalization	67,198	2,348
Price-to-Book DataStream data available and ratio > 0	63,914	3,284
Other data filters		
Number of words in Form 10-K (including exhibits) >= 2,000	63,750	164
Number of words in Form 10-K (excluding exhibits) >= 2,000	62,757	993
Observations in final sample		
Firm-year sample	62,757	
Number of unique firms	6,659	-
Average number of years per firm	9	-

Notes: The table presents the data filters and sample selection process.

Table 15. Distribution of Form 10-K filings across industries and fiscal years for stock return examination.

Panel A: Distribution of Form 10-K filings across industries

Fama-French industry	Filings (Num.)	Filings (%)	Filings (∑%)	Fama-French industry	Filings (Num.)	Filings (%)	Filings (∑%)
Business Services	7,767	12.38	12.38	Apparel	796	1.27	86.95
Trading	4,348	6.93	19.31	Banking	726	1.16	88.11
Electronic Equipment	4,185	6.67	25.98	Electrical Equipment	703	1.12	89.23
Pharmaceutical Products	4,028	6.42	32.40	Personal Services	706	1.12	90.35
Retail	3,210	5.11	37.51	Entertainment	676	1.08	91.43
Insurance	2,656	4.23	41.74	Real Estate	618	0.98	92.41
Petroleum and Natural Gas	2,478	3.95	45.69	Business Supplies	605	0.96	93.37
Utilities	2,453	3.91	49.60	Printing and Publishing	561	0.89	94.26
Medical Equipment	2,344	3.74	53.34	Almost Nothing	531	0.85	95.11
Computers	2,335	3.72	57.06	Rubber and Plastic Products	443	0.71	95.82
Machinery	2,201	3.51	60.57	Recreation	402	0.64	96.46
Wholesale	2,102	3.35	63.92	Aircraft	307	0.49	96.95
Measuring and Control Equipment	1,574	2.51	66.43	Non-Metallic and Industrial Metal Minin	269	0.43	97.38
Transportation	1,572	2.50	68.93	Beer & Liquor	248	0.40	97.78
Communication	1,527	2.43	71.36	Fabricated Products	193	0.31	98.09
Chemicals	1,280	2.04	73.40	Shipping Containers	195	0.31	98.40
Restaraunts, Hotels, Motels	1,208	1.92	75.32	Agriculture	168	0.27	98.67
Healthcare	1,197	1.91	77.23	Textiles	168	0.27	98.94
Food Products	1,017	1.62	78.85	Shipbuilding, Railroad Equipment	156	0.25	99.19
Construction Materials	919	1.46	80.31	Precious Metals	146	0.23	99.42
Construction	874	1.39	81.70	Defense	130	0.21	99.63
Automobiles and Trucks	872	1.39	83.09	Coal	101	0.16	99.79
Consumer Goods	817	1.30	84.39	Candy & Soda	71	0.11	99.90
Steel Works Etc	812	1.29	85.68	Tobacco Products	62	0.10	100.00

Panel B: Distribution of Form 10-K filings across fiscal years

Fiscal year	Filings (Num.)	Filings (%)	Filings (∑%)	Fiscal year	Filings (Num.)	Filings (%)	Filings (∑%)
1993	255	0.41	0.41	2005	3,088	4.92	48.75
1994	409	0.65	1.06	2006	3,105	4.95	53.69
1995	1,075	1.71	2.77	2007	3,099	4.94	58.63
1996	1,891	3.01	5.78	2008	3,045	4.85	63.48
1997	2,555	4.07	9.86	2009	3,124	4.98	68.46
1998	2,830	4.51	14.36	2010	3,118	4.97	73.43
1999	3,068	4.89	19.25	2011	3,068	4.89	78.32
2000	3,080	4.91	24.16	2012	3,109	4.95	83.27
2001	3,111	4.96	29.12	2013	3,243	5.17	88.44
2002	3,093	4.93	34.05	2014	3,263	5.20	93.64
2003	3,054	4.87	38.91	2015	3,231	5.15	98.79
2004	3,083	4.91	43.83	2016	760	1.21	100.00

Notes: The table illustrates the number of annual reports in the data sample across Fama-French industries (Panel A) and across fiscal years (Panel B).

3.3.2 Regression model, dependent variable, and independent variables

Regression model

To examine the presence and success of accounting narrative obfuscation in the footnotes sections of annual reports on Form 10-K, the following regression model was estimated using OLS[18]:

$$\begin{aligned} Y_{i;t} = \beta_0 &+ \beta_1 * \Delta\, Tone_{Net;Full\ Report;i;t} + \beta_2 * \Delta\, Tone_{Net;Footnotes;i;t} \\ &+\beta_3 * Low\ ROA\ Dummy\ (D)_{i;t} + \beta_4 * \Delta\, Tone_{Net;Footnotes;i;t} * D_{i;t} \\ &+\beta_5 * Ln\ MVE_{i;t} + \beta_6 * PtB_{i;t} + \beta_7 * ROA_{i;t} + \beta_8 * \Delta\, ROA_{i;t} \\ &+\beta_9 * Leverage_{i;t} + \sum Year\ Fixed\ Effects \\ &+ \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \tag{13}$$

Where: the dependent variable ($Y_{i;t}$) is a firm's abnormal stock return around its Form 10-K filing. $\Delta\, Tone_{Net;Footnotes;i;t}$ is the change in the overall disclosure tone in the footnotes section of a firm's annual report as measured by the number of negative and positive words in this text section. The control variables are described below. *Year Fixed Effects* and *Industry Fixed Effects* represent the included fixed effects.

Dependent variable

To test whether capital market participants react to changes in disclosure tone, a firm's stock market performance in the days around the Form 10-K filing was calculated. The measure of a firm's short-term stock market performance in the days around the SEC submission is the cumulative abnormal stock return (CAR) over a four-day holding period starting one day before the initial filing date of the annual report (market-adjusted model - MAM)[19]:

$$CAR_{i;[-1,+2];MAM} = \sum_{t-1}^{+2} \left(Return_{i;t} - Return_{S\&P\ 500;t} \right) \tag{14}$$

[18] Note that linear regression models are most commonly used in the accounting and finance literature when examining the effect of textual information on firm characteristics (Kearney and Liu 2014, 177). Further note that the study used a linear regression model over various machine learning models (e.g. neural networks) to better examine the relationship between input (independent) and output (dependent) variables.

[19] Note that the study focused on short-term capital market effects following the Form 10 K submission to avoid biases due to other effects not being captured by the model (e.g. other corporate information not being disclosed in the annual report, other corporate events, market trends).

Where: $Return_{i;t}$ is the return of a certain stock (i) on day (t) calculated as the price of the stock (i) on day (t) divided by the price of the stock (i) on the previous day minus one and $Return_{S\&P\ 500;t}$ is the return of the market index (S&P 500) on day (t) calculated as the level of the market index on day (t) divided by the level of the market index on the previous day minus one.

In addition, a market model (MM) was used to calculate a firm's cumulative abnormal stock return (CAR) during the event period ([t-1, t+2]):

$$CAR_{i;[-1,+2];MM} = \sum_{t-1}^{+2}\left(Return_{i;t} - (\hat{\alpha}_i + \hat{\beta}_i * Return_{S\&\ 500;t})\right) \tag{15}$$

Where: $\hat{\alpha}_{\mathrm{i}}$ (intercept) and $\hat{\beta}_{\mathrm{i}}$ (slope) are estimated using an OLS regression of the daily returns of the company (i) on the daily returns of the market index (S&P 500) during the non-event period ([t-250, t-21]).

Lastly, a 4-factor model (4 Factor) was used to obtain a firm's cumulative abnormal stock return (CAR) during the event period ([t-1, t+2]):

$$CAR_{i;[-1,+2];4\ Factor} = \sum_{t-1}^{t+2}(Return_{i;t} - Return_{i;t;4\ Factor}) \tag{16}$$

Where: $Return_{i;t;4\ Factor}$ is the expected return of a company (i) on day (t) based on a 4-factor model. Using the data provided by Kenneth R. French daily stock returns were regressed on the excess market returns (MKT-RF), a size factor (small-minus-big), a value factor (high-minus-low), and a momentum factor (up-minus-down) over an estimation window ([t-250, t-21]) relative to the Form 10-K filing dates to estimate the corresponding factor loadings for the model (Amel-Zadeh and Faasse 2016, 16-17).

Note that the study followed prior research studies in the field of textual analysis (e.g. Amel-Zadeh and Faasse 2016) in defining the non-event period to avoid biases caused by abnormal trading behaviour before the release date of the annual report. Further note that dividends and other capital gains being paid (total return) were accounted for.

To assess the timeliness of the market reaction in more detail, stock returns over multiple time horizons (holding periods) were calculated starting one day before the SEC submission. Note that dividends and other capital gains being paid during the different holding periods (total return) were included in the analysis.

Main variable of interest

The present study explores the presence and success of accounting narrative obfuscation in financial statements by analysing whether and how investors react to changes in the disclosure tone. To quantify the disclosure tone in the notes to the financial statements, the Loughran and McDonald business dictionary and its word lists (e.g. *"negative"*, *"positive"*, *"uncertainty"*, *"litigious"*, *"constraining"*) were used. Following prior work (e.g. Henry and Leone 2016), an equal weighting scheme was applied in this study to measure the linguistic tone in annual reports. For each footnote section in the data sample, the annual change in disclosure tone was calculated.

The measure of the disclosure tone in the footnotes sections of annual reports is the overall or net linguistic tone (negative minus positive words) scaled by the text length of the footnotes section in a particular year minus the overall or net textual sentiment of the same disclosure outlet in a firm's previous Form 10-K filing:

$$\Delta\, Tone_{Net;Footnotes;i;t} = \frac{Negative_{i;t} - Positive_{i;t}}{Total_{i;t}} - \frac{Negative_{i;t-1} - Positive_{i;t-1}}{Total_{i;t-1}} \quad (17)$$

Where: $\Delta\, Tone_{Net;Footnotes;i;t}$ is the change in the overall or net textual sentiment in the notes to the financial statements (current filing versus previous submission), $Negative_{i;t}$ is the number of negative words, $Positive_{i;t}$ is the number of positive words and $Total_{i;t}$ is the number of all words appearing in this report section (*"Footnotes"*).

Note that the Loughran and McDonald business dictionary (excluding stop words) was used to measure the text length of the footnotes sections in financial statements. Furthermore, this process of quantifying disclosure tone is objective and replicable (Rogers et al. 2011, 2162; Loughran and McDonald 2016, 1200).

Control variables

To control for common firm fundamentals in each regression, the natural logarithm of the market value of equity and price-to-book ratio were included as control variables. In addition, a firm's leverage ratio, return on assets, and the change in return on assets were included as control variables. Furthermore, the change in the disclosure tone of the entire filing (Form 10-K submission including exhibits) was also included as a control variable to proxy for textual characteristics in the entire SEC submission (Loughran and McDonald 2016, 1218). Finally, for each industry the firm-year observations with the lowest profitability (ROA) were identified (25 percent quantile) to proxy for companies with low corporate earnings. The corresponding dummy variable (binary variable) takes on a value of 1 if a firm-year observation belongs to this category and 0 if it does not (Low ROA Dummy). Note that the study included the change in return on assets as well as the change in the disclosure tone in the overall filing as control variables to address concerns related to omitted variable bias. All regressions used year and industry fixed effects. Standard errors were clustered by firms, and all included variables were winsorized at the 1% level. A detailed overview of all variables is presented in the appendix.

3.3.3 Summary statistics and correlations

This section presents the summary statistics for the market, disclosure, and accounting variables and the correlations between the key variables of interest in the study. The mean (median) cumulative abnormal stock return (market model) of a company over a four-day holding period around the filing date has a slightly negative value of -0.1 (-0.1) percent. Over the same holding period, the cumulative abnormal stock return (market model) at the 10th percentile is -6.52 percent compared to a value of 6.27 percent at the 90th percentile. Panel A of Table 16 provides the summary statistics for cumulative abnormal stock returns over different holding periods around the SEC submission (dependent variables).

The mean (median) change in overall textual sentiment in the entire report is 0.00 (0.03) percent. The mean (median) change in overall textual sentiment in the footnotes section is 0.04 (0.00) percent. The mean (median) market value of a firm included in the analysis is $553M ($539M). The average price-to-book ratio is 3.37 (2.04). The return on assets has a mean (median) value of 2.61 (4.45) percent. The mean (median) change in return on assets is 0.32 (0.02) percent. The mean (median) leverage ratio of the companies included in the data sample is 48.41 (49.18) percent. Panel B of Table 16 presents the summary statistics for the market, disclosure, and accounting variables used as control variables in the analysis (independent variables).

Form 10-K textual sentiment generally relates to a firm's abnormal stock return in the expected manner. A positive change in overall textual sentiment (more pessimistic financial statements) is correlated with a lower abnormal stock return. The overall disclosure tone in financial statements also generally relates to other variables in the expected manner. A positive change in net disclosure tone is correlated with a lower market capitalization, lower price-to-book ratio, lower return on assets, and a negative change in return on assets. Conversely, a more pessimistic disclosure tone in Form 10-K filings is correlated with a lower leverage ratio. Table 17 provides the correlation coefficients for the change in the disclosure tone in annual reports and firm characteristics.

Table 16. Dependent and independent variables for stock return examination.

Panel A: Dependent variables				
Variable	Num.	Mean	Median	St. Dev.
$\mathrm{CAR}_{[-1,1];\mathrm{MAM}}$	31,981	0.0004	0.0000	0.0559
$\mathrm{CAR}_{[-1,2];\mathrm{MAM}}$	31,981	0.0006	0.0000	0.0638
$\mathrm{CAR}_{[-1,3];\mathrm{MAM}}$	31,981	0.0005	-0.0001	0.0697
$\mathrm{CAR}_{[-1,4];\mathrm{MAM}}$	31,981	0.0010	0.0004	0.0748
$\mathrm{CAR}_{[-1,5];\mathrm{MAM}}$	31,981	0.0014	0.0004	0.0796
$\mathrm{CAR}_{[-1,1];\mathrm{MM}}$	31,981	-0.0010	-0.0008	0.0555
$\mathrm{CAR}_{[-1,2];\mathrm{MM}}$	31,981	-0.0010	-0.0010	0.0632
$\mathrm{CAR}_{[-1,3];\mathrm{MM}}$	31,981	-0.0013	-0.0012	0.0693
$\mathrm{CAR}_{[-1,4];\mathrm{MM}}$	31,981	-0.0012	-0.0010	0.0748
$\mathrm{CAR}_{[-1,5];\mathrm{MM}}$	31,981	-0.0011	-0.0013	0.0797
$\mathrm{CAR}_{[-1,1];4\ \mathrm{Factor}}$	31,981	-0.0013	-0.0013	0.0552
$\mathrm{CAR}_{[-1,2];4\ \mathrm{Factor}}$	31,981	-0.0013	-0.0018	0.0627
$\mathrm{CAR}_{[-1,3];4\ \mathrm{Factor}}$	31,981	-0.0013	-0.0017	0.0685
$\mathrm{CAR}_{[-1,4];4\ \mathrm{Factor}}$	31,981	-0.0010	-0.0018	0.0741
$\mathrm{CAR}_{[-1,5];4\ \mathrm{Factor}}$	31,981	-0.0011	-0.0016	0.0789
Panel B: Independent variables				
Variable	Num.	Mean	Median	St. Dev.
$\Delta\ \mathrm{Tone}_{\mathrm{Net;Full\ Report}}$	53,848	0.0004	0.0003	0.0068
$\Delta\ \mathrm{Tone}_{\mathrm{Net;Footnotes}}$	35,587	0.0004	0.0000	0.0068
Low ROA Dummy (D)	62,757	0.2500	0.0000	0.4330
$\Delta\ \mathrm{Tone}_{\mathrm{Net;Footnotes}} * (\mathrm{D})$	35,587	0.0002	0.0000	0.0035
Ln MVE	62,757	20.1315	20.1050	1.9395
PtB	62,757	3.3737	2.0400	4.6094
ROA	62,757	0.0261	0.0445	0.1173
Δ ROA	62,757	0.0032	0.0002	0.0683
Leverage	62,757	0.4841	0.4918	0.2316

Notes: The table provides an overview of stock returns over various holding periods around the Form 10-K submission and common firm fundamentals included in the analysis as control variables.

Table 17. Correlations between stock returns, disclosure tone, and firm fundamentals.

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1: $CAR_{[-1,2];MAM}$	1	0.97***	0.92***	0.00	-0.01**	-0.02***	-0.01**	0.00	-0.05***	0.04***	0.01*	0.04***
2: $CAR_{[-1,2];MM}$	0.94***	1	0.94***	0.00	-0.01**	-0.02***	-0.01*	0.00	-0.06***	0.04***	-0.02***	0.04***
3: $CAR_{[-1,2];4\ Factor}$	0.88***	0.92***	1	0.00	-0.01*	-0.01**	-0.01*	0.00	-0.04***	0.03***	-0.01*	0.03***
4: $\Delta\ Tone_{Net;Full\ Report}$	-0.01	0.00	0.00	1	0.12***	0.04***	0.06***	-0.02***	-0.01	-0.04***	-0.07***	-0.01***
5: $\Delta\ Tone_{Net;Footnotes}$	-0.01**	-0.01*	-0.01**	0.15***	1	0.03***	0.50***	-0.01*	-0.02***	-0.02***	-0.04***	-0.01
6: Low ROA Dummy (D)	-0.03***	-0.03***	-0.03***	0.05***	0.04***	1	0.09***	-0.26***	-0.02***	-0.55***	-0.14***	0.00
7: $\Delta\ Tone_{Net;Footnotes}$ * (D)	-0.01	-0.01	0.00	0.08***	0.48***	0.11***	1	-0.03***	-0.02***	-0.04***	-0.04***	0.00
8: Ln MVE	0.02***	0.02***	0.02***	-0.02***	-0.02***	-0.26***	-0.03***	1	0.18***	0.28***	0.02***	0.23***
9: PtB	-0.03***	-0.06***	-0.05***	-0.03***	-0.04***	-0.16***	-0.04***	0.38***	1	-0.04***	0.09***	0.07***
10: ROA	0.04***	0.03***	0.03***	-0.05***	-0.04***	-0.63***	-0.07***	0.30***	0.29***	1	0.15***	0.07***
11: Δ ROA	0.02***	-0.01***	-0.01***	-0.09***	-0.06***	-0.18***	-0.04***	0.04***	0.12***	0.20***	1	-0.01
12: Leverage	0.03***	0.03***	0.03***	-0.02***	-0.01	0.00	0.00***	0.24***	-0.03***	-0.05***	0.00	1

Notes: The table provides correlation results for stock returns, the change in tone, and firm fundamentals. Pearson correlation coefficients are presented in the upper triangle. Spearman correlation coefficients are presented in the lower triangle.

3.4 Empirical results

Inconsistent with *Hypothesis 3a*, the regression coefficient for $\Delta\, Tone_{Net;Footnotes}$ is significantly negative. Following naïve expectations, a positive change in $\Delta\, Tone_{Net;Footnotes}$ (more pessimistic tone in the notes to the financial statements) is negatively associated with short-term stock returns. Contrary to the idea that company officials can successfully *"bury"* negative corporate information in the footnotes section of an annual report on Form 10-K (management obfuscation by disclosure outlet), the regression results confirm that capital market participants process the textual information (accounting narratives) in this part of the annual report and react accordingly. Consistent with prior research studies (e.g. Henselmann and Hering 2017), this result suggests that management obfuscation via the footnotes section in an annual report as a disclosure outlet to hide adverse corporate information has indeed *"failed"*, even when using risk-adjusted stock returns (4-factor model). Adding to this notion, the regression results also reveal that market participants more strongly react to changes in the disclosure tone in this report section and in a more timely fashion than to linguistic changes in the entire Form 10-K filing ($\Delta\, Tone_{Net;Full\ Report}$).

However, possibly, a firm's ability to hide negative information in the notes to the financial statements depends on its accounting profitability. Faced with longer and more complex footnote disclosures when considering low profitable companies (Li 2008, 234-236), investors may underreact or ignore the negative information content in the notes to the financial statements (*"incomplete revelation hypothesis"*) (Grossman and Stiglitz 1980, 404-405; Bloomfield 2002, 234-235; Li 2008, 224; Devos and Sarkar 2015, 5; Miller 2010, 2108, 2138; Amel-Zadeh and Fasse 2016, 2). Inconsistent with this notion and *Hypothesis 3b*, the regression coefficient for $\Delta\, Tone_{Net;Footnotes}$ * *Low ROA Dummy* is insignificant (negative). This result suggests that the negative market reaction to a positive change in $\Delta\, Tone_{Net;Full\ Report}$ (more pessimistic tone in the notes to the financial statements) is not reduced (less negative market impact) when considering firms with low corporate profitability (one would expect a significant positive sign on $\Delta\, Tone_{Net;Footnotes}$ * *Low ROA Dummy*). In other words, investors do not underreact to more pessimistic disclosures in the footnotes sections of firms even if corporate earnings are low.

Inconsistent with *Hypothesis 4a* and the idea that company officials use meaningless positive statements to frame negative disclosures in the footnotes, the regression results show that a negative change in $\Delta\, Tone_{Net;Footnotes}$ (more optimistic disclosure tone in the notes to the financial statements) is associated with higher subsequent stock returns around the Form 10-K filing. Consistent with prior research (e.g. Czerney et al. 2017; Henselmann and Hering 2017) and the idea of an increased litigation risk (overly optimistic financial statements) (e.g. Rogers et al. 2011), this result indicates that company officials do not disclose meaningless positive statements to mitigate the effect of pessimistic corporate disclosures in the footnotes sections of annual reports (management obfuscation by word choice). Interestingly, this result is not line with previous empirical findings showing an opposite behavior (e.g. Amel-Zadeh and Fasse 2016; Seebeck et al. 2018; Huang et al. 2014). The lack of obfuscation in the notes to the financial statements might stem from the higher net benefits of tone management in other corporate documents less subject to evidentiary use in litigations, such as earnings press releases (Huang et al. 2014, 1091). Rogers et al. (2011) also noted this aspect: *"Why would investors respond to an optimistic tone if there are no enforcement mechanisms to lend credibility to tone?"* (Rogers et al. 2011).

Again, possibly, companies with a lower accounting profitability are more likely to use meaningless positive statements in the footnotes sections of Form 10-K filings to mitigate the effect of pessimistic corporate disclosures made in this part of the report. If companies with lower corporate earnings use positive statements in the footnotes sections to frame negative information, one would expect that a negative change in $\Delta\, Tone_{Net;Footnotes}$ (more optimistic disclosure tone in the notes to the financial statements) is associated with lower abnormal stock returns for these firms (positive sign on $\Delta\, Tone_{Net;Footnotes}$* *Low ROA Dummy*). To put it in another way, if $\Delta\, Tone_{Net;Footnotes}$* *Low ROA Dummy* carries a positive sign (loading) on abnormal stock returns a higher amount of positive information would decrease the magnitude of the coefficient (since it is defined as the change in negative minus positive tone) and also the abnormal stock return (since it has a positive loading) suggesting that positive information is used to frame negative news. However, the (insignificant) regression results suggest that for low profitable firms, a more optimistic footnotes section is associated with higher abnormal stock returns around the SEC submission (negative sign on $\Delta\, Tone_{Net;Footnotes}$ * *Low ROA Dummy*). To put in an another way, since $\Delta\, Tone_{Net;Footnotes}$* *Low ROA Dummy* (in fact) carries a negative sign (loading) on abnormal stock returns a higher amount of positive information decreases the magnitude of the coefficient (since it is defined as the change in negative minus positive tone) but increases the abnormal stock return (since it has a negative loading) suggesting that the positive statements are credible. Therefore, *Hypothesis 4b* cannot be confirmed.

In general, these results suggest that management obfuscation in financial statements, especially in the footnotes sections of annual reports, is not successful. The results also indicate that company officials disclose truthful narrative information in the footnotes sections of annual reports representing accurate corporate disclosures. Furthermore, the results show that investors do not underreact to textual changes in the footnotes sections when considering firms with low corporate earnings, disregarding the ability of company officials to successfully hide negative news in this Form 10-K disclosure outlet. Managers also do not seem to frame negative information with positive statements in the notes to the financial statements, regardless of their firm's corporate profitability. On the contrary, despite including various control variables to avoid problems associated with omitted variables (omitted variable bias) it might be the case that certain variables not included in the study explain the empirical findings. For instance, a firm's corporate governance might influence the disclosure behavior of a firm in regulatory filings such as annual reports while at the same time affecting a firm's stock price and stock market performance (despite standardized filing, disclosure, and auditing requirements). This holds true for other variables possibly associated with the main variable of interest and the dependent variable (e.g. credit rating changes). Please also note the low explanatory power for the dependent variable which might be explained by its variability itself and various market environments and market forces over time. In a broader context, the (fast) incorporation of textual information into a firm's stock market performance following the submission of the annual report by capital market participants do lend support for the idea of efficient capital markets (efficient markets hypothesis - EHM) and the assumption that all relevant corporate (quantitative and qualitive) information is been incorporated into the stock price of a company. Table 18 to Table 20 provide the regression results.

Table 18. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market-adjusted model).

Variable	$CAR_{[-1,1];MAM}$	$CAR_{[-1,2];MAM}$	$CAR_{[-1,3];MAM}$	$CAR_{[-1,4];MAM}$	$CAR_{[-1,5];MAM}$
Δ $Tone_{Net;Full\ Report}$	-0.0227	-0.0164	0.0301	-0.0105	-0.0112
	(-0.498)	(-0.314)	(0.531)	(-0.173)	(-0.172)
Δ $Tone_{Net;Footnotes}$	**-0.0992****	**-0.0870**	**-0.0919**	**-0.0967**	**-0.0941**
	(-2.062)	**(-1.567)**	**(-1.529)**	**(-1.533)**	**(-1.410)**
Low ROA Dummy (D)	-0.000710	-0.00149	-0.000859	-0.00118	-0.00129
	(-0.669)	(-1.240)	(-0.661)	(-0.839)	(-0.852)
Δ $Tone_{Net;Footnotes}$*(D)	**-0.0372**	**-0.134**	**-0.116**	**-0.104**	**-0.104**
	(-0.306)	**(-0.927)**	**(-0.757)**	**(-0.645)**	**(-0.603)**
Ln MVE	-0.000297	-0.000689***	-0.000666***	-0.000953***	-0.00125***
	(-1.546)	(-3.074)	(-2.801)	(-3.729)	(-4.592)
PtB	-0.000421***	-0.000622***	-0.000715***	-0.000803***	-0.000887***
	(-4.152)	(-5.204)	(-5.537)	(-5.988)	(-6.057)
ROA	0.0211***	0.0225***	0.0272***	0.0270***	0.0270***
	(3.592)	(3.378)	(3.736)	(3.448)	(3.235)
Δ ROA	0.00421	0.00845	0.0102	0.0103	0.00394
	(0.543)	(0.972)	(1.019)	(0.969)	(0.345)
Leverage	0.00688***	0.0108***	0.0125***	0.0162***	0.0171***
	(3.853)	(5.297)	(5.593)	(6.761)	(6.712)
Constant	0.00278	0.00928	0.00420	0.00310	0.00881
	(0.379)	(1.223)	(0.460)	(0.291)	(0.791)
N	31,981	31,981	31,981	31,981	31,981
R^2	0.00726	0.0111	0.0142	0.0143	0.0157
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For the dependent variable the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

Table 19. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market model).

Variable	$CAR_{[-1,1];MM}$	$CAR_{[-1,2];MM}$	$CAR_{[-1,3];MM}$	$CAR_{[-1,4];MM}$	$CAR_{[-1,5];MM}$
$\Delta\ Tone_{Net;Full\ Report}$	0.00233	0.00608	0.0586	0.0256	0.0281
	(0.0515)	(0.118)	(1.041)	(0.423)	(0.429)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$	**-0.0786**	**-0.0712**	**-0.0643**	**-0.0789**	**-0.0566**
	(-1.625)	**(-1.293)**	**(-1.066)**	**(-1.237)**	**(-0.833)**
Low ROA Dummy (D)	-0.00113	-0.00177	-0.00109	-0.00145	-0.00159
	(-1.073)	(-1.490)	(-0.846)	(-1.037)	(-1.062)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$***(D)**	**-0.0320**	**-0.118**	**-0.113**	**-0.0622**	**-0.0723**
	(-0.262)	**(-0.831)**	**(-0.740)**	**(-0.387)**	**(-0.420)**
Ln MVE	-0.000254	-0.000724***	-0.000894***	-0.00123***	-0.00159***
	(-1.329)	(-3.249)	(-3.752)	(-4.799)	(-5.776)
PtB	-0.000648***	-0.000900***	-0.00106***	-0.00122***	-0.00137***
	(-6.249)	(-7.204)	(-7.830)	(-8.527)	(-8.510)
ROA	0.0179***	0.0191***	0.0248***	0.0248***	0.0244***
	(3.052)	(2.864)	(3.369)	(3.130)	(2.883)
Δ ROA	-0.0179**	-0.0200**	-0.0261***	-0.0342***	-0.0447***
	(-2.297)	(-2.315)	(-2.617)	(-3.204)	(-3.875)
Leverage	0.00715***	0.0110***	0.0135***	0.0173***	0.0186***
	(4.033)	(5.471)	(6.050)	(7.197)	(7.173)
Constant	0.00148	0.00965	0.01000	0.0114	0.0191*
	(0.211)	(1.286)	(1.102)	(1.064)	(1.718)
N	31,981	31,981	31,981	31,981	31,981
R^2	0.00943	0.0139	0.0184	0.0190	0.0200
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For the dependent variable the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

Table 20. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (4-factor model).

Variable	$CAR_{[-1,1];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,3];4\ Factor}$	$CAR_{[-1,4];4\ Factor}$	$CAR_{[-1,5];4\ Factor}$
$\Delta\ Tone_{Net;Full\ Report}$	-0.00466	0.00632	0.0283	0.0164	0.0345
	(-0.103)	(0.124)	(0.508)	(0.271)	(0.529)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$	**-0.0826***	**-0.0612**	**-0.0424**	**-0.0191**	**0.0177**
	(-1.699)	**(-1.107)**	**(-0.707)**	**(-0.297)**	**(0.257)**
Low ROA Dummy (D)	-0.000695	-0.00120	-0.000271	-0.000932	-0.000737
	(-0.664)	(-1.014)	(-0.213)	(-0.674)	(-0.495)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$ ***(D)**	**-0.0268**	**-0.127**	**-0.156**	**-0.149**	**-0.154**
	(-0.219)	**(-0.875)**	**(-1.015)**	**(-0.915)**	**(-0.900)**
Ln MVE	-0.000128	-0.000507**	-0.000666***	-0.00111***	-0.00138***
	(-0.670)	(-2.339)	(-2.876)	(-4.459)	(-5.229)
PtB	-0.000438***	-0.000602***	-0.000705***	-0.000785***	-0.000914***
	(-4.610)	(-5.233)	(-5.953)	(-5.992)	(-6.313)
ROA	0.0121**	0.0123*	0.0160**	0.0130	0.0153*
	(2.064)	(1.822)	(2.210)	(1.625)	(1.834)
Δ ROA	-0.0138*	-0.0104	-0.0185*	-0.0258**	-0.0340***
	(-1.738)	(-1.182)	(-1.895)	(-2.401)	(-2.976)
Leverage	0.00526***	0.00769***	0.00974***	0.0133***	0.0139***
	(3.022)	(3.903)	(4.460)	(5.656)	(5.507)
Constant	-0.00207	0.00906	0.00788	0.0109	0.0194*
	(-0.317)	(1.301)	(0.881)	(1.013)	(1.823)
N	31,981	31,981	31,981	31,981	31,981
R^2	0.00430	0.00625	0.00710	0.00837	0.00900
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For the dependent variable the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

3.5 Robustness tests

3.5.1 Sample selection

To test whether the empirical results of this study are robust, a series of robustness tests was conducted. As a first robustness test, the study applied additional and more restrictive data filters to the data sample and overall sample selection process. From the selected sample, firm-year observations of the financial industry were excluded (SIC codes 6021-6799). Substantial accounting differences between financial and non-financial firms as well as differences in their regulatory framework might affect the empirical results (Ditter and Scherr 2015, 60). In addition, the study removed firms with a low market capitalization (10 percent quantile of market capitalization) from the data set.

By eliminating low-priced firms, the study addressed concerns that the regression results were influenced by bid-ask bounces around the filing date of an annual report while investigating the market reaction to the disclosure tone in the footnotes sections of annual reports (Loughran and McDonald 2011a, 40; Loughran and McDonald 2015, 3). Applying these additional and more restrictive data filters to the data sample, the analysis was rerun.

The regression results did not confirm the prior findings. The regression coefficient on $\Delta\, Tone_{Net;Footnotes}$ is insignificantly negative. Using this restricted data sample, the analysis suggests that market participants do not seem to react to more pessimistic disclosures in the notes to the financial statements. The observed under reaction by investors indicates that managers might be able to hide negative information in the footnotes sections of annual reports on Form 10-K (management obfuscation by disclosure outlet). Interestingly, the regression coefficient on $\Delta\, Tone_{Net;Footnotes} *\ Low\ ROA\ Dummy$ is positive. This result implies that managers of low profitable companies might use positive statements to frame negative news in the footnotes sections (management obfuscation by word choice). However, the regression result is insignificant. Table 21 to Table 23 present the regression results.

Table 21. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market-adjusted model) excluding firm-year observation of the financial industry and low market capitalization.

Variable	$CAR_{[-1,1];MAM}$	$CAR_{[-1,2];MAM}$	$CAR_{[-1,3];MAM}$	$CAR_{[-1,4];MAM}$	$CAR_{[-1,5];MAM}$
$\Delta\ Tone_{Net;Full\ Report}$	-0.0149	0.0115	0.0656	0.00167	0.00962
	(-0.300)	(0.204)	(1.062)	(0.0254)	(0.135)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$	**-0.0682**	**-0.0374**	**-0.0432**	**-0.0452**	**-0.0538**
	(-1.287)	**(-0.621)**	**(-0.647)**	**(-0.652)**	**(-0.716)**
Low ROA Dummy (D)	-0.000824	-0.00145	-0.000601	-0.00155	-0.00184
	(-0.698)	(-1.086)	(-0.415)	(-0.991)	(-1.112)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$***(D)**	**0.00269**	**-0.0258**	**-0.0692**	**-0.0995**	**-0.0959**
	(0.0205)	**(-0.168)**	**(-0.420)**	**(-0.576)**	**(-0.516)**
Ln MVE	-5.97e-05	-0.000306	-0.000367	-0.000612**	-0.000829***
	(-0.273)	(-1.191)	(-1.321)	(-2.051)	(-2.669)
PtB	-0.000520***	-0.000774***	-0.000883***	-0.000979***	-0.00108***
	(-4.769)	(-6.025)	(-6.217)	(-6.538)	(-6.758)
ROA	0.0196***	0.0214***	0.0258***	0.0235***	0.0236**
	(3.040)	(2.924)	(3.167)	(2.675)	(2.551)
Δ ROA	0.00932	0.0144	0.0148	0.0166	0.0137
	(1.065)	(1.466)	(1.307)	(1.362)	(1.046)
Leverage	0.00433**	0.00911***	0.0116***	0.0154***	0.0154***
	(2.157)	(4.023)	(4.626)	(5.709)	(5.474)
Constant	0.000235	0.00479	0.000996	0.000629	0.00520
	(0.0315)	(0.611)	(0.105)	(0.0578)	(0.457)
N	25,010	25,010	25,010	25,010	25,010
R^2	0.00849	0.0133	0.0159	0.0157	0.0172
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words).

Table 22. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market model) excluding firm-year observation of the financial industry and low market capitalization.

Variable	$CAR_{[-1,1];MM}$	$CAR_{[-1,2];MM}$	$CAR_{[-1,3];MM}$	$CAR_{[-1,4];MM}$	$CAR_{[-1,5];MM}$
Δ $Tone_{Net;Full\ Report}$	0.00360	0.0296	0.0893	0.0354	0.0508
	(0.0732)	(0.531)	(1.457)	(0.542)	(0.710)
Δ $\mathbf{Tone_{Net;Footnotes}}$	**-0.0556**	**-0.0277**	**-0.0260**	**-0.0344**	**-0.0252**
	(-1.047)	**(-0.464)**	**(-0.387)**	**(-0.493)**	**(-0.331)**
Low ROA Dummy (D)	-0.00125	-0.00183	-0.00111	-0.00204	-0.00228
	(-1.069)	(-1.387)	(-0.773)	(-1.316)	(-1.379)
Δ $\mathbf{Tone_{Net;Footnotes}}$*(D)	**0.0505**	**0.0304**	**-0.0147**	**-0.0235**	**-0.0243**
	(0.384)	**(0.201)**	**(-0.0888)**	**(-0.135)**	**(-0.130)**
Ln MVE	-7.46e-05	-0.000359	-0.000531*	-0.000841***	-0.00111***
	(-0.342)	(-1.414)	(-1.905)	(-2.823)	(-3.531)
PtB	-0.000774***	-0.00109***	-0.00128***	-0.00147***	-0.00165***
	(-7.017)	(-8.545)	(-8.938)	(-9.558)	(-9.897)
ROA	0.0173***	0.0178**	0.0242***	0.0227**	0.0231**
	(2.710)	(2.463)	(2.958)	(2.574)	(2.471)
Δ ROA	-0.0147*	-0.0160	-0.0226**	-0.0277**	-0.0359***
	(-1.678)	(-1.641)	(-2.013)	(-2.299)	(-2.741)
Leverage	0.00539***	0.0103***	0.0137***	0.0180***	0.0189***
	(2.714)	(4.607)	(5.492)	(6.693)	(6.677)
Constant	-0.000913	0.00463	0.00405	0.00651	0.0129
	(-0.125)	(0.593)	(0.424)	(0.587)	(1.124)
N	25,010	25,010	25,010	25,010	25,010
R^2	0.0120	0.0174	0.0221	0.0226	0.0236
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words).

Table 23. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (4-factor model) excluding firm-year observation of the financial industry and low market capitalization.

Variable	$CAR_{[-1,1];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,3];4\ Factor}$	$CAR_{[-1,4];4\ Factor}$	$CAR_{[-1,5];4\ Factor}$
$\Delta\ Tone_{Net;Full\ Report}$	-0.00148	0.0335	0.0720	0.0350	0.0552
	(-0.0304)	(0.613)	(1.210)	(0.539)	(0.785)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$	**-0.0608**	**-0.0211**	**0.000632**	**0.0180**	**0.0308**
	(-1.152)	**(-0.352)**	**(0.00960)**	**(0.258)**	**(0.402)**
Low ROA Dummy (D)	-0.000951	-0.00129	-0.000788	-0.00167	-0.00240
	(-0.819)	(-0.992)	(-0.561)	(-1.108)	(-1.476)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$***(D)**	**0.0592**	**0.00461**	**-0.0680**	**-0.0923**	**-0.0725**
	(0.458)	**(0.0307)**	**(-0.417)**	**(-0.539)**	**(-0.396)**
Ln MVE	0.000134	-7.62e-05	-0.000206	-0.000648**	-0.000756**
	(0.619)	(-0.306)	(-0.763)	(-2.250)	(-2.479)
PtB	-0.000515***	-0.000741***	-0.000834***	-0.000938***	-0.00108***
	(-5.094)	(-6.216)	(-6.507)	(-6.646)	(-7.040)
ROA	0.0102	0.00953	0.0126	0.00899	0.0109
	(1.602)	(1.319)	(1.597)	(1.017)	(1.178)
Δ ROA	-0.0105	-0.00595	-0.0132	-0.0167	-0.0222*
	(-1.192)	(-0.604)	(-1.207)	(-1.385)	(-1.714)
Leverage	0.00309	0.00644***	0.00925***	0.0128***	0.0127***
	(1.580)	(2.925)	(3.772)	(4.836)	(4.521)
Constant	-0.00364	0.00233	0.00213	0.00561	0.0112
	(-0.544)	(0.313)	(0.227)	(0.496)	(1.011)
N	25,010	25,010	25,010	25,010	25,010
R^2	0.00548	0.00731	0.00816	0.00916	0.0102
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words).

3.5.2 Disclosure length

The results could be influenced by how the overall text length of financial statements was measured in the study (Loughran and McDonald business dictionary excluding stop words). Loughran and McDonald (2011a) and Loughran and McDonald (2014) originally counted the total number of words (more than two characters) appearing in an annual report (including stop words) to determine the entire text length of a specific Form 10-K filing (Loughran and McDonald 2011b, 2; Loughran and McDonald 2014, 1669). To examine whether the results were sensitive to the document length of an annual report, a modified word count for the overall text length of a Form 10-K filing and its footnotes section was obtained (Loughran and McDonald dictionary including stop words) for each submission in the data sample, and the analysis was rerun.

The regression results can only confirm the findings to a certain extent. Based on the Loughran and McDonald business dictionary (including stop words), more pessimistic footnote disclosures generally relate to a lower abnormal stock return around the filing date of an annual report. However, most regression results for the different holding periods are once again insignificant underscoring the potential for managers to hide negative information (management obfuscation by disclosure outlet). However, the results again do not support the notion that managers use positive statements in the notes to the financial statements to frame negative corporate information (management obfuscation by word choice). Table 24 to Table 26 present the regression results.

Table 24. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (marked-adjusted model) using the Loughran and McDonald business dictionary including stop words.

Variable	$CAR_{[-1,1];MAM}$	$CAR_{[-1,2];MAM}$	$CAR_{[-1,3];MAM}$	$CAR_{[-1,4];MAM}$	$CAR_{[-1,5];MAM}$
$\Delta\ Tone_{Net;Full\ Report}$	-0.0447	-0.0245	0.0718	-0.00711	-0.00204
	(-0.499)	(-0.239)	(0.647)	(-0.0602)	(-0.0159)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$	**-0.175****	**-0.148**	**-0.154**	**-0.171**	**-0.171**
	(-2.015)	**(-1.471)**	**(-1.417)**	**(-1.496)**	**(-1.415)**
Low ROA Dummy (D)	-0.000714	-0.00150	-0.000857	-0.00118	-0.00129
	(-0.672)	(-1.244)	(-0.659)	(-0.839)	(-0.852)
$\Delta\ Tone_{Net;Footnotes}$*(D)	-0.0614	-0.246	-0.237	-0.202	-0.205
	(-0.277)	(-0.929)	(-0.844)	(-0.692)	(-0.653)
Ln MVE	-0.000297	-0.000689***	-0.000667***	-0.000953***	-0.00125***
	(-1.546)	(-3.075)	(-2.804)	(-3.730)	(-4.594)
PtB	-0.000421***	-0.000622***	-0.000715***	-0.000803***	-0.000887***
	(-4.152)	(-5.204)	(-5.537)	(-5.988)	(-6.058)
ROA	0.0211***	0.0225***	0.0273***	0.0271***	0.0271***
	(3.593)	(3.381)	(3.739)	(3.451)	(3.238)
Δ ROA	0.00423	0.00850	0.0103	0.0104	0.00401
	(0.545)	(0.979)	(1.026)	(0.975)	(0.352)
Leverage	0.00689***	0.0108***	0.0125***	0.0162***	0.0171***
	(3.854)	(5.299)	(5.594)	(6.763)	(6.714)
Constant	0.00278	0.00928	0.00421	0.00310	0.00882
	(0.378)	(1.223)	(0.461)	(0.291)	(0.791)
N	31,981	31,981	31,981	31,981	31,981
R^2	0.00725	0.0111	0.0142	0.0143	0.0157
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (including stop words).

Table 25. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (marked model) using the Loughran and McDonald business dictionary including stop words.

Variable	$CAR_{[-1,1];MM}$	$CAR_{[-1,2];MM}$	$CAR_{[-1,3];MM}$	$CAR_{[-1,4];MM}$	$CAR_{[-1,5];MM}$
Δ $Tone_{Net;Full\ Report}$	0.00759	0.0250	0.135	0.0718	0.0834
	(0.0854)	(0.248)	(1.219)	(0.607)	(0.648)
Δ $Tone_{Net;Footnotes}$	**-0.139**	**-0.121**	**-0.107**	**-0.143**	**-0.107**
	(-1.593)	**(-1.212)**	**(-0.978)**	**(-1.240)**	**(-0.872)**
Low ROA Dummy (D)	-0.00114	-0.00178	-0.00109	-0.00146	-0.00160
	(-1.080)	(-1.497)	(-0.848)	(-1.041)	(-1.065)
Δ $Tone_{Net;Footnotes}$*(D)	**-0.0478**	**-0.212**	**-0.223**	**-0.116**	**-0.140**
	(-0.215)	**(-0.821)**	**(-0.799)**	**(-0.399)**	**(-0.447)**
Ln MVE	-0.000254	-0.000725***	-0.000895***	-0.00123***	-0.00159***
	(-1.330)	(-3.252)	(-3.755)	(-4.801)	(-5.778)
PtB	-0.000648***	-0.000900***	-0.00106***	-0.00122***	-0.00137***
	(-6.248)	(-7.203)	(-7.830)	(-8.527)	(-8.511)
ROA	0.0179***	0.0191***	0.0249***	0.0248***	0.0245***
	(3.053)	(2.866)	(3.372)	(3.133)	(2.885)
Δ ROA	-0.0178**	-0.0199**	-0.0260***	-0.0341***	-0.0446***
	(-2.293)	(-2.306)	(-2.607)	(-3.196)	(-3.867)
Leverage	0.00715***	0.0110***	0.0135***	0.0173***	0.0186***
	(4.033)	(5.473)	(6.051)	(7.199)	(7.175)
Constant	0.00148	0.00965	0.0100	0.0114	0.0191*
	(0.210)	(1.287)	(1.103)	(1.065)	(1.719)
N	31,981	31,981	31,981	31,981	31,981
R^2	0.00942	0.0139	0.0184	0.0190	0.0200
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (including stop words).

Table 26. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (4-factor model) using the Loughran and McDonald business dictionary including stop words.

Variable	$CAR_{[-1,1];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,3];4\ Factor}$	$CAR_{[-1,4];4\ Factor}$	$CAR_{[-1,5];4\ Factor}$
$\Delta\ Tone_{Net;Full\ Report}$	-0.00895	0.0177	0.0698	0.0457	0.0898
	(-0.101)	(0.177)	(0.640)	(0.386)	(0.702)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$	**-0.148***	**-0.100**	**-0.0728**	**-0.0359**	**0.0279**
	(-1.682)	**(-1.002)**	**(-0.669)**	**(-0.308)**	**(0.223)**
Low ROA Dummy (D)	-0.000701	-0.00120	-0.000273	-0.000935	-0.000742
	(-0.670)	(-1.018)	(-0.214)	(-0.675)	(-0.498)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$***(D)**	**-0.0370**	**-0.232**	**-0.297**	**-0.278**	**-0.286**
	(-0.166)	**(-0.879)**	**(-1.062)**	**(-0.943)**	**(-0.919)**
Ln MVE	-0.000128	-0.000507**	-0.000667***	-0.00111***	-0.00138***
	(-0.670)	(-2.340)	(-2.878)	(-4.460)	(-5.230)
PtB	-0.000437***	-0.000602***	-0.000705***	-0.000785***	-0.000914***
	(-4.610)	(-5.231)	(-5.953)	(-5.992)	(-6.314)
ROA	0.0121**	0.0123*	0.0160**	0.0130	0.0153*
	(2.064)	(1.824)	(2.213)	(1.627)	(1.835)
Δ ROA	-0.0138*	-0.0103	-0.0185*	-0.0258**	-0.0339***
	(-1.735)	(-1.176)	(-1.889)	(-2.397)	(-2.969)
Leverage	0.00526***	0.00769***	0.00974***	0.0133***	0.0139***
	(3.022)	(3.904)	(4.461)	(5.657)	(5.509)
Constant	-0.00208	0.00907	0.00789	0.0109	0.0194*
	(-0.318)	(1.302)	(0.882)	(1.014)	(1.824)
N	31,981	31,981	31,981	31,981	31,981
R^2	0.00429	0.00623	0.00711	0.00838	0.00901
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (including stop words).

3.5.3 Table content

The results of this study could be further influenced by the fact that HTML tables and their corresponding content were excluded from the analysis. EDGAR filers might use HTML table tags (<Table>) to structure and disclose certain textual information in annual reports on Form 10-K filed with the SEC. Thus, the exclusion of tables in financial statements and their corresponding content might affect the results. To address concerns that tables contain relevant narrative information and complex or technical disclosures affecting investors' behaviour regarding financial markets, tables and their corresponding content were included and the analysis was rerun.

The regression results show that investors react to disclosure tone changes in the notes to the financial statements. When tables and their corresponding content were included, the regression coefficients on $\Delta\ Tone_{Net;Footnotes}$ were significantly negative. This result suggests that managers do not have the ability to hide negative information in the footnotes sections of annual reports (management obfuscation by disclosure outlet). Table 27 to Table 29 present the regression results.

Table 27. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market-adjusted model) including table content.

Variable	$CAR_{[-1,1];MAM}$	$CAR_{[-1,2];MAM}$	$CAR_{[-1,3];MAM}$	$CAR_{[-1,4];MAM}$	$CAR_{[-1,5];MAM}$
$\Delta\ Tone_{Net;Full\ Report}$	-0.0386	-0.0234	0.0290	-0.0201	-0.0257
	(-0.924)	(-0.489)	(0.558)	(-0.362)	(-0.426)
$\Delta\ \mathbf{Tone}_{\mathbf{Net;Footnotes}}$	**-0.102*** **	**-0.0839***	**-0.0813***	**-0.0975***	**-0.0968***
	(-2.583)	**(-1.854)**	**(-1.661)**	**(-1.881)**	**(-1.757)**
Low ROA Dummy (D)	-0.00135	-0.00211**	-0.00225**	-0.00253**	-0.00281**
	(-1.559)	(-2.165)	(-2.100)	(-2.208)	(-2.280)
$\Delta\ \mathbf{Tone}_{\mathbf{Net;Footnotes}}$*(D)	**0.0941**	**0.0157**	**-0.00122**	**0.0845**	**0.0838**
	(1.004)	**(0.143)**	**(-0.0105)**	**(0.703)**	**(0.660)**
Ln MVE	-0.000179	-0.000493***	-0.000599***	-0.000897***	-0.00112***
	(-1.122)	(-2.686)	(-3.013)	(-4.199)	(-5.011)
PtB	-0.000378***	-0.000512***	-0.000582***	-0.000646***	-0.000737***
	(-5.097)	(-5.841)	(-6.120)	(-6.311)	(-6.516)
ROA	0.0168***	0.0161***	0.0188***	0.0196***	0.0182***
	(3.439)	(2.895)	(3.147)	(3.057)	(2.630)
Δ ROA	0.0101	0.0137*	0.0164**	0.0161*	0.0111
	(1.613)	(1.950)	(2.070)	(1.919)	(1.236)
Leverage	0.00673***	0.00930***	0.0109***	0.0129***	0.0138***
	(4.535)	(5.539)	(5.959)	(6.578)	(6.637)
Constant	-0.00117	0.00415	0.00381	0.00569	0.0104
	(-0.201)	(0.661)	(0.531)	(0.648)	(1.150)
N	46,441	46,441	46,441	46,441	46,441
R^2	0.00613	0.00825	0.0109	0.0115	0.0131
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words).

Table 28. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market-adjusted model) including table content.

Variable	$CAR_{[-1,1];MM}$	$CAR_{[-1,2];MM}$	$CAR_{[-1,3];MM}$	$CAR_{[-1,4];MM}$	$CAR_{[-1,5];MM}$
Δ $Tone_{Net;Full Report}$	-0.00443	0.00858	0.0700	0.0322	0.0372
	(-0.107)	(0.182)	(1.355)	(0.581)	(0.618)
Δ $Tone_{Net;Footnotes}$	**-0.0864****	**-0.0758***	**-0.0630**	**-0.0845**	**-0.0704**
	(-2.200)	**(-1.682)**	**(-1.275)**	**(-1.606)**	**(-1.258)**
Low ROA Dummy (D)	-0.00144*	-0.00203**	-0.00202*	-0.00229**	-0.00245**
	(-1.677)	(-2.117)	(-1.920)	(-2.010)	(-2.001)
Δ $Tone_{Net;Footnotes}$*(D)	**0.0878**	**0.0173**	**-0.0135**	**0.0944**	**0.0650**
	(0.952)	**(0.162)**	**(-0.117)**	**(0.788)**	**(0.513)**
Ln MVE	-0.000189	-0.000588***	-0.000836***	-0.00118***	-0.00147***
	(-1.184)	(-3.233)	(-4.230)	(-5.548)	(-6.512)
PtB	-0.000577***	-0.000762***	-0.000897***	-0.00103***	-0.00117***
	(-7.562)	(-8.285)	(-8.948)	(-9.306)	(-9.353)
ROA	0.0152***	0.0151***	0.0189***	0.0196***	0.0189***
	(3.106)	(2.749)	(3.175)	(3.037)	(2.735)
Δ ROA	-0.0108*	-0.0137**	-0.0183**	-0.0262***	-0.0347***
	(-1.715)	(-1.972)	(-2.319)	(-3.115)	(-3.834)
Leverage	0.00698***	0.00966***	0.0117***	0.0138***	0.0149***
	(4.749)	(5.801)	(6.460)	(7.009)	(7.088)
Constant	-0.00151	0.00547	0.00970	0.0134	0.0200**
	(-0.263)	(0.867)	(1.320)	(1.488)	(2.167)
N	46,441	46,441	46,441	46,441	46,441
R^2	0.00765	0.0104	0.0140	0.0150	0.0166
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words).

Table 29. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (4-factor model) including table content.

Variable	$CAR_{[-1,1];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,3];4\ Factor}$	$CAR_{[-1,4];4\ Factor}$	$CAR_{[-1,5];4\ Factor}$
$\Delta\ Tone_{Net;Full\ Report}$	-0.00699	0.00951	0.0442	0.0289	0.0435
	(-0.168)	(0.203)	(0.864)	(0.519)	(0.725)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$	**-0.0899****	**-0.0783***	**-0.0584**	**-0.0685**	**-0.0294**
	(-2.280)	**(-1.750)**	**(-1.205)**	**(-1.320)**	**(-0.524)**
Low ROA Dummy (D)	-0.000988	-0.00160*	-0.00144	-0.00208*	-0.00184
	(-1.152)	(-1.667)	(-1.386)	(-1.854)	(-1.531)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$***(D)**	**0.0781**	**-0.0197**	**-0.0330**	**0.0606**	**-0.0215**
	(0.849)	**(-0.185)**	**(-0.294)**	**(0.506)**	**(-0.170)**
Ln MVE	-8.05e-05	-0.000368**	-0.000606***	-0.00107***	-0.00125***
	(-0.502)	(-2.052)	(-3.134)	(-5.119)	(-5.685)
PtB	-0.000442***	-0.000575***	-0.000680***	-0.000748***	-0.000870***
	(-6.240)	(-6.761)	(-7.694)	(-7.450)	(-7.796)
ROA	0.0115**	0.0103*	0.0136**	0.0111*	0.0118*
	(2.349)	(1.842)	(2.329)	(1.733)	(1.744)
Δ ROA	-0.00880	-0.00679	-0.0129*	-0.0213**	-0.0271***
	(-1.372)	(-0.958)	(-1.645)	(-2.487)	(-3.007)
Leverage	0.00570***	0.00711***	0.00925***	0.0113***	0.0118***
	(3.936)	(4.354)	(5.204)	(5.846)	(5.729)
Constant	-0.00493	0.00374	0.00587	0.0127	0.0196**
	(-0.900)	(0.625)	(0.820)	(1.383)	(2.197)
N	46,441	46,441	46,441	46,441	46,441
R^2	0.00401	0.00514	0.00619	0.00733	0.00801
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words).

3.6 Sensitivity tests

3.6.1 Type of disclosure tone

To validate the empirical results of this study, the stock market reaction to changes in the disclosure tone was examined using different word lists to quantify the textual sentiment in financial statements and the footnotes sections of annual reports. If company officials do not have the ability to hide negative news in the footnotes sections of annual reports, other types of disclosure tone in the notes to the financial statements should be associated with a firm's abnormal stock return around its Form 10-K filing in the expected manner. Based on the Loughran and McDonald business dictionary, other types of linguistic tone and the market reaction to tone changes in the footnotes of annual reports were analysed (*"uncertainty"*, *"litigious"*, *"constraining"*).

The regression results again showed that capital market participants do not react to disclosure tone changes in the footnotes sections of annual reports. Despite showing the expected negative sign, a positive change in $\Delta\, Tone_{Unc;Full\ Report}$ (more uncertain information in the footnotes section) is not significantly associated with a lower abnormal stock return around the filing date of an annual report. These results suggest that company officials do have the ability to hide negative corporate information in the footnotes sections of annual reports on Form 10-K (management obfuscation by disclosure outlet). Table 30 to Table 32 present the regression results.

Table 30. Regression coefficients on different disclosure tones (changes) in the footnotes sections for abnormal stock return (market-adjusted model).

Variable	$CAR_{[-1,2];MAM}$	$CAR_{[-1,2];MAM}$	$CAR_{[-1,2];MAM}$
$\Delta\ Tone_{Unc;Full\ Report}$	-0.0755		
	(-0.859)		
$\boldsymbol{\Delta\ Tone_{Unc;Footnotes}}$	**-0.0699**		
	(-1.305)		
$\boldsymbol{\Delta\ Tone_{Unc;Footnotes}*(D)}$	**0.0603**		
	(0.406)		
$\Delta\ Tone_{Lit;Full\ Report}$		-0.00717	
		(-0.400)	
$\boldsymbol{\Delta\ Tone_{Lit;Footnotes}}$		**0.0632***	
		(1.892)	
$\boldsymbol{\Delta\ Tone_{Lit;Footnotes}*(D)}$		**-0.124**	
		(-1.285)	
$\Delta\ Tone_{Con;Full\ Report}$			-0.0349
			(-0.380)
$\boldsymbol{\Delta\ Tone_{Con;Footnotes}}$			**0.0213**
			(0.351)
$\boldsymbol{\Delta\ Tone_{Con;Footnotes}*(D)}$			**-0.0505**
			(-0.334)
Low ROA Dummy (D)	-0.00167	-0.00169	-0.00163
	(-1.396)	(-1.415)	(-1.364)
Ln MVE	-0.000691***	-0.000690***	-0.000694***
	(-3.084)	(-3.077)	(-3.097)
PtB	-0.000619***	-0.000617***	-0.000619***
	(-5.185)	(-5.167)	(-5.181)
ROA	0.0225***	0.0225***	0.0225***
	(3.373)	(3.373)	(3.374)
Δ ROA	0.00904	0.00914	0.00905
	(1.043)	(1.054)	(1.044)
Leverage	0.0108***	0.0108***	0.0108***
	(5.311)	(5.304)	(5.310)
Constant	0.00932	0.00918	0.00936
	(1.224)	(1.206)	(1.230)
N	31,981	31,981	31,981
R^2	0.011	0.011	0.0109
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for uncertain, litigious, and constraining disclosure tones (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For the dependent variable the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

Table 31. Regression coefficients on different disclosure tones (changes) in the footnotes sections for abnormal stock return (market model).

Variable	$CAR_{[-1,2];MM}$	$CAR_{[-1,2];MM}$	$CAR_{[-1,2];MM}$
$\Delta\ Tone_{Unc;Full\ Report}$	-0.0659		
	(-0.753)		
$\Delta\ \mathbf{Tone}_{\mathbf{Unc;Footnotes}}$	**-0.0604**		
	(-1.133)		
$\Delta\ \mathbf{Tone}_{\mathbf{Unc;Footnotes}}$*(D)	**0.0696**		
	(0.472)		
$\Delta\ Tone_{Lit;Full\ Report}$		-0.0114	
		(-0.638)	
$\Delta\ \mathbf{Tone}_{\mathbf{Lit;Footnotes}}$		**0.0623***	
		(1.882)	
$\Delta\ \mathbf{Tone}_{\mathbf{Lit;Footnotes}}$*(D)		**-0.105**	
		(-1.088)	
$\Delta\ Tone_{Con;Full\ Report}$			-0.0168
			(-0.185)
$\Delta\ \mathbf{Tone}_{\mathbf{Con;Footnotes}}$			**0.0189**
			(0.316)
$\Delta\ \mathbf{Tone}_{\mathbf{Con;Footnotes}}$*(D)			**-0.0334**
			(-0.222)
Low ROA Dummy (D)	-0.00192	-0.00193	-0.00188
	(-1.627)	(-1.639)	(-1.598)
Ln MVE	-0.000725***	-0.000723***	-0.000727***
	(-3.254)	(-3.245)	(-3.265)
PtB	-0.000897***	-0.000896***	-0.000897***
	(-7.192)	(-7.173)	(-7.186)
ROA	0.0190***	0.0190***	0.0190***
	(2.856)	(2.856)	(2.857)
Δ ROA	-0.0196**	-0.0196**	-0.0196**
	(-2.279)	(-2.273)	(-2.276)
Leverage	0.0111***	0.0110***	0.0111***
	(5.479)	(5.471)	(5.477)
Constant	0.00967	0.00951	0.00970
	(1.285)	(1.265)	(1.290)
N	31,981	31,981	31,981
R^2	0.0138	0.0139	0.0138
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for uncertain, litigious, and constraining disclosure tones (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For the dependent variable the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

Table 32. Regression coefficients on different disclosure tones (changes) in the footnotes sections for abnormal stock return (4-factor model).

Variable	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$
$\Delta\ Tone_{Unc;Full\ Report}$	-0.0791		
	(-0.901)		
$\boldsymbol{\Delta\ Tone_{Unc;Footnotes}}$	**-0.0637**		
	(-1.182)		
$\boldsymbol{\Delta\ Tone_{Unc;Footnotes}*(D)}$	**0.0564**		
	(0.376)		
$\Delta\ Tone_{Lit;Full\ Report}$		-0.00411	
		(-0.228)	
$\boldsymbol{\Delta\ Tone_{Lit;Footnotes}}$		**0.0701****	
		(2.114)	
$\boldsymbol{\Delta\ Tone_{Lit;Footnotes}*(D)}$		**-0.0668**	
		(-0.700)	
$\Delta\ Tone_{Con;Full\ Report}$			0.0755
			(0.825)
$\boldsymbol{\Delta\ Tone_{Con;Footnotes}}$			**0.0194**
			(0.330)
$\boldsymbol{\Delta\ Tone_{Con;Footnotes}*(D)}$			**0.0290**
			(0.198)
Low ROA Dummy (D)	-0.00135	-0.00135	-0.00134
	(-1.149)	(-1.151)	(-1.141)
Ln MVE	-0.000507**	-0.000504**	-0.000509**
	(-2.340)	(-2.327)	(-2.349)
PtB	-0.000599***	-0.000599***	-0.000598***
	(-5.217)	(-5.206)	(-5.207)
ROA	0.0122*	0.0122*	0.0122*
	(1.813)	(1.812)	(1.813)
Δ ROA	-0.0100	-0.0100	-0.00996
	(-1.146)	(-1.142)	(-1.138)
Leverage	0.00770***	0.00770***	0.00769***
	(3.911)	(3.910)	(3.905)
Constant	0.00907	0.00889	0.00909
	(1.299)	(1.274)	(1.304)
N	31,981	31,981	31,981
R^2	0.00618	0.00622	0.00614
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for uncertain, litigious, and constraining disclosure tones (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For the dependent variable the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

3.6.2 Measurement of disclosure tone

Capital market participants might react to disclosure tone changes (time series) given the specific tone level (cross section) of an annual report on Form 10-K and its disclosure outlets based on the idea that tone levels alone do not reflect the importance of new corporate disclosures in regulatory filings (Miwa 2020, 5). Therefore, changes in disclosure tone might affect important stock characteristics (Miwa 2020, 2). To examine whether investors react to disclosure tone changes in the footnotes sections of annual reports in the context of tone levels, the following regression model was estimated using OLS:

$$\begin{aligned} Y_{i;t} = {} & \beta_0 + \beta_1 * \Delta\, Tone_{Net;Full\ Report;i;t} + \beta_2 * \Delta\, Tone_{Net;Footnotes;i;t} \\ & + \beta_3 * Low\ ROA\ Dummy\ (D)_{i;t} + \beta_4 * \Delta\, Tone_{Net;Footnotes;i;t} * D_{i;t} \\ & + \beta_5 * Tone_{Net;Full\ Report;i;t} + \beta_6 * Tone_{Net;Footnotes;i;t} + \beta_7 * Ln\ MVE_{i;t} \\ & + \beta_8 * PtB_{i;t} + \beta_9 * ROA_{i;t} + \beta_{10} * \Delta\, ROA_{i;t} + \beta_{11} * Leverage_{i;t} \\ & + \sum Year\ Fixed\ Effects + \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \tag{18}$$

Where: the dependent variable ($Y_{i;t}$) is a firm's abnormal stock return around its Form 10-K filing. $\Delta\, Tone_{Net;Full\ Report;i;t}$ and $\Delta\, Tone_{Net;Footnotes;i;t}$ represent the change in the overall disclosure tone in the entire annual report and footnotes section of a firm's Form 10-K filing (time series). $Tone_{Net;Full\ Report;i;t}$ and $Tone_{Net;Footnotes;i;t}$ are the overall disclosure tone (ratio of negative minus positive words) in the entire report and footnotes sections of a submission (cross section). The control variables are described in the appendix. *Year Fixed Effects* and *Industry Fixed Effects* represent the included fixed effects.

Despite controlling for the tone level (cross section) of the overall annual report and in the footnotes section, the regression coefficients for $\Delta\, Tone_{Net;Footnotes}$ for all holding periods are insignificantly negative (market model and 4-factor model). This result suggests that market participants do underreact to disclosure tone changes in the footnotes sections of annual reports, regardless of the tone level of the overall filing and notes to the financial statements (management obfuscation by disclosure outlet). Again, the regression results do not support the idea that managers frame negative information with positive statements in this section of the Form 10-K filing (management obfuscation by word choice). Table 33 to Table 35 present the regression results.

Table 33. Regression coefficients on net disclosure tone changes in the footnotes sections for abnormal stock return (market-adjusted model) after controlling for net tone levels.

Variable	$CAR_{[-1,1];MAM}$	$CAR_{[-1,2];MAM}$	$CAR_{[-1,3];MAM}$	$CAR_{[-1,4];MAM}$	$CAR_{[-1,5];MAM}$
$\Delta\ Tone_{Net;Full\ Report}$	-0.0577	-0.0444	0.0236	-0.0101	-0.00858
	(-1.140)	(-0.767)	(0.376)	(-0.151)	(-0.118)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$	**-0.0895***	**-0.0856**	**-0.0859**	**-0.0917**	**-0.0904**
	(-1.714)	**(-1.422)**	**(-1.323)**	**(-1.335)**	**(-1.242)**
Low ROA Dummy (D)	-0.000853	-0.00163	-0.000873	-0.00117	-0.00126
	(-0.801)	(-1.345)	(-0.667)	(-0.824)	(-0.834)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}*(D)}$	**-0.0354**	**-0.134**	**-0.115**	**-0.103**	**-0.103**
	(-0.291)	**(-0.925)**	**(-0.750)**	**(-0.639)**	**(-0.599)**
$Tone_{Net;Full\ Report}$	0.0729	0.0587	0.0132	-0.00107	-0.00575
	(1.632)	(1.122)	(0.232)	(-0.0173)	(-0.0859)
$Tone_{Net;Footnotes}$	-0.0206	-0.00443	-0.0117	-0.00942	-0.00676
	(-0.563)	(-0.105)	(-0.264)	(-0.197)	(-0.136)
Ln MVE	-0.000302	-0.000696***	-0.000666***	-0.000951***	-0.00125***
	(-1.575)	(-3.106)	(-2.799)	(-3.720)	(-4.585)
PtB	-0.000415***	-0.000617***	-0.000714***	-0.000803***	-0.000887***
	(-4.093)	(-5.169)	(-5.531)	(-5.987)	(-6.055)
ROA	0.0212***	0.0227***	0.0273***	0.0270***	0.0270***
	(3.617)	(3.398)	(3.733)	(3.442)	(3.227)
Δ ROA	0.00386	0.00810	0.0102	0.0104	0.00401
	(0.496)	(0.930)	(1.015)	(0.972)	(0.351)
Leverage	0.00672***	0.0106***	0.0125***	0.0162***	0.0171***
	(3.760)	(5.210)	(5.579)	(6.749)	(6.703)
Constant	0.00236	0.00893	0.00412	0.00311	0.00885
	(0.321)	(1.174)	(0.451)	(0.291)	(0.793)
N	31,981	31,981	31,981	31,981	31,981
R^2	0.00729	0.0111	0.0141	0.0142	0.0156
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change and level) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For the dependent variable the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic (p < 1%). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

Table 34. Regression coefficients for the disclosure tone changes in the footnotes sections for abnormal stock return (market model) after controlling for net tone levels.

Variable	$CAR_{[-1,1];MM}$	$CAR_{[-1,2];MM}$	$CAR_{[-1,3];MM}$	$CAR_{[-1,4];MM}$	$CAR_{[-1,5];MM}$
$\Delta\ Tone_{Net;Full\ Report}$	-0.0271	-0.0190	0.0488	0.0215	0.0235
	(-0.541)	(-0.334)	(0.780)	(0.322)	(0.323)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$	**-0.0718**	**-0.0710**	**-0.0572**	**-0.0758**	**-0.0564**
	(-1.376)	**(-1.192)**	**(-0.877)**	**(-1.097)**	**(-0.767)**
Low ROA Dummy (D)	-0.00126	-0.00189	-0.00111	-0.00146	-0.00161
	(-1.186)	(-1.586)	(-0.862)	(-1.040)	(-1.075)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}*(D)}$	**-0.0307**	**-0.118**	**-0.112**	**-0.0616**	**-0.0722**
	(-0.252)	**(-0.831)**	**(-0.731)**	**(-0.383)**	**(-0.420)**
$Tone_{Net;Full\ Report}$	0.0613	0.0526	0.0202	0.00832	0.00982
	(1.384)	(1.014)	(0.354)	(0.134)	(0.145)
$Tone_{Net;Footnotes}$	-0.0149	-0.00197	-0.0142	-0.00618	-0.000603
	(-0.415)	(-0.0480)	(-0.325)	(-0.131)	(-0.0122)
Ln MVE	-0.000259	-0.000731***	-0.000894***	-0.00123***	-0.00159***
	(-1.356)	(-3.280)	(-3.751)	(-4.796)	(-5.781)
PtB	-0.000643***	-0.000896***	-0.00106***	-0.00122***	-0.00137***
	(-6.199)	(-7.176)	(-7.821)	(-8.520)	(-8.500)
ROA	0.0181***	0.0192***	0.0249***	0.0248***	0.0245***
	(3.075)	(2.883)	(3.370)	(3.129)	(2.883)
Δ ROA	-0.0182**	-0.0203**	-0.0261***	-0.0342***	-0.0448***
	(-2.331)	(-2.349)	(-2.617)	(-3.198)	(-3.871)
Leverage	0.00701***	0.0109***	0.0135***	0.0173***	0.0186***
	(3.946)	(5.384)	(6.027)	(7.170)	(7.141)
Constant	0.00112	0.00934	0.00988	0.0113	0.0190*
	(0.159)	(1.242)	(1.088)	(1.059)	(1.709)
N	31,981	31,981	31,981	31,981	31,981
R^2	0.00943	0.0139	0.0183	0.0189	0.0199
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change and level) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For the dependent variable the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

Table 35. Regression coefficients for the disclosure tone changes in the footnotes sections for abnormal stock return (4-factor model) after controlling for net tone levels.

Variable	$CAR_{[-1,1];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,3];4\ Factor}$	$CAR_{[-1,4];4\ Factor}$	$CAR_{[-1,5];4\ Factor}$
$\Delta\ Tone_{Net;Full\ Report}$	-0.0321	-0.0205	0.0381	0.0178	0.0447
	(-0.637)	(-0.362)	(0.613)	(0.265)	(0.620)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$	**-0.0843**	**-0.0628**	**-0.0571**	**-0.0366**	**0.00290**
	(-1.602)	**(-1.051)**	**(-0.881)**	**(-0.526)**	**(0.0390)**
Low ROA Dummy (D)	-0.000834	-0.00133	-0.000267	-0.000978	-0.000732
	(-0.794)	(-1.124)	(-0.208)	(-0.704)	(-0.490)
$\boldsymbol{\Delta\ Tone_{Net;Footnotes}}$***(D)**	**-0.0270**	**-0.127**	**-0.159**	**-0.152**	**-0.157**
	(-0.221)	**(-0.877)**	**(-1.033)**	**(-0.934)**	**(-0.915)**
$Tone_{Net;Full\ Report}$	0.0575	0.0562	-0.0198	-0.00210	-0.0205
	(1.293)	(1.099)	(-0.353)	(-0.0346)	(-0.315)
$Tone_{Net;Footnotes}$	0.00141	0.00134	0.0284	0.0333	0.0288
	(0.0396)	(0.0332)	(0.658)	(0.720)	(0.594)
Ln MVE	-0.000136	-0.000515**	-0.000669***	-0.00112***	-0.00139***
	(-0.710)	(-2.373)	(-2.886)	(-4.481)	(-5.238)
PtB	-0.000433***	-0.000598***	-0.000707***	-0.000785***	-0.000915***
	(-4.560)	(-5.199)	(-5.963)	(-5.984)	(-6.313)
ROA	0.0123**	0.0125*	0.0160**	0.0130	0.0153*
	(2.093)	(1.846)	(2.209)	(1.632)	(1.831)
Δ ROA	-0.0142*	-0.0107	-0.0186*	-0.0260**	-0.0340***
	(-1.780)	(-1.222)	(-1.894)	(-2.409)	(-2.969)
Leverage	0.00510***	0.00753***	0.00975***	0.0133***	0.0139***
	(2.926)	(3.811)	(4.456)	(5.618)	(5.499)
Constant	-0.00241	0.00873	0.00799	0.0109	0.0195*
	(-0.368)	(1.249)	(0.889)	(1.008)	(1.826)
N	31,981	31,981	31,981	31,981	31,981
R^2	0.00430	0.00623	0.00705	0.00833	0.00895
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change and level) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For the dependent variable the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

3.7 Additional tests

Again, despite being more technical than other sections (less read by investors) (Clatworthy and Jones 2001, 314; Sutton et al. 2009, 3-6; Bedard et al. 2012, 25-26; Amel-Zadeh and Faasse 2016, 4, 10-11; Czerney et al. 2017, 2) *"Item 8"* the notes to the financial statements contains important (additional) information regarding a firm's financial performance and future risks (Leder 2003, 17; Heidari and Felden 2015, 1; Amel-Zadeh and Faasse 2016, 25; Czerney et al. 2017, 8-9; Lee 2012, 1159). However, besides narrative disclosures explaining a firm's current financial performance the *"Footnotes"* section might also contains important information regarding the future profitability of a firm not being displayed in the current financial figures. For instance, information about acquisitions and divestitures being made or being planned might not only explain the present financial situation of a firm but also its future corporate profitability. The same holds true for information about corporate debts or financing costs going forward. Furthermore, information about potential future risks such as commitments and contingencies not currently affecting a firm's corporate earnings might be predictive for the profitability of a company in the future.

Assuming that company officials try to obfuscate (substantial) negative corporate information in financial statements by using the footnotes section as a disclosure outlet (*Hypotheses 3a and 3b*) one would expect that a positive change in $\Delta\, Tone_{Net;Footnotes}$ (more pessimistic disclosure tone in the notes to the financial statements) is significantly negatively associated with a firm's future accounting profitability (regardless of how market participants react to tone changes in this section). Adding to this notion, if company officials try to shift (incremental) negative information content from other prominent report sections (e.g. *"MD&A"*) to the footnotes section (Li 2008, 236), one would expect that a positive change in $\Delta\, Tone_{Net;Footnotes}$ demonstrates a stronger negative association with a firm's future return on assets than a positive change in $\Delta\, Tone_{Net;Full\ Report}$ (more pessimistic disclosure tone in the entire Form 10-K submission).

Besides choosing the notes to the financial statements as a particular disclosure outlet in the annual report on Form 10-K to display pessimistic information (management obfuscation by disclosure outlet), managers might also use positive statements to frame negative corporate disclosures in the footnotes section (management obfuscation by word choice) (*Hypotheses 4a and 4b*). If true, one would expect that a negative change in $\Delta\, Tone_{Net;Footnotes}$ (more optimistic disclosure tone in the notes to the financial statements) is negatively associated with a firm's future accounting profitability (regardless of how market participants react to tone changes in this section).

However, the regression results suggest the opposite. Despite showing the expected negative sign, the regression coefficient on $\Delta\, Tone_{Net;Footnotes}$ is insignificant. A more negative disclosure tone in the notes to the financial statements is not predictive for lower corporate earnings in the future (management obfuscation by disclosure outlet). In contrast, the regression coefficient on $\Delta\, Tone_{Net;Full\ Report}$ is significantly negative and therefore predictive for a lower return on assets in the future. Interestingly, a positive change in $\Delta\, Tone_{Net;Footnotes}$ (more pessimistic disclosure tone in the notes to the financial statements) for low profitable firms is positively associated with corporate earnings in the future. However, a detailed analysis of the overall or net disclosure tone (negative tone, positive tone) again does not provide empirical evidence that managers frame negative information in the footnotes sections of annual reports (management obfuscation by word choice). Overall, the results suggest that the notes to the financial statements do not contain incremental information at least in the context of future corporate earnings and that managers do not hide or frame negative information in the footnotes sections of annual reports on Form 10-K. Table 36 presents the regression results.

Table 36. Regression coefficients on different disclosure tones in the footnotes sections for return on assets and change in returns on assets.

Panel A: Future return on assets			
Variable	ROA_{t+1}	ROA_{t+1}	ROA_{t+1}
$\Delta\ Tone_{Net;Full\ Report}$	-0.0957**		
	(-2.353)		
$\Delta\ \mathbf{Tone_{Net;Footnotes}}$	**-0.0165**		
	(-0.445)		
$\Delta\ \mathbf{Tone_{Net;Footnotes}}$ * (D)	**0.303*** **		
	(2.590)		
$\Delta\ Tone_{Neg;Full\ Report}$		-0.108**	
		(-2.462)	
$\Delta\ \mathbf{Tone_{Neg;Footnotes}}$		**-0.0229**	
		(-0.724)	
$\Delta\ \mathbf{Tone_{Neg;Footnotes}}$ * (D)		**0.243** **	
		(2.373)	
$\Delta\ Tone_{Pos;Full\ Report}$			0.0450
			(0.356)
$\Delta\ \mathbf{Tone_{Pos;Footnotes}}$			**-0.0416**
			(-0.537)
$\Delta\ \mathbf{Tone_{Pos;Footnotes}}$ * (D)			**0.0400**
			(0.155)
Low ROA Dummy (D)	0.00473***	0.00209***	0.00493***
	(4.754)	(10.03)	(4.993)
Ln MVE	0.00209***	-0.000887***	0.00208***
	(10.02)	(-5.273)	(10.01)
PtB	-0.000887***	0.800***	-0.000889***
	(-5.270)	(91.83)	(-5.278)
ROA	0.800***	-0.0211**	0.800***
	(91.85)	(-2.198)	(91.86)
Δ ROA	-0.0211**	0.0221***	-0.0209**
	(-2.192)	(10.75)	(-2.173)
Leverage	0.0222***	0.00209***	0.0222***
	(10.76)	(10.03)	(10.77)
Constant	-0.0333***	-0.0331***	-0.0332***
	(-5.286)	(-5.259)	(-5.246)
N	33,023	33,023	33,023
R^2	0.742	0.742	0.742
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Change in future return on assets			
Variable	Δ ROA_{t+1}	Δ ROA_{t+1}	Δ ROA_{t+1}
Δ $Tone_{Net;Full\ Report}$	-0.0549		
	(-1.421)		
Δ $Tone_{Net;Footnotes}$	**-0.0522**		
	(-1.400)		
Δ $Tone_{Net;Footnotes}$ * (D)	**0.323*** **		
	(2.870)		
Δ $Tone_{Neg;Full\ Report}$		-0.0697*	
		(-1.693)	
Δ $Tone_{Neg;Footnotes}$		**-0.0471**	
		(-1.500)	
Δ $Tone_{Neg;Footnotes}$ * (D)		**0.270*** **	
		(2.742)	
Δ $Tone_{Pos;Full\ Report}$			-0.0178
			(-0.149)
Δ $Tone_{Pos;Footnotes}$			**-0.0274**
			(-0.361)
Δ $Tone_{pos;Footnotes}$ * (D)			**0.103**
			(0.417)
Low ROA Dummy (D)	0.00694***	0.00693***	0.00715***
	(6.961)	(6.945)	(7.199)
Ln MVE	0.00164***	0.00164***	0.00163***
	(7.647)	(7.659)	(7.626)
PtB	-0.000728***	-0.000728***	-0.000729***
	(-2.971)	(-2.971)	(-2.975)
ROA	-0.160***	-0.160***	-0.160***
	(-16.63)	(-16.63)	(-16.61)
Δ ROA	-0.00557	-0.00564	-0.00545
	(-0.511)	(-0.518)	(-0.501)
Leverage	0.0211***	0.0211***	0.0211***
	(8.753)	(8.748)	(8.750)
Constant	-0.0271***	-0.0270***	-0.0270***
	(-4.200)	(-4.186)	(-4.152)
N	33,023	33,023	33,023
R^2	0.120	0.120	0.120
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for overall or net disclosure tone (change) for the footnotes sections of annual reports on Form 10-K. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

3.8 Conclusion

This study examined the presence and success of accounting narrative obfuscation in financial statements. Based on a large sample of Form 10-K filings, the study showed that management obfuscation in financial statements and especially the footnotes sections is either *"not present"* (management obfuscation by word choice) or *"has failed"* (management obfuscation by disclosure outlet). The regression results indicated that capital market participants process the important corporate information disclosed in the footnotes and that company officials do not successfully *"bury"* negative corporate information in this disclosure outlet (significant market reaction to changes in textual characteristics). Consistent with the idea of an increased litigation risk, the study found no evidence that managers use positive statements in the footnotes sections to frame or obfuscate negative corporate disclosures to mitigate their effect on financial markets. Thus, it is concluded that company officials seem to disclose truthful information in the footnotes sections of annual reports on Form 10-K and that capital market participants incorporate this information in firms' stock market valuations (*"efficient markets hypothesis"*). *"The efficient markets hypothesis does not distinguish where and in what form information is disclosed to the market - whether as a narrative, recognized in the financial statements or disclosed in the footnotes of the 10-K"* (Amel-Zadeh and Faasse 2016, 1). *"It's so important to read the footnotes. The footnotes to financial statements are packed with information"* (SEC 2007). *"The tone of footnote disclosures, for example, could help investors to understand a company's financial position and performance better"* (Czerney et al. 2017.1). *"For investors the message is clear: there is value in reading companies' annual reports, and especially understanding the cause of changes to narrative disclosures and footnotes"* (Amel-Zadeh and Faasse 2016, 30).

This study is subject to a number of limitations. First, the applied information extraction algorithm to utilize narrative information disclosed in regulatory filing might contain errors (e.g. parsing errors). This concern was mitigated by using an information extraction procedure tested for accuracy. Second, regulatory changes and corresponding changes in the information content of the footnotes section over time might affect the regression results. This concern could not be mitigated in this study. Third, other regulatory filings or disclosures and particular corporate events not subject to Form 10-K filing requirements might affect the empirical results of this study. These concerns were mitigated by analysing the market reaction to disclosure tone changes over a short period around the SEC submission.

Future research avenues could examine the presence and success of accounting narrative obfuscation in other parts of the annual report on Form 10-K (e.g. *"Exhibits"*). In addition, future studies could apply or combine different textual analysis techniques (e.g. sentiment, readability, similarity) to explore whether company officials successfully *"burying"* negative corporate information in financial statements. Furthermore, the presence and success of management obfuscation in other regulatory filings with the SEC could be subject to further empirical investigations (e.g. quarterly reports on Form 10-Q, current reports on Form 8-K).

Chapter 4

Form 10-K textual analysis and corporate bankruptcy

Contents

4.0 Introduction 165

4.1 Related literature 167

4.2 Hypothesis development 175

4.3 Research design 177

4.3.1 Sample selection and description 177

4.3.2 Model specification 181

4.3.3 Descriptive statistics 183

4.4 Empirical results 185

4.5 Robustness tests 187

4.5.1 Sample selection 187

4.5.2 Disclosure length 188

4.5.3 Table content 189

4.6 Sensitivity tests 190

4.6.1 Type of disclosure tone 190

4.6.2 Measurement of disclosure tone 192

4.7 Additional tests 199

4.7.1 Relevance of text sections 199

4.7.2 Firm fundamentals 204

4.7.3 All available bankrupt and non-bankrupt firm-year observations 208

4.7.4 Financial distress 211

4.8 Conclusion 214

"This research has shown that unique information useful in the explanation of bankruptcy is contained in the narratives of annual reports" (Tennyson et al. 1990, S. 406).

4.0 Introduction

The economic losses associated with corporate bankruptcy are extremely high for all stakeholders in a firm (e.g. stockholders, bondholders, employees, suppliers, auditors, communities, taxpayers, regulators). Given the immense costs of bankruptcy researchers try to identify indicators (red flags) of corporate failure to provide investors methods to minimize the risk of their investments (Lopatta et al. 2017, 316). Over decades, academics have developed various quantitative models consisting of financial ratios and market-based variables to predict future firm bankruptcies (e.g. Altman Z-Score). Each model has its own advantages and disadvantages (Pusch 2019, 60), and few significant improvements have been made over time (Lopatta et al. 2017, 316). In fact, *"despite all efforts, no comprehensive model has been established to date"* (Lopatta et al. 2017, 316). One explanation for this might be inappropriate accounting and auditing standards especially of failing firms affecting the predictive ability of current quantitative bankruptcy prediction models (Mayew et al. 2015, 1622; Lopatta et al. 2017, 316). Another reason might be disregarding qualitative or soft information when assessing a firm's default risk (e.g. credit analyst's judgement of a firm's market position based on experience) (Lehmann 2003, 6; Godbillon-Camus and Godlewski 2005; Henselmann and Scherr 2012, 3). Given the limitations of existing models, new insights into the explanatory ability of corporate information in the context of firm failures can be of great value not only for academics, but also for capital market participants and practitioners, and regulatory authorities.

The majority of bankruptcy models today are solely based on quantitative data, since hard information (financial ratios) is easy to collect (Henselmann and Scherr 2012, 3). However, besides quantitative (financial) data, annual reports also contain vast volume of qualitative (textual) information (Li 2010a, 143; Bannier et al. 2017, 4; Lo et al. 2017, 2; Lehavy et al. 2011, 1089). Neglecting soft information (narrative disclosures) can lead to the loss of useful and important corporate insights (Henselmann and Scherr 2012, 3; Li 2010a, 143-144; Bannier et al. 2017, 5). The future of a firm and its ability to continue may be communicated in the narrative sections of the Form 10-K filing (Tennyson et al. 1990, 392). Content analysis, included in the literature on qualitative information (Loughran and McDonald 2011a, 37; Loughran and McDonald 2016, 1188), is a methodology that classifies narrative segments (e.g. words, sentences, paragraphs) via categories or frequency counts (Tennyson et al. 1990, 394) to identify important and value-relevant information about a firm's future in text documents (Grant and

Conlon 2006, 119). In recent years, a research stream has emerged to measure the disclosure tone (textual sentiment) in regulatory filings (e.g. annual reports), media articles (e.g. newspapers), and internet messages (e.g. board postings) (Kearney and Liu 2014, 173-174). As a research tool, textual analysis provides a method for converting narrative (qualitative) disclosures into a quantitative measure (e.g. sentiment score) that can be related to underlying economic circumstances or firm characteristics (Tennyson et al. 1990, 392; Loughran and McDonald 2011a, 37). To this end, Loughran and McDonald (2011a) developed a series of word lists based on annual reports (e.g. *"negative"*, *"positive"*, *"uncertain"*, *"litigious"*, *"constraining"*) to measure the information content (textual sentiment) of Form 10-K filings (Loughran and McDonald 2011a, 35-37; Loughran and McDonald 2015, 2). The corresponding domain-specific business dictionary and its word lists became the predominant method in the accounting and finance literature to quantify the information content (disclosure tone) of financial documents.

Previous literature investigated the effect of the disclosure tone in regulatory filings on a firm's future stock characteristics (e.g. Li 2006; Feldman et al. 2008; Li 2008; Loughran and McDonald 2011a; Jegadeesh and Wu 2013; Kravet and Muslu 2013; Campbell et al. 2014; Chouliaras 2015; Hope et al. 2016; Henry and Leone 2016; Amel-Zadeh and Faasse 2016). However, far less is known about the explanatory ability of textual sentiment in the context of corporate bankruptcy. Based on a large sample of annual reports on Form 10-K filed with the United States Securities and Exchange Commission (SEC), this study investigates the association between textual information (linguistic tone) and corporate bankruptcy. It estimates a regression model with a firm's bankruptcy as the dependent variable and textual sentiment of accounting narratives as the independent variable. The regression results provide evidence that the tone of narrative disclosures in annual reports is associated with future corporate bankruptcy.

The study contributes to the literature in several ways. First, it shows that narrative information in Form 10-K filings is associated with future firm bankruptcies. Second, the study connects various types of disclosure tone to firm failures in the expected manner. Third, the results empirically confirm that specific subsections of the annual report on Form 10-K and their narrative disclosures are related to the future demise of corporations. Finally, the study illustrates the effect of different research design choices on empirical results in the field of textual analysis. Besides academia, the results of this study are also important for investors, practitioners, and regulatory authorities. Providing evidence for the explanatory power of narrative information in the context of future firm bankruptcies, investors (e.g. stockholders, bondholders) can use the results to better evaluate the financial soundness of their investments. In addition, practitioners (e.g. financial auditors) can use the empirical findings about narrative data to improve their analytical reviews and their going concern decisions. Furthermore, regulatory authorities (e.g. exchange commission, standard setters) can incorporate the results of this study to improve their economic forecasts and risk monitoring activities as well as their reporting requirements.

The remainder of Chapter 4 proceeds as follows. The next section presents the related literature. Section 4.2 develops the hypotheses. Section 4.3 illustrates the sample selection process and the research methodology. Section 4.4 presents the empirical results. Sections 4.5 and 4.6 report the results of the robustness and sensitivity tests. The additional tests are presented in Section 4.7, and finally, Section 4.8 concludes the Chapter.

4.1 Related literature

Because of the enormous economic losses, predicting potential bankruptcy is important not only for researchers, but also for investors (e.g. stockholders, bondholders), practitioners (e.g. financial auditors), and regulatory authorities (e.g. exchange commission, standard setters). Despite decades of intensive research to predict corporate bankruptcies, none of the developed models has prevailed. Actually, researchers are still trying to find an appropriate model and predictive variables to adequately forecast bankruptcy (Pusch 2019, 60). Most studies use accounting data (financial ratios), and very few employ narrative disclosures (textual information) to predict future corporate bankruptcies (Henselmann and Scherr 2012, 3; Pusch 2019, 61; Lopatta et al. 2017, 318; Yang et al. 2018; Gandhi et al. 2019). This has created several research gaps in the literature, which this study aims to address. The following literature review is not a comprehensive overview of all studies conducted.[20] Instead, the two relevant streams of research in the field of bankruptcy prediction and their most important findings are presented (quantitative and qualitative bankruptcy research).

[20] A comprehensive overview of studies on bankruptcy prediction is presented in Bellovary et al. (2007) and Kumar and Ravi (2007).

Quantitative models - Statistical models using financial ratios[21]

Beaver (1966) first examined the predictive ability of financial ratios (cash-flow-, profitability-, leverage-, liquidity-, and turnover-ratios) on future corporate bankruptcies (univariate discriminant analysis) (Beaver 1966, 78, 100). The author found that failed firms have lower cash flows and fewer liquid assets than non-failed firms (Beaver 1966, 80, 82). In addition, they tend to have more debt than non-failed firms before filing for bankruptcy (Beaver 1966, 80, 82). The results highlight a difference in the various financial ratios of failed and healthy firms for up to five years before filing for bankruptcy, with an increasing difference as corporate failure approaches (Beaver 1966, 81-82). Interestingly, the results also show that the average asset size of failed firms increases over time despite the future discontinuance of the firm (Beaver 1966, 81-82). Not examining ex-post financial ratios or differences, but their (real) predictive ability, an analysis indicated that the net liquid-asset flow is most important for predicting future corporate bankruptcies (cash flow to total debt ratio, net income to total assets ratio, total debt to total assets ratio) (Beaver 1966, 85-86, 101). The author concluded that *"the evidence indicates that ratio analysis can be useful in the prediction of failure for at least five years before failure"* (Beaver 1966, 102).

Another bankruptcy prediction model was developed by Altman (1968) using multiple financial ratios (MDA - multivariate discriminant analysis) (Altman 1968, 589) to analyse various firm characteristics and their interactions (Altman 1968, 592). Ultimately, five financial variables (working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value of equity to book value of total debt, sales to total assets) were selected to form the best discriminant function (Z-Score) to predict future corporate bankruptcies (Altman 1968, 594). The author highlighted significant differences in the financial ratios between bankrupt and non-bankrupt firms (Altman 1968, 596). The empirical results indicated that bankruptcy can be predicted accurately for up to two years before the actual failure occurs (deteriorating trend of financial ratios) (Altman 1968, 599-600, 604-605). However, the results also showed that the accuracy of predicting corporate bankruptcies is significantly reduced when considering the same financial ratios from more remote years prior to the actual event (Altman 1968, 604). Based on the accuracy one and two years before bankruptcy, the

[21] Besides traditional methods, researchers use machine learning techniques to predict corporate bankruptcies including decision trees (e.g. Frydman et al. 1985; Gepp et al. 2010), support vector machines (e.g. Shin et al. 2005), neural networks (e.g. Odom and Sharda 1990; Wilson and Sharda 1994; Zhang et al. 1999), and hybrid models (e.g. Pendharkar and Rodger 2004) (Pusch 2019, 61-62).

author concluded that the designed model is an *"accurate forecaster"* of corporate failure (Altman 1968, 604, 609).

Ohlson (1980) addressed the problems associated with MDA in the context of bankruptcy prediction by examining the probability of failure using a logit model (Ohlson 1980, 111-112).[22] A profile analysis (not prediction analysis) revealed that failing firms have significant different financial ratios than non-failing firms in the years prior to bankruptcy (Ohlson 1980, 119). For instance, failing firms have a smaller asset size, higher liabilities, smaller working capital, higher current liabilities, lower net income, and lower operating cash flows than non-failing firms in the year prior to bankruptcy (Ohlson 1980, 119). The prediction results of the logit models showed that size, leverage, profitability, and liquidity are statistically significant factors when assessing the probability of corporate bankruptcy (Ohlson 1980, 120-123). Interestingly, the results also indicated that the profit margin and ratio of assets with little or no cash value (intangibles, deferred assets) are not predictive for corporate failure (Ohlson 1980, 123). Comparing the prediction error rates of a MDA model and a logit model, the author concluded that *"many reasonable procedures will lead to results which will not differ too much"* (Ohlson 1980, 129).

Westgaard and van der Wijst (2001) examined the expected default frequency (EDF) in the context of corporate credit portfolios (Westgaard and van der Wijst 2001, 339). Specifically, they investigated the impact and predictive ability of different firm characteristics such as size, age, and industry classification on corporate bankruptcies using a logistic regression model (Westgaard and van der Wijst 2001, 340). The analysis showed large differences in default rates across different industries ("Real estate and service" - 1.08% default rate, "Hotel and restaurant" - 7.75% default rate) (Westgaard and van der Wijst 2001, 342-343). A feature analysis (not regression analysis) revealed that non-bankrupt firms have a higher cash flow to debt ratio, a higher financial coverage ratio, higher liquidity rate, and higher equity ratio (Westgaard and van der Wijst 2001, 345). The regression results of the logit model negatively associated the cash ratio, size, liquidity, financial coverage, and equity ratio with a firm's probability of going into bankruptcy (Westgaard and van der Wijst 2001, 346-347). Analysing the *"predictive power"* (hold out sample), the study further confirmed the ability of the developed logit model to estimate the expected default frequency (Westgaard and van der Wijst 2001, 347). Based on

[22] As Ohlson (1980) noted, the MDA model is associated with various issues such as the statistical requirements imposed on the distributional properties of the predictors, output of the MDA model and its application, and problems related to the firm matching procedure in the MDA model (Ohlson 1980, 111-112).

their results, the authors concluded that the expected default frequency of a firm is influenced by cash-, liquidity-, and equity-ratios as well as its size and age (Westgaard and van der Wijst 2001, 348).

In other work, Zmijewski (1984) explored methodological issues related to bankruptcy prediction models using a probit model (Zmijewski 1984, 65-67, 74-75). The author examined the effect of choice-based sample selection (higher probability of selecting bankrupt firms into the data sample) (Zmijewski 1984, 60) and the bias resulting from a complete data sample selection criterion (lower probability of selecting bankrupt firms into the data sample) (Zmijewski 1984, 62) on financial distress models. The empirical results confirmed the existence of a choice-based and data availability (data collection) bias in bankruptcy prediction models (Zmijewski 1984, 63, 68-74, 76-80). However, they also showed that both issues do not significantly change the overall classification and prediction rates of financial distress (Zmijewski 1984, 63, 77, 80).

Shumway (2001) developed a hazard model with the time spent by a firm in the non-bankrupt group as the dependent variable to account for a firm's changing risk of bankruptcy over time (Shumway 2001, 102). Re-examining the results of Altman (1968) and Zmijewski (1984) with a hazard model, the study determined that numerous variables are unrelated to a firm's bankruptcy probability (Shumway 2001, 102, 117-120). Using a hazard model only identified EBIT/TA (earnings before interest and taxes to total assets) and ME/TA (market equity to total assets) (Shumway 2001, 117-118) as well as NI/TA (net income to total assets) and TL/TA (total liabilities to total assets) (Shumway 2001, 119-120) as significant predictors for corporate bankruptcy. Implementing market and accounting variables into a hazard model, the author increased the accuracy of predicting corporate bankruptcies (Shumway 2001, 121-23). Based on the results, it was concluded that *"bankruptcy forecasts can be improved dramatically by conditioning on market-driven variables"* (Shumway 2001, 123).

Furthermore, Chava and Jarrow (2004) attempted to improve bankruptcy prediction models by including industry effects and adding monthly estimation intervals to the hazard model (Chava and Jarrow 2004, 538-539). The results confirmed industry effects as an important component in bankruptcy prediction for both private and public companies (Chava and Jarrow 2004, 538-539, 555-558, 567). In addition, the results also showed that monthly estimation intervals (monthly accounting and market variables) can significantly improve the accuracy of bankruptcy prediction models (Chava and Jarrow 2004, 539, 564-566, 567). The authors concluded that hazard models have superior forecasting performance in the context of bankruptcy prediction (Chava and Jarrow 2004, 567).

Qualitative models - Statistical models using narrative information

Tennyson et al. (1990) were among the first to investigate the usefulness of narrative information in explaining corporate bankruptcy (Tennyson et al. 1990, 391). Specifically, the authors examined the information content of a firm's management analysis (MA) in the annual report and the president's letter (PL) (Tennyson et al. 1990, 396) to explain bankruptcy using a multivariate logistic regression model (MLR) (Tennyson et al. 1990, 394).[23] A first analysis revealed that the information content of narrative disclosures differs when considering future bankrupt and non-bankrupt firms. While healthy firms focus on internal operations, growth, and expansion, bankrupt firms concentrate on the external environment (Tennyson et al. 1990, 400-401). The analysis further indicated that both sources of narrative information (MA, PL) are useful in explaining future corporate bankruptcies (Tennyson et al. 1990, 400-401). More important, the empirical results confirmed that narrative disclosures provide incremental information in explaining future corporate failure even when controlling for firm fundamentals (financial ratios) (Tennyson et al. 1990, 402-404). Based on their results, the authors concluded that annual reports contain unique narrative information to explain the future demise of companies (Tennyson et al. 1990, 406).

Magnusson et al. (2005) used self-organizing maps as quantitative data and collocational networks as qualitative data to examine the language of quarterly reports as an indicator of change in a firm's financial situation (Magnusson et al. 2005, 561). Assuming that actions of corporate management as well as expectations are reflected in a firm's narrative disclosures need some time to materialise the authors expect a delayed representation in the financial performance data of a company (Magnusson et al. 2005, 562). Consistent with this idea, the analysis showed that a change in textual disclosures is followed by a change in the financial disclosures the following quarter (Magnusson et al. 2005, 571-572). Interestingly, the results also highlight that a change in textual information is preceded by a smaller anticipatory change in the qualitative information before a change in the firm fundamentals occurs (Magnusson et al. 2005, 571-572). Based on their results, the authors confirmed their *"original hypothesis that changes in the*

[23] Tennyson et al. (1990) noted that a multivariate logistic regression model (MLR) enables an assessment of the significance of individual independent variables, enhancing the interpretation of the information content of each variable (Tennyson et al. 1990, 394, 399).

qualitative data (text) preceded changes in the quantitative data (financial figures)" (Magnusson et al. 2005, 572).

Shirata and Sakagami (2008) analysed the non-financial (qualitative) information in annual reports to differentiate between going concern and non-going concern companies using key words (Shirata and Sakagami 2008, 1). In a first analysis, the authors used a classification and regression tree to discriminate between bankrupt and non-bankrupt firms to identify important financial variables in the context of firm bankruptcy (Shirata and Sakagami 2008, 7). The quantitative analysis revealed that certain firm fundamentals are able to discriminate between bankrupt and non-bankrupt companies (e.g. retained earnings to total liabilities and owners' equity, net income before taxes to total liabilities and owners' equity, inventory turnover period, interest expenses to sales) (Shirata and Sakagami 2008, 8). The qualitative analysis confirmed significant differences in narrative disclosures between future bankrupt and non-bankrupt companies (Shirata and Sakagami 2008, 12). In particular, terms such as *"dividends"*, *"profit appropriation"*, and *"retained earnings"* appeared more often in the Form 10-K filings of non-bankrupt companies (Shirata and Sakagami 2008, 12-13). Based on their results, the authors concluded that it is *"possible to assess whether a company will continue or discontinue operating based on qualitative data, without looking at financial numbers"* (Shirata and Sakagami 2008, 15).

Other interesting work includes Cecchini et al. (2010), who developed dictionaries from annual reports on Form 10-K (Management Discussion and Analysis section - MD&A) to discriminate between bankrupt and non-bankrupt firms using a support vector machine (SVM) as a classification method (Cecchini et al. 2010, 164, 171).[24] In general, the empirical results showed that firms in poor financial health tend to have lower market values and lower returns on assets (Cecchini et al. 2010, 170-171). The analysis (automated ontology creation) further demonstrated that textual references (tokens) to quantitative financial data in financial statements have the highest discriminative power for bankruptcy (*"net income"*, *"gross margin"*, *"research and development"*) (Cecchini et al. 2010, 167, 170-171). Noteworthy is that the word pair *"risk uncertainty"* had a low relevance for bankruptcy (Cecchini et al. 2010, 170). The refinement of the SVM model indicated the importance of salient accounting phrases in the bankruptcy do-

[24] According to Cecchini et al. (2010) a support vector machine model (SVM) produces a linear discriminant function while balancing the risk of misclassifying the training set and using overly complex hypothesis spaces (*"overfitting"* of the machine learning method). In addition, this machine learning method is able to work with large feature spaces (word phrases or input variables) and small datasets (Cecchini et al. 2010, 171).

main (*“gross margin decline”*, *“company record charge”*, *“inflation impact company”*) (Cecchini et al. 2010, 171-172). The empirical (prediction) results showed that using the information contained in the MD&A section is useful in discriminating between future bankrupt and non-bankrupt firms (Cecchini et al. 2010, 172). The authors further confirmed that combining narrative disclosures with financial ratios can improve the accuracy of predicting future corporate bankruptcies (Cecchini et al. 2010, 173). Based on the results, the authors concluded that the *“text of the MD&A contains information that is complementary to the quantitative information”* (Cecchini et al. 2010, 174).

Also to predict bankruptcy, Shirata et al. (2011) extracted key phrases from (Japanese) annual reports (dividend section) (Shirata et al. 2011, 31, 35). To extract specific phrases to discriminate between bankrupt and non-bankrupt companies, they analysed the context of annual report sentences via a morphological analysis and the conditional probability method (Shirata et al. 2011, 36). The analysis showed that non-bankrupt companies use keywords such as *“dividend”*, *“execute”*, and *“dividend policy”* more frequently than non-bankrupt firms. In contrast, phrases like *“research and development”* and *“corporate value”* are rarely used by bankrupt firms (Shirata et al. 2011, 37-40). In addition, while firms heading for bankruptcy discuss restrictions to dividend payments, non-bankrupt firms state information about retained earnings in the context of capital investments in their annual reports (Shirata et al. 2011, 40-41). Based on their results, the authors concluded that *“narrative information can improve the bankruptcy prediction power”* (Shirata et al. 2011, 42).

Henselmann and Scherr (2013) examined the ability of XBRL (eXtensible Business Reporting Language) tags in annual reports as a supplement of bankruptcy prediction. Using negative XBRL tags as an aggregate of pessimistic textual information in the footnotes sections of Form 10-K filings, the authors found that the number of negative tags (*“Red Flags”*) was positively associated with a firm’s future bankruptcy (Henselmann and Scherr 2013, 61). They concluded that the qualitative information disclosed in annual reports and XBRL structured data in particular has the power to predict firm bankruptcy (Henselmann and Scherr 2013, 63).

Mayew et al. (2015) investigated the explanatory power of the disclosure tone in Form 10-K filings on a firm’s ability to continue as a going-concern. The authors noted that negative and positive textual sentiment in the *“MD&A”* section of an annual report is indicative of the risk of firm bankruptcy (Mayew et al. 2015, 1632-1633). The authors concluded that qualitative information in regulatory filings and the disclosure tone in particular provides additional incremental information apart from financial variables (Mayew et al. 2015, 1633, 1645).

Lopatta et al. (2017) analysed the predictive ability of textual sentiment on a firm's future bankruptcy. The authors confirmed that the annual reports of failing firms contain significantly more negative and litigious words than the Form 10-K filings of non-failing firms (Lopatta et al. 2017, 332). Importantly, the regression results revealed that negative disclosure tone in annual report filings is associated with future firm bankruptcies (Lopatta et al. 2017, 334, 338). Based on their results, the authors concluded that regulatory filings and language contain information content beyond a company's financial figures (Lopatta et al. 2017, 338-339).

Yang et al. (2018) also extracted qualitative information (e.g. key words, phrases, expressions, sentences, topics) from the MD&A section of annual reports to identify differences in textual disclosures between bankrupt and non-bankrupt companies (Yang et al. 2018, 46). Their analysis showed that several words appear exclusively in annual reports of failing firms (*"end"*, *"net"*, *"include"*, *"period"*, *"facility"*) while others only appear in the Form 10-K filings of successful firms (*"business"*, *"risk"*, *"primarily"*, *"product"*, *"investment"*) (Yang et al. 2018, 48-49). Since considering frequent words in isolation may not provide valuable information about a company's likelihood to go bankrupt in the future, the authors focused on concept links (co-occurrence words associated with high-frequency words) to identify disclosure differences between the two groups of firms (Yang et al. 2018, 48-49). For instance, the concept link for *"capital"* is base, asset, increase, additional, finance, future, period, credit, and estimate (Yang et al. 2018, 50). However, a further analysis of topics showed that certain themes merely capture industry characteristics rather than providing sufficient information to observe bankruptcy tendencies (Yang et al. 2018, 52-53). Therefore, the authors concluded that *"hundreds of topic and textual features can be extracted through text mining, and it is usually difficult to foresee what features are useful"* (Yang et al. 2018, 54).

Loughran and McDonald (2019) examined the negative annual report sentiment of banks in the United States as a proxy for financial distress (Gandhi et al. 2019, 424-425). Descriptive statistics of Form 10-K disclosures show that negative annual report sentiment is positively correlated with future stock delistings and provisions for loan losses and negatively associated with dividend payments and returns on assets (Gandhi et al. 2019, 430). The empirical results showed a significant relationship between the amount of negative language in Form 10-K filings and future firm delistings (Gandhi et al. 2019, 430-431). In addition, the results indicated that more pessimistic annual report sentiment is predictive for a lower probability of subsequent dividend payments to shareholders (Gandhi et al. 2019, 431-432). Furthermore, the study empirically confirmed that a negative tone in annual reports is linked to higher subsequent loan loss provisions (Gandhi et al. 2019, 432-433). Finally, a significant relationship between negative Form 10-K textual sentiment and lower returns on assets in the future was found (Gandhi et al. 2019, 433). Based on their results, the authors concluded that annual report sentiment *"can supplement existing analytical techniques as an additional check on the health of individual banks or the general well-being of the banking sector"* (Gandhi et al. 2019, 434).

Overall, various studies to predict future corporate bankruptcies have been conducted using both quantitative and qualitative data. Again, thus far, only a minority of academic research has used narrative information in the context of bankruptcy prediction (Henselmann and Scherr 2012, 3; Pusch 2019, 61; Lopatta et al. 2017, 318). In particular, a detailed analysis of the explanatory power of different types of disclosure tone (and their measurement) and importance of different text sections in the annual report on Form 10-K is still missing in the literature. This study aims to address these research gaps in the literature.

4.2 Hypothesis development

Assuming that company officials (corporate insiders) better understand a company than others, the tone of narrative disclosures embedded in a firm's annual report might contain additional *"hard-to-quantify"* information about the current situation of the company or its future performance (Kearney and Liu 2014, 172-173; Elrod 2009, 3-4; Gandhi et al. 2019, 425-426). Research in the field of textual analysis can determine a connection between the linguistic tone in annual reports and important firm characteristics (e.g. Amel-Zadeh and Faasse 2016). Besides the explanatory power of disclosure tone regarding stock characteristics and firm fundamentals, Form 10-K textual sentiment might also be associated with a corporation's future demise.

The management literature identifies three consecutive phases or periods of corporate crisis that lead to bankruptcy: a strategy crisis, performance crisis, and management crisis. Therefore, it is assumed that filing for bankruptcy is the result of a long-term process that should be reflected in a firm's corporate disclosures (Lopatta et al. 2017, 320). *"Except in rare circumstances, the going concern assumption becomes questionable some time before its questionability becomes apparent by virtue of obvious and concrete indicators"* (Boritz 1991, 27). Consistent with the idea that language in regulatory filings can be predictive for a firm's future, empirical research has associated the disclosure tone in annual reports with future corporate earnings (e.g. Gandhi et al. 2019) and found the likelihood to continue as a going concern (Mayew et al. 2015). Thus, the following hypothesis is proposed:

Hypothesis 5. There is a positive relationship between negative Form 10-K textual sentiment and future corporate bankruptcy given the unique insights of corporate management into a firm's future corporate earnings (accounting profitability) not being reflected in any present firm fundamentals.

Company officials have strong incentives to disclose positive information in regulatory filings to present their firm's performance (and their own) in the most favourable way (e.g. management compensation) (Twedt and Rees 2011, 21; Kothari et al. 2009, 1641-1643; Campbell et al. 2014, 401). Managers might also use positive words in annual reports to frame negative corporate information. *"Positive words, on the other hand, in addition to their positive usage, are just as frequently used to frame a negative statement"* (Loughran and McDonald 2016, 1217). Consistent with this notion, most empirical studies in the field of textual analysis (e.g. Tetlock 2007; Kothari et al. 2009; Loughran and McDonald 2011a; Loughran and McDonald 2015) have been unable to connect positive disclosure tone and a firm's future stock market performance (Tetlock et al. 2008, 1442; Twedt and Rees 2011, 25; Loughran and McDonald 2011a, 38; Loughran and McDonald 2016, 1217-1218; Gandhi et al. 2019, 425; Loughran and McDonald 2015, 8).

In contrast, shareholder litigation is one example of an important mechanism to limit opportunistic disclosure behaviour by a firm's management in corporate reports (Rogers et al. 2011, 2157; Gandhi et al. 2019, 425). *"Why would investors respond to an optimistic tone if there are no enforcement mechanisms to lend credibility to tone?"* (Rogers et al. 2011, 2158). Therefore, optimistic corporate disclosures in regulatory filings might contain truthful information about the current financial situation of a firm and more importantly its expected future performance and accounting profitability (incremental information). Consistent with this notion, a limited number of empirical studies have connected positive linguistic tone to a firm's future market performance (e.g. Jegadeesh and Wu 2013). Furthermore, it might be possible that both negative and positive information (decreasing and increasing effects on future corporate earnings) in annual reports when considered as an aggregate (negative minus positive tone) reveal valuable insights into a firm's future financial performance. Despite the notion that a firm's management might disclose meaningless positive information either to present itself in a favorable light or to frame negative corporate news the present study expects that positive information being disclosed in annual reports represents trustworthy incremental corporate information especially in light of available litigation mechanisms. Therefore, the following hypotheses are proposed:

Hypothesis 6a. There is a negative relationship between positive Form 10-K textual sentiment and future corporate bankruptcy given the credibility and relevance of positive information in the context of a firm's future earnings not being reflected in any current financial figure.

Hypothesis 6b. There is a negative relationship between overall (net) Form 10-K textual sentiment (negative minus positive tone) and future corporate bankruptcy given the credibility and relevance of corporate statements in the context of future earnings not being communicated by other disclosures.

4.3 Research design

4.3.1 Sample selection and description

Using the SEC EDGAR *"Form Index Files"* all annual reports filed with the SEC between 1993 and 2016 were downloaded (Form 10-K, Form 10-K405). Note that the time period was chosen due to data availability reasons on various information platforms (e.g. SEC EDGAR database, commercial data providers, public research repositories). An overview of annual report submissions filed with the SEC is presented in the appendix.[25] For each annual report filing, the text version of the submission was retrieved from the EDGAR database (*"Complete Submission Text File"*). The text version contains all information disclosed within a particular SEC submission (e.g. core Form 10-K document, exhibits) (Bodnaruk et al. 2015, 643; Loughran and McDonald 2011b, 1). To utilize all textual information disclosed in annual reports, an information extraction algorithm (regular expressions) was applied.[26] Furthermore, in addition to the entire Form 10-K submission (core document including exhibits), all individual subsections contained in a Form 10-K filing were extracted (non-relevant documents were deleted). Based on the extraction results, word counts were obtained to quantify the disclosure tone (textual sentiment) in the financial statements (Loughran and McDonald business dictionary).

[25] Please note that the Form 10-K sample is identical to the one used in Henselmann and Hering (2017).

[26] Note that the information extraction algorithm applied in this study is based on the *"Annual Report Algorithm"* and *"Items Algorithm"* designed by Hering (2016). Furthermore, the parsing procedures in this study are identical to those used in Henselmann and Hering (2017). For a detailed overview of the parsing rules applied in this study, please refer to the appendix.

The study followed prior empirical works (e.g. Cao and Narayanamoorthy 2014; Correia et al. 2012; Jiang et al. 2012; Lopatta et al. 2017) and used the UCLA LoPucki Bankruptcy Research Database (BRD). The database provides information on 1,118 public US corporations that filed for bankruptcy between 1980 and 2018.

For all firms, the data sample in this study was restricted to one Form 10-K filing per year. The sample for bankrupt firms was further restricted to submissions filed between seven years before the bankruptcy and the last Form 10-K filing before the bankruptcy. In addition, annual reports filed after a specific firm declared bankruptcy were removed (possible influence on narrative disclosures). The time span was chosen because early bankruptcy research (e.g. Altman et al. 1977; Beaver 1966) reports explanatory power of prediction models up to five years before bankruptcy (Lopatta et al. 2017, 324). The remaining firm-year observations (see other data filters) constituted the pool of annual reports from bankrupt and non-bankrupt firms. Following prior bankruptcy as well as corporate fraud research (e.g. Beaver 1966; Altman 1968; Goel et al. 2010; Henselmann and Scherr 2013; Purda and Skillicorn 2015; Lopatta et al. 2017; Pusch 2019) the study applies a paired-sample design to ensure that for each bankrupt firm-year observation the best possible firm-year observation from the control sample (non-bankrupt) is matched. The paired-sample design was selected to help provide a control over factors that might blur the relationship between explanatory variables (ratios) and corporate failure (differences among industries prevent the direct comparison of firms from different industries) (Beaver 1966, 74; Lopatta et al. 2017, 325). The matching process for each bankrupt firm-year observation is based on industry classification (Fama-French 48 industry classification) and market capitalization (size proxy) to identify a healthy (biggest) control observation. Note that the matching process within the paired-sample design identifies exact one bankrupt and one non-bankrupt firm-year observation for each fiscal year and industry category. If no match is found or more than one company filed for bankruptcy in the same industry and fiscal year all corresponding observations are excluded from the analysis. The final sample consisted of 476 firm-year observations of which 238 are associated with failing firms (238 non-failing firm-year observations - peer group). This corresponds to approximately 0,02% of the overall Form 10-K sample. Table 37 provides an overview of the sample selection process.

The majority of annual reports in the data sample (firm-year observations associated with failing and non-failing firms) were filed by companies in the *“Apparel”* industry (5.88 percent), followed by those in *“Machinery”* (5.88%), *“Recreation”* (4.62%), and *“Healthcare”*

(4.62%). In general, the distribution of annual reports among certain industries in the data sample is unequal (Fama-French 48 industry classification). While the distribution of firm-year observations across industries is unequal, the distribution of Form 10-K filings over time is more balanced. Table 38 presents the distribution of annual report submissions by certain industries and fiscal years.

Table 37. Sample selection process for corporate bankruptcy examination.

	Firm-year observations	**Observations removed**
All observations (Form 10-K & Form 10-K405)	189,998	
Less		
Missing 10-K report (database error)	189,997	1
Removing duplicates CIK and fiscal-year	184,089	5,908
Removing late filers (>100 days after accounting period)	163,838	20,251
Missing SIC (from filing)	162,759	1,079
Merge with Capital IQ data		
Missing market value	107,901	54,858
Other data filters		
Number of words in Form 10-K (including exhibits) >0	107,776	125
Merge with ULCA LoPucki bankruptcy data		
Removing 10-Ks filed after bankruptcy	106,838	938
Removing 10-Ks filed more than 6 years before bankruptcy	105,592	1,246
Removing multiple bankruptcy cases per industry and year	21,244	84,348
Removing bankruptcy cases without control observations (highest market value per industry and year)	476	20,768
Observations in final sample		
Total		476
Of which associated with failing firms		238
Of which associated with control firms		238

Notes: The table presents the sample selection process and data filters applied to the overall data sample.

Table 38. Distribution of Form 10-K filings across industries and fiscal years for corporate bankruptcy examination.

Panel A: Distribution of Form 10-K filings across industries

Fama-French Industry	Filings (Num.)	Filings (%)	Filings (∑%)	Fama-French Industry	Filings (Num.)	Filings (%)	Filings (∑%)
Machinery	28	5.88	5.88	Petroleum and Nat. Gas	10	2.1	74.34
Apparel	28	5.88	11.76	Fabricated Products	10	2.1	76.44
Healthcare	22	4.62	16.38	Steel Works Etc	10	2.1	78.54
Recreation	22	4.62	21.00	Textiles	10	2.1	80.64
Wholesale	20	4.2	25.20	Pharmaceutical Products	10	2.1	82.74
Construction	20	4.2	29.40	Almost Nothing	8	1.68	84.42
Consumer Goods	20	4.2	33.60	Shipping Containers	8	1.68	86.10
Real Estate	18	3.78	37.38	Non-Metallic and Industrial Metal Minin	8	1.68	87.78
Restaurants, Hotels, Motels	18	3.78	41.16	Automobiles and Trucks	8	1.68	89.46
Aircraft	16	3.36	44.52	Electrical Equipment	8	1.68	91.14
Agriculture	16	3.36	47.88	Printing and Publishing	8	1.68	92.82
Food Products	14	2.94	50.82	Trading	6	1.26	94.08
Business Supplies	12	2.52	53.34	Transportation	6	1.26	95.34
Measuring and Control Equipment	12	2.52	55.86	Banking	4	0.84	96.18
Electronic Equipment	12	2.52	58.38	Computers	4	0.84	97.02
Personal Services	12	2.52	60.90	Communication	4	0.84	97.86
Chemicals	12	2.52	63.42	Construction Materials	4	0.84	98.70
Entertainment	12	2.52	65.94	Retail	2	0.42	99.12
Insurance	10	2.1	68.04	Coal	2	0.42	99.54
Business Services	10	2.1	70.14	Precious Metals	2	0.42	100.00
Utilities	10	2.1	72.24	-	-	-	-

Panel B: Distribution of Form 10-K filings across fiscal years

Fiscal year	Filings (Num.)	Filings (%)	Filings (∑%)	Fiscal year	Filings (Num.)	Filings (%)	Filings (∑%)
1993	12	2.52	2.52	2005	26	5.46	57.98
1994	20	4.20	6.72	2006	14	2.94	60.92
1995	18	3.78	10.50	2007	24	5.04	65.97
1996	14	2.94	13.45	2008	20	4.20	70.17
1997	18	3.78	17.23	2009	24	5.04	75.21
1998	20	4.20	21.43	2010	22	4.62	79.83
1999	28	5.88	27.31	2011	16	3.36	83.19
2000	28	5.88	33.19	2012	12	2.52	85.71
2001	22	4.62	37.82	2013	14	2.94	88.66
2002	26	5.46	43.28	2014	18	3.78	92.44
2003	24	5.04	48.32	2015	30	6.30	98.74
2004	20	4.20	52.52	2016	6	1.26	100.00

Notes: The table illustrates the number of annual reports in the data sample across Fama-French industries (Panel A) and across fiscal years (Panel B).

4.3.2 Model specification

To examine whether the disclosure tone in annual reports can be associated with a firm's future demise and can therefore be served as a valid supplement in efforts to forecast future corporate bankruptcies, the following logistic regression model was estimated[27]:

$$Y_{i;t+1} = \beta_0 + \beta_1 * Tone_{Type;Full\ Report;i;t} + \varepsilon_{i;t} \tag{19}$$

Where: the dependent variable ($Y_{i;t+1}$) is a binary variable equal to one if a firm files for bankruptcy in the future or zero if it does not. $Tone_{Type;Full\ Report;i;t}$ is the textual sentiment in the entire annual report on Form 10-K (submission including exhibits) as measured by the number of pre-defined words in the SEC filing.

Faced with a large variety of possible variables (e.g. type of textual sentiment, measurement of textual sentiment, sentiment in specific Form 10-K subsections) to be included in more sophisticated models (multivariate analysis) the study followed earlier bankruptcy research (e.g. Beaver 1966; Henselmann and Scherr 2013) and examined the explanatory (discriminatory) power of the disclosure tone in regulatory filings as a stand-alone variable (univariate analysis). The aim of this approach is to optimize the model through a selection of best discriminant tone variables. Note that this study examined the explanatory power of the disclosure tone in regulatory filings also in the presence of firm fundamentals or control variables (multivariate analysis).

To quantify the linguistic tone in annual reports, the Loughran and McDonald business dictionary and its word lists (e.g. *"negative"*, *"positive"*) is being used. Following prior work (e.g. Henry and Leone 2016), an equal weighting scheme was applied to quantify the tone of narrative information in annual reports.

The first measure for the disclosure tone in Form 10-K filings is the ratio of negative words (negative sentiment) within a particular SEC submission:

$$Tone_{Neg;Full\ Report;i;t} = \frac{Negative_{Full\ Report;i;t}}{Total_{Full\ Report;i;t}} \tag{20}$$

[27] Note that linear regression models are most commonly used in the accounting and finance literature when examining the effect of textual information on firm characteristics (Kearney and Liu 2014, 177). Further note that the study used a logistic regression model over various machine learning models (e.g. neural networks) to better examine the relationship between input (independent) and output (dependent) variables.

Where: $Tone_{Neg;Full\,Report;i;t}$ is the negative textual sentiment (negative linguistic tone) in the entire annual report on Form 10-K (submission including exhibits), $Negative_{Full\,Report;i;t}$ is the number of negative words in the Form 10-K filing, and $Total_{Full\,Report;i;t}$ is the number of all words appearing in the SEC submission.

The second measure for the disclosure tone in Form 10-K filings is the ratio of positive words (positive sentiment) in an annual report on Form 10-K filed with the SEC:

$$Tone_{Pos;Full\,Report;i;t} = \frac{Positive_{Full\,Report;i;t}}{Total_{Full\,Report;i;t}} \tag{21}$$

Where: $Tone_{Pos;Full\,Report;i;t}$ is the positive textual sentiment (positive linguistic tone) in the entire annual report on Form 10-K (submission including exhibits), $Positive_{Full\,Report;i;t}$ is the number of positive words in the Form 10-K filing, and $Total_{Full\,Report;i;t}$ is the number of all words appearing in the SEC submission.

The difference between negative and positive disclosure tone is the third measure of linguistic tone (overall or net sentiment) in Form 10-K filings made with the SEC:

$$Tone_{Net;Full\,Report;i;t} = Tone_{Neg;Full\,Report;i;t} - Tone_{Pos;Full\,Report;i;t} \tag{22}$$

Where: $Tone_{Net;Full\,Report;i;t}$ is the overall or net textual sentiment (negative minus positive linguistic tone) in the entire annual report on Form 10-K (submission including exhibits), $Tone_{Neg;Full\,Report;i;t}$ is the negative textual sentiment in the annual report, and $Tone_{Pos;Full\,Report;i;t}$ is the positive textual sentiment in the Form 10-K filing as defined above.

Note that the study used the Loughran and McDonald dictionary (excluding stop words) to determine the text length of a certain Form 10-K report, as this method of quantifying disclosure tone in regulatory filings is objective and replicable (Rogers et al. 2011, 2162; Loughran and McDonald 2016, 1200). Furthermore, all included variables were winsorized at the 1% level.

4.3.3 Descriptive statistics

This section presents the descriptive statistics for the disclosure tone of bankrupt and non-bankrupt firms. On average (median), annual report filings of bankrupt firms state 3.30 (3.25) percent negative information. Form 10-K filings of non-bankrupt firms on average (median) contain 3.06 (2.91) percent negative words. In addition, while failing firms on average state only 1.24 (1.18) percent positive information in their annual reports, non-failing companies disclose 1.38 (1.34) percent optimistic statements in Form 10-K filings. Furthermore, bankrupt firms disclose more litigious and constraining words than non-bankrupt firms. While the former disclose 3.78% litigious and 1.61% constraining words, the latter only state 3.57% and 1.28%, respectively. Note that the means of negative, positive, net (overall), and constraining disclosure tone show significant differences between bankrupt and non-bankrupt companies. In general, these results indicate that firms' financial future is reflected in the tone of accounting narratives disclosed in their annual reports filed with the SEC. Table 39 provides descriptive statistics for different types of disclosure tones in annual reports on Form 10-K.

Table 39. Descriptive statistics for disclosure tones.

Panel A: Bankrupt firm sample (N=238) and non-bankrupt firm sample - control sample (N=238)

	Bankrupt firm sample (N=238)					Non-bankrupt firm sample (N=238)					Test statistics	
Variable	Mean	SD	Min.	Med.	Max.	Mean	SD	Min.	Med.	Max.	t-value	p-value
$Tone_{Negative;Full\ Report;[-6,0]}$	0.0330	0.0096	0.0126	0.0325	0.0709	0.0307	0.0102	0.0124	0.0291	0.0685	-2.5903	0.0099***
$Tone_{Positive;Full\ Report;[-6,0]}$	0.0124	0.0031	0.0076	0.0118	0.0233	0.0138	0.0036	0.0078	0.0134	0.0251	4.4321	0.0000***
$Tone_{Net;Full\ Report;[-6,0]}$	0.0206	0.0101	-0.0024	0.0203	0.0593	0.0169	0.0115	-0.0056	0.0151	0.0561	-3.7820	0.0002***
$Tone_{Uncertainty;Full\ Report;[-6,0]}$	0.0217	0.0067	0.0091	0.0218	0.0352	0.0220	0.0058	0.0100	0.0216	0.0355	0.5677	0.5705
$Tone_{Litigious;Full\ Report;[-6,0]}$	0.0378	0.0208	0.0092	0.0320	0.0853	0.0357	0.0180	0.0097	0.0316	0.0777	-1.1810	0.2382
$Tone_{Constraining;FullReport;[-6,0]}$	0.0161	0.0043	0.0080	0.0154	0.0274	0.0128	0.0036	0.0068	0.0125	0.0255	-9.1325	0.0000***

Panel B: Form 10-K disclosure tone prior to firm bankruptcy

	Bankrupt firm sample (N=238)											
Variable	-7	-6	-5	-4	-3	-2	-1	-	-	-	-	-
$Tone_{Negative;Full\ Report;[X]}$	2.78%	2.86%	3.09%	3.08%	3.18%	3.40%	3.86%	-	-	-	-	-
$Tone_{Positive;Full\ Report;[X]}$	1.27%	1.20%	1.31%	1.23%	1.26%	1.17%	1.27%	-	-	-	-	-
$Tone_{Net;Full\ Report;[X]}$	1.51%	1.66%	1.77%	1.86%	1.92%	2.23%	2.59%	-	-	-	-	-
$Tone_{Uncertainty;Full\ Report;[X]}$	2.16%	2.15%	2.12%	2.19%	2.18%	2.10%	2.25%	-	-	-	-	-
$Tone_{Litigious;Full\ Report;[X]}$	3.44%	3.69%	3.81%	3.54%	3.70%	4.17%	3.72%	-	-	-	-	-
$Tone_{Constraining;Full\ Report;[X]}$	1.59%	1.55%	1.51%	1.57%	1.59%	1.63%	1.72%	-	-	-	-	-

Notes: The table provides descriptive statistics for different types of disclosure tones in Form 10-K filings. Panel A illustrates the Form 10-K disclosure tones of bankrupt and non-bankrupt firms. Panel B illustrates the Form 10-K disclosure tones of bankrupt firms over time.

4.4 Empirical results

The regression coefficient on negative disclosure tone is significantly positive. Consistent with prior empirical findings (e.g. Henselmann and Scherr 2013; Lopatta et al. 2017) a higher level of negative Form 10-K textual sentiment is associated with a firm's bankruptcy in the following year (*Hypothesis 5*). In addition, this result holds true when considering pessimistic information disclosed in annual reports over a longer period before filing for bankruptcy. However, considering the explanatory power of negative disclosure tone in annual reports two years before filing for bankruptcy, no negative disclosure tone was found to be associated with a firm's future development. Nevertheless, these results provide evidence that annual reports on Form 10-K and negative disclosure tone in particular contain important and value-relevant information not only for researchers, but also for investors, auditors, and regulators. Panel A of Table 40 presents the estimated regression coefficients for negative Form 10-K disclosure tone.

The regression coefficient for positive disclosure tone is significantly negative. In general, a higher level of positive Form 10-K textual sentiment is related to the survival of a firm (*Hypothesis 6a*). However, considering the explanatory power of positive information one year prior to the bankruptcy, the results do not indicate any relationship between optimistic tone and the survival of a firm. This result is interesting in light of potential management obfuscation in regulatory filings. Company officials might intentionally disclose the same amount of positive information as in previous filings rather than present the appropriate economic situation of their firms. Despite this assumption, the regression results as well as prior empirical results (e.g. Shirata and Sakagami 2008; Shirata et al. 2011) show that positive Form 10-K disclosure tone can be a supplement for predicting future corporate bankruptcies. Panel B of Table 40 presents the estimated regression coefficients for positive Form 10-K disclosure tone.

The regression coefficient on overall (net) disclosure tone is significantly positive (negative minus positive disclosure tone). A higher level of overall (net) Form 10-K textual sentiment is associated with a firm's bankruptcy in the following year (*Hypothesis 6b*). In addition, overall or net textual sentiment in Form 10-K filings is also associated with future firm bankruptcies when considered over a longer period prior to filing for bankruptcy. Again, these results confirm that the overall (net) disclosure tone in Form 10-K filings contains important and value-relevant information for researchers, investors, auditors, and regulators. Panel C of Table 40 presents the estimated regression coefficients on the overall (net) Form 10-K disclosure tone.

Table 40. Regression coefficients on different Form 10-K disclosure tones (negative, positive, net) for corporate bankruptcy.

Panel A: Negative disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
$\textbf{Tone}_{\textbf{Neg;Full Report}}$	+	**24.43****	**66.48*****	**15.50**	**22.94**	**11.35**	**17.92**	**-2.407**	**258.3**
		(2.544)	**(2.724)**	**(0.728)**	**(0.900)**	**(0.527)**	**(0.605)**	**(-0.0722)**	**(1.388)**
Constant		-0.777**	-2.353***	-0.515	-0.707	-0.343	-0.540	0.0690	-6.454
		(-2.442)	(-2.686)	(-0.700)	(-0.865)	(-0.498)	(-0.583)	(0.0694)	(-1.352)
N		476	102	94	76	78	62	54	8
Pseudo R^2		0.0102	0.0657	0.00413	0.00790	0.00260	0.00437	6.96e-05	0.244
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel B: Positive disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
$\textbf{Tone}_{\textbf{Pos;Full Report}}$	**-**	**-121.9*****	**-63.57**	**-203.5*****	**-179.6*****	**-138.4****	**-15.42**	**-125.3**	**-222.2**
		(-4.237)	**(-0.922)**	**(-2.591)**	**(-2.583)**	**(-2.079)**	**(-0.222)**	**(-1.225)**	**(-1.223)**
Constant		1.589***	0.826	2.535**	2.466**	1.821**	0.203	1.557	3.312
		(4.137)	(0.901)	(2.556)	(2.531)	(2.031)	(0.214)	(1.204)	(1.192)
N		476	102	94	76	78	62	54	8
Pseudo R^2		0.0295	0.00609	0.0606	0.0732	0.0459	0.000576	0.0219	0.174
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel C: Overall (negative – positive) disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
$\textbf{Tone}_{\textbf{Net;Full Report}}$	+	**32.73*****	**57.26*****	**30.30**	**48.43***	**23.98**	**14.05**	**12.27**	**345.9**
		(3.649)	**(2.730)**	**(1.450)**	**(1.875)**	**(1.199)**	**(0.562)**	**(0.406)**	**(1.067)**
Constant		-0.611***	-1.287**	-0.626	-0.815*	-0.406	-0.238	-0.197	-4.028
		(-3.210)	(-2.546)	(-1.312)	(-1.665)	(-0.999)	(-0.482)	(-0.354)	(-0.949)
N		476	102	94	76	78	62	54	8
Pseudo R^2		0.0217	0.0633	0.0172	0.0371	0.0139	0.00373	0.00221	0.377
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Notes: The table shows the regression coefficients for different Form 10-K disclosure tones for future firm bankruptcies. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (goodness-of-fit, precision). The Hosmer-Lemeshow test for goodness-of-fit yielded a p-value > 5% when using negative and positive textual sentiment and a p-value < 5% when using overall or net textual sentiment. The lowest precision of all baseline regressions were 57.77%. Based on these results the study used an logistic regression (logit) model to examine the dependent variable.

4.5 Robustness tests

4.5.1 Sample selection

To test whether the empirical results of this study are robust, a series of robustness tests was conducted. As a first robustness test, an additional data filter to the data sample and overall sample selection process was applied. From the selected sample, firm-year observations of the financial industry (SIC codes 6021 - 6799) were excluded. Substantial accounting differences between financial and non-financial entities as well as differences in their regulatory framework might affect the corresponding narrative disclosures and therefore the empirical results (Ditter and Scherr 2015, 60). Based on this restricted data sample, the analysis was rerun. The regression results confirmed the findings. As before, different types of disclosure tones in the annual reports on Form 10-K are associated with future firm bankruptcies in the expected manner. While there is a positive relationship between negative tone in annual reports and future firm bankruptcies, the ratio of positive statements is negatively associated with the future demise of corporations. The overall (net) disclosure tone in annual reports is also positively associated with future firm bankruptcies. Table 41 presents the estimated regression coefficients for the disclosure tones in Form 10-K.

Table 41. Regression coefficients on different Form 10-K disclosure tones for corporate bankruptcy excluding firm-year observations of the financial industry.

Years to bankruptcy	Exp. sign	1-7	1-7	1-7
$\mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	+	**22.64****		
		(2.152)		
$\mathbf{Tone}_{\mathbf{Pos;Full\ Report}}$	-		**-120.7*****	
			(-4.090)	
$\mathbf{Tone}_{\mathbf{Net;Full\ Report}}$	+			**32.46*****
				(3.336)
Constant		-0.713**	1.590***	-0.592***
		(-2.071)	(3.991)	(-2.939)
N		436	436	436
Pseudo R^2		0.00789	0.0300	0.0195
Model		Logit	Logit	Logit

Notes: The table shows the regression coefficients for different Form 10-K disclosure tones for future firm bankruptcies. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words).

4.5.2 Disclosure length

The results could be influenced by how the overall text length of annual reports on Form 10-K (Loughran and McDonald business dictionary excluding stop words) was measured in this study. Loughran and McDonald (2011a) and Loughran and McDonald (2014) originally counted the total number of words (more than two characters) appearing in an annual report (including stop words) to determine the entire text length of a certain Form 10-K filing (Loughran and McDonald 2011b, 2; Loughran and McDonald 2014, 1669). To examine whether the results presented in this study are sensitive to the document length of an annual report, a modified word count was obtained for the overall text length of a Form 10 K filing (Loughran and McDonald business dictionary including stop words) for each submission in the data sample, and the analysis was rerun. The regression results confirmed the findings. Based on the Loughran and McDonald business dictionary (including stop words), there is a positive association between negative Form 10-K disclosure tone and a firm's future corporate bankruptcy. In addition, the regression results again show a negative relationship between positive textual sentiment in annual reports and future firm bankruptcies. Furthermore, overall (net) annual report sentiment is again positively associated with future firm bankruptcies. Table 42 presents the estimated regression coefficients for the disclosure tones in Form 10-K.

Table 42. Regression coefficients on different Form 10-K disclosure tones for corporate bankruptcy using the Loughran and McDonald business dictionary including stop words.

Years to bankruptcy	Exp. sign	1-7	1-7	1-7
$\mathbf{Tone_{Neg;Full\ Report}}$	+	**31.58***		
		(1.785)		
$\mathbf{Tone_{Pos;Full\ Report}}$	-		**-251.1*****	
			(-4.737)	
$\mathbf{Tone_{Net;Full\ Report}}$	+			**54.09*****
				(3.230)
Constant		-0.528*	1.723***	-0.530***
		(-1.707)	(4.623)	(-2.828)
N		476	476	476
Pseudo R^2		0.00492	0.0376	0.0168
Model		Logit	Logit	Logit

Notes: The table shows the regression coefficients for different Form 10-K disclosure tones for future firm bankruptcies. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (including stop words).

4.5.3 Table content

The results of this study could further be influenced by the fact that HTML tables and their corresponding content were excluded from the analysis. EDGAR filers might use HTML table tags (<Table>) to structure and disclose certain textual information in annual reports on Form 10-K filed with the SEC. Thus, the exclusion of tables in financial statements and their information content might affect the results. To address concerns that tables contain relevant narrative information when examining the explanatory power of disclosure tone on future firm bankruptcies, tables and their corresponding content were included and the analysis was rerun. The regression results confirmed the findings. When tables and their corresponding content were included, the regression coefficients for negative, positive, and overall (net) disclosure tone for the annual report on Form 10-K were unchanged. While negative and overall (net) textual sentiment is positively associated with future firm bankruptcies, positive statements in annual reports are negatively associated with the future demise of firms. Table 43 presents the estimated regression coefficients for the disclosure tones in Form 10-K. Based on these results, the study concludes that all presented findings seem unaffected by the series of robustness tests.

Table 43. Regression coefficients on different Form 10-K disclosure tones for corporate bankruptcy including table content.

Years to bankruptcy	Exp. sign	1-7	1-7	1-7
$\mathbf{Tone_{Neg;Full\ Report}}$	+	**44.59***** **(4.071)**		
$\mathbf{Tone_{Pos;Full\ Report}}$	-		**-135.5***** **(-4.492)**	
$\mathbf{Tone_{Net;Full\ Report}}$	+			**51.70***** **(5.074)**
Constant		-1.400*** (-3.954)	1.754*** (4.394)	-0.950*** (-4.574)
N		476	476	476
Pseudo R^2		0.0279	0.0333	0.0448
Model		Logit	Logit	Logit

Notes: The table shows the regression coefficients for different Form 10-K disclosure tones for future firm bankruptcies. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words).

4.6 Sensitivity tests

4.6.1 Type of disclosure tone

Few studies in the accounting and finance literature have been able to connect different types of linguistic tone in annual reports on Form 10-K to important firm characteristics in the expected manner (e.g. Henselmann and Hering 2017). Apart from negative, positive, and overall (net) textual sentiment, possibly, other (more specific) types of disclosure tone are also associated with future corporate bankruptcies (e.g. *"uncertainty"*, *"litigious"*, *"constraining"*). To examine the explanatory power of other types of disclosure tone in Form 10-K filings in the context of a firm's future bankruptcy, the analysis in this study was rerun using the Loughran and McDonald business dictionary and its word lists

The regression results showed that the explanatory power of other types of textual sentiment in financial statements is limited. For instance, the regression coefficient on uncertain textual sentiment is not significant. This result is consistent with prior empirical findings (e.g. Cecchini et al. 2010). Legal aspects in annual reports seem to be associated with future firm failures only one year prior to filing for bankruptcy. This result is also consistent with prior empirical findings (e.g. Lopatta et al. 2017). However, litigious statements in Form 10-K filings cannot be connected to future firm bankruptcies over a longer period. Actually, there is a significant positive association between constraining textual information and future firm bankruptcies. Table 44 presents the estimated regression coefficients for uncertain, litigious, and constraining text sentiments in Form 10-K.

Table 44. Regression coefficients on different Form 10-K disclosure tones (uncertainty, litigious, constraining) for corporate bankruptcy.

Panel A: Uncertain disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
$Tone_{Unc;Full\ Report}$	+	**-8.364**	**-50.09**	**-35.80**	**29.50**	**-1.831**	**-7.937**	**52.01**	**119.7**
		(-0.569)	**(-1.452)**	**(-1.123)**	**(0.752)**	**(-0.0506)**	**(-0.200)**	**(1.044)**	**(0.971)**
Constant		0.183	1.166	0.779	-0.627	0.0402	0.170	-1.076	-2.287
		(0.547)	(1.408)	(1.077)	(-0.725)	(0.0487)	(0.192)	(-1.010)	(-0.927)
N		476	102	94	76	78	62	54	8
Pseudo R^2		0.000490	0.0153	0.00982	0.00541	2.37e-05	0.000468	0.0150	0.0996
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel B: Litigious disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
$Tone_{Lit;Full\ Report}$	+	**5.579**	**20.19***	**13.10**	**3.700**	**0.0779**	**-0.0231**	**-13.19**	**-7.393**
		(1.180)	**(1.802)**	**(1.212)**	**(0.331)**	**(0.00658)**	**(-0.00172)**	**(-0.884)**	**(-0.245)**
Constant		-0.205	-0.686	-0.514	-0.134	-0.00276	0.000881	0.516	0.269
		(-1.044)	(-1.610)	(-1.091)	(-0.288)	(-0.00579)	(0.00154)	(0.801)	(0.206)
N		476	102	94	76	78	62	54	8
Pseudo R^2		0.00212	0.0244	0.0115	0.00104	4.00e-07	3.46e-08	0.0106	0.00548
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel C: Constraining disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
$Tone_{Con;Full\ Report}$	+	**221.2*** **	**215.5*** **	**245.8*** **	**211.4*** **	**353.7*** **	**156.5** **	**179.3** **	**238.8**
		(7.726)	**(3.600)**	**(3.520)**	**(3.302)**	**(3.729)**	**(2.036)**	**(2.334)**	**(0.854)**
Constant		-3.151***	-3.282***	-3.533***	-2.915***	-4.893***	-2.202**	-2.478**	-3.091
		(-7.633)	(-3.576)	(-3.499)	(-3.243)	(-3.680)	(-1.993)	(-2.292)	(-0.885)
N		476	102	94	76	78	62	54	8
Pseudo R^2		0.119	0.121	0.133	0.132	0.191	0.0543	0.0921	0.143
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Notes: The table shows the regression coefficients for different Form 10-K disclosure tones for future firm bankruptcies. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (goodness-of-fit, precision). The Hosmer-Lemeshow test for goodness-of-fit yielded a p-value > 5% when using litigious and constraining textual sentiment and a p-value < 5% when using uncertain textual sentiment. The lowest precision of all baseline regressions were 49.58% (uncertain sentiment) The highest precision of all baseline regressions were 68.91% (constraining sentiment). Based on these results the study used an logistic regression (logit) model to examine the dependent variable.

4.6.2 Measurement of disclosure tone

The accounting and finance literature uses various methods to quantify the disclosure tone in text documents. For instance, numerous studies employed the ratio of certain word categories (e.g. *"negative"*, *"positive"*) in relation to the entire text length to measure the disclosure tone of narrative information (tone level) (e.g. Loughran and McDonald 2011a; Gandhi et al. 2019). Other empirical works examined the effect of tone changes on future firm characteristics (tone differences) (e.g. Chouliaras 2015; Henry and Leone 2016).

To examine the explanatory power of disclosure tone in annual reports on a firm's future bankruptcy based on different tone measurements, the following logistic regression model was estimated:

$$Y_{i;t+1} = \beta_0 + \beta_1 * \Delta\, Tone_{Type;Full\ Report;i;t} + \varepsilon_{i;t} \qquad (23)$$

Where: the dependent variable ($Y_{i;t+1}$) is a binary variable equal to one if a firm files for bankruptcy or zero if it does not. $\Delta\, Tone_{Type;Full\ Report;i;t}$ is the change in textual sentiment in the entire annual report on Form 10-K (submission including exhibits) as measured by the difference in disclosure tone between the current and previous filing.

Interestingly, while the level of the disclosure tone in annual reports across companies is associated with future firm bankruptcies, the change in Form 10-K textual sentiment is not associated with their future demise. Despite showing the expected sign, the regression coefficients for different types of disclosure tone are insignificant. Based on these results, the study concludes that the change in Form 10-K textual sentiment is not related to future firm bankruptcies. Table 45 presents the estimated regression coefficients for different Form 10-K disclosure tones.

Table 45. Regression coefficients on different Form 10-K disclosure tones (changes) for corporate bankruptcy.

Panel A: Negative disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ **Tone**$_{\mathbf{Neg;Full\ Report}}$	+	**15.92**	**43.87**	**7.452**	**57.82**	**-44.01**	**40.04**	**-43.75**	**236.8**
		(1.125)	**(1.358)**	**(0.250)**	**(1.587)**	**(-1.157)**	**(0.986)**	**(-0.911)**	**(1.138)**
Constant		-0.0146	-0.0889	-0.00359	-0.0913	0.0184	0.0470	0.0580	-0.343
		(-0.152)	(-0.412)	(-0.0164)	(-0.366)	(0.0783)	(0.166)	(0.202)	(-0.421)
N		438	96	84	70	74	52	52	8
Pseudo R^2		0.00210	0.0144	0.000538	0.0280	0.0135	0.0139	0.0119	0.155
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel B: Positive disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ **Tone**$_{\mathbf{Pos;Full\ Report}}$	-	**-12.85**	**50.66**	**-62.31**	**-51.03**	**-33.91**	**35.26**	**-64.29**	**207.0**
		(-0.312)	**(0.567)**	**(-0.674)**	**(-0.430)**	**(-0.317)**	**(0.352)**	**(-0.475)**	**(0.518)**
Constant		0.00126	0.00376	-0.0234	0.0387	-0.00485	-0.0226	0.00660	0.0720
		(0.0132)	(0.0184)	(-0.106)	(0.151)	(-0.0208)	(-0.0792)	(0.0237)	(0.0982)
N		438	96	84	70	74	52	52	8
Pseudo R^2		0.000160	0.00243	0.00395	0.00192	0.000983	0.00173	0.00315	0.0252
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel C: Overall (negative – positive) disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ **Tone**$_{\mathbf{Net;Full\ Report}}$	+	**13.63**	**32.44**	**14.63**	**52.13**	**-32.35**	**11.45**	**-27.98**	**197.1**
		(1.055)	**(1.127)**	**(0.499)**	**(1.576)**	**(-0.931)**	**(0.342)**	**(-0.666)**	**(1.018)**
Constant		-0.0109	-0.0689	-0.0113	-0.0397	0.0170	0.0206	0.0341	-0.384
		(-0.113)	(-0.322)	(-0.0514)	(-0.162)	(0.0723)	(0.0725)	(0.121)	(-0.459)
N		438	96	84	70	74	52	52	8
Pseudo R^2		0.00184	0.00977	0.00215	0.0277	0.00866	0.00163	0.00627	0.111
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel D: Uncertain disclosure tone									
Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ **Tone**$_{\text{Unc;Full Report}}$	+	**-6.082**	**-34.73**	**14.44**	**-13.12**	**44.35**	**-95.44**	**53.45**	**-127.8**
		(-0.270)	**(-0.726)**	**(0.255)**	**(-0.214)**	**(0.754)**	**(-1.357)**	**(0.987)**	**(-0.574)**
Constant		0.00208	0.00655	0.000407	0.00717	-0.0361	0.0222	-0.0254	-0.101
		(0.0217)	(0.0320)	(0.00187)	(0.0297)	(-0.152)	(0.0783)	(-0.0902)	(-0.135)
N		438	96	84	70	74	52	52	8
Pseudo R^2		0.000120	0.00401	0.000560	0.000474	0.00562	0.0273	0.0139	0.0324
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel E: Litigious disclosure tone									
Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ **Tone**$_{\text{Lit;Full Report}}$	+	**0.991**	**6.623**	**-1.022**	**8.877**	**-14.16**	**7.914**	**-8.942**	**37.94**
		(0.216)	**(0.676)**	**(-0.0904)**	**(0.740)**	**(-1.179)**	**(0.625)**	**(-0.732)**	**(0.803)**
Constant		0.000665	0.0129	0.000194	0.00395	-0.00566	0.0202	0.00622	-0.294
		(0.00695)	(0.0628)	(0.000890)	(0.0165)	(-0.0241)	(0.0720)	(0.0223)	(-0.360)
N		438	96	84	70	74	52	52	8
Pseudo R^2		7.68e-05	0.00347	7.02e-05	0.00573	0.0140	0.00550	0.00757	0.0682
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel F: Constraining disclosure tone									
Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ **Tone**$_{\text{Con;Full Report}}$	+	**-1.887**	**-38.78**	**19.64**	**31.03**	**23.25**	**-51.60**	**-23.93**	**239.8**
		(-0.0836)	**(-0.774)**	**(0.339)**	**(0.517)**	**(0.416)**	**(-0.800)**	**(-0.427)**	**(0.951)**
Constant		0.000243	0.0218	0.00566	-0.0116	0.00752	-0.00722	0.0117	-0.339
		(0.00255)	(0.105)	(0.0258)	(-0.0481)	(0.0322)	(-0.0258)	(0.0418)	(-0.414)
N		438	96	84	70	74	52	52	8
Pseudo R^2		1.15e-05	0.00456	0.000991	0.00278	0.00170	0.00908	0.00254	0.125
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Notes: The table shows the regression coefficients for different Form 10-K disclosure tones (change) for future firm bankruptcies. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (goodness-of-fit, precision). The Hosmer-Lemeshow test for goodness-of-fit yielded a p-value > 5% for all types of textual sentiment except when using constraining textual sentiment (p-value < 5%). The lowest precision of all baseline regressions were 47.49% (change uncertain sentiment) The highest precision of all baseline regressions were 53.42% (change negative sentiment). Based on these results the study used an logistic regression (logit) model to examine the dependent variable.

Capital market research has shown that disclosure tone changes can be predictive for important firm characteristics in a capital market setting (Miwa 2020, 2). Despite showing no explanatory power on future firm bankruptcies, possibly, disclosure tone changes might have explanatory power regarding future firm developments after controlling for the specific tone level of narrative information. To further examine the possible explanatory power of disclosure tone changes on future corporate bankruptcies and account for the tone level of accounting narratives, the following logistic regression model was estimated:

$$Y_{i;t+1} = \beta_0 + \beta_1 * \Delta\, Tone_{Type;Full\ Report;i;t} + \beta_2 *\ Tone_{Type;Full\ Report;i;t} + \varepsilon_{i;t} \qquad (24)$$

Where: the dependent variable ($Y_{i;t+1}$) is a binary variable equal to one if a firm files for bankruptcy or zero if it does not. $\Delta\, Tone_{Type;Full\ Report;i;t}$ is the change in textual sentiment in the entire annual report on Form 10-K (submission including exhibits) as measured by the difference in disclosure tone between the current and previous filing. $Tone_{Type;Full\ Report;i;t}$ is the textual sentiment in the entire annual report on Form 10-K (submission including exhibits) as measured by the ratio of a specific disclosure tone type (e.g. *"negative"*, *"positive"*, *"uncertainty"*, *"litigious"*, *"constraining"*).

In general, the regression results showed that disclosure tone changes in annual reports on Form 10-K have no significant association with future firm bankruptcies even when accounting for the tone level of qualitative information. While the amount (tone level) of negative, positive, overall (net), and constraining textual sentiment can be connected to future firm bankruptcies in the expected manner, their corresponding changes cannot be (tone differences). Based on these results, the study again concludes that the change in Form 10-K textual sentiment is not associated with future firm bankruptcies. Table 46 presents the estimated regression coefficients for different disclosure tones in Form 10-K.

Table 46. Regression coefficients on different Form 10-K disclosure tones (changes) after controlling for tone levels.

Panel A: Negative disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ Tone$_{Neg;Full Report}$	+	**-2.246**	**-2.720**	**-6.548**	**47.40**	**-54.07**	**37.73**	**-61.73**	**-38.70**
		(-0.143)	**(-0.0711)**	**(-0.201)**	**(1.189)**	**(-1.292)**	**(0.856)**	**(-1.209)**	**(-0.113)**
Tone$_{Neg;Full Report}$	+	30.55***	71.74***	26.71	19.22	14.44	4.323	45.58	287.1
		(2.752)	(2.579)	(1.077)	(0.655)	(0.588)	(0.134)	(1.222)	(0.897)
Constant		-0.989***	-2.540***	-0.893	-0.685	-0.439	-0.0858	-1.290	-7.112
		(-2.703)	(-2.622)	(-1.048)	(-0.729)	(-0.541)	(-0.0833)	(-1.133)	(-0.931)
N		438	96	84	70	74	52	52	8
Pseudo R^2		0.0153	0.0758	0.0109	0.0325	0.0169	0.0142	0.0334	0.245
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel B: Positive disclosure tone

Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ Tone$_{Pos;Full Report}$	**-**	**65.29**	**101.6**	**-12.82**	**146.5**	**29.97**	**138.8**	**-22.17**	**295.4**
		(1.415)	**(1.040)**	**(-0.131)**	**(0.997)**	**(0.250)**	**(1.089)**	**(-0.156)**	**(0.712)**
Tone$_{Pos;Full Report}$	-	-141.8***	-119.0	-160.2*	-234.1***	-84.92	-139.8	-136.1	-252.2
		(-4.422)	(-1.496)	(-1.943)	(-2.863)	(-1.255)	(-1.471)	(-1.443)	(-1.242)
Constant		1.851***	1.536	2.011*	3.134***	1.099	1.785	1.782	3.850
		(4.338)	(1.471)	(1.891)	(2.838)	(1.213)	(1.426)	(1.422)	(1.231)
N		438	96	84	70	74	52	52	8
Pseudo R^2		0.0355	0.0197	0.0398	0.104	0.0172	0.0347	0.0349	0.222
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel C: Overall (negative – positive) disclosure tone									
Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ **Tone**$_{\textbf{Net;Full Report}}$	+	**-10.48**	**-13.31**	**-4.380**	**20.72**	**-45.58**	**-1.332**	**-41.80**	**-1,472**
		(-0.723)	**(-0.382)**	**(-0.138)**	**(0.542)**	**(-1.199)**	**(-0.0356)**	**(-0.952)**	**(-1.066)**
$\text{Tone}_{\text{Net;Full Report}}$	+	40.48***	68.07***	37.31	54.20*	19.56	21.64	49.11	3,117
		(3.918)	(2.723)	(1.564)	(1.750)	(0.895)	(0.761)	(1.497)	(1.050)
Constant		-0.770***	-1.515***	-0.773	-0.974*	-0.347	-0.357	-0.783	-36.76
		(-3.575)	(-2.671)	(-1.453)	(-1.666)	(-0.741)	(-0.627)	(-1.268)	(-1.044)
N		438	96	84	70	74	52	52	8
Pseudo R^2		0.0297	0.0784	0.0251	0.0640	0.0167	0.00994	0.0396	0.595
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel D: Uncertain disclosure tone									
Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ **Tone**$_{\textbf{Unc;Full Report}}$	+	**-8.584**	**-21.76**	**21.94**	**-30.85**	**52.30**	**-115.7**	**12.81**	**-613.6**
		(-0.359)	**(-0.436)**	**(0.371)**	**(-0.477)**	**(0.848)**	**(-1.506)**	**(0.203)**	**(-0.958)**
$\text{Tone}_{\text{Unc;Full Report}}$	+	5.202	-36.80	-15.70	37.07	-19.17	34.07	76.61	331.3
		(0.310)	(-0.988)	(-0.433)	(0.872)	(-0.455)	(0.702)	(1.282)	(1.309)
Constant		-0.113	0.859	0.345	-0.806	0.392	-0.713	-1.682	-7.078
		(-0.295)	(0.968)	(0.418)	(-0.836)	(0.404)	(-0.657)	(-1.273)	(-1.218)
N		438	96	84	70	74	52	52	8
Pseudo R^2		0.000278	0.0114	0.00217	0.00840	0.00765	0.0342	0.0377	0.334
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel E: Litigious disclosure tone									
Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ $\mathbf{Tone_{Lit;Full\ Report}}$	+	**-3.091**	**-8.668**	**-6.235**	**9.013**	**-13.74**	**12.14**	**-15.62**	**-**
		(-0.559)	**(-0.664)**	**(-0.519)**	**(0.632)**	**(-0.916)**	**(0.772)**	**(-1.002)**	**-**
$\mathrm{Tone_{Lit;Full\ Report}}$	+	8.117	28.01*	17.49	-0.246	-0.750	-8.883	13.14	-
		(1.328)	(1.787)	(1.364)	(-0.0176)	(-0.0469)	(-0.455)	(0.699)	-
Constant		-0.297	-0.961*	-0.675	0.0130	0.0209	0.360	-0.490	-
		(-1.219)	(-1.656)	(-1.249)	(0.0229)	(0.0341)	(0.451)	(-0.643)	-
N		438	96	84	70	74	52	52	-
Pseudo R^2		0.00300	0.0290	0.0165	0.00574	0.0140	0.00838	0.0145	-
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel F: Constraining disclosure tone									
Years to bankruptcy	Exp. sign	1-7	1	2	3	4	5	6	7
Δ $\mathbf{Tone_{Con;Full\ Report}}$	+	**-156.8*** **	**-286.8*** **	**-61.72**	**-222.7** **	**-205.6** **	**-161.7* **	**-104.7**	**9.242**
		(-4.916)	**(-3.171)**	**(-0.922)**	**(-2.291)**	**(-2.285)**	**(-1.850)**	**(-1.474)**	**(0.0159)**
$\mathrm{Tone_{Con;Full\ Report}}$	+	312.6***	404.5***	312.8***	344.0***	618.7***	193.4**	218.1**	233.7
		(8.537)	(4.423)	(3.627)	(3.553)	(4.125)	(2.091)	(2.552)	(0.416)
Constant		-4.476***	-5.886***	-4.485***	-4.943***	-8.675***	-2.672**	-3.133**	-3.033
		(-8.446)	(-4.431)	(-3.609)	(-3.505)	(-4.083)	(-2.052)	(-2.495)	(-0.462)
N		438	96	84	70	74	52	52	8
Pseudo R^2		0.167	0.239	0.165	0.187	0.317	0.0779	0.121	0.143
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit

Notes: The table shows the regression coefficients for different Form 10-K disclosure tones and disclosure tone changes for future firm bankruptcies. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). For the dependent variable and corresponding regression model diagnostics were conducted (goodness-of-fit, precision, multicollinearity of predictors). The Hosmer-Lemeshow test for goodness-of-fit yielded a p-value < 5% for all baseline regressions except when using litigious textual sentiment (p-value < 5%). The lowest precision of all baseline regressions were 51.14% (change uncertain sentiment) The highest precision of all baseline regressions were 72.37% (change constraining sentiment). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an logistic regression (logit) model to examine the dependent variable.

4.7 Additional tests

4.7.1 Relevance of text sections

Apart from the entire annual report, numerous studies examine the information content and value-relevance of specific text sections within the overall filing (e.g. Loughran and McDonald 2011a; Amel-Zadeh and Faasse 2016; Henselmann and Hering 2017). Annual reports on Form 10-K contain more than 20 text sections (*"Items"*), each assigned to certain aspects of the company (e.g. *"Item 1 - Business"*, *"Item 1A - Risk Factors"*, *"Item 3 - Legal Proceedings"*, *"Item 5 - Shareholder Matters"*) (SEC 2009). However, the majority of narrative information within a Form 10-K filing is disclosed in a few text sections (e.g. *"Item 7 - Management's Discussion and Analysis of Financial Condition and Results of Operations"*, *"Item 8 - Financial Statements and Supplementary Data"*) (Hering 2016, 48-49; Henselmann and Hering 2017, 99; Cazier and Pfeiffer, 2016, 8; Li 2008, 227; Amel-Zadeh and Faasse 2016, 4, 12). For instance, *"Item 7"* the *"MD&A"* section provides important forward-looking information (e.g., corporate earnings, cash flows) and contextual disclosures (Tavcar 1998, 10, 24-25; Humpherys et al. 2011, 585; Czyzewski and Wilkinson 2014, 99-100; Rogers and Grant 1997, 17-18; Amel-Zadeh and Faasse 2016, 3, 7; SEC 2003; SEC 2003b). *"Item 8"*, *"Footnotes"*, contains important narrative disclosures regarding a firm's financial performance and future risks (Leder 2003, 17; Heidari and Felden 2015, 1; Amel-Zadeh and Faasse 2016, 25; Czerney et al. 2017, 8-9). Consistently, research has provided evidence that certain *"Items"* in annual reports are more important to investors than others (Sutton et al. 2009, 5; Bedard et al. 2012, 25-26).

To examine the explanatory power of the disclosure tone in specific text sections of the annual report regarding a firm's future bankruptcy, the following logistic regression model was estimated:

$$Y_{i;t+1} = \beta_0 + \beta_1 * Tone_{Type;Full\ Report;i;t} + \beta_2 * Tone_{Type;Section\ X;i;t} + \varepsilon_{i;t} \quad (25)$$

Where: the dependent variable ($Y_{i;t+1}$) is a binary variable equal to one if a firm files for bankruptcy or zero if it does not. $Tone_{Type;Full\ Report;i;t}$ is the specific textual sentiment or linguistic tone in the entire annual report on Form 10-K (submission including exhibits). $Tone_{Type;Section\ X;i;t}$ is the specific textual sentiment or linguistic tone of an individual text section within the Form 10-K filing.

Note that the study also included the textual sentiment of the entire annual report filing (Form 10-K submission including exhibits) as a control variable to proxy for textual characteristics in the entire SEC submission as suggested in the literature (Loughran and McDonald 2016, 1218).

In general, the regression results revealed that the disclosure tone of the Form 10-K filing in its entirety (submission including exhibits) has a higher explanatory power for future firm bankruptcies than the linguistic tone of specific annual report subsections. However, the regression coefficients on negative disclosure tone for *"Item 1 - Business"* and *"Item 7 - Management's Discussion and Analysis of Financial Condition and Results of Operations"* are significantly positive. The tone of narrative information in *"Item 1"* and *"Item 7"* seems to have more explanatory power on future firm bankruptcies than the textual sentiment of the overall filing. This result is consistent with prior empirical findings about the importance of specific text sections within Form 10-K filings in explaining future corporate bankruptcy (e.g. Tennyson et al. 1990; Cecchini et al. 2010; Mayew et al. 2015; Yang et al. 2018). Interestingly, the regression coefficients for *"Item 1A - Risk Factors"* and *"Item 8 - Financial Statements and Supplementary Data"* are either not significant or smaller in magnitude compared to the regression coefficients for the overall filing. Table 47 presents the estimated regression coefficients on different disclosure tones for specific annual report subsections in Form 10-K.

Table 47. Regression coefficients on different disclosure tones in specific Form 10-K sections for corporate bankruptcy.

Panel A: Negative disclosure tone							
Variable	Exp. sign	1-7	1-7	1-7	1-7	1-7	1-7
$Tone_{Neg;Full\ Report}$	+	9.642	80.47*	13.82	30.49**	14.06	33.39**
		(0.678)	(1.915)	(1.098)	(2.205)	(0.920)	(2.292)
$\mathbf{Tone_{Neg;Item\ 1}}$	+	**21.20***					
		(1.787)					
$\mathbf{Tone_{Neg;Item\ 1A}}$	+		**-12.95**				
			(-1.035)				
$\mathbf{Tone_{Neg;Item\ 3}}$	+			**2.427**			
				(1.250)			
$\mathbf{Tone_{Neg;Item\ 5}}$	+				**-5.318**		
					(-0.408)		
$\mathbf{Tone_{Neg;Item\ 7}}$	+					**31.50*****	
						(3.668)	
$\mathbf{Tone_{Neg;Item\ 8}}$	+						**1.987**
							(0.229)
Constant		-0.635	-1.928	-0.651*	-0.867**	-0.955**	-1.008**
		(-1.587)	(-1.380)	(-1.676)	(-2.015)	(-2.150)	(-2.344)
N		278	68	296	260	254	288
Pseudo R^2		0.0162	0.0440	0.00974	0.0142	0.0574	0.0158
Model		Logit	Logit	Logit	Logit	Logit	Logit

Panel B: Positive disclosure tone							
Variable	Exp. sign	1-7	1-7	1-7	1-7	1-7	1-7
$Tone_{Pos;Full\ Report}$	-	-180.3***	-31.91	-93.87***	-93.20**	-128.2***	-119.0***
		(-4.022)	(-0.480)	(-2.616)	(-2.344)	(-3.084)	(-3.238)
$\mathbf{Tone_{Pos;Item\ 1}}$	-	**99.05*****					
		(4.711)					
$\mathbf{Tone_{Pos;Item\ 1A}}$	-		**-25.60**				
			(-0.848)				
$\mathbf{Tone_{Pos;Item\ 3}}$	-			**-13.72**			
				(-0.567)			
$\mathbf{Tone_{Pos;Item\ 5}}$	-				**-0.421**		
					(-0.0220)		
$\mathbf{Tone_{Pos;Item\ 7}}$	-					**23.50**	
						(1.640)	
$\mathbf{Tone_{Pos;Item\ 8}}$	-						**-6.115**
							(-0.296)
Constant		0.810	0.960	1.240**	1.184**	1.405***	1.559***
		(1.533)	(1.076)	(2.576)	(2.282)	(2.622)	(3.206)
N		278	68	296	260	254	288
Pseudo R^2		0.0874	0.0190	0.0180	0.0161	0.0343	0.0294
Model		Logit	Logit	Logit	Logit	Logit	Logit

Panel C: Overall (negative – positive) disclosure tone							
Variable	Exp. sign	1-7	1-7	1-7	1-7	1-7	1-7
$Tone_{Net;Full\ Report}$	+	34.92***	58.96*	20.79*	35.83***	16.81	41.19***
		(2.614)	(1.842)	(1.790)	(2.781)	(1.170)	(3.045)
$\mathbf{Tone_{Net;Item\ 1}}$	+	**-10.31**					
		(-1.033)					
$\mathbf{Tone_{Net;Item\ 1A}}$	+		**-10.00**				
			(-0.757)				
$\mathbf{Tone_{Net;Item\ 3}}$	+			**2.179**			
				(1.105)			
$\mathbf{Tone_{Net;Item\ 5}}$	+				**-6.162**		
					(-0.541)		
$\mathbf{Tone_{Net;Item\ 7}}$	+					**48.76*** **	
						(3.811)	
$\mathbf{Tone_{Net;Item\ 8}}$	+						**5.290**
							(0.426)
Constant		-0.585**	-0.724	-0.564**	-0.584**	-0.639**	-0.720***
		(-2.310)	(-1.082)	(-2.189)	(-2.356)	(-2.491)	(-2.936)
N		278	68	296	260	254	288
Pseudo R^2		0.0188	0.0383	0.0154	0.0227	0.0760	0.0304
Model		Logit	Logit	Logit	Logit	Logit	Logit

Panel D: Uncertain disclosure tone							
Variable	Exp. sign	1-7	1-7	1-7	1-7	1-7	1-7
$Tone_{Unc;Full\ Report}$	+	-36.74*	18.01	-38.43*	-40.29*	-36.52	-19.52
		(-1.711)	(0.335)	(-1.850)	(-1.772)	(-1.519)	(-0.929)
$\mathbf{Tone_{Unc;Item\ 1}}$	+	**20.84**					
		(1.549)					
$\mathbf{Tone_{Unc;Item\ 1A}}$	+		**8.562**				
			(0.754)				
$\mathbf{Tone_{Unc;Item\ 3}}$	+			**5.827**			
				(1.273)			
$\mathbf{Tone_{Unc;Item\ 5}}$	+				**-7.293**		
					(-0.858)		
$\mathbf{Tone_{Unc;Item\ 7}}$	+					**2.070**	
						(0.280)	
$\mathbf{Tone_{Unc;Item\ 8}}$	+						**-6.480**
							(-0.776)
Constant		0.291	-0.978	0.554	0.885*	0.660	0.462
		(0.623)	(-0.619)	(1.244)	(1.908)	(1.464)	(1.101)
N		278	68	296	260	254	288
Pseudo R^2		0.0114	0.00737	0.0122	0.0115	0.00680	0.00487
Model		Logit	Logit	Logit	Logit	Logit	Logit

Panel E: Litigious disclosure tone							
Variable	Exp. sign	1-7	1-7	1-7	1-7	1-7	1-7
$Tone_{Lit;Full\ Report}$	+	7.093	-0.597	5.057	9.352	10.73*	7.906
		(1.142)	(-0.0338)	(0.889)	(1.498)	(1.675)	(1.322)
$\mathbf{Tone_{Lit;Item\ 1}}$	+	**-6.885**					
		(-0.623)					
$\mathbf{Tone_{Lit;Item\ 1A}}$	+		**-39.68***				
			(-1.706)				
$\mathbf{Tone_{Lit;Item\ 3}}$	+			**0.378**			
				(0.162)			
$\mathbf{Tone_{Lit;Item\ 5}}$	+				**-4.677**		
					(-0.589)		
$\mathbf{Tone_{Lit;Item\ 7}}$	+					**-16.31****	
						(-2.033)	
$\mathbf{Tone_{Lit;Item\ 8}}$	+						**-13.01*****
							(-2.660)
Constant		-0.183	0.886	-0.254	-0.333	-0.170	-0.0156
		(-0.647)	(1.318)	(-0.649)	(-1.158)	(-0.562)	(-0.0575)
N		278	68	296	260	254	288
Pseudo R^2		0.00365	0.0348	0.00199	0.00651	0.0170	0.0210
Model		Logit	Logit	Logit	Logit	Logit	Logit

Panel F: Constraining disclosure tone							
Variable	Exp. sign	1-7	1-7	1-7	1-7	1-7	1-7
$Tone_{Con;Full\ Report}$	+	143.3***	139.3*	139.6***	136.0***	141.4***	140.7***
		(4.381)	(1.947)	(4.480)	(4.075)	(4.166)	(4.484)
$\mathbf{Tone_{Con;Item\ 1}}$	+	**14.30**					
		(0.636)					
$\mathbf{Tone_{Con;Item\ 1A}}$	+		**21.81**				
			(0.564)				
$\mathbf{Tone_{Con;Item\ 3}}$	+			**-26.97****			
				(-2.574)			
$\mathbf{Tone_{Con;Item\ 5}}$	+				**20.48****		
					(2.465)		
$\mathbf{Tone_{Con;Item\ 7}}$	+					**12.12**	
						(0.999)	
$\mathbf{Tone_{Con;Item\ 8}}$	+						**14.62****
							(1.996)
Constant		-2.121***	-2.526**	-1.744***	-2.140***	-2.056***	-2.145***
		(-4.477)	(-2.307)	(-3.900)	(-4.479)	(-4.270)	(-4.672)
N		278	68	296	260	254	288
Pseudo R^2		0.0651	0.0668	0.0789	0.0784	0.0639	0.0696
Model		Logit	Logit	Logit	Logit	Logit	Logit

Notes: The table shows the regression coefficients for different disclosure tones for specific Form 10-K text sections. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). Note that model diagnostics for each corresponding baseline regression are discussed in Section 4.4 Empirical results.

4.7.2 Firm fundamentals

Besides having explanatory power for future firm bankruptcies on a stand-alone basis (univariate analysis), the question arises as to whether narrative information and tone thereof can also serve as a supplement in bankruptcy prediction when accounting for firm fundamentals (multivariate analysis). Prior empirical work (e.g. Lopatta et al. 2017) showed that certain firm fundamentals (e.g. liquidity) can be associated with a firm's future demise. To examine the relationship between disclosure tone and future corporate bankruptcy in the context of firm fundamentals in more detail, the following logistic regression model was estimated:

$$\begin{aligned} Y_{i;t+1} = \beta_0 + \beta_1 * Tone_{Type;Full\ Report;i;t} + \beta_2 * Ln\ MVE_{i;t} + \beta_3 * PtB_{i;t} \\ + \beta_4 * ROA_{i;t} + \beta_5 * Liquidity_{i;t} + \beta_6 * Cash_{i;t} + \varepsilon_{i;t} \end{aligned} \quad (26)$$

Where: the dependent variable ($Y_{i;t+1}$) is a binary variable equal to one if a firm files for bankruptcy in the future or zero if it does not. $Tone_{Type;Full\ Report;i;t}$ is the textual sentiment in the entire annual report on Form 10-K (submission including exhibits) as measured by the number (ratio) of pre-defined words in the SEC filing (e.g. *"positive"*, *"negative"*, *"uncertain"*, *"litigious"*, *"constraining"*).

To examine the explanatory power of disclosure tone on future firm bankruptcies in the context of firm fundamentals, various control variables were included in the analysis, which demonstrated discriminate power between bankrupt and non-bankrupt companies (Lopatta et al. 2017, 323). Assuming that larger firms have a lower bankruptcy probability (Lopatta et al. 2017, 323, 330, 332), first, the natural logarithm of the market value of equity was included (*Ln MVE*) in the analysis. Second, showing that bankrupt firms tend to have a lower market value in relation to their book value (Lopatta et al. 2017, 330, 332), a commonly used market-based variable found in the bankruptcy literature (Lopatta et al. 2017, 323) was included, namely the price-to-book ratio (*PtB*). Third, based on the idea that more profitable firms are less likely to file for bankruptcy (Lopatta et al. 2017, 323, 330, 332), the return on assets of a firm was included in the analysis (*ROA*). Fourth, showing a negative association with firm bankruptcies (Lopatta et al. 2017, 323, 330, 332), the current ratio calculated as total current assets divided by total current liabilities was included in the analysis (*Liquidity*). Finally, a cash ratio was incorporated into the regression calculated as a firm's cash and short-term investments divided by its total assets (*Cash)*, assuming that firms with more cash are less likely to file for bankruptcy (Lopatta et al. 2017, 323).

The regression results showed that the majority of included firm fundamentals are associated with future firm failures in the expected manner. Firms filing for bankruptcy tend to be smaller (negative sign on *Ln MVE*) and less profitable (negative sign on *ROA*), and have a lower price-to-book ratio (negative sign on *PtB*). Interestingly, contrary to naïve expectations, a higher liquidity ratio and bigger cash reserve is positively associated with a firm's future bankruptcy (positive sign on *Liquidity*, positive sign on *Cash*). More important, the regression results confirmed that the majority of disclosure tone types do not have explanatory power for future firm bankruptcies after controlling for firm fundamentals. Contrary to naïve expectations, negative and uncertain narrative disclosures in Form 10-K are negatively (not positively) related to future firm bankruptcies. Despite showing the expected negative (positive) sign, positive (litigious) accounting narratives are not significantly related to future firm bankruptcies when financial figures are included in the analysis. However, the results also reveal that constraining Form 10-K information is significantly positively related to future firm bankruptcies. Based on this result, the study concludes that (specific) narrative information in annual reports and tone thereof has explanatory power for future firm bankruptcies even after controlling for firms' fundamentals. Note that other variables not included in the regression (analysis) might explain this result (omitted variable bias). Despite standardized filing, disclosure, and auditing requirements in regulatory filings a firm's corporate governance might influence the disclosure behaviour in annual reports on Form 10-K while at the same time affecting the probability of a company filing for bankruptcy. This effect holds true for all other presented empirical results. Table 48 presents the estimated regression coefficients for different disclosure tones in Form 10-K.

Table 48. Regression coefficients on different Form 10-K disclosure tones for corporate bankruptcy after controlling for firm fundamentals.

Panel A: Negative, positive, and overall (net) disclosure tone

Variable	Exp. sign	1-7	1-7	1-7	1-7	1-7	1-7	1-7
$\mathbf{Tone}_{\mathbf{Neg;Full\ Report}}$	+		21.54**	-21.50				
			(2.276)	(-0.595)				
$\mathbf{Tone}_{\mathbf{Pos;Full\ Report}}$	-				-114.9***	-57.06		
					(-3.984)	(-0.531)		
$\mathbf{Tone}_{\mathbf{Net;Full\ Report}}$	+						29.24***	-13.32
							(3.319)	(-0.399)
Ln MVE	-	-2.242***		-2.263***		-2.246***		-2.255***
		(-7.995)		(-7.875)		(-7.915)		(-7.917)
PtB	-	-0.0580		-0.0582*		-0.0545		-0.0589*
		(-1.642)		(-1.657)		(-1.505)		(-1.675)
ROA	-	-14.09***		-14.71***		-14.18***		-14.45***
		(-2.618)		(-2.668)		(-2.618)		(-2.642)
Liquidity	-	0.0680		0.0617		0.0714		0.0634
		(0.367)		(0.341)		(0.383)		(0.348)
Cash	-	3.886		4.327		3.648		4.221
		(0.944)		(1.035)		(0.894)		(1.001)
Constant		48.35***	-0.687**	49.50***	1.491***	49.13***	-0.551***	48.90***
		(8.003)	(-2.183)	(7.625)	(3.888)	(7.723)	(-2.916)	(7.797)
N		482	482	482	482	482	482	482
Pseudo R^2		0.849	0.00800	0.849	0.0256	0.849	0.0175	0.849
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel B: Uncertain, litigious, and constraining disclosure tone								
Variable	Exp. sign	1-7	1-7	1-7	1-7	1-7	1-7	1-7
$\mathbf{Tone}_{\mathbf{Unc;Full\ Report}}$	+		-2.486	-0.743				
			(-0.168)	(-0.0169)				
$\mathbf{Tone}_{\mathbf{Lit;Full\ Report}}$	+				2.344	2.232		
					(0.502)	(0.163)		
$\mathbf{Tone}_{\mathbf{Con;Full\ Report}}$	+						192.6***	153.1**
							(7.378)	(2.500)
Ln MVE	-	-2.242***		-2.242***		-2.242***		-2.301***
		(-7.995)		(-7.991)		(-7.988)		(-7.618)
PtB	-	-0.0580		-0.0580		-0.0578		-0.0526
		(-1.642)		(-1.638)		(-1.611)		(-1.146)
ROA	-	-14.09***		-14.09***		-14.10***		-13.71**
		(-2.618)		(-2.612)		(-2.614)		(-2.545)
Liquidity	-	0.0680		0.0683		0.0704		0.00677
		(0.367)		(0.367)		(0.377)		(0.0383)
Cash	-	3.886		3.892		3.844		3.366
		(0.944)		(0.942)		(0.930)		(0.842)
Constant		48.35***	0.0541	48.37***	-0.0879	48.27***	-2.796***	47.39***
		(8.003)	(0.162)	(7.869)	(-0.445)	(7.965)	(-7.243)	(7.501)
N		482	482	482	482	482	482	482
Pseudo R^2		0.849	4.22e-05	0.849	0.000377	0.849	0.102	0.859
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit

Notes: The table shows the regression coefficients for different Form 10-K disclosure tones for future firm bankruptcies after controlling for firm fundamentals. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). Noteworthy is the increased number of observations through the reiteration of the matching process based on industry classification (Fama-French 48 industry classification) and market capitalization (size proxy). Please note that the matching process within the paired-sample design identifies exact one bankrupt and one non-bankrupt firm-year observation for each fiscal year and industry category. If no match is found or more than one company filed for bankruptcy in the same industry and fiscal year all corresponding observations are excluded from the analysis. After the inclusion of firm fundamentals and the exclusion of observations missing the corresponding control variables the overall number of observations included in this analysis increased (reiteration of the matching process). For the dependent variable and corresponding regression model diagnostics were conducted (goodness-of-fit, precision, multicollinearity of predictors). The Hosmer-Lemeshow test for goodness-of-fit yielded a p-value > 5% for all regressions were textual sentiment and firm fundamentals were included. The precision of all regressions were between 95.85% and 96.47%. The average variance inflation factor (VIF) was below 6 (highest for Ln MVE and included sentiment variable). Based on these results the study used an logistic regression (logit) model to examine the dependent variable.

4.7.3 All available bankrupt and non-bankrupt firm-year observations

The paired-sample design and the exclusion of bankrupt firm-year observations without a control or peer group match might lead to biased results concerning the relevance of narrative disclosures in the context of corporate bankruptcy. To examine the association between textual sentiment and future corporate bankruptcies in more detail the study followed prior empirical work (e.g. Mayew et al. 2015) and included all available bankrupt and non-bankrupt firm-year observations. Based on all available observations (1,590 bankrupt and 78,723 non-bankrupt firm-year observations) the study estimated the following logistic regression model:

$$\begin{aligned} Y_{i;t+1} = {} & \beta_0 + \beta_1 * Tone_{Type;Full\ Report;i;t} + \beta_2 * Ln\ MVE_{i;t} + \beta_3 * PtB_{i;t} \\ & + \beta_4 * ROA_{i;t} + \beta_5 * Liquidity_{i;t} + \beta_6 * Cash_{i;t} \\ & + \sum Year\ Fixed\ Effects + \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \tag{27}$$

Where: the dependent variable ($Y_{i;t+1}$) is a binary variable equal to one if a firm files for bankruptcy in the future or zero if it does not. $Tone_{Type;Full\ Report;i;t}$ is the textual sentiment in the entire annual report on Form 10-K (submission including exhibits) as measured by the number (ratio) of pre-defined words in the SEC filing (e.g. *"positive"*, *"negative"*, *"uncertain"*, *"litigious"*, *"constraining"*). Year Fixed Effects and Industry Fixed Effects represent the included fixed effects.

The regression results show that most disclosure tone types are associated with future firm bankruptcies in the expected manner when all available firm-year observations are included in the analysis. While negative, net (overall), litigious, and constraining textual sentiment is positively associated with future firm failure positive disclosure tone is negatively associated with a firm filing for bankruptcy in the future. More importantly, the results also provide evidence that narrative disclosures and their textual sentiment can provide additional information value in explaining future firm failure beyond the financial figures being disclosed by a company. Overall, the results underscore the importance of narrative information in financial statements and the validity to serve as supplements in bankruptcy prediction models. Table 49 presents the estimated regression coefficients for different disclosure tones in Form 10-K based on all available bankrupt and non-bankrupt firm year-observations in the data sample.

Table 49. Regression coefficients on different Form 10-K disclosure tones after controlling for firm fundamentals including all available observations (bankrupt and non-bankrupt firm-year observations).

Panel A: Negative, positive, and overall (net) disclosure tone

Variable	Exp. sign	1-7	1-7	1-7	1-7	1-7	1-7	1-7
$Tone_{Neg;Full\ Report}$	+		45.90***	48.76***				
			(10.04)	(10.61)				
$Tone_{Pos;Full\ Report}$	**-**				-44.43***	-33.79**		
					(-2.959)	(-2.200)		
$Tone_{Net;Full\ Report}$	+						45.50***	46.77***
							(10.25)	(10.43)
Ln MVE	-	-0.0423**		-0.0384**		-0.0357*		-0.0299
		(-2.258)		(-2.026)		(-1.870)		(-1.574)
PtB	-	-0.00792**		-0.00754*		-0.00792**		-0.00754*
		(-2.056)		(-1.875)		(-2.060)		(-1.881)
ROA	-	0.238**		0.272**		0.212**		0.232**
		(2.457)		(2.529)		(2.202)		(2.212)
Liquidity	-	-0.0164		-0.0131		-0.0170		-0.0139
		(-0.954)		(-0.790)		(-0.987)		(-0.835)
Cash	-	-2.654***		-3.017***		-2.565***		-2.874***
		(-4.180)		(-4.654)		(-4.043)		(-4.456)
Constant		-3.120***	-5.193***	-4.338***	-3.586***	-2.845***	-4.643***	-3.902***
		(-3.452)	(-6.220)	(-4.818)	(-4.147)	(-3.103)	(-5.586)	(-4.352)
N		80,313	80,313	80,313	80,313	80,313	80,313	80,313
Pseudo R^2		0.0896	0.0952	0.105	0.0827	0.0905	0.0966	0.106
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm	Firm
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit

Panel B: Uncertain, litigious, and constraining disclosure tone								
Variable	Exp. sign	1-7	1-7	1-7	1-7	1-7	1-7	1-7
$\mathbf{Tone}_{\mathbf{Unc;Full\ Report}}$	+		-4.272	2.238				
			(-0.505)	(0.263)				
$\mathbf{Tone}_{\mathbf{Lit;Full\ Report}}$	+				13.97***	13.48***		
					(7.572)	(7.336)		
$\mathbf{Tone}_{\mathbf{Con;Full\ Report}}$	+						126.2***	120.0***
							(14.86)	(13.75)
Ln MVE	-	-0.0423**		-0.0425**		-0.0535***		-0.0463**
		(-2.258)		(-2.265)		(-2.844)		(-2.330)
PtB	-	-0.00792**		-0.00792**		-0.00762**		-0.00699*
		(-2.056)		(-2.053)		(-1.967)		(-1.760)
ROA	-	0.238**		0.236**		0.241**		0.153
		(2.457)		(2.440)		(2.454)		(1.523)
Liquidity	-	-0.0164		-0.0164		-0.0145		-0.0144
		(-0.954)		(-0.955)		(-0.858)		(-0.848)
Cash	-	-2.654***		-2.667***		-2.531***		-2.159***
		(-4.180)		(-4.218)		(-4.021)		(-3.441)
Constant		-3.120***	-4.058***	-3.146***	-4.724***	-3.497***	-5.487***	-4.372***
		(-3.452)	(-4.805)	(-3.468)	(-5.675)	(-3.895)	(-6.671)	(-4.865)
N		80,313	80,313	80,313	80,313	80,313	80,313	80,313
Pseudo R^2		0.0896	0.0811	0.0897	0.0875	0.0956	0.107	0.113
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm	Firm
Model		Logit	Logit	Logit	Logit	Logit	Logit	Logit

Notes: The table shows the regression coefficients for different Form 10-K disclosure tones for future firm bankruptcies after controlling for firm fundamentals. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). Please note that the for the analysis the entire sample of bankrupt and non-bankrupt firm-year observations were used (no paired-sample design). For the dependent variable and corresponding regression model diagnostics were conducted (goodness-of-fit, precision, multicollinearity of predictors). The Hosmer-Lemeshow test for goodness-of-fit yielded a p-value > 5% for all regressions were textual sentiment and firm fundamentals were included. The average precision of all regressions were 98.02%. The average variance inflation factor (VIF) was below 7 (highest for Ln MVE and included sentiment variable). Based on these results the study used an logistic regression (logit) model to examine the dependent variable.

4.7.4 Financial distress

Over several decades, the accounting and finance literature has developed numerous models to discriminate between bankrupt and non-bankrupt companies. Most models consist only of quantitative variables (e.g. financial ratios, market variables). A commonly used quantitative model in the literature to predict future firm bankruptcies or financial distress more generally is the Altman Z-Score. As a linear combination (discriminant function) of different financial ratios, the Altman Z-Score is able to differentiate between healthy firms (high Z-Score) and companies in (future) financial distress (low Z-Score). To further examine the explanatory power of textual information in the context of a firm's future financial condition, this study analysed the association between Form 10-K textual sentiment and the Altman Z-Score as a proxy for firms' future financial condition (alternative dependent variable). Therefore, the following linear regression model was estimated (not a logistic regression model):

$$\begin{aligned} Z - Score_{i;t+1} = \beta_0 + \beta_1 * Tone_{Type;Full\ Report;i;t} + \beta_2 * Z - Score_{i;t} \\ + \sum Year\ Fixed\ Effects \\ + \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \tag{28}$$

Where: the dependent variable ($Z - Score_{i;t+1}$) is a firm's future Altman Z-Score (next fiscal year). $Tone_{Type;Full\ Report;i;t}$ is the textual sentiment in the entire annual report on Form 10-K (submission including exhibits). $Z - Score_{i;t}$ is a firm's actual Z-Score (present fiscal year). Year Fixed Effects and Industry Fixed Effects represent the included fixed effects.

Note that the analysis includes all firm-year observations in the data sample except for the financial industry (SIC codes 6021 - 6799) (no pairwise matching process). Furthermore, all included variables were winsorized at the 1% level.

In general, the regression results showed that Form 10-K textual information is not associated with a firm's future Z-Score or financial situation after controlling for a firm's present Z-Score. One would expect a higher amount of negative (positive) information to be associated with a lower (higher) Altman Z-Score. Regardless, negative and positive annual report information could not be connected to a firm's future Z-Score in the expected manner. This result holds true for litigious and constraining information. This result is rather interesting since prior research was able to connect negative textual sentiment in annual reports to various proxies of financial distress (e.g. Gandhi et al. 2019). However, a higher amount of uncertain information is associated with a lower subsequent Z-Score, even after controlling for a firm’s present financial situation. Table 50 presents the estimated regression coefficients for different disclosure tones in Form 10-K.

Table 50. Regression coefficients on different Form 10-K disclosure tones for future Z-Scores.

Variable	Exp. sign	$Z-Score_{t+1}$	$Z-Score_{t+1}$	$Z-Score_{t+1}$	$Z-Score_{t+1}$	$Z-Score_{t+1}$	$Z-Score_{t+1}$
$\mathbf{Tone_{Neg;Full\ Report}}$	-	**0.351***					
		(1.911)					
$\mathbf{Tone_{Pos;Full\ Report}}$	+		**-2.805*** ***				
			(-6.117)				
$\mathbf{Tone_{Net;Full\ Report}}$	-			**0.659*** ***			
				(3.947)			
$\mathbf{Tone_{Unc;Full\ Report}}$	-				**-2.181*** ***		
					(-7.290)		
$\mathbf{Tone_{Lit;Full\ Report}}$	-					**0.282*** ***	
						(3.332)	
$\mathbf{Tone_{Con;Full\ Report}}$	-						**1.035*** ***
							(2.608)
$Z-Score_t$	+	0.875***	0.874***	0.875***	0.874***	0.875***	0.875***
		(210.0)	(210.6)	(210.6)	(209.6)	(210.8)	(210.6)
Constant		0.149***	0.192***	0.150***	0.191***	0.145***	0.146***
		(4.458)	(5.689)	(4.512)	(5.615)	(4.382)	(4.397)
N		55,539	55,539	55,539	55,539	55,539	55,539
R^2		0.844	0.844	0.844	0.844	0.844	0.844
Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering		Firm	Firm	Firm	Firm	Firm	Firm
Model		Linear	Linear	Linear	Linear	Linear	Linear

Notes: The table shows the regression coefficients for different Form 10-K disclosure tones for a firm's future Z-Score after controlling for its current Z-Score. Word counts (sentiment variables) were obtained using the Loughran and McDonald business dictionary (excluding stop words). Note that the analysis included all firm-year observations in the data sample except for the financial industry (SIC codes 6021 - 6799) (no pairwise matching process). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For the dependent variable the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

4.8 Conclusion

This study examined the explanatory power of the disclosure tone in annual reports regarding a firm's future bankruptcy. Based on a large sample of Form 10-K filings made with the SEC, the analysis revealed that narrative disclosures in annual reports are associated with future corporate bankruptcies. In particular, evidence was provided that a higher level of negative Form 10-K textual sentiment is positively associated with future firm bankruptcies. Besides the explanatory power of negative corporate disclosures, the study indicated that other types of textual sentiment are associated with future corporate bankruptcies in the expected manner (e.g. *"positive"* disclosure tone, *"constraining"* disclosure tone). Furthermore, it was found that textual information in specific annual report subsections is associated with future firm bankruptcies (e.g. *"Item 1 - Business", "Item 7 - Management's Discussion and Analysis of Financial Condition and Results of Operations"*). Based on these results, the study concludes that regulatory filings and disclosure tone in particular contain important and value-relevant information not only for researchers, but also for capital market participants, practitioners, and regulatory authorities.

This study has several limitations. First, the applied information extraction algorithm to utilize narrative information disclosed in regulatory filings might contain errors (e.g. parsing errors). This concern was mitigated by using an information extraction procedure tested for accuracy. Second, regulatory changes and corresponding changes in the information content of individual subsections over time might affect the regression results (e.g. *"Item 1A - Risk Factors"*). This concern was mitigated by the fact that several findings of this study are not based on individual subsections of the annual report, but on the Form 10-K filing in its entirety. Third, other regulatory filings or disclosures and particular corporate events not included in the analysis might have explanatory power in the context of a firm's bankruptcy. This concern could not be mitigated in this study.

Future research avenues could analyse the explanatory power of specific disclosure topics within subsections or overall filing for future corporate bankruptcies (e.g. research and development, competition). In addition, the effect of the disclosure tone in Form 10-K filings could be examined based on other important firm characteristics not discussed in this study (e.g. revenue growth, cost of capital). Furthermore, the importance of the disclosure tone in other regulatory filings with the SEC on future corporate bankruptcies should be further investigated (e.g. quarterly reports on Form 10-Q, current reports on Form 8-K).

Chapter 5

The effect of annual report readability on financial markets

Contents

5.0 Introduction ... 217

5.1 Related literature ... 220

5.2 Hypothesis development ... 226

5.3 Research design ... 229

5.3.1 Sample selection and description ... 229

5.3.2 Regression model, dependent variables, and independent variables ... 232

5.3.3 Summary statistics and correlations ... 237

5.4 Empirical results ... 241

5.5 Robustness and sensitivity tests ... 249

5.6 Additional tests ... 262

5.7 Conclusion ... 266

"In addition to using plain language, investors don't want more information than they need, so less is more. Whether they use the term or not, information overload is a real concern to investors" (SEC 2007a).

5.0 Introduction

Publicly traded companies are required to file annual reports on Form 10-K with the United States Securities and Exchange Commission (SEC). Annual reports are the primary source of information for shareholders, bondholders, and other capital market participants such as financial analysts (Ertugrul et al. 2017, 811). In particular, Form 10-K filings provide capital market participants with comprehensive information about a firm's business operations and financial condition (*"interrelated financial items"*) (Loughran and McDonald 2014, 1643; Chouliaras 2015, 1; Lehavy et al. 2011, 1088; Gan and Qiu 2018, 2). Besides quantitative information (financial figures), annual reports contain a substantial amount of qualitative or textual information (narrative disclosures) (Li 2010a, 143; Bannier et al. 2017, 4; Lo et al. 2017, 2; Lehavy et al. 2011, 1089; Lo et al. 2017, 2). For instance, textual information (e.g. accounting narratives) might provide a useful and valuable context for understanding a firm's financial data. In addition, narrative disclosures may provide important non-public corporate information to a firm's stakeholders (Li 2010a, 143-144; Bannier et al. 2017, 5). The amount of narrative disclosures in Form 10-K filings has steadily increased over time (Li 2008, 228) following changes in financial and reporting regulations (e.g. changes in segment disclosures, employee stock option reporting, Sarbanes-Oxley disclosures) (Lehavy et al. 2011, 1087). *"The clarity of this large component of mandatory disclosure is crucial to understanding and to interpreting the information contained in the report"* (Lo et al. 2017, 2). Specifically, the readability of narrative disclosures in regulatory filings is vital in ensuring the effective communication of value-relevant information between a company and the capital market (Ertugrul et al. 2017, 811; Loughran and McDonald 2014, 1644; Lo et al. 2017, 2). In this regard, Gan and Qiu (2018) note that *"the readability of 10-K files affects the effectiveness of information dissemination among investors"* (Gan and Qiu 2018, 2).

Nevertheless, financial researchers and regulatory authorities (SEC) struggle to define and measure the *"readability"* of mandated disclosures (Loughran and McDonald 2014, 1643). Note that readability and its measure has evolved predominantly in the process of grade-leveling school text books, insurance contracts, and the understandability of instructions in military applications (Loughran and McDonald 2014, 1643). Most studies in the field of textual analysis use the Fog Index to measure (quantify) the readability of text documents (e.g. Li 2008;

Loughran and McDonald 2009; Miller 2010; Lehavy et al. 2011; Lawrence 2013). The Fog Index[28] is defined as a linear function of the average sentence length and proportion of complex words (words with more than two syllables).[29] Lower (higher) values on this index indicate that certain texts are more (less) readable (Loughran and McDonald 2014, 1644, 1647, 1655; Bonsall et al. 2017, 333).[30] Despite being considered by the SEC as an appropriate measure to *"judge the level of compliance with the plain English rules"* (SEC 2007b), the Fog Index as a proxy for textual complexity is complicated, error-prone (misleading), and less replicable than other readability scores (e.g. document file size) (Loughran and McDonald 2014, 1644-1645; Ertugurl et al. 2017, 815-816).[31] For example, the Fog Index suggests that an increased number of complex (multi-syllable) words decrease the readability of financial documents (accounting for half of the measure's inputs). However, common business terms are frequently used in regulatory filings such as annual reports and easy to comprehend for average investors (Ertugurl et al. 2017, 815; Bonsall et al. 2017, 330, 333). *"By its very nature, business text has an extremely high percentage of complex words - one of Fog's two components - that are well understood by investors and analysts"* (Loughran and McDonald 2014, 1644). Showing that syllable counts are a poor measure of readability in the context of firms' business disclosure Loughran and McDonald (2014) found that 52 *"complex"* words (e.g. financial, company, interest, agreement, including, operations, period, related, management, consolidated, information, services, provided, pursuant, following, securities, approximately, reference, operating, material, capital, expenses, corporation, outstanding, additional, effective) account for more than 25 percent of all complex words appearing in Form 10-K filings (more than two syllables) (Loughran and McDonald 2014, 1645, 1656). *"Even if we ignore the most common multisyllable words, few of the remaining complex words are ones that an average reader would stumble*

---

[28] Fog Index = 0.4 * (average number of words per sentence + percent of complex words).

[29] The Fog Index estimates the years of formal education a reader of average intelligence would need to understand the text in the first reading. The relation between the Fog Index and reading ease (textual complexity) is as follows: Fog >= 18 means the text is unreadable, 14-18 (difficult), 12-14 (ideal), 10-12 (acceptable), and 8-10 (childish) (Li 2008, 225; Ertugrul et al. 2017, 815). Lehavy et al. (2011) noted that the Fog Index has also been widely used in social science research to examine the relation between the readability of written information and decisions or outcomes (Lehavy et al. 2011, 1088-1089).

[30] According to Loughran and McDonald (2014), the Kincaid Index (Flesch-Kincaid Index) and Flesch Reading Ease Index are two other readability measures often used in the accounting and finance literature (Loughran and McDonald 2014, 1644). The Kincaid Index rates text by US grade school level (Index = (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59). The Flesch Reading Ease Index rates text on a 100-point scale (Index = 206.835 - (1.015 * words per sentence) - (84.6 * syllables per word)). A higher Flesch Reading Ease Index indicates a more readable text document (Li 2008, 225).

[31] For Lehavy et al. (2011), the Fog Index offers several important advantages. First, it is an objective readability measure. Second, it is not based on analyst surveys or opinion. Third, it can be calculated for any narrative disclosure. Fourth, it measures the readability of a variety of financial items. Fifth, the Fog Index allows examining the overall syntactic complexity of corporate information beyond its specific content (Lehavy et al. 2011, 1089).

over" (Loughran and McDonald 2014, 1645). In addition, measuring the average words per sentence (other component of the Fog Index) in the context of financial disclosures is substantially less precise (Form 10-K structure, table content, HTML code) than measuring sentence length in non-financial areas (plain text) (Loughran and McDonald 2014, 1645). Furthermore, writing style as the central focus of the Fog Index is just one dimension of readability and at the same time less differentiated in financial documents (versus books from various grade levels) (Loughran and McDonald 2014, 1646). Finally, research provides empirical evidence that that the Fog Index (traditional readability measure) is not associated with important firm characteristics (e.g. volatility) when controlling for other (alternative) readability measures (document file size) (Loughran and McDonald 2014, 1660).

Despite their various limitations (e.g. human judgement, computational cost, comparability), numerous studies uses traditional readability measures to examine the effect of textual complexity on firm characteristics (e.g. Li 2008; Loughran and McDonald 2009; Miller 2010; Lehavy et al. 2011; Lawrence 2013; Loughran and McDonald 2014). Interestingly, alternative measures for readability such as the document file size are less examined in the accounting and finance literature despite being straightforward, less prone to measurement error, replicable, and strongly correlated with traditional readability measures (Loughran and McDonald 2014, 1644-1645). To address this research gap, this study uses the document file size established by Loughran and McDonald (2014) as an alternative readability measure to investigate the effect of textual complexity in regulatory filings on various important firm characteristics. A regression model is estimated using an ordinary least squares (OLS) with a firm's market performance, market liquidity, and market risk as dependent variables and several proxies for firm characteristics as independent variables. The results show that the document file size of an annual report as an alternative measure of its complexity (readability) can be connected to various important firm characteristics in the expected manner.

The study makes several contributions to the growing literature on textual analysis. First, the study shows that the document files size of an annual report as an alternative measure of its readability is associated with various stock characteristics. Second, it provides empirical evidence that the document file size of an annual report is predictive for a firm's future corporate earnings. Third, it associates document file size as an alternative measure for annual report readability and future dividend payments to investors. Finally, it illustrates the effect of different research design choices on empirical results in the field of textual analysis. Besides academia, the empirical results are important for investors and regulatory authorities. The results about the effects of textual complexity on a variety firm characteristics help investors to better understand the implications of textual complexity on financial markets. For instance, they may contribute to improve stock trading strategies based on textual information and its features. The results may also help financial regulators to improve the information requirements of public traded companies and their risk monitoring activities on financial markets.

The remainder of Chapter 5 is organized as follows. The next section presents the related literature. Section 5.2 develops the hypotheses. Section 5.3 illustrates the sample selection process and research methodology. Section 5.4 presents the empirical results. Section 5.5 shows the robustness and sensitivity tests. The additional tests are presented in Section 5.6, and finally, Section 5.7 concludes the Chapter.

5.1 Related literature

Li (2008) examined the relation between annual report readability and firm performance (Li 2008, 221). The author used the Fog Index and document length (log number of words) as measures for Form 10-K readability (Li 2008, 225-226). Li found that firms with lower corporate earnings tend to file Form 10-K filings that are more difficult to read (higher Fog Index) and longer (more words) (Li 2008, 222; 233-235; 244). In addition, the study associated an increase (decrease) in corporate earnings with Form 10-K filings that are easier (more difficult) to read than previous years' reports (Li 2008, 222; 234; 236). Furthermore, Li provided empirical evidence that annual report readability is related to earnings persistence. In particular, he showed that profitable firms with more complicated annual reports have lower earnings persistence (Li 2008, 222; 235; 237-239; 244). Despite finding no *"significant evidence that firms make their annual reports more difficult to read in order to hide more persistent bad news"* (Li 2008, 239), the author concluded that a *"clear"* correlation exists between the linguistic features of annual reports and firm performance (Li 2008, 222), and that managers may indeed structure annual reports opportunistically (lengthy disclosures) to delay the incorporation of adverse information into stock prices (Li 2008, 243-245).[32]

You and Zhand (2009) studied the immediate and delayed stock market reaction to annual reports on Form 10-K filed with the SEC in the context of their complexity (You and Zhand 2009, 560, 578). Using the word count (text length) as a simple measure for annual report complexity

[32] Lee (2012) noted that *"an important but untested presumption underlying Li's conclusion is that market participants are fooled by or at least react less completely to information that is made more obscure"* (Lee 2012, 1139).

(readability) (You and Zhand 2009, 561, 578), they found that capital market participants underreact to the information content in Form 10-K filings (You and Zhand 2009, 560). While low complex (short) annual reports are not associated with a delayed market reaction (no under-reaction), complex (long) Form 10-K filings are associated with subsequent long-term stock returns (under-reaction) (You and Zhand 2009, 579-580). The results demonstrated that a substantial amount of Form 10-K information in a complex report is not incorporated into the stock price around the filing date, but over the subsequent year after the submission (You and Zhand 2009, 579-580). Based on these results, the authors concluded that the complexity of textual information affects the extent to which investors are able to incorporate corporate disclosures into stock prices (You and Zhand 2009, 585).

Miller (2010) investigated the effect of Form 10-K reporting complexity (length, Fog Index) on investors' trading behaviour (Miller 2010, 2108). He found an association between the complexity of annual reports (length) and trading activities. In particular, longer Form 10-K filings were associated with lower levels of aggregate trading volume (Miller 2010, 2108, 2119, 2123-2125, 2138). Interestingly, less readable annual reports (higher Fog Index) are not significantly related to lower abnormal trading volumes in the immediate days following the SEC submissions, especially when considering large investors (Miller 2010, 2123-2124, 2127-2129). Contrary to the idea of accounting narrative obfuscation by company officials, the author concluded that large investors seem to prefer more data in financial statements written in a more technical style (less readable) (Miller 2010, 2136).

Lehavy et al. (2011) examined the effect of annual report readability on the behaviour of sell-side financial analysts. In particular, they investigated the relation between Form 10-K textual complexity and analyst following, analyst forecast revision response time, the information content of analysts' reports, and properties of analyst earnings forecasts (Lehavy et al. 2011, 1089). Using the Fog Index, the study first examined the effect of annual report readability on analyst following and the properties of their earnings forecasts (Lehavy et al. 2011, 1093). The authors determined a positive and significant association between annual report readability and the number of analysts covering a specific firm (Lehavy et al. 2011, 1089, 1102-1103, 1112). They also found that financial analysts covering firms with less readable annual reports take longer to release their first report after the Form 10-K filing is made with the SEC (Lehavy et al. 2011, 1090, 1106-1107, 1112). Furthermore, the study suggested that investors find analysts' reports more informative for firms with less readable corporate disclosures (Lehavy et al. 2011, 1090, 1108-1109, 1112). Finally, the study provided empirical evidence that analyst forecasts are negatively affected by less readable annual reports (forecast dispersion, forecast accuracy) (Lehavy et al. 2011, 1090, 1110-1112). The authors concluded that the overall linguistic complexity of Form 10-K filings influences analyst behaviour beyond the effects of the content of the annual report (Lehavy et al. 2011, 1112).

Lee (2012) investigated the association between quarterly report readability (Form 10-Q) and the information efficiency of stock prices (post-earnings announcement drift - PEAD) (Lee 2012, 1137). The author used document length and the Fog Index to measure the readability of Form 10-Q submissions (Lee 2012, 1138). Lee found an association between Form 10-Q readability and the amount of information impounded into stock prices (Lee 2012, 1138). In particular, Lee showed that less (more) earnings-related information is incorporated into a firm's stock market valuation around the filing date (post-filing drift) of a quarterly report when the Form 10-Q filing is longer and more textually complex (Lee 2012, 1138-1139, 1154-1158, 1166). The analysis further revealed that the negative effect of low readability is more pronounced for firms with lower financial analyst coverage and smaller institutional ownership, suggesting that professional and sophisticated capital market participants (e.g. securities analysts, money managers, institutional investors) are less affected by less readable corporate disclosures (Lee 2012, 1138-1139, 1159-1160, 1166). Lee concluded that more difficult-to-read corporate disclosures prolong the price discovery process and lengthen the price drift (after earnings announcements), thus negatively affecting capital market efficiency (Lee 2012, 1138-1139, 1166).

Lawrence (2013) investigated the effect of annual report readability (Fog Index, length) on investors' holdings and performance (Lawrence 2013, 131, 135, 144). In particular, the author examined whether investors' shareholdings increase for firms that submit clearer and more concise corporate disclosures. The results indicated that investors invest more in firms with clearer and more concise disclosures (Lawrence 2013, 131, 137-138, 142-144). In addition, the study looked at whether personal traits and corporate disclosures interact to affect individuals' shareholdings, showing that professional investors are more likely to invest in firms with more complex annual reports than non-professional investors (Lawrence 2013, 131, 138-140, 144). Furthermore, the effect of clearer and more concise financial disclosures on individuals' investment performance was explored. The results indicated that individuals' investor performance increases with clearer and more concise Form 10-K disclosures (Lawrence 2013, 131, 140-142, 144). Based on these results, the author concluded that high-quality corporate disclosures can reduce investors' information disadvantage in financial markets (Lawrence 2013, 131).

Loughran and McDonald (2014) evaluated different measures of readability in a financial context. First, they explored whether the Fog Index as a measure for annual report readability has explanatory power regarding a firm's stock price volatility surrounding the submission date. They found that a higher Fog Index (less readable Form 10-K filing) was significantly associated with a higher level of volatility after an annual report was filed with the SEC (Loughran and McDonald 2014, 1653-1654). They also determined the implications of the different components of the Fog Index (sentence length, word complexity). While the length of sentences is associated with post-filing volatility, the proportion of complex words appears to have no relation to a firm's risk profile on financial markets (Loughran and McDonald 2014, 1654-1655). However, the authors showed that the two components of the Fog Index are not able to measure clarity in business writing (percent of complex words) and are prone to error (average words per sentence) (Loughran and McDonald 2014, 1655-1657). Therefore, they proposed the document file size of a Form 10-K filing as an alternative or *"omnibus measure for readability"* (Loughran and McDonald 2014, 1659). Their analysis significantly related document size (log file size) to subsequent firm volatility after an annual report was submitted to the SEC (after controlling for the Fog Index) (Loughran and McDonald 2014, 1660). Finally, after examining other proxies for a firm's information environment (e.g. earnings surprises, analyst dispersion) and alternative readability measures (e.g. common words, log number of words), the authors recommended using document file size as a simple proxy when measuring the readability of a financial text (Loughran and McDonald 2014, 1661-1663, 1665-1667).

Ajina et al. (2016) investigated the relationship between earnings management and annual report readability (Ajina et al. 2016, 509-510). Using discretionary accruals as a measure for earnings management and the Fog Index as a measure for readability (Ajina et al. 2016, 509, 512-513), they found a positive and significant relation between the level of discretionary accounting adjustments and annual report complexity (Ajina et al. 2016, 509, 515). Based on their results, they concluded that managers faced with declining corporate earnings attempt to hide future profits by increasing the complexity of the annual report (Ajina et al. 2016, 515).

Similarly, Lo et al. (2017) explored how annual report readability (Fog Index) varies in the context of earnings management (Lo et al. 2017, 1). They observed that firms with managed earnings have *"MD&A"* sections (*"Item 7 - Management's Discussion and Analysis of Financial Condition and Results of Operations"*) that are more textually complex (Lo et al. 2017, 1), suggesting that managers strategically use corporate disclosures to mislead investors (Lo et al. 2017, 2). In particular, they found that firms that meet or just beat the prior year's earnings submit more complex (less readable) *"MD&A"* sections in their annual report filings (Lo et al. 2017, 2, 10-12). Furthermore, annual reports with managed earnings (financial restatements) or fraudulent Form 10-K filings (financial misstatements) contain more textually complex *"MD&A"* sections (Lo et al. 2017, 18, 23). The authors concluded that firms likely to have managed earnings to meet or beat the prior year's earnings (benchmark) disclose less readable *"MD&A"* sections in Form 10-K filings (Lo et al. 2017, 24).

Extending this, Bonsall et al. (2017) proposed a new measure for annual report readability (*"Bog Index"*) (Bonsall et al. 2017, 329). In contrast to other readability measures, which are based on writing clarity (e.g. Fog Index) or disclosure quantity (e.g. document file size), the *"Bog Index"* captures plain English attributes of corporate disclosures (Bonsall et al. 2017, 329). In addition to including sentence length, passive voice, weak verbs, overused words, complex words, and jargon, the *"Bog Index"* determines word complexity by word familiarity based on a proprietary list of more than 200,000 words (Bonsall et al. 2017, 330). Derived from a software program (StyleWriter), the proposed readability measure is calculated using a pre-programmed algorithm (no human discretion) by capturing the plain English writing attributes recommended by the SEC (Bonsall et al. 2017, 330, 333). An experiment demonstrated that participants who receive a more readable disclosure (based on the *"Bog Index"*) find the specific text significantly easier to read than those who receive a less readable document. This suggests that the proposed *"Bog Index" '* accurately captures financial statement readers` evaluation of readability (Bonsall et al. 2017, 330, 335-338). Examining the capital market outcomes of different measures for annual report readability, they also found a significant association between Form 10-K readability and a firm`s subsequent volatility (Bonsall et al. 2017, 331, 341-342, 346). In particular, the analysis revealed that the magnitude of the association between the *"Bog Index"* and a firm`s future stock return volatility is significantly higher than those of other readability measures (Bonsall et al. 2017, 331, 342-343, 346). Unable to connect the proposed *"Bog Index"* to earnings forecasts from financial analysts, Bonsall et al. only found an association between quantity-based measures of Form 10-K readability (document file size, length) and future earnings (Bonsall et al. 2017, 331, 342-344, 346-347). Further analysis revealed that only the document file size related to the actual text of annual reports (core document and exhibits, excluding non-textual components) is significantly associated with earnings forecasts (Bonsall et al. 2017, 331, 343-346). The authors concluded that the quantity of disclosure, not the clarity of disclosure, affects sophisticated capital market participants (Bonsall et al. 2017, 343), and that both quantity-based and clarity-based measures of readability are valid (Bonsall et al. 2017, 346).

Ertugrul et al. (2017) investigated the impact of a firm's annual report readability and disclosure tone on its borrowing costs. They identified a significant association between annual report readability (document file size) and a firm's loan spread (loan rate minus London Interbank Offered Rate: LIBOR) (Ertugrul et al. 2017, 813, 817-818, 820-821). In addition, they confirmed that specific types of disclosure tone in Form 10-K filings (weak modal words, uncertain words) are related to a firm's borrowing costs (Ertugrul et al. 2017, 813, 822-823). Furthermore, the results showed that less readable Form 10-K filings and an ambiguous tone negatively affect a firm's (non-price) debt contract terms (loan maturity, security requirements, collateral) (Ertugrul et al. 2017, 813, 817-818, 823-825). Finally, the study established a link between annual report readability as well as tone and a firm's future crash risk (Ertugrul et al. 2017, 814, 817-818, 823-824, 827-831). The authors concluded that shareholders of firms with less readable and more ambiguous Form 10-K filings demonstrate less transparent disclosure and higher external financing costs (Ertugrul et al. 2017, 811, 814, 832).

Boubaker et al. (2018) examined the effects of annual report readability on a firm's stock market liquidity. The authors observed that less readable annual reports are negatively associated with proxies for a firm's market liquidity (Boubaker et. al. 2018, 4). Using the Fog Index as a measure of readability (Boubaker et. al. 2018, 10), the regression results showed that a higher level of textual complexity is positively related to a firm's illiquidity ratio (yearly average ratio of daily absolute returns to daily trading volume) after annual report submission (Boubaker et. al. 2018, 14-15, 35). This result regarding the negative effect of annual report readability on a firm's market liquidity holds true for different proxies for textual complexity (LIX Index, Flesch Reading Ease Index, document length, document file size) and market liquidity (bid-ask spread, zero-return days) (Boubaker et. al. 2018, 15-16, 21-22, 35, 40). The authors concluded that firms with hard-to-read narratives experience low stock liquidity (Boubaker et. al. 2018, 23).

Gan and Qiu (2018) examined the predictive ability of Form 10-K readability (document file size change) on future stock returns. The authors found that changes in Form 10-K readability predict future long-term stock returns after an annual report is filed with the SEC (Gan and Qiu 2018, 3, 10, 15, 28, 35). The authors also investigated the explanatory power of readability changes on future stock returns in different information environments. The results indicated that changes in annual report readability have greater return predictability for smaller firms, with lower levels of analysts' coverage, lower institutional ownership, and higher idiosyncratic volatility (Gan and Qiu 2018, 4, 6, 18-19, 39).[33] A detailed analysis showed that the return predictability of readability changes in Form 10-K filings is driven by managerial disclosure discretion (disclosure obfuscation), not by informing investors about the potential risk associated with a firm's business operations (Gan and Qiu 2018, 7, 18, 24-25, 28, 38, 42). In addition to changes in document file size as a measure for Form 10-K readability, the authors explored the explanatory power of word count changes in annual reports. They found a negative relationship between word count changes in specific subsections of annual reports (e.g. *"Item 3 - Legal Proceedings"*, *"Item 5 - Market for Registrant's Common Equity and Related Stockholder Matter"*, *"Item 7 - Management's Discussion and Analysis of Financial Condition and Results of Operations"*, *"Item 8 - Financial Statements and Supplementary Data"*) and future stock returns (Gan and Qiu 2018, 7, 25, 43). Finally, the authors found a negative relationship between document file size changes and a firm`s future accounting profitability (future change in return on assets) (Gan and Qiu 2018, 16, 36). The authors concluded that a Form 10-K file size change (not the document file size itself) is predictive for future stock returns and future cash flows (Gan and Qiu 2018, 28).

Overall, prior research suggests an association between annual report readability and important future firm characteristics. However, these studies either relied on common readability measures including their various limitations or examined only a limited set of dependent variables. This study attempts to address this specific research gap.

5.2 Hypothesis development

Market performance

The *"management obfuscation hypothesis"* argues that managers have incentives to obfuscate adverse corporate information (e.g. poor performance, negative news) to prevent its incorporation into stock prices (Bloomfield 2002, 238; Li 2008, 224; Devos and Sarkar 2015, 5; Boubaker et. al. 2018, 7). In this context, the *"incomplete revelation hypothesis"* assumes that information that is costlier to extract (e.g. complex information, intransparent disclosures) is less completely reflected in market prices (Grossman and Stiglitz 1980, 404-405; Bloomfield 2002, 234-235;

[33] According to Gan and Qiu (2018), *"an alternative explanation is that managers spend extra 10-K length to elaborate on the negative development of the firm and disclose potential risks associated with its business operations. Although the managers disclose extra information on potential risks to alert investors, investors somehow underreact to these additional risk disclosures (e.g. due to inattention), which then results in a significant delay in the incorporation of the negative value-relevant information into stock prices"* (Gan and Qiu 2018, 4-5).

Li 2008, 224-225; Devos and Sarkar 2015, 5; Miller 2010, 2108, 2138; Lawrence 2013, 131; Amel-Zadeh and Fasse 2016, 2). *"Thus, it appears that the readability and the excessive length of some financial statements are serious barriers to individual investors' extraction of relevant information from financial disclosures"* (Lawrence 2013, 132). Therefore, a firm's management might use lengthy corporate disclosures (less readable) to obfuscate (*"bury"*) information related to negative earnings in their annual reports (Loughran and McDonald 2014, 1646; Ertugrul 2017, 815, 822; Gan and Qiu 2018, 2, 10). Li (2008) noted this aspect: *"Therefore, the length of an annual report could be used strategically by managers in order to make an annual report less transparent and to hide adverse information from investors"* (Li 2008, 225). However, *"when managers use longer 10-Ks to obfuscate bad news, the resulting delay in stock price reactions leads to a significant return predictability. By contrast, when managers use shorter 10-Ks to disclose clear good news message, stock prices should respond more quickly and fully, and the future return predictability should be attenuated"* (Gan and Qiu 2018, 4).

Research confirms that firms with lower corporate earnings submit less readable Form 10-K filings (e.g. Li 2008). Shareholders might interpret more complex corporate disclosures as a negative sign (lower current and future earnings) and assign lower values to firms with longer and more complex (less readable) Form 10-K filings. Thus, the following hypothesis is proposed:

Hypothesis 7. There is a negative relationship between annual report complexity (low readability) and a firm's future stock market performance.

Market liquidity

Prior research has shown that capital market participants are susceptible to biases given the way information is conveyed to them (e.g. Mullainathan and Shleifer 2005; Merkl-Davies and Brennan 2007), and that the complexity of corporate disclosure affects the speed of information diffusion (e.g. Hong and Stein 1999). Assuming that reading and interpreting a large volume of textual information disclosed in annual reports requires considerable (cognitive) efforts, lengthy (complex) corporate disclosures might represent an additional barrier to investor willingness and ability to process the given information (Boubaker et. al. 2018, 3-4, 7; Lee 2012, 1140-1141; Lawrence 2013, 131). Lee (2012) highlighted this aspect: *"A more difficult-to-read document requires a higher cognitive effort to process, which hampers the user's ability to search for and extract useful information in an efficient manner"* (Lee 2012, 1137). Thus, this negatively affects investors' trading behaviour on financial markets. Based on the idea that *"difficult-to-read disclosure creates another layer of information asymmetry between sophisticated* *and unsophisticated information users"* (Lee 2012, 1139), less readable annual reports should therefore be associated with lower liquidity in a firm's stock.

However, lengthy (complex) corporate disclosures might also have a positive effect on a firm's market liquidity by reducing information asymmetries between investors and other capital market participants (higher disclosure quality). This is noted by Lawrence (2013): *"The excessive length finding is not completely obvious, as more financial disclosure information is often considered to be better than less financial disclosure information"* (Lawrence 2013, 132). Despite this notion, prior studies using common or traditional readability measures (e.g. Fog Index) confirm that disclosure complexity in regulatory filings is negatively associated with proxies for a firm's market liquidity on financial markets. Therefore, the following hypothesis is proposed:

Hypothesis 8. There is a negative relationship between annual report complexity (low readability) and a firm's future market liquidity.

Market risk

The announcement of corporate earnings is associated with substantial information uncertainty because of the unknown accounting choices underlying reported earnings and factors affecting the persistence of corporate profits (Lee 2012, 1140). The literature argues that regulatory filings such as annual reports on Form 10-K filed with the SEC are supposed to help investors resolve this information uncertainty (e.g. Lee 2012). Given investors limited ability to process complex corporate disclosures (Boubaker et. al. 2018, 3-4, 7; Lee 2012, 1140-1141), less readable annual reports should exacerbate information uncertainty on financial markets. Furthermore, prior research noted that less readable corporate disclosures reduce investors' belief in the accuracy of the presented information (e.g. Rennekamp 2012). Unable to value a firm properly, less readable annual reports should then increase a firm's overall market risk (Loughran and McDonald 2014, 1648, 1654). Thus, the following hypothesis is proposed:

Hypothesis 9. There is a positive relationship between annual report complexity (low readability) and a firm's future market risk.

5.3 Research design

5.3.1 Sample selection and description

Based on the SEC EDGAR *"Form Index Files"* all annual reports (Form 10-K and Form 10-K405) filed with the SEC between 1993 and 2016 were downloaded. Note that the time period was chosen due to data availability reasons on various information platforms (e.g. SEC EDGAR database, commercial data providers, public research repositories). The number of annual reports filed with the SEC is illustrated in the appendix.[34] For each submission, the *"Complete Submission Text File"* (file extension *.txt) was retrieved from the EDGAR database (Electronic Data Gathering and Retrieval system). Based on the Loughran and McDonald 10X File Summaries, the document file size for each submission was obtained (gross file size, net file size) to quantify the readability of annual reports on Form 10-K filed with the SEC (entire submission including exhibits).

For the analysis, several databases were used to match stock market data and accounting variables. For each annual report filed with the SEC, the Central Index Key (CIK) number was extracted from the document name on the SEC server. The CIK number was used to obtain the International Securities Identification Number (ISIN) and accounting variables for each financial statement (filer) from the Standard & Poor's Capital IQ database. Based on the ISIN, stock market returns were retrieved for each company around its annual report filing as well as for the Standard & Poor's 500 stock market index (S&P 500) from the Thomson Reuters DataStream platform. To include a firm-year observation in the data sample, it was required that accounting data (return on assets, leverage ratio) and stock market data (market capitalization, price-to-book ratio) were available in addition to an exchange listing on the New York Stock Exchange (NYSE), National Association of Securities Dealers Automated Quotations (NASDAQ), or American Stock Exchange (NYSE MKT). Table 51 provides an overview of the sample selection process and number of annual reports included in the analysis.

The majority of annual reports in the data sample was filed by companies in the *"Business Services"* industry (12.20%), followed by those in *"Trading"* (6.96%), *"Electronic Equipment"* (6.57% percent), and *"Pharmaceutical Products"* (6.32%). In general, the distribution of financial statements among certain industries in the data sample is unequal (Fama-French 48 industry classification). While the distribution of firm-year observations across industries is

[34] Note that the Form 10-K sample is identical to the one used in Henselmann and Hering (2017).

unequal, the distribution of annual reports on Form 10-K over time is more balanced. On average, more than 3,000 annual reports from each year were included in the analysis. Table 52 presents the distribution of annual submissions by certain industries and fiscal years.

Table 51. Sample selection process for stock characteristics examination.

	Firm-year observations	Observations removed
All observations (Form 10-K & Form 10-K405)	189,998	
Less		
Missing 10-K report (database error)	189,997	1
Removing duplicates CIK and fiscal-year	184,089	5,908
Removing late filers (>100 days after accounting period)	163,838	20,251
Missing SIC (from filing)	162,759	1,079
Merge with Loughran and McDonald 10X Summaries		
Missing gross file size & net file size (readability proxies)	162,751	8
Merge with Capital IQ data		
Missing ISIN	118,822	43,929
Missing return on assets	112,675	6,147
Missing leverage	99,443	13,232
Missing NYSE, AMEX or NASDAQ exchange listing	71,962	27,481
Merge with DataStream data		
Missing market capitalization	69,324	2,638
Price-to-Book DataStream data available and ratio > 0	65,831	3,493
Observations in final sample		
Firm-year sample	65,831	
Number of unique firms	6,796	-
Average number of years per firm	9	-

Notes: The table presents the sample selection process and number of annual reports included in the analysis.

Table 52. Distribution of Form 10-K filings across industries and fiscal years for stock characteristics examination.

Panel A: Distribution of Form 10-K filings across industries

Fama-French industry	Filings (Num.)	Filings (%)	Filings (∑%)	Fama-French industry	Filings (Num.)	Filings (%)	Filings (∑%)
Business Services	8,034	12.20	12.20	Apparel	847	1.29	86.78
Trading	4,580	6.96	19.16	Electrical Equipment	769	1.17	87.95
Electronic Equipment	4,326	6.57	25.73	Banking	760	1.15	89.10
Pharmaceutical Products	4,161	6.32	32.05	Personal Services	740	1.12	90.22
Retail	3,414	5.19	37.24	Entertainment	716	1.09	91.31
Insurance	2,751	4.18	41.42	Business Supplies	658	1.00	92.31
Petroleum and Natural Gas	2,592	3.94	45.36	Real Estate	657	1.00	93.31
Utilities	2,539	3.86	49.22	Printing and Publishing	601	0.91	94.22
Computers	2,453	3.73	52.95	Almost Nothing	578	0.88	95.10
Medical Equipment	2,424	3.68	56.63	Rubber and Plastic Products	465	0.71	95.81
Machinery	2,344	3.56	60.19	Recreation	421	0.64	96.45
Wholesale	2,196	3.34	63.53	Aircraft	332	0.50	96.95
Measuring and Control Equipment	1,650	2.51	66.04	Non-Metallic and Industrial Metal Minin	278	0.42	97.37
Transportation	1,649	2.50	68.54	Beer & Liquor	260	0.39	97.76
Communication	1,594	2.42	70.96	Fabricated Products	210	0.32	98.08
Chemicals	1,327	2.02	72.98	Shipping Containers	207	0.31	98.39
Restaurants and Hotels	1,258	1.91	74.89	Agriculture	175	0.27	98.66
Healthcare	1,246	1.89	76.78	Textiles	175	0.27	98.93
Food Products	1,138	1.73	78.51	Shipbuilding, Railroad Equipment	162	0.25	99.18
Construction Materials	1,022	1.55	80.06	Precious Metals	157	0.24	99.42
Automobiles and Trucks	918	1.39	81.45	Defense	134	0.20	99.62
Construction	910	1.38	82.83	Coal	110	0.17	99.79
Consumer Goods	883	1.34	84.17	Candy & Soda	73	0.11	99.90
Steel Works Etc	870	1.32	85.49	Tobacco Products	67	0.10	100.00

Panel B: Distribution of Form 10-K filings across fiscal years

Fiscal year	Filings (Num.)	Filings (%)	Filings (∑%)	Fiscal year	Filings (Num.)	Filings (%)	Filings (∑%)
1993	433	0.66	0.66	2005	3,164	4.81	50.49
1994	598	0.91	1.57	2006	3,173	4.82	55.31
1995	1,434	2.18	3.74	2007	3,160	4.80	60.11
1996	2,508	3.81	7.55	2008	3,108	4.72	64.84
1997	2,849	4.33	11.88	2009	3,150	4.78	69.62
1998	3,036	4.61	16.49	2010	3,151	4.79	74.41
1999	3,235	4.91	21.41	2011	3,098	4.71	79.11
2000	3,235	4.91	26.32	2012	3,150	4.78	83.90
2001	3,248	4.93	31.26	2013	3,276	4.98	88.87
2002	3,196	4.85	36.11	2014	3,293	5.00	93.88
2003	3,141	4.77	40.88	2015	3,264	4.96	98.83
2004	3,164	4.81	45.69	2016	767	1.17	100.00

Notes: The table illustrates the number of annual reports in the data sample across Fama-French industries (Panel A) and across fiscal years (Panel B).

5.3.2 Regression model, dependent variables, and independent variables

Regression model

To examine the effect of annual report readability on various firm characteristics, the following regression model was estimated using ordinary least squares (OLS)[35]:

$$\begin{aligned} Y_{i;t} = \beta_0 + \beta_1 * File\ Size_{Gross;i;t} + \beta_2 * Ln\ MVE_{i;t} + \beta_3 * PtB_{i;t} + \beta_4 * ROA_{i;t} \\ + \sum_X \beta_X * Other\ Controls_{i;t} + \sum Year\ Fixed\ Effects \\ + \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \tag{29}$$

Where: the dependent variable ($Y_{i;t}$) is a firm's subsequent market performance, market liquidity, or market risk after an annual report on Form 10-K was filed with the SEC. $File\ Size_{Gross;i;t}$ is the gross file size of a firm's annual report (readability measure). The control variables are described below. *Year Fixed Effects* and *Industry Fixed Effects* represent the included fixed effects.

Dependent variables

To analyse the effect of annual report readability on important firm characteristics, various market variables were obtained (dependent variables)[36].

For a firm's market performance in the days around the SEC submission, the cumulative abnormal stock return (CAR) was calculated over a four-day holding period starting one day before the initial filing date of the annual report (market-adjusted model - MAM):

$$CAR_{i;t-1,t+2;MAM} = \sum_{t-1}^{+2} \left(Return_{i;t} - Return_{S\&P\ 500;t} \right) \tag{30}$$

Where: $Return_{\,i;t}$ is the return of a certain stock (i) on day (t) calculated as the price of the stock (i) on day (t) divided by the price of the stock (i) on the previous day minus one and $Return_{S\&P\ 500;t}$ is the return of the market index (S&P 500) on day (t) calculated as the level

[35] Note that linear regression models are most commonly used in the accounting and finance literature when examining the effect of textual information on firm characteristics (Kearney and Liu 2014, 177). Further note that the study used a linear regression model over various machine learning models (e.g. neural networks) to better examine the relationship between input (independent) and output (dependent) variables.

[36] Note that the study focused on short-term capital market effects following the Form 10 K submission to avoid biases due to other effects not being captured by the model (e.g. other corporate information not being disclosed in the annual report, other corporate events, market trends).

of the market index on day (t) divided by the level of the market index on the previous day minus one.

In addition, a market model (MM) was used to calculate a firm's cumulative abnormal stock return (CAR) during the event period ([t-1, t+2]):

$$CAR_{i;t-1,t+2;MM} = \sum_{t-1}^{+2} \left(Return_{i;t} - (\hat{\alpha}_i + \hat{\beta}_i * Return_{S\&\,500;t}) \right) \tag{31}$$

Where: $\hat{\alpha}_\mathrm{i}$ (intercept) and $\hat{\beta}_\mathrm{i}$ (slope) are estimated using an OLS regression of the daily returns of the company (i) on the daily returns of the market index (S&P 500) during the non-event period ([t-250, t-21]).

Furthermore, a 4-factor model (4 Factor) was used to obtain a firm's cumulative abnormal stock return (CAR) during the event period ([t-1, t+2]):

$$CAR_{i;[-1,+2];4\,Factor} = \sum_{t-1}^{t+2} (Return_{i;t} - Return_{i;t;4\,Factor}) \tag{32}$$

Where: $Return_{i;t;4\,Factor}$ is the expected return of a company (i) on day (t) based on a 4-factor model. Using the data provided by Kenneth R. French daily stock returns were regressed on the excess market returns (MKT-RF), a size factor (small-minus-big), a value factor (high-minus-low), and a momentum factor (up-minus-down) over an estimation window ([t-250, t-21]) relative to the Form 10-K filing dates to estimate the corresponding factor loadings for the model (Amel-Zadeh and Faasse 2016, 16-17).

For a firm's market liquidity in the days around the Form 10-K filing, the abnormal trading volume (ATV) was calculated:

$$ATV_{i;[-1,+3];Ave} = \sum_{t-1}^{t+3} \left(Share\ Turnover_t - \emptyset\ Share\ Turnober_{[\mathrm{t-156,t-6}]} \right) \tag{33}$$

Where: $ATV_{i;[-1,+3],Ave}$ is the abnormal trading volume of a certain stock (i) calculated as the cumulative daily share turnover (proportion of traded shares) during the event period ([t-1, t+3]) minus the average share turnover during the non-event period ([t-156, t-6]). Note that the abnormal trading volume is also calculated by adjusting for the median share turnover during the non-event period ($ATV_{i;[-1,+3];Med}$) (Ditter 2015, 122, 146, 148; Atiase and Bamber 1994, 313-

314; Bamber et al. 1997, 585, 588-589). Further note that the starting date of the event period was chosen since prior research (e.g. Morse 1981; Bamber 1987) suggests that the bulk of the trading volume reaction occurs on days -1 and 0 (Atiase and Bamber 1994, 313) and is well established in the literature (e.g. Kravet and Muslu 2013; Blankespoor et al. 2014; Ditter 2015).

As an additional measure for a firm's market liquidity during the event window starting one day before the initial filing date of the annual report, the abnormal bid-ask spread (ASpread) was calculated:

$$ASpread_{i;[-1,+3];Ave} = \sum_{t-1}^{t+3} \left(Spread_t - \emptyset\, Spread_{[t-156,t-6]} \right) \quad (34)$$

Where: $ASpread_{\,i;[-1,+3];Ave}$ is the abnormal bid-ask spread of a certain stock (i) calculated as the cumulative daily bid-ask spread during the event period ([t-1, t+3]) less the average bid-ask spread during the non-event period ([t-156, t-6]). Note that the abnormal bid-ask spread is also calculated by adjusting for the median bid-ask spread during the non-event period ($ASpread_{\,i;[-1,+3];Med}$) (Ditter 2015, 122, 146, 148; Bushee et al. 2010, 6-7; Blankespoor et al. 2014, 1475-1476, 1478).

Furthermore, the effects of annual report readability on a firm's market risk were explored by obtaining the abnormal stock return volatility (ARV) during a five-day event window around the Form 10-K submission:

$$ARV_{i;t-1,t+3;MAM} = \sum_{t-1}^{t+3} \left| \left(Return_{i;t} - Return_{S\&P\ 500;t} \right) \right| \quad (35)$$

Where: $ARV_{\,i;t}$ is the abnormal stock return volatility of a company (i) calculated as the sum of the absolute daily abnormal stock returns during the event period ([t-1, t+3]) (Ditter 2015, 122, 147; Bailey et al. 2003, 2493-2494). The abnormal stock returns were computed using the S&P 500 stock market index as a benchmark portfolio (MAM).

Again, a MM was used to calculate a firm's abnormal stock return volatility (ARV) during the event period ([t-1, t+3]):

$$ARV_{i;t-1,t+3;MM} = \sum_{t-1}^{t+3} \left| Return_{i;t} - \left(\hat{\alpha}_i + \hat{\beta}_i * Return_{S\&P\ 500;t} \right) \right| \quad (36)$$

Where: $\hat{\alpha}_{\mathrm{i}}$ (intercept) and $\hat{\beta}_{\mathrm{i}}$ (slope) are estimated using an OLS regression of the daily returns of the company (i) on the daily returns of the market index (S&P 500) during the non-event period ([t-250, t-21]). Note that the study included dividends and other capital gains paid during the different event periods (total return).

Finally, a 4-factor model (4 Factor) was implemented to obtain the abnormal stock return volatility (ARV) of a company during the event period ([t-1, t+3]):

$$ARV_{i;[-1,+3];4\,Factor} = \sum_{t-1}^{t+3} |Return_{i;t} - Return_{i;t;4\,Factor}| \tag{37}$$

Where: $Return_{i;t;4\,Factor}$ is the expected return of a company (i) on day (t) based on a 4-factor model. Using the data provided by Kenneth R. French daily stock returns were regressed on the excess market returns (MKT-RF), a size factor (small-minus-big), a value factor (high-minus-low), and a momentum factor (up-minus-down) over an estimation window ([t-250, t-21]) relative to the Form 10-K filing dates to estimate the corresponding factor loadings for the model (Amel-Zadeh and Faasse 2016, 16-17).

Note that the study followed prior research studies in the field of textual analysis (e.g. Amel-Zadeh and Faasse 2016) in defining the non-event period to avoid biases caused by abnormal trading behaviour before the release date of the annual report. Further note that dividends and other capital gains being paid (total return) were accounted for in tests.

Main variable of interest

The study explored the effect of annual report readability (disclosure complexity) on firm characteristics following the SEC submission. Commonly applied readability measures (e.g. Fog Index[37]) are poorly specified in financial applications and difficult to measure (Loughran and McDonald 2014, 1643). However, readability measures based on the document file size are easy to compute, comparable, and independent of human judgement of the information content of the annual report (Gan and Qiu 2018, 9). In addition, traditional readability measures focus on only one dimension of readability, which is less differentiated in financial documents (writing style). Therefore, the present study used the document file size of a firm's annual report to quantify the disclosure complexity of its Form 10-K filing, because document file size *"can be*

[37] As noted by Bonsall et al. (2017), most readability studies in accounting and finance use the Fog Index as a proxy for readability when examining the effect of financial reporting quality (Bonsall et al. 2017, 330).

viewed as an omnibus measure capturing the many dimensions of readability" (Loughran and McDonald 2014, 1646).[38] [39] [40]

Note that document file size as a measure for Form 10-K readability is objective and replicable, addressing the call in the literature. In addition, this proxy for annual report disclosure complexity is correlated with commonly used readability indices (e.g. Fog Index, log number of words) (Loughran and McDonald 2014, 1646; Li 2008, 230).[41] [42] Furthermore, the study used the gross file size of a Form 10-K filing (total number of characters) as provided by the Loughran and McDonald 10X File Summaries.[43] [44]

Control variables

To control for common firm fundamentals in each regression, the natural logarithm of the market value of equity (calculated as the share price one day before the Form 10-K filing multiplied by the number of ordinary shares in issue) and price-to-book ratio (calculated as the share price one day before the submission divided by the book value per share) were included as control variables. In addition, a firm's return on assets (calculated as a firm's earnings before interest divided by the average of total assets) was included in each regression.

To examine a firm's market performance, the change in return on assets and its leverage ratio were included as control variables. For examining a firm's market liquidity and market risk, the previous literature was followed (e.g. Ditter 2015; Blankespoor et al. 2014; You and Zhang 2008) by including the average daily share turnover and stock return volatility in the prior fiscal

---

38 According to Bonsall et al. (2017), *"quantity-based readability measures are based on the notion of overwriting, where documents are written in a manner that is too detailed and long for readers to easily process*" (Bonsall et al. 2017, 333).

39 For Bonsall et al. (2017), *"measuring the quantity of disclosure in a 10-K setting inadvertently includes separate exhibits that are unrelated to the annual 10-K filing requirements (e.g. compensation contracts, supplier/customer agreements, or bond indentures) but are attached to the filing"* (Bonsall et al. 2017, 333).

40 Loughran and McDonald (2014) stated that a firm's Form 10-K document file size could also simply represent firm complexity rather than annual report readability (textual complexity) (Loughran and McDonald 2014, 1646).

41 In contrast, the change in Form 10-K file size has a low correlation with other readability measures (e.g. Fog Index) (Gan and Qiu 2018, 9).

42 Lawrence (2013) provided empirical evidence that the Fog Index and text length of Form 10-K filings are complements rather than substitutes for annual report readability (Lawrence 2013, 138).

43 As noted by Gan and Qiu (2018), the *"gross file size is the total number of characters in the original filing, which also include ASCII-encoded characters and characters attributable to HTML, XBRL and/or XML encodings"* (Gan and Qiu 2018, 11).

44 For Bonsall et al. (2017), *"a vast amount of the variation in Form 10-K file size over time is driven by the inclusion of content unrelated to the underlying text in the 10-K (e.g. HTML, XML, PDFs)"* (Bonsall et al. 2017, 229).

period as control variables in addition to a firm's beta-factor as proxy for its market risk (Ditter 2015, 123).

All regressions employed year and industry fixed effects. In particular, the study follows prior work (e.g. Boubaker et. al. 2018) by including year dummies to control for time trends and market-wide shocks driving a company's market liquidity over time. Standard errors were clustered by firms. All included variables were winsorized at the 1% level (except document file size).

5.3.3 Summary statistics and correlations

This section presents the summary statistics for the market, disclosure, and accounting variables and the correlations between the key variables of interest in the study. The mean (median) cumulative abnormal stock return of a company over a four-day holding period around the filing date has a slightly positive value of 0.03 (-0.01) percent. Over the same holding period, the cumulative abnormal stock return at the 10th percentile is -6.65 percent compared to a value of 6.75 percent at the 90th percentile. Panel A of Table 53 provides the summary statistics for cumulative abnormal stock returns over different holding periods following the SEC submission.

The mean (median) document file size of an annual report included in the analysis is 5.507M (1.366M) characters. The mean (median) market value of a firm included in the analysis is $3,270M ($518M). The average price-to-book ratio is 3.38 (2.04). The return on assets has a mean (median) value of 2.66 (4.49) percent. The mean (median) leverage ratio of the companies included in our data sample is 48.35 (49.11) percent. Panel B of Table 53 presents the summary statistics for the market, disclosure, and accounting variables used as control variables in the analysis (independent variables).

Annual report readability generally does not relate to stock characteristics in the expected manner. A higher gross document file size (lower readability) is correlated with a higher abnormal stock return and a higher abnormal trading volume. In addition, more complex annual reports are correlated with a lower abnormal bid-ask spread and a lower abnormal stock return volatility. Furthermore, annual report also does not relate to other variables in the expected manner. A higher gross document file size is correlated with a higher market capitalization and a higher price-to-book ratio. On the contrary, less readable Form 10-K filings are correlated a lower return on assets and a higher leverage ratio. Table 54 provides the correlations for Form 10-K readability and firm characteristics.

Table 53. Dependent and independent variables for stock characteristics examination.

Panel A: Dependent variables				
Variable	Num.	Mean	Median	St. Dev.
$\mathrm{CAR}_{[-1,2];\mathrm{MAM}}$	52,484	0.0010	0.0001	0.0637
$\mathrm{CAR}_{[-1,2];\mathrm{MM}}$	52,484	-0.0006	-0.0008	0.0633
$\mathrm{CAR}_{[-1,2];4\ \mathrm{Factor}}$	52,484	-0.0009	-0.0014	0.0629
$\mathrm{ATV}_{[-1,3];\mathrm{Ave}}$	55,263	0.0040	-0.0005	0.0285
$\mathrm{ATV}_{[-1,3];\mathrm{Med}}$	55,263	0.0117	0.0032	0.0298
$\mathrm{ASpread}_{[-1,3];\mathrm{Ave}}$	29,768	-0.0007	-0.0007	0.0587
$\mathrm{ASpread}_{[-1,3];\mathrm{Med}}$	29,768	0.0098	0.0002	0.0591
$\mathrm{ARV}_{[-1,3],\mathrm{MAM}}$	53,121	0.1054	0.0780	0.0886
$\mathrm{ARV}_{[-1,3],\mathrm{MM}}$	53,121	0.1032	0.0753	0.0890
$\mathrm{ARV}_{[-1,3],4\ \mathrm{Factor}}$	53,121	0.1043	0.0762	0.0889

Panel B: Independent variables				
Variable	Num.	Mean	Median	St. Dev.
$\mathrm{File\ Size}_{\mathrm{Gross}}$	65,831	5,506,905	1,365,625	10,600,000
Ln MVE	65,831	20.1017	20.0656	1.9423
PtB	65,831	3.3748	2.0400	4.6452
ROA	65,831	0.0266	0.0449	0.1180
Δ ROA	63,914	0.0032	0.0002	0.0676
Leverage	65,831	0.4835	0.4911	0.2316
Turnover	63,770	0.0073	0.0051	0.0071
Volatility	59,141	0.0313	0.0265	0.0191
Market Risk	59,141	0.8819	0.8537	0.5819

Notes: The table provides an overview of stock characteristics around the Form 10-K submission and common firm fundamentals included in the analysis as control variables.

Table 54. Correlations between stock characteristics, annual report readability, and firm fundamentals.

Panel A: Correlation coefficients between cumulative abnormal stock returns, annual report readability, and firm fundamentals

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1: $\text{CAR}_{[-1,2];\text{MAM}}$	1.00	0.98***	0.95***	0.00	0.00	-0.01**	0.03***	0.03***	0.01*	0.00	0.01**	0.01
2: $\text{CAR}_{[-1,2];\text{MM}}$	0.96***	1.00	0.97***	0.01	-0.01	-0.03***	0.03***	0.00	0.01	-0.01	-0.01	0.00
3: $\text{CAR}_{[-1,2];\text{4 Factor}}$	0.91***	0.95***	1.00	0.01	-0.01	-0.03***	0.03***	0.01	0.01	-0.01	-0.01	-0.01
4: File $\text{Size}_{\text{Gross}}$	0.02***	0.02***	0.02***	1.00	0.38***	-0.01	0.04***	-0.01	0.23***	0.01	-0.30***	0.02***
5: Ln MVE	0.03***	0.01**	0.02**	0.43***	1.00	0.12***	0.31***	0.01**	0.26***	0.30***	-0.53***	0.27***
6: PtB	-0.01**	-0.04***	-0.04***	0.07***	0.31***	1.00	-0.07***	0.07***	0.16***	0.02***	-0.02***	0.01*
7: ROA	0.05***	0.04***	0.04***	-0.02***	0.35***	0.26***	1.00	0.13***	0.04***	0.05***	-0.39***	0.03***
8: Δ ROA	0.04***	0.01*	0.02**	0.01	0.05***	0.11***	0.19***	1.00	-0.01	0.01	0.02***	0.01
9: Leverage	0.01**	0.01*	0.01*	0.26***	0.27***	0.05***	-0.05***	0.01*	1.00	0.07***	-0.08***	0.07***
10: Turnover	0.01	0.00	0.01	0.08***	0.45***	0.12***	0.14***	0.00	0.09***	1.00	0.16***	0.39***
11: Volatility	-0.02**	-0.03***	-0.03***	-0.42***	-0.57***	-0.17***	-0.31***	-0.03***	-0.17***	0.10***	1.00	0.14***
12: Market Risk	0.01	0.00	-0.01	0.08***	0.24***	0.06***	0.02***	0.02***	0.04***	0.45***	0.23***	1.00

Panel B: Correlation coefficients between abnormal trading volumes, annual report readability, and firm fundamentals

Variable	1	2	3	4	5	6	7	8	9	10	11	-
1: $\text{ATV}_{[-1,3];\text{Ave}}$	1.00	0.95***	0.00	0.03***	0.02***	0.04***	0.02***	0.03***	0.05***	-0.03***	0.06***	-
2: $\text{ATV}_{[-1,3];\text{Med}}$	0.93***	1.00	-0.02**	0.02***	0.03***	0.00	0.03***	0.02***	0.25***	0.06***	0.10***	-
3: File $\text{Size}_{\text{Gross}}$	0.04***	0.02***	1.00	0.38***	-0.01	0.04***	-0.01	0.23***	0.01	-0.30***	0.02***	-
4: Ln MVE	0.05***	0.05***	0.43***	1.00	0.12***	0.31***	0.01**	0.26***	0.30***	-0.53***	0.27***	-
5: PtB	0.00	0.03***	0.07***	0.31***	1.00	-0.07***	0.07***	0.16***	0.02***	-0.02***	0.01*	-
6: ROA	0.05***	0.03***	-0.02***	0.35***	0.26***	1.00	0.13***	0.04***	0.05***	-0.39***	0.03***	-
7: Δ ROA	0.03***	0.03***	0.01	0.05***	0.11***	0.19***	1.00	-0.01	0.01	0.02***	0.01	-
8: Leverage	0.04***	0.03***	0.26***	0.27***	0.05***	-0.05***	0.01*	1.00	0.07***	-0.08***	0.07***	-
9: Turnover	-0.05***	0.13***	0.08***	0.45***	0.12***	0.14***	0.00	0.09***	1.00	0.16***	0.39***	-
10: Volatility	-0.08***	0.00	-0.42***	-0.57***	-0.17***	-0.31***	-0.03***	-0.17***	0.10***	1.00	0.14***	-
11: Market Risk	0.02***	0.08***	0.08***	0.24***	0.06***	0.02***	0.02***	0.04***	0.45***	0.23***	1.00	-

Panel D: Correlation coefficients between abnormal bid-ask spreads, annual report readability, and firm fundamentals											
Variable	1	2	3	4	5	6	7	8	9	10	11	-
1: $ASpread_{[-1,3];Ave}$	1.00	0.84***	-0.01	-0.01	0.01	-0.01*	-0.03***	0.01	0.01	-0.01**	0.03***	-
2: $ASpread_{[-1,3];Med}$	0.75***	1.00	-0.04***	-0.20***	0.00	-0.08***	-0.01**	-0.01**	-0.10***	0.14***	-0.11***	-
3: File $Size_{Gross}$	-0.06***	-0.06***	1.00	0.38***	-0.01	0.04***	-0.01	0.23***	0.01	-0.30***	0.02***	-
4: Ln MVE	0.04***	-0.13***	0.43***	1.00	0.12***	0.31***	0.01**	0.26***	0.30***	-0.53***	0.27***	-
5: PtB	0.01	-0.07***	0.07***	0.31***	1.00	-0.07***	0.07***	0.16***	0.02***	-0.02***	0.01*	-
6: ROA	0.03***	-0.05***	-0.02***	0.35***	0.26***	1.00	0.13***	0.04***	0.05***	-0.39***	0.03***	-
7: Δ ROA	-0.05***	-0.06***	0.01	0.05***	0.11***	0.19***	1.00	-0.01	0.01	0.02***	0.01	-
8: Leverage	-0.01*	-0.02***	0.26***	0.27***	0.05***	-0.05***	0.01*	1.00	0.07***	-0.08***	0.07***	-
9: Turnover	0.10***	-0.04***	0.08***	0.45***	0.12***	0.14***	0.00	0.09***	1.00	0.16***	0.39***	-
10: Volatility	-0.03***	0.05***	-0.42***	-0.57***	-0.17***	-0.31***	-0.03***	-0.17***	0.10***	1.00	0.14***	-
11: Market Risk	0.05***	-0.04***	0.08***	0.24***	0.06***	0.02***	0.02***	0.04***	0.45***	0.23***	1.00	-

Panel : Correlation coefficients between abnormal stock return volatilities, annual report readability, and firm fundamentals												
Variable	1	2	3	4	5	6	7	8	9	10	11	12
1: $ARV_{[-1,3],MAM}$	1.00	0.99***	0.97***	-0.19***	-0.45***	-0.03***	-0.25***	-0.01**	-0.05***	0.02***	0.57***	-0.01
2: $ARV_{[-1,3],MM}$	0.97***	1.00	0.98***	-0.19***	-0.44***	-0.03***	-0.25***	-0.01**	-0.05***	0.03***	0.58***	0.01
3: $ARV_{[-1,3],4\ Factor}$	0.94***	0.96***	1.00	-0.19***	-0.45***	-0.03***	-0.26***	-0.01	-0.05***	0.02***	0.58***	-0.02***
4: File $Size_{Gross}$	-0.28***	-0.27***	-0.27***	1.00	0.38***	-0.01	0.04***	-0.01	0.23***	0.01	-0.30***	0.02***
5: Ln MVE	-0.50***	-0.49***	-0.49***	0.43***	1.00	0.12***	0.31***	0.01**	0.26***	0.30***	-0.53***	0.27***
6: PtB	-0.16***	-0.15***	-0.16***	0.07***	0.31***	1.00	-0.07***	0.07***	0.16***	0.02***	-0.02***	0.01*
7: ROA	-0.25***	-0.25***	-0.25***	-0.02***	0.35***	0.26***	1.00	0.13***	0.04***	0.05***	-0.39***	0.03***
8: Δ ROA	-0.03***	-0.03***	-0.03***	0.01	0.05***	0.11***	0.19***	1.00	-0.01	0.01	0.02***	0.01
9: Leverage	-0.11***	-0.11***	-0.11***	0.26***	0.27***	0.05***	-0.05***	0.01*	1.00	0.07***	-0.08***	0.07***
10: Turnover	-0.04***	-0.02***	-0.03***	0.08***	0.45***	0.12***	0.14***	0.00	0.09***	1.00	0.16***	0.39***
11: Volatility	0.61***	0.62***	0.62***	-0.42***	-0.57***	-0.17***	-0.31***	-0.03***	-0.17***	0.10***	1.00	0.14***
12: Market Risk	0.02***	0.05***	0.03***	0.08***	0.24***	0.06***	0.02***	0.02***	0.04***	0.45***	0.23***	1.00

Notes: The table provides correlation results for stock characteristics, Form 10-K readability (gross file size) and firm fundamentals. Pearson correlation coefficients are presented in the upper triangles. Spearman correlation coefficients are presented in the lower triangles.

5.4 Empirical results

Overall, the results provide empirical evidence that Form 10-K readability has an effect on a firm's market performance, market liquidity, and market risk. The regression coefficient on $File\ Size_{Gross}$ for a firm's market performance is significantly negative. Consistent with naïve expectations, less readable annual reports on Form 10-K (higher document file size) are negatively associated with subsequent short-term stock returns. In general, this result is consistent with prior empirical work documenting the negative effect of annual report disclosure complexity on future stock returns (e.g. Gan and Qiu 2018). Investors seem to interpret lengthy (complex) annual reports as a negative sign and assign lower market values to firms with less readable Form 10-K filings (lower market performance). Based on this result, *Hypothesis 7* is confirmed.

Besides the predictive ability of a firm's market performance, the analysis revealed that disclosure complexity in annual reports on Form 10-K has a negative effect on a company's trading activity. The regression coefficient for $File\ Size_{Gross}$ for a firm's abnormal trading volume is significantly negative. Faced with less readable (lengthy) corporate disclosures, investors seem less likely to trade a firm's stock around the SEC submission (lower market liquidity). This result is consistent with prior empirical findings connecting Form 10-K complexity to lower levels of trading volume (e.g. Miller 2010). In addition, the regression coefficient on $File\ Size_{Gross}$ for a firm's abnormal bid-ask spread is also significantly positive. In line with prior research (e.g. Boubaker et al. 2018) less readable annual reports seem to increase investors' inability to agree on a fair market price for a firm's shares (lower market liquidity). Taken together, these results provide evidence that annual report readability has an effect on a firm's market liquidity after an annual report was filed with the SEC. Based on these results, the *Hypothesis 8* is confirmed.

While a company's market performance and market liquidity is negatively affected by less readable annual reports, its market risk is positively affected. Consistent with prior empirical findings (e.g. Loughran and McDonald 2014; Bonsall et al. 2017) the regression coefficient for $File\ Size_{Gross}$ for a firm's abnormal stock return volatility is significantly positive. Given investors limited ability to process complex corporate disclosures, less readable (lengthy) regulatory filings such as annual reports on Form 10-K seem to increase information uncertainty regarding financial markets (higher market risk). Based on this result, *Hypothesis 9* is confirmed.

When interpreting the results note that the document file size ($\text{File Size}_{\text{Gross}}$) is measured by the number of characters appearing in the annual report. Therefore, the magnitude of the coefficients is small. For instance, assuming an average document file size of 5,506,905 characters (univariate analysis) the expected cumulative abnormal return of a company during the event period using the market adjusted model ($\text{CAR}_{[-1,2];\text{MAM}}$) is -0.0589 percent ($-7.11 * 10^{-11} * 5{,}506{,}905 - 0.000197$) or -1.1101 percent ($-5.47 * 10^{-11} * 5{,}506{,}905 - 0.0108$) when using a market model ($\text{CAR}_{[-1,2];\text{MM}}$) and -0.3600 percent ($0 * 5{,}506{,}905 - 0.00360$) when using a 4-factor model ($\text{CAR}_{[-1,2];4\ \text{Factor}}$). Increasing the document file size by 10 percent to 6,057,596 characters ($5{,}506{,}905 * 1.1$) would correspond with an expected cumulative abnormal return of -0.0628 percent (market adjusted model) or with -1.1131 percent (market model) and with -0.3600 percent (4-factor model).

Also note that other variables not included in the study might explain the empirical findings (omitted variable bias). For instance, a firm's corporate governance might influence the disclosure behaviour of a firm in regulatory filings such as annual reports while at the same time affecting a firm's stock characteristics (despite standardized filing, disclosure, and auditing requirements). This holds true for other variables possibly associated with the main variable of interest and the dependent variable (e.g. credit rating changes). Further note the differences in explanatory power for each dependent variable (stock characteristic) which might be explained by their variability itself and various market environments and market forces over time. In a broader context, the (fast) incorporation of textual features into a firm's stock characteristics following the SEC by capital market participants do lend support for the idea of efficient capital markets (efficient markets hypothesis - EHM) and the assumption that all relevant corporate (quantitative and qualitive) information is been processed and incorporated on modern financial markets. The estimated OLS coefficients for annual report readability (document file size) are presented in the Tables 55 to 58.

Table 55. Regression coefficients on Form 10-K readability (gross file size) for abnormal stock return (market-adjusted model, market model, and 4-factor model).

Panel A: Cumulative abnormal stock return (Marked-adjusted model)			
Variable	$CAR_{[-1,2];MAM}$	$CAR_{[-1,2];MAM}$	$CAR_{[-1,2];MAM}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-7.11 * 10^{-11}}$***	$\mathbf{-9.31 * 10^{-11}}$***
		(-3.108)	**(-3.836)**
Ln MVE	-0.000581***		-0.000472***
	(-3.303)		(-2.603)
PtB	-0.000475***		-0.000490***
	(-6.161)		(-6.351)
ROA	0.0259***		0.0255***
	(6.051)		(5.950)
Δ ROA	0.0133**		0.0133**
	(2.048)		(2.057)
Leverage	0.0101***		0.0105***
	(6.403)		(6.616)
Constant	0.00570	-0.000197	0.00323
	(0.960)	(-0.0399)	(0.537)
N	52,484	52,484	52,484
R^2	0.00847	0.00531	0.00858
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Cumulative abnormal stock return (Market model)			
Variable	$CAR_{[-1,2];MM}$	$CAR_{[-1,2];MM}$	$CAR_{[-1,2];MM}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-5.47 * 10^{-11}}$**	$\mathbf{-7.88 * 10^{-11}}$***
		(-2.435)	**(-3.326)**
Ln MVE	-0.000596***		-0.000503***
	(-3.417)		(-2.802)
PtB	-0.000712***		-0.000725***
	(-8.970)		(-9.117)
ROA	0.0249***		0.0246***
	(5.845)		(5.759)
Δ ROA	-0.0136**		-0.0136**
	(-2.125)		(-2.117)
Leverage	0.00999***		0.0103***
	(6.382)		(6.562)
Constant	-0.00391	-0.0108**	-0.00600
	(-0.651)	(-2.170)	(-0.988)
N	52,484	52,484	52,484
R^2	0.0105	0.00631	0.0106
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel C: Cumulative abnormal stock return (4-factor model)			
Variable	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$
$\mathbf{File\ Size_{Gross}}$		**-0**	**-0**
		(-0.637)	**(-1.261)**
Ln MVE	-0.000445***		-0.000412**
	(-2.577)		(-2.314)
PtB	-0.000537***		-0.000542***
	(-7.176)		(-7.226)
ROA	0.0166***		0.0165***
	(3.886)		(3.851)
Δ ROA	-0.00449		-0.00447
	(-0.688)		(-0.685)
Leverage	0.00744***		0.00754***
	(4.828)		(4.889)
Constant	0.00171	-0.00360	0.000952
	(0.295)	(-0.745)	(0.162)
N	52,484	52,484	52,484
R^2	0.00490	0.00271	0.00490
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability. Form 10-K readability scores (gross file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

Table 56. Regression coefficients on Form 10-K readability (gross file size) for abnormal trading volume (average and median adjusted trading volume).

Panel A: Abnormal trading volume (average adjusted)			
Variable	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$
$\mathbf{File\ Size_{Gross}}$		**-0*** **	$\mathbf{-5.91 * 10^{-11}}$***
		(-3.047)	**(-3.739)**
Ln MVE	$-5.11 * 10^{-5}$		2.90e-05
	(-0.479)		(0.268)
PtB	0.000164***		0.000155***
	(4.434)		(4.206)
ROA	0.00779***		0.00755***
	(4.392)		(4.260)
Turnover	-0.0669		-0.0652
	(-1.255)		(-1.223)
Volatility	-0.0359***		-0.0350***
	(-3.284)		(-3.189)
Beta	0.00190***		0.00185***
	(5.371)		(5.242)
Constant	0.00287	0.00305	0.00121
	(0.817)	(1.098)	(0.341)
N	55,263	55,263	55,263
R^2	0.0235	0.0210	0.0238
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Abnormal trading volume (median adjusted)			
Variable	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{File\ Size_{Gross}}$		**-0****	$\mathbf{-7.10 * 10^{-11}}$***
		(-2.366)	**(-4.431)**
Ln MVE	-0.000821***		-0.000725***
	(-7.588)		(-6.583)
PtB	0.000207***		0.000198***
	(5.439)		(5.180)
ROA	0.00518***		0.00490***
	(2.855)		(2.701)
Turnover	1.010***		1.012***
	(18.92)		(18.94)
Volatility	0.0402***		0.0413***
	(3.552)		(3.647)
Beta	$4.05 * 10^{-5}$		$-1.50 * 10^{-5}$
	(0.112)		(-0.0418)
Constant	0.0189***	0.00557*	0.0169***
	(5.637)	(1.694)	(4.976)
N	55,263	55,263	55,263
R^2	0.0764	0.0256	0.0767
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability. Form 10-K readability scores (gross file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic (p < 1%). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

Table 57. Regression coefficients on Form 10-K readability (gross file size) for abnormal bid-ask spread (average and median adjusted bid-ask spread).

Panel A: Abnormal bid-ask spread (average adjusted)			
Variable	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$	$ASpread_{[-1,3];Ave}$
$\mathbf{File\ Size_{Gross}}$		**0**	**0***
		(0.453)	**(1.915)**
Ln MVE	-0.00182***		-0.00193***
	(-5.715)		(-5.959)
PtB	2.87e-05		$3.77 * 10^{-5}$
	(0.288)		(0.379)
ROA	-0.0102*		-0.00986*
	(-1.808)		(-1.742)
Turnover	0.107**		0.107**
	(2.005)		(2.014)
Volatility	-0.305***		-0.306***
	(-4.466)		(-4.489)
Beta	0.00685***		0.00692***
	(6.084)		(6.117)
Constant	0.126***	0.0883***	0.128***
	(14.91)	(18.45)	(15.05)
N	29,768	29,768	29,768
R^2	0.0168	0.0126	0.0169
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (median adjusted)			
Variable	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$	$ASpread_{[-1,3];Med}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-2.94 * 10^{-10}}$***	$\mathbf{6.27 * 10^{-11}}$**
		(-5.569)	**(2.531)**
Ln MVE	-0.00373***		-0.00388***
	(-11.10)		(-11.37)
PtB	$8.52 * 10^{-5}$		$9.82 * 10^{-5}$
	(0.826)		(0.952)
ROA	-0.00343		-0.00294
	(-0.577)		(-0.493)
Turnover	-0.373***		-0.373***
	(-6.838)		(-6.835)
Volatility	0.313***		0.310***
	(4.903)		(4.872)
Beta	-0.00822***		-0.00812***
	(-7.447)		(-7.322)
Constant	0.178***	0.113***	0.180***
	(19.18)	(17.08)	(19.35)
N	29,768	29,768	29,768
R^2	0.0630	0.0202	0.0630
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability. Form 10-K readability scores (gross file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic (p < 1%). The average variance inflation factor (VIF) was below 3 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

Table 58. Regression coefficients on Form 10-K readability (gross file size) for abnormal stock return volatility (market-adjusted model, market model, and 4-factor model).

Panel A: Abnormal stock return volatility (Market adjusted model)			
Variable	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-7.28 * 10^{-10}}$***	$\mathbf{1.77 * 10^{-10}}$***
		(-5.680)	**(5.130)**
Ln MVE	-0.0111***		-0.0113***
	(-38.86)		(-38.48)
PtB	0.000567***		0.000591***
	(6.139)		(6.383)
ROA	-0.0498***		-0.0490***
	(-9.616)		(-9.447)
Turnover	0.512***		0.507***
	(7.454)		(7.383)
Volatility	1.562***		1.559***
	(37.01)		(36.94)
Beta	0.00890***		0.00905***
	(9.797)		(9.941)
Constant	0.260***	0.0655***	0.265***
	(31.75)	(7.037)	(31.94)
N	53,121	53,121	53,121
R^2	0.376	0.173	0.377
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Abnormal stock return volatility (Market model)			
Variable	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-6.96 * 10^{-10}}$***	$\mathbf{1.78 * 10^{-10}}$***
		(-5.646)	**(4.991)**
Ln MVE	-0.0104***		-0.0106***
	(-35.69)		(-35.37)
PtB	0.000618***		0.000642***
	(6.606)		(6.845)
ROA	-0.0487***		-0.0479***
	(-9.366)		(-9.197)
Turnover	0.548***		0.543***
	(7.994)		(7.923)
Volatility	1.666***		1.663***
	(38.49)		(38.43)
Beta	0.0101***		0.0103***
	(11.10)		(11.24)
Constant	0.246***	0.0685***	0.251***
	(29.24)	(7.166)	(29.45)
N	53,121	53,121	53,121
R^2	0.371	0.162	0.371
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel C: Abnormal stock return volatility (4-factor model)			
Variable	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-7.08 * 10^{-10}}$***	$\mathbf{2.16 * 10^{-10}}$***
		(-5.663)	**(5.325)**
Ln MVE	-0.0107***		-0.0109***
	(-36.04)		(-35.79)
PtB	0.000453***		0.000483***
	(5.168)		(5.482)
ROA	-0.0463***		-0.0453***
	(-9.043)		(-8.841)
Turnover	0.452***		0.445***
	(6.636)		(6.546)
Volatility	1.743***		1.739***
	(40.04)		(39.97)
Beta	0.00677***		0.00694***
	(7.434)		(7.618)
Constant	0.257***	0.0730***	0.263***
	(29.77)	(7.404)	(30.06)
N	53,121	53,121	53,121
R^2	0.381	0.161	0.382
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability. Form 10-K readability scores (gross file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic (p < 1%). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

5.5 Robustness and sensitivity tests

As a first robustness test, additional and more restrictive data filters were applied to the data sample and overall sample selection process. From the selected sample, firm-year observations of the financial industry were excluded (SIC codes 6021-6799). Substantial accounting differences between financial and non-financial entities as well as differences in their regulatory framework might affect the corresponding narrative disclosures and therefore the empirical results (Ditter and Scherr 2015, 60). In addition, the study removed firms with a low market capitalization (10 percent quantile of market capitalization) from the data set. By eliminating low-priced firms, the study addresses concerns that the regression results are influenced by bid-ask bounces around the filing date of an annual report while investigating the effect of Form 10-K readability on important firm characteristics (Loughran and McDonald 2011a, 40; Loughran and McDonald 2015, 3). Applying these additional and more restrictive data filters to the data sample, the analysis was rerun.

The regression results confirmed the findings. The regression coefficient on $File\ Size_{Gross}$ for a firm's subsequent stock market performance is significantly negative. While a firm's subsequent market liquidity is again negatively affected by the file size of its annual report (negative sign on abnormal trading volume), the market risk is positively affected (positive sign on abnormal stock return volatility). Tables 59 to 62 present the estimated OLS coefficients on gross document file size (readability).

The accounting and finance literature uses various methods to measure the readability of Form 10-K filings made with the SEC. For instance, numerous studies used indices based on textual complexity (e.g. Fog Index, Flesch Index, Flesch-Kincade Index) to determine the readability of regulatory filings (e.g. Li 2008). Recent empirical works also provide evidence that different submission size measurements (e.g. gross file size, net file size) are associated with important firm characteristics (e.g. Loughran and McDonald 2014). While the gross file size of a Form 10-K submission measures the document size as is (no parsing procedure), the net file size accounts for non-readable or non-relevant elements (e.g. mark-up tags, ASCII-encoded graphics, tables).[45] To examine whether the presented empirical results are sensitive to the specific measurement of document file size, a corresponding test was run.

[45] As noted by Gan and Qiu (2018), the *"net file size is the total number of characters in a 10-K filing after the Stage One Parse process, in which all 10-K text filings and their variants are distilled into cleaned text files. This process substantially decreases the file sizes by excluding extraneous material such as HTML, ASCII-encoded segments, and tables"* (Gan and Qiu 2018, 11).

The regression results did not confirm the prior findings. Despite showing the expected negative sign, the regression coefficient on $File\ Size_{Net}$ for a firm's subsequent stock market performance is insignificant. In addition, the regression coefficient on $File\ Size_{Net}$ for a firm's abnormal trading volume is significantly positive (insignificantly positive for abnormal bid-ask spread). However, the regression coefficient on $File\ Size_{Net}$ for a firm's abnormal stock return volatility is significantly positive. Tables 63 to 66 present the estimated OLS coefficients on net document file size (readability).

Table 59. Regression coefficients on Form 10-K readability (gross file size) for abnormal stock return (market-adjusted model and market model) excluding firm-year observations of the financial industry and low market capitalization.

Panel A: Cumulative abnormal stock return (Marked-adjusted model)			
Variable	$CAR_{[-1,2];MAM}$	$CAR_{[-1,2];MAM}$	$CAR_{[-1,2];MAM}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-9.33 * 10^{-11}}$*******	$\mathbf{-1.42 * 10^{-10}}$*******
		(-3.333)	**(-4.338)**
Ln MVE	-0.000192		-2.94e-05
	(-0.957)		(-0.143)
PtB	-0.000488***		-0.000507***
	(-5.579)		(-5.789)
ROA	0.0251***		0.0247***
	(5.357)		(5.264)
Δ ROA	0.0174**		0.0173**
	(2.449)		(2.442)
Leverage	0.00918***		0.00972***
	(5.198)		(5.481)
Constant	0.000297	0.00196	-0.00335
	(0.0472)	(0.395)	(-0.527)
N	41,294	41,294	41,294
R^2	0.0102	0.00697	0.0105
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Cumulative abnormal stock return (Market model)			
Variable	$CAR_{[-1,2];MM}$	$CAR_{[-1,2];MM}$	$CAR_{[-1,2];MM}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-6.47 * 10^{-11}}$******	$\mathbf{-1.19 * 10^{-10}}$*******
		(-2.394)	**(-3.936)**
Ln MVE	-0.000150		-1.28e-05
	(-0.755)		(-0.0633)
PtB	-0.000759***		-0.000775***
	(-8.621)		(-8.788)
ROA	0.0243***		0.0239***
	(5.231)		(5.152)
Δ ROA	-0.0105		-0.0106
	(-1.499)		(-1.505)
Leverage	0.00970***		0.0102***
	(5.569)		(5.806)
Constant	-0.0109*	-0.00882*	-0.0139**
	(-1.701)	(-1.739)	(-2.161)
N	41,294	41,294	41,294
R^2	0.0130	0.00841	0.0132
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel C: Cumulative abnormal stock return (4-factor model)			
Variable	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$
$\mathbf{File\ Size_{Gross}}$		**0**	$\mathbf{-6.86 * 10^{-11}}$**
		(-0.923)	**(-2.552)**
Ln MVE	$4.99 * 10^{-5}$		0.000128
	(0.254)		(0.639)
PtB	-0.000551***		-0.000560***
	(-6.679)		(-6.779)
ROA	0.0152***		0.0150***
	(3.325)		(3.278)
Δ ROA	-0.00246		-0.00249
	(-0.349)		(-0.352)
Leverage	0.00664***		0.00690***
	(3.872)		(4.009)
Constant	-0.00818	-0.00388	-0.00995
	(-1.302)	(-0.769)	(-1.569)
N	41,294	41,294	41,294
R^2	0.00516	0.00295	0.00521
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability (excluding SIC codes 6021-6799). Form 10-K readability scores (gross file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits).

Table 60. Regression coefficients on Form 10-K readability (gross file size) for abnormal trading volume (average and median adjusted) excluding firm-year observations of the financial industry and low market capitalization.

Panel A: Abnormal trading volume (average adjusted)			
Variable	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-5.17 * 10^{-11}}$**	$\mathbf{-5.85 * 10^{-11}}$***
		(-2.526)	**(-2.845)**
Ln MVE	-0.000117		-4.20e-05
	(-0.857)		(-0.306)
PtB	0.000226***		0.000220***
	(4.865)		(4.735)
ROA	0.00772***		0.00753***
	(3.411)		(3.333)
Turnover	-0.0459		-0.0443
	(-0.784)		(-0.755)
Volatility	-0.0277*		-0.0271
	(-1.668)		(-1.632)
Beta	0.00232***		0.00228***
	(5.356)		(5.269)
Constant	0.00358	0.00295	0.00205
	(0.905)	(1.014)	(0.514)
N	43,106	43,106	43,106
R^2	0.0258	0.0236	0.0260
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Abnormal trading volume (median adjusted)			
Variable	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{File\ Size_{Gross}}$		**0****	$\mathbf{-6.44 * 10^{-11}}$***
		(-2.123)	**(-3.121)**
Ln MVE	-0.000940***		-0.000857***
	(-6.892)		(-6.248)
PtB	0.000270***		0.000263***
	(5.697)		(5.561)
ROA	0.00597***		0.00576**
	(2.593)		(2.507)
Turnover	0.972***		0.974***
	(16.81)		(16.82)
Volatility	0.0765***		0.0771***
	(4.465)		(4.497)
Beta	0.000118		$7.40 * 10^{-5}$
	(0.271)		(0.170)
Constant	0.0204***	0.00545	0.0188***
	(5.375)	(1.587)	(4.891)
N	43,106	43,106	43,106
R^2	0.0757	0.0249	0.0758
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability (excluding SIC codes 6021-6799). Form 10-K readability scores (gross file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits).

Table 61. Regression coefficients on Form 10-K readability (gross file size) for abnormal bid-ask spread (average and median adjusted) excluding firm-year observations of the financial industry and low market capitalization.

Panel A: Abnormal bid-ask spread (average adjusted)			
Variable	$\text{ASpread}_{[-1,3];\text{Ave}}$	$\text{ASpread}_{[-1,3];\text{Ave}}$	$\text{ASpread}_{[-1,3];\text{Ave}}$
$\mathbf{File\ Size_{Gross}}$		**0**	**0**
		(0.212)	**(0.269)**
Ln MVE	-0.000893***		-0.000901***
	(-5.133)		(-5.112)
PtB	$1.85 * 10^{-5}$		$1.90 * 10^{-5}$
	(0.334)		(0.344)
ROA	-0.00351		-0.00348
	(-1.160)		(-1.151)
Turnover	0.101***		0.101***
	(3.694)		(3.692)
Volatility	-0.228***		-0.228***
	(-6.376)		(-6.375)
Beta	0.00450***		0.00451***
	(6.951)		(6.946)
Constant	0.107***	0.0875***	0.108***
	(22.11)	(27.28)	(21.98)
N	23,351	23,351	23,351
R^2	0.0239	0.0166	0.0239
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (median adjusted)			
Variable	$\text{ASpread}_{[-1,3];\text{Med}}$	$\text{ASpread}_{[-1,3];\text{Med}}$	$\text{ASpread}_{[-1,3];\text{Med}}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-1.09 * 10^{-10}}$***	**0****
		(-3.859)	**(2.237)**
Ln MVE	-0.00185***		-0.00193***
	(-10.06)		(-10.30)
PtB	$5.56 * 10^{-5}$		$6.11 * 10^{-5}$
	(0.976)		(1.071)
ROA	$5.09 * 10^{-6}$		0.000236
	(0.00163)		(0.0755)
Turnover	-0.158***		-0.158***
	(-5.672)		(-5.678)
Volatility	0.0871**		0.0865**
	(2.486)		(2.467)
Beta	-0.00368***		-0.00364***
	(-5.708)		(-5.631)
Constant	0.149***	0.116***	0.150***
	(24.19)	(22.92)	(24.32)
N	23,351	23,351	23,351
R^2	0.0463	0.0196	0.0464
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability (excluding SIC codes 6021-6799). Form 10-K readability scores (gross file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits).

Table 62. Regression coefficients on Form 10-K readability (gross file size) for abnormal return volatility (market-adjusted model and market model) excluding firm-year observations of the financial industry and low market capitalization.

Panel A: Abnormal stock return volatility (Market adjusted model)			
Variable	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-6.56 * 10^{-10}}$***	$\mathbf{1.43 * 10^{-10}}$***
		(-4.001)	**(3.549)**
Ln MVE	-0.00949***		-0.00967***
	(-32.57)		(-32.37)
PtB	0.000570***		0.000584***
	(5.926)		(6.065)
ROA	-0.0441***		-0.0437***
	(-8.333)		(-8.232)
Turnover	0.525***		0.521***
	(7.852)		(7.800)
Volatility	1.622***		1.621***
	(34.45)		(34.44)
Beta	0.00929***		0.00939***
	(9.848)		(9.942)
Constant	0.225***	0.0645***	0.229***
	(26.93)	(7.482)	(27.11)
N	41,625	41,625	41,625
R^2	0.355	0.168	0.355
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Abnormal stock return volatility (Market model)			
Variable	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-6.38 * 10^{-10}}$***	$\mathbf{1.44 * 10^{-10}}$***
		(-3.984)	**(3.503)**
Ln MVE	-0.00904***		-0.00923***
	(-30.63)		(-30.45)
PtB	0.000620***		0.000634***
	(6.383)		(6.519)
ROA	-0.0438***		-0.0433***
	(-8.317)		(-8.216)
Turnover	0.574***		0.570***
	(8.597)		(8.545)
Volatility	1.703***		1.702***
	(35.43)		(35.41)
Beta	0.00981***		0.00991***
	(10.28)		(10.37)
Constant	0.216***	0.0668***	0.220***
	(25.32)	(7.559)	(25.52)
N	41,625	41,625	41,625
R^2	0.353	0.160	0.353
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel C: Abnormal stock return volatility (4-factor model)			
Variable	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-6.20 * 10^{-10}}$***	$\mathbf{1.84 * 10^{-10}}$***
		(-4.008)	**(3.772)**
Ln MVE	-0.00915***		-0.00938***
	(-31.02)		(-30.92)
PtB	0.000441***		0.000460***
	(4.820)		(5.012)
ROA	-0.0403***		-0.0396***
	(-7.906)		(-7.778)
Turnover	0.497***		0.492***
	(7.532)		(7.462)
Volatility	1.731***		1.729***
	(36.35)		(36.34)
Beta	0.00658***		0.00671***
	(7.013)		(7.144)
Constant	0.224***	0.0710***	0.229***
	(26.05)	(8.038)	(26.35)
N	41,625	41,625	41,625
R^2	0.356	0.158	0.356
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability (excluding SIC codes 6021-6799). Form 10-K readability scores (gross file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits).

Table 63. Regression coefficients on Form 10-K readability (net file size) for abnormal stock return (market-adjusted model, market model, and 4-factor model).

Panel A: Cumulative abnormal stock return (Marked-adjusted model)			
Variable	$CAR_{[-1,2];MAM}$	$CAR_{[-1,2];MAM}$	$CAR_{[-1,2];MAM}$
$\mathbf{File\ Size_{Net}}$		$\mathbf{-2.01 * 10^{-10}}$	$\mathbf{-5.05 * 10^{-10}}$
		(-0.192)	**(-0.445)**
Ln MVE	-0.000581***		-0.000560***
	(-3.303)		(-3.052)
PtB	-0.000475***		-0.000477***
	(-6.161)		(-6.174)
ROA	0.0259***		0.0258***
	(6.051)		(5.966)
Δ ROA	0.0133**		0.0133**
	(2.048)		(2.052)
Leverage	0.0101***		0.0103***
	(6.403)		(6.400)
Constant	0.00570	$-6.20 * 10^{-5}$	0.00536
	(0.960)	(-0.0126)	(0.894)
N	52,484	52,484	52,484
R^2	0.00847	0.00523	0.00846
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Cumulative abnormal stock return (Market model)			
Variable	$CAR_{[-1,2];MM}$	$CAR_{[-1,2];MM}$	$CAR_{[-1,2];MM}$
$\mathbf{File\ Size_{Net}}$		$\mathbf{1.51 * 10^{-10}}$	$\mathbf{-1.01 * 10^{-10}}$
		(0.144)	**(-0.0883)**
Ln MVE	-0.000596***		-0.000591***
	(-3.417)		(-3.253)
PtB	-0.000712***		-0.000712***
	(-8.970)		(-8.950)
ROA	0.0249***		0.0249***
	(5.845)		(5.790)
Δ ROA	-0.0136**		-0.0136**
	(-2.125)		(-2.123)
Leverage	0.00999***		0.0100***
	(6.382)		(6.327)
Constant	-0.00391	-0.0108**	-0.00397
	(-0.651)	(-2.161)	(-0.657)
N	52,484	52,484	52,484
R^2	0.0105	0.00627	0.0105
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel C: Cumulative abnormal stock return (4-factor model)			
Variable	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$	$CAR_{[-1,2];4\ Factor}$
$\mathbf{File\ Size_{Net}}$		$\mathbf{4.49 * 10^{-10}}$	$\mathbf{3.02 * 10^{-10}}$
		(0.432)	**(0.269)**
Ln MVE	-0.000445***		-0.000458**
	(-2.577)		(-2.536)
PtB	-0.000537***		-0.000535***
	(-7.176)		(-7.134)
ROA	0.0166***		0.0167***
	(3.886)		(3.876)
Δ ROA	-0.00449		-0.00451
	(-0.688)		(-0.691)
Leverage	0.00744***		0.00737***
	(4.828)		(4.728)
Constant	0.00171	-0.00369	0.00192
	(0.295)	(-0.763)	(0.327)
N	52,484	52,484	52,484
R^2		0.00490	0.00271
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients on annual report readability (entire data sample). Form 10-K readability scores (net file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

Table 64. Regression coefficients on Form 10-K readability (net file size) for abnormal trading volume (average and median adjusted).

Panel A: Abnormal trading volume (average adjusted)			
Variable	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$	$ATV_{[-1,3];Ave}$
$\mathbf{File\ Size_{Net}}$		$\mathbf{2.44 * 10^{-9}}$***	$\mathbf{2.51 * 10^{-9}}$***
		(4.726)	**(4.675)**
Ln MVE	$-5.11 * 10^{-5}$		-0.000166
	(-0.479)		(-1.512)
PtB	0.000164***		0.000173***
	(4.434)		(4.687)
ROA	0.00779***		0.00851***
	(4.392)		(4.783)
Turnover	-0.0669		-0.0711
	(-1.255)		(-1.337)
Volatility	-0.0359***		-0.0381***
	(-3.284)		(-3.480)
Beta	0.00190***		0.00190***
	(5.371)		(5.372)
Constant	0.00287	0.00250	0.00457
	(0.817)	(0.924)	(1.300)
N	55,263	55,263	55,263
R^2	0.0235	0.0213	0.0240
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Abnormal trading volume (median adjusted)			
Variable	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$	$ATV_{[-1,3];Med}$
$\mathbf{File\ Size_{Net}}$		$\mathbf{4.49 * 10^{-9}}$***	$\mathbf{2.80 * 10^{-9}}$***
		(7.294)	**(5.088)**
Ln MVE	-0.000821***		-0.000949***
	(-7.588)		(-8.455)
PtB	0.000207***		0.000217***
	(5.439)		(5.714)
ROA	0.00518***		0.00599***
	(2.855)		(3.290)
Turnover	1.010***		1.005***
	(18.92)		(18.89)
Volatility	0.0402***		0.0377***
	(3.552)		(3.327)
Beta	$4.05 * 10^{-5}$		$4.11 * 10^{-5}$
	(0.112)		(0.114)
Constant	0.0189***	0.00451	0.0208***
	(5.637)	(1.418)	(6.183)
N	55,263	55,263	55,263
R^2	0.0764	0.0270	0.0770
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients on annual report readability (entire data sample). Form 10-K readability scores (net file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

Table 65. Regression coefficients on Form 10-K readability (net file size) for abnormal bid-ask spread (average and median adjusted).

Panel A: Abnormal bid-ask spread (average adjusted)			
Variable	$\text{ASpread}_{[-1,3];\text{Ave}}$	$\text{ASpread}_{[-1,3];\text{Ave}}$	$\text{ASpread}_{[-1,3];\text{Ave}}$
$\text{File Size}_{\text{Net}}$		$\mathbf{2.32 * 10^{-9}}$**	$\mathbf{3.11 * 10^{-9}}$***
		(2.061)	**(3.116)**
Ln MVE	-0.00182***		-0.00197***
	(-5.715)		(-6.065)
PtB	$2.87 * 10^{-5}$		$3.97 * 10^{-5}$
	(0.288)		(0.399)
ROA	-0.0102*		-0.00942*
	(-1.808)		(-1.665)
Turnover	0.107**		0.102*
	(2.005)		(1.922)
Volatility	-0.305***		-0.309***
	(-4.466)		(-4.524)
Beta	0.00685***		0.00682***
	(6.084)		(6.060)
Constant	0.126***	0.0876***	0.128***
	(14.91)	(18.46)	(15.11)
N	29,768	29,768	29,768
R^2	0.0168	0.0127	0.0170
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Abnormal bid-ask spread (median adjusted)			
Variable	$\text{ASpread}_{[-1,3];\text{Med}}$	$\text{ASpread}_{[-1,3];\text{Med}}$	$\text{ASpread}_{[-1,3];\text{Med}}$
$\text{File Size}_{\text{Net}}$		$\mathbf{-1.31 * 10^{-8}}$***	$\mathbf{-1.38 * 10^{-9}}$
		(-8.569)	**(1.356)**
Ln MVE	-0.00373***		-0.00380***
	(-11.10)		(-11.09)
PtB	$8.52 * 10^{-5}$		$9.01 * 10^{-5}$
	(0.826)		(0.873)
ROA	-0.00343		-0.00309
	(-0.577)		(-0.516)
Turnover	-0.373***		-0.375***
	(-6.838)		(-6.882)
Volatility	0.313***		0.311***
	(4.903)		(4.865)
Beta	-0.00822***		-0.00823***
	(-7.447)		(-7.465)
Constant	0.178***	0.117***	0.178***
	(19.18)	(17.04)	(19.22)
N	29,768	29,768	29,768
R^2	0.0630	0.0204	0.0630
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients on annual report readability (entire data sample). Form 10-K readability scores (net file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic (p < 1%). The average variance inflation factor (VIF) was below 3 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

Table 66. Regression coefficients on Form 10-K readability (net file size) for abnormal return volatility (market-adjusted model, market model, and 4-factor model).

Panel A: Abnormal stock return volatility (Market adjusted model)			
Variable	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$	$ARV_{[-1,3];MAM}$
$File\ Size_{Gross}$		**$-1.22 * 10^{-8}$*****	**$9.86 * 10^{-9}$*****
		(-6.583)	**(7.813)**
Ln MVE	-0.0111***		-0.0115***
	(-38.86)		(-39.19)
PtB	0.000567***		0.000600***
	(6.139)		(6.492)
ROA	-0.0498***		-0.0469***
	(-9.616)		(-9.012)
Turnover	0.512***		0.494***
	(7.454)		(7.225)
Volatility	1.562***		1.553***
	(37.01)		(36.81)
Beta	0.00890***		0.00893***
	(9.797)		(9.845)
Constant	0.260***	0.0695***	0.266***
	(31.75)	(7.359)	(32.55)
N	53,121	53,121	53,121
R^2	0.376	0.170	0.377
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel B: Abnormal stock return volatility (Market model)			
Variable	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$	$ARV_{[-1,3];MM}$
$File\ Size_{Gross}$		**$-1.04 * 10^{-8}$*****	**$-1.00 * 10^{-8}$*****
		(-5.597)	**(7.832)**
Ln MVE	-0.0104***		-0.0108***
	(-35.69)		(-36.20)
PtB	0.000618***		0.000651***
	(6.606)		(6.957)
ROA	-0.0487***		-0.0457***
	(-9.366)		(-8.761)
Turnover	0.548***		0.530***
	(7.994)		(7.756)
Volatility	1.666***		1.657***
	(38.49)		(38.29)
Beta	0.0101***		0.0102***
	(11.10)		(11.15)
Constant	0.246***	0.0720***	0.253***
	(29.24)	(7.447)	(30.10)
N	53,121	53,121	53,121
R^2	0.371	0.159	0.372
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Panel C: Abnormal stock return volatility (4-factor model)			
Variable	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$	$ARV_{[-1,3];4\ Factor}$
$\mathbf{File\ Size_{Gross}}$		$\mathbf{-1.20 * 10^{-8}}$***	$\mathbf{-1.05 * 10^{-8}}$***
		(-6.326)	**(7.996)**
Ln MVE	-0.0107***		-0.0111***
	(-36.04)		(-36.50)
PtB	0.000453***		0.000489***
	(5.168)		(5.565)
ROA	-0.0463***		-0.0432***
	(-9.043)		(-8.407)
Turnover	0.452***		0.433***
	(6.636)		(6.384)
Volatility	1.743***		1.733***
	(40.04)		(39.84)
Beta	0.00677***		0.00680***
	(7.434)		(7.482)
Constant	0.257***	0.0769***	0.264***
	(29.77)	(7.691)	(30.60)
N	53,121	53,121	53,121
R^2	0.381	0.158	0.382
Fixed Effects	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients on annual report readability (entire data sample). Form 10-K readability scores (net file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For each dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine each dependent variable.

5.6 Additional tests

The majority of empirical studies (e.g. Loughran and McDonald 2014) connect the readability of regulatory filings and a firm's future stock market characteristics (e.g. stock return, volatility). Naturally, a question arises regarding the implications of annual report readability on other firm characteristics.

To explore the implications of Form 10-K readability on important firm characteristics relevant to investors, bondholders, and other capital market participants, a series of corresponding tests were performed. Therefore, the following OLS regression model was estimated:

$$\begin{aligned} Y_{i;t} = \beta_0 + \beta_1 * File\ Size_{Type;i;t} + \beta_2 * Ln\ MVE_{i;t} + \beta_3 * PtB_{i;t} + \beta_4 * ROA_{i;t} \\ + \sum_X \beta_X * Other\ Controls_{i;t} + \sum Year\ Fixed\ Effects \\ + \sum Industry\ Fixed\ Effects + \varepsilon_{i;t} \end{aligned} \tag{38}$$

Where: the dependent variable ($Y_{i;t}$) is a firm's future accounting profitability and dividend payments to investors. $File\ Size_{Type;i;t}$ is the file size of a firm's annual report submitted to the SEC (gross file size, net file size).

Note that a firm's future accounting profitability is its future return on assets (ROA) calculated as the future earnings before interest divided by the future average of total assets. Furthermore, a firm's future dividend payments to investors is its future dividend yield (DY) calculated as the future dividends per share divided by the future price per share (fiscal year end), and its future payout ratio (PR) calculated as the future dividends per share divided by the future earnings per share.

The regression results showed an association between annual report readability and a firm's future accounting profitability. Consistent with naïve expectations and prior empirical research (e.g. Gan and Qiu 2018), less readable Form 10-K filings are significantly associated with lower future corporate earnings (negative signs on $File\ Size_{Gross}$ and $File\ Size_{Net}$). This result holds true when using the net file size as a measure for the readability of annual reports on Form 10-K filed with the SEC. In addition, the regression results revealed that less readable annual reports are associated with a lower payout ratio in the future (negative sign on $File\ Size_{Net}$). Tables 67 to 69 present the estimated OLS coefficients on document file size (readability) readability.

Table 67. Regression coefficients on Form 10-K readability (gross and net file size) for return on assets.

Variable	ROA_{t+1}	ROA_{t+1}	ROA_{t+1}	ROA_{t+1}	ROA_{t+1}	ROA_{t+1}
$\text{File Size}_{\text{Gross}}$		$\mathbf{6.04 * 10^{-10}}$***	$\mathbf{-1.03 * 10^{-10}}$***			
		(5.182)	**(-4.413)**			
$\text{File Size}_{\text{Net}}$					$\mathbf{-6.95 * 10^{-9}}$***	$\mathbf{-1.06 * 10^{-8}}$***
					(-2.869)	**(-11.69)**
Ln MVE	0.00238***		0.00250***	0.00238***		0.00282***
	(13.49)		(13.89)	(13.49)		(15.50)
PtB	-0.000766***		-0.000781***	-0.000766***		-0.000816***
	(-5.677)		(-5.786)	(-5.677)		(-6.052)
ROA	0.771***		0.770***	0.771***		0.768***
	(128.9)		(128.8)	(128.9)		(127.9)
Δ ROA	-0.0185***		-0.0185***	-0.0185***		-0.0180***
	(-2.643)		(-2.649)	(-2.643)		(-2.578)
Leverage	0.0273***		0.0277***	0.0273***		0.0295***
	(14.53)		(14.66)	(14.53)		(15.53)
Constant	-0.0431***	0.0734***	-0.0457***	-0.0431***	0.0742***	-0.0503***
	(-7.596)	(4.956)	(-7.996)	(-7.596)	(5.050)	(-9.017)
N	59,239	59,239	59,239	59,239	59,239	59,239
R^2	0.753	0.182	0.753	0.753	0.181	0.753
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability. Form 10-K readability scores (gross file size, net file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

Table 68. Regression coefficients on Form 10-K readability (gross and net file size) for dividend yield.

Variable	DY_{t+1}	DY_{t+1}	DY_{t+1}	DY_{t+1}	DY_{t+1}	DY_{t+1}
$\mathbf{File\ Size_{Gross}}$		$\mathbf{7.37 * 10^{-11}}$***	$\mathbf{1.15 * 10^{-11}}$			
		(3.473)	**(1.589)**			
$\mathbf{File\ Size_{Net}}$					$\mathbf{5.06 * 10^{-10}}$	$\mathbf{-2.80 * 10^{-10}}$
					(0.743)	**(-1.210)**
Ln MVE	0.000218***		0.000205***	0.000218***		0.000230***
	(8.229)		(7.570)	(8.229)		(8.106)
PtB	$5.21 * 10^{-5}$***		$5.39 * 10^{-5}$***	$5.21 * 10^{-5}$***		$5.08 * 10^{-5}$***
	(3.786)		(3.900)	(3.786)		(3.665)
ROA	0.00416***		0.00420***	0.00416***		0.00407***
	(10.54)		(10.64)	(10.54)		(10.18)
Δ ROA	0.000193		0.000192	0.000193		0.000205
	(0.289)		(0.288)	(0.289)		(0.306)
Leverage	-0.000910***		-0.000950***	-0.000910***		-0.000850***
	(-3.940)		(-4.121)	(-3.940)		(-3.612)
DY	0.879***		0.879***	0.879***		0.879***
	(147.4)		(147.5)	(147.4)		(147.1)
Δ DY	-0.182***		-0.182***	-0.182***		-0.182***
	(-14.15)		(-14.14)	(-14.15)		(-14.14)
Constant	-0.00446***	0.0196***	-0.00417***	-0.00446***	0.0194***	-0.00465***
	(-4.157)	(6.897)	(-3.835)	(-4.157)	(6.838)	(-4.310)
N	50,693	50,693	50,693	50,693	50,693	50,693
R^2	0.805	0.328	0.805	0.805	0.328	0.805
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability. Form 10-K readability scores (gross file size, net file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic ($p < 1\%$). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

Table 69. Regression coefficients on Form 10-K readability (gross and net file size) for payout ratio.

Variable	PR_{t+1}	PR_{t+1}	PR_{t+1}	PR_{t+1}	PR_{t+1}	PR_{t+1}
File Size$_{Gross}$		$\mathbf{1.98 * 10^{-9}}$***	$\mathbf{1.64 * 10^{-10}}$*			
		(4.888)	**(1.727)**			
File Size$_{Net}$					$\mathbf{-1.72 * 10^{-8}}$**	$\mathbf{-1.04 * 10^{-8}}$***
					(-2.528)	**(-4.869)**
Ln MVE	0.00324***		0.00308***	0.00324***		0.00367***
	(10.61)		(9.829)	(10.61)		(11.45)
PtB	0.000214**		0.000236**	0.000214**		0.000168
	(2.106)		(2.309)	(2.106)		(1.642)
ROA	0.0562***		0.0566***	0.0562***		0.0536***
	(13.36)		(13.47)	(13.36)		(12.76)
Δ ROA	-0.00924*		-0.00920*	-0.00924*		-0.00868*
	(-1.762)		(-1.755)	(-1.762)		(-1.655)
Leverage	-0.00677***		-0.00737***	-0.00677***		-0.00467**
	(-3.033)		(-3.269)	(-3.033)		(-2.061)
PR	0.902***		0.902***	0.902***		0.900***
	(168.8)		(168.8)	(168.8)		(166.7)
Δ PR	-0.217***		-0.217***	-0.217***		-0.216***
	(-16.36)		(-16.37)	(-16.36)		(-16.25)
Constant	-0.0263	0.312***	-0.0225	-0.0263	0.314***	-0.0325*
	(-1.456)	(5.240)	(-1.236)	(-1.456)	(5.372)	(-1.792)
N	40,420	40,420	40,420	40,420	40,420	40,420
R²	0.791	0.259	0.792	0.791	0.255	0.792
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Error Clustering	Firm	Firm	Firm	Firm	Firm	Firm

Notes: The table shows the OLS regression coefficients for annual report readability. Form 10-K readability scores (gross file size, net file size) were obtained using the Loughran and McDonald 10X Summaries (submission including exhibits). For the dependent variable and corresponding regression model diagnostics were conducted (normality of residuals, homoscedasticity of residuals, multicollinearity of predictors). For all dependent variables the results showed that the residuals follow a fairly normal distribution. The White's test and the Breusch-Pagan test suggested that the residuals are heteroscedastic (p < 1%). The average variance inflation factor (VIF) was below 2 (also for each individual independent variable). Based on these results the study used an ordinary least square regression (OLS) model to examine the dependent variable.

5.7 Conclusion

This study examined the effect of annual report readability on important future firm characteristics using document file size as an alternative measure for textual complexity. Based on a large sample of Form 10-K filings, the analysis revealed that the document file size of an annual report or its readability is associated with a firm's market performance, market liquidity, and market risk. In particular, the study provided empirical evidence that less readable annual reports are negatively associated with future stock returns after an annual report is filed with the SEC. In addition, less readable Form 10-K filings are negatively associated with firms' market liquidity in the immediate days following annual report submission (lower abnormal trading volume, higher abnormal bid-ask spread). Furthermore, the study confirmed that less readable Form 10-K filings increase the market risk of firms on financial markets (higher abnormal stock return volatility). Finally, it verified that the document file size of an annual report filed with the SEC is predictive for a firm's future corporate earnings (return on assets) and dividend payments (payout ratio). Based on these results, it is concluded that the readability of regulatory filings measured by their document file size affects important firm characteristics.

The present study is subject to a number of limitations. First, despite controlling for common firm fundamentals, other annual report characteristics might affect the empirical results (e.g. disclosure tone, disclosure similarity). Second, despite focusing on short-term capital market effects, other regulatory filing or corporate information might bias the regression results (e.g. current reports, news articles). Third, despite being correlated with common readability indices, other measures or proxies for annual report readability might be more appropriate in exploring the effect of disclosure complexity on financial markets (e.g. Fog Index).

Future research avenues could examine the effect of alternative readability measures on other important firm characteristics not discussed in this study (e.g. revenue growth, cost of capital). In addition, studies could explore the explanatory power of alternative readability measures on corporate bankruptcies. Furthermore, the effect of annual report readability on bond prices (bond yields) is a potential subject for further research. Finally, other regulatory filings and their readability in the context of a firm's market performance, market liquidity, and market risk should be investigated (e.g. quarterly reports on Form 10-Q, current reports on Form 8-K).

Chapter 6

Conclusion

Contents

6.0 Summary of main findings 269

6.0.1 Capital market effect of annual report sentiment 269

6.0.2 Management obfuscation in financial statements 269

6.0.3 Explanatory power of disclosure tone on corporate bankruptcy 271

6.0.4 Capital market effect of annual report readability 272

6.1 Limitations and future research avenues 273

6.0 Summary of main findings

This dissertation contributes to the growing field of textual analyses that analyzes the information content and importance of narrative disclosures in corporate documents. The remainder of this Chapter presents an overview of the empirical findings of the four research studies presented in this dissertation.

6.0.1 Capital market effect of annual report sentiment

Due to the growing amount of narrative information in regulatory filings over time, a large body of the accounting and finance literature investigates the importance and value-relevance of textual information on financial markets. The majority of research in the field of textual analysis focuses solely on traditional stock characteristics important to shareholders (e.g. market performance, accounting profitability) (Kearney and Liu 2014, 172). This research documents that narrative information in corporate reports and tone thereof affects companies' future market performance. However, the effect of narrative disclosure tone on other important firm (stock) characteristics (e.g. market liquidity, market risk) is far less examined, especially considering certain subsections of the annual report on Form 10-K filed with the SEC. To this end, this study extends previous research in the field of textual analysis that associated disclosure tone and a firm's stock market performance as well as other firm variables (e.g. Li 2006; Feldman et al. 2008; Li 2010a; Loughran and McDonald 2011a; Kravet and Muslu 2013; Jegadeesh and Wu 2013; Campbell et al. 2014; Chouliaras 2015; Hope et al. 2016; Henry and Leone 2016; Amel-Zadeh and Faasse 2016; Henselmann and Hering 2017; Gandhi et al. 2017) by exploring the effect of textual sentiment in each Form 10-K subsection on a firm's future market liquidity and market risk.

The main findings can be summarized as follows:

1. Negative textual sentiment in annual reports is positively associated with a firm's subsequent market liquidity (higher market liquidity) and positively associated with its market risk (higher market risk).

2. The regression results showed that certain types of corporate information are associated with a firm's market liquidity and market risk in the expected manner. More litigious and constraining corporate information is significantly positively associated with a firm's abnormal trading volume (higher market liquidity) and abnormal stock return volatility (higher market risk).

3. Contrary to the effect of negative news (cross section tone measurement) on a firm's market liquidity and market risk, a change in negative Form 10-K disclosures (time series tone measurement) is not significantly associated with a firm's market liquidity nor its market risk.

4. The regression results revealed a positive relationship between negative linguistic tone in *"Item 1 - Business Description"* and a firm's market liquidity as well as its market risk following SEC submission.

5. Further analysis showed that different research design choices (sample selection process, measurement of disclosure length, inclusion of table content) in the field of textual analysis can have significant effects on empirical results.

Overall, the study documented the effect of narrative disclosures and their linguistic tone in annual reports on Form 10-K filed with the SEC on the market liquidity and the market risk of exchange listed companies. The study illustrated that market participants react to specific types of corporate information and specific report sections in regulatory filings. Besides the empirical findings, the study provided evidence that corporate reports, and textual sentiment in particular, have information content and affect various important stock characteristics.

6.0.2 Management obfuscation in financial statements

The *"management obfuscation hypothesis"* contends that company officials have incentives to obfuscate negative corporate information in regulatory filings to prevent and/or delay its incorporation into stock prices (Bloomfield 2002, 238; Li 2008, 224; Devos and Sarkar 2015, 5). In this context, a stream of literature investigates the presence and success of management obfuscation (accounting narrative obfuscation) in corporate reports (e.g. Bloomfield 2002; Li 2008; Devos and Sarkar 2015). Based on the idea that managers might use lengthy and complex narrative disclosures to obfuscate adverse corporate information, research provides evidence that regulatory filings and specific sections within corporate reports are more likely to be used as a strategic deterrence to investors than others (footnotes) (Li 2008, 236; Sutton et al. 2009, 3; Bedard et al. 2012, 25-26). However, despite the empirical results of dozens of studies, the findings are mixed regarding whether managers are trying and able to hide negative information (quantitative and qualitative) in corporate reports (e.g. Li 2008; Miller 2010; Goel and Gangolly 2012; Huan et al. 2014; Devos and Sarkar 2015; Amel-Zadeh and Fasse 2016; Henselmann and Hering 2017; Seebeck et al. 2018). An ever-increasing amount of corporate information and technical disclosures in regulatory filings renders the question regarding whether managers can in fact hide negative news from the public and their shareholders still relevant today. Rather than investigating the general idea of management obfuscation in regulatory filings using text characteristics (e.g. length, readability, tone), this study focused on the limitations of previous research and various research design choices potentially affecting empirical results and corresponding inferences.

The main findings can be summarized as follows:

1. Companies cannot hide negative information in financial statements even if corporate profitability is low.

2. Even when faced with low corporate earnings, managers do not use positive statements to frame negative news in corporate reports to mislead investors.

3. Despite their relevance for stock characteristics, the information content of the notes to the financial statements is not predictive for future corporate earnings.

4. The results also show that different research design choices (sample selection process, measurement of disclosure length, inclusion of table content) in the field of textual analysis can have significant effects on research results.

Despite being lengthy and complex, investors incorporate the information content of the notes to the financial statements into a firm's stock market valuation. The results illustrated that company officials cannot hide negative corporate information in the footnotes of Form 10-K filings (management obfuscation by disclosure outlet). In addition, the results indicated that managers do not seem to mislead investors by disclosing overly optimistic textual information in regulatory filings. Overall, the results demonstrated that annual reports on Form 10-K and the footnotes in particular contain value-relevant information for capital market participants.

6.0.3 Explanatory power of disclosure tone on corporate bankruptcy

Given the enormous financial losses associated with failures, the prediction of corporate bankruptcies represents an important research objective in the accounting and finance literature. Over the years, numerous studies have developed bankruptcy prediction models solely based on quantitative variables (financial figures) (e.g. Beaver 1966; Altman 1968; Ohlson 1980). Despite these efforts, researchers are still trying to develop models to adequately predict future bankruptcies (Pusch 2019, 60). Furthermore, only a minority of academic research utilizes the information content and predictive ability of textual information in the context of future firm failures (Henselmann and Scherr 2012, 3; Pusch 2019, 61). To address this research gap, this study used narrative information disclosed in annual reports on Form 10-K filed with the SEC and investigated whether this kind of corporate information can be associated with future corporate bankruptcies.

The main findings can be summarized as follows:

1. Negative annual report information is positively associated with future firm bankruptcies.
2. Positive annual report information is negatively associated with future firm failures.
3. Negative textual disclosures in *'Item 1 - Business Description'* and *'Item 7 - Management's Discussion and Analysis of Financial Condition and Results of Operations'* are associated with the future demise of companies.
4. While the tone level of textual information in corporate reports is associated with corporate bankruptcy, the change in linguistic tone is not.
5. When controlling for firm fundamentals, only constraining corporate information is associated with future firm failures.

Besides providing evidence that regulatory reports and their textual disclosures have important information content for various capital market participants (e.g. shareholders, bondholders), this study illustrated how textual information can be used to better forecast future corporate bankruptcies. Overall, the study showed that textual information can be used alone and alongside financial ratios to better explain future firm failures. In particular, the study illustrated that different types of textual information can be connected to a firm's future economic development in the expected manner, even after controlling for a firm's present financial performance.

6.0.4 Capital market effect of annual report readability

Besides examining the information content of narrative disclosures and its effect on financial markets, researchers have also investigated the effect of textual complexity on firm characteristics. The implications of textual complexity or readability of corporate information using traditional measures for textual complexity (Fog Index) is well examined in the literature (e.g. Li 2008; Miller 2010; Lee 2012; Loughran and McDonald 2014; Boubaker et al. 2018). However, traditional measures for readability are associated with various significant disadvantages (e.g.

human judgement, computational costs, comparability) when analysing the effect of textual complexity on capital markets. Given the limitations of commonly used measures for readability, the question arises as to whether alternative complexity measures can be connected to firm characteristics in the expected manner. To answer this question, this study evaluated the document file size of a Form 10-K filing as an alternative measure for annual report readability and investigated whether it can be connected to important firm variables.

The findings can be summarized as follows:

1. Complex annual reports are negatively associated with a firm's subsequent stock market performance (lower stock return).
2. Complex annual reports are negatively associated with a firm's subsequent market liquidity (lower market liquidity).
3. Complex annual reports are positively associated with a firm's subsequent market risk (higher market risk).
4. Complex annual reports are negatively associated with a firm's future corporate earnings (lower corporate earnings).
5. Complex annual reports are negatively associated with a firm's future dividend payment behaviour (lower payout ratio).
6. Research design choices in the field of textual analysis (sample selection process, submission size measurement) can have significant effects on empirical results.

Overall, the study illustrated the effect of textual complexity in corporate reports on financial markets. In particular, the study documented the negative effect of hard to read corporate information and the future consequences for a firm's shareholders. Besides the empirical findings, the study also showed that annual reports on Form 10-K filed with the SEC and their features have an effect on financial markets.

6.1 Limitations and future research avenues

This dissertation contains four distinct research papers, each highlighting the importance and relevance of narrative disclosures in financial statements for various firm stakeholders and regulatory authorities. The findings provide empirical evidence that textual information in regulatory filings has an effect on various important firm characteristics (e.g. market liquidity, market risk, market performance) and can be used to better forecast a firm's future (e.g.

corporate failure, corporate earnings, dividend payments). However, the field of textual analysis and therefore the four research papers forming this dissertation are associated with different disadvantages. In this context, numerous limitations need to be considered when interpreting the results presented here. Despite being tested for validity, the information extraction algorithms used to extract textual information might contain errors. Thus, the empirical findings regarding the importance and effects of narrative information disclosed in financial statements could be biased. In addition, the loss of context when analysing narrative information using quantitative measures as a general drawback associated with textual analysis might impact the presented findings. Furthermore, the effects of other relevant corporate information disclosed or published in various information sources such as newspapers, interviews, content on the Internet, and other corporate reports might impact the empirical results. Finally, despite the statistical significance of the results presented in this dissertation, the economic significance of the empirical findings in real-world applications such as trading strategies might be limited.

Future research avenues in the field of textual analysis (e.g. sentiment, readability, similarity) have various opportunities not only to find answers to important research questions, but also to progress the methodologies used to find these answers. For instance, researchers should disclose, evaluate, and improve any information extraction method they use to allow their results to be validated and replicated, thus adding knowledge to the overall field of research. Instead of relying on commonly used word lists and weighting schemes, researchers should try to develop field-specific dictionaries and improved or advanced term-weighting schemes (Kearney and Liu 2014, 181; Loughran and McDonald 2015, 9-10; Loughran and McDonald 2016, 1223-1224). In this context, textual analysis and the related literature may also adopt methods of machine learning (ML) and artificial intelligence (AI) to extract in-depth meaning and context from various narrative disclosures (Loughran and McDonald 2016; 1223-1224) not only in English, but also in other languages (e.g. German; Chinese) (Loughran and McDonald 2016; 1225; Kearney and Liu 2014, 184). In addition, the importance of narrative disclosures in other corporate reports (e.g. quarterly reports on Form 10-Q) and other information sources (e.g. social media, news articles, board postings) (Loughran and McDonald 2016; 1223-1224) might be subject to further empirical investigation. Specifically, future research could focus on the effects and implications of narrative corporate information on important firm characteristics not discussed in this dissertation (e.g. cost of capital, credit ratings) (Li 2010a, 156). In addition, the effects of textual information on markets other than the stock market and its corresponding variables should be examined in the future (e.g. bond market, commodity market) (Kearney and Liu 2014, 184). Furthermore, the implications of textual information in regulatory filings on financial instruments other than stocks could also be explored (e.g. derivatives). In sum, as Loughran and McDonald (2016) note *“with increasing computational power and an explosion of digital text available for research, there is much yet to be done.”* (Loughran and McDonald 2016, 1226).

References

Alexander, S. S. 1961. Price movements in speculative markets: Trends or random walks. *Industrial Management Review* 2 (2): 7-26.

Altman, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23 (4): 589-609.

Altman, E. I., R. G. Haldeman, and P. Narayanan, 1977. ZETA analysis: A new model to identify bankruptcy risk of corporation. Journal of Banking and Finance 1 (1): 29-54.

Amel-Zadeh, A., and J. Faasse. 2016. The information content of 10-K narratives: Comparing MD&A and footnote disclosures. Working paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2807546.

Ariel, R. A. 1990. High stock returns before holidays: Existence and evidence on possible causes. *The Journal of Finance* 45 (5): 1611-1626.

Atiase, R. K., and L. S. Bamber. 1993. Trading volume reactions to annual accounting earnings announcements. *Journal of accounting and Economics* 17, 309-329.

Bailey, W., H. Li, C. X. Mao, and R. Zhong. 2003. Regulation fair disclosure and earnings information: Market, analyst, and corporate responses. *The Journal of Finance* 58 (6): 2487-2514.

Ball, R., and P. Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6 (2): 159-178.

Ball, R., and S. P. Kothari. 1989. Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns. *Journal of Financial Economics* 25, 51-74.

Balvers, R. J., T. F. Cosimano, and B. McDonald. 1990. Predicting stock returns in an efficient market. *The Journal of Finance* 45 (4): 1109-1128.

Bamber, L. S. 1987. Unexpected earnings, firm size, and trading volume around quarterly earnings announcements. *The Accounting Review* 62 (3): 510-532.

Bamber, L. S., O. E. Barron, and T. L. Stober. 1997. Trading volume and different aspects of disagreement coincident with earnings announcements. *The Accounting Review* 72, 4, 575-597.

Bannier, C., T. Pauls, and A. Walter. 2017. Content analysis of business-specific text document: Introducing a German dictionary. Working Paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2961820.

Bauer, G. 2004. A Taxonomy of Market Efficiency. Financial System Review articles. Available online on URL: https://www.bankofcanada.ca/wp-content/uploads/2012/01/fsr-1204-bauer.pdf.

Beaver, W. H. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research* 4: 71-111.

Bedard, J. C., S. G. Sutton, V. Arnold, and J. R. Phillips. 2012. Another piece of the "expectations gap": What do investors know about auditor involvement with information in the annual report? *Current Issues in Auditing* 6 (1): 17-30.

Bellovary, J. L., D. E. Giacomino, and M. D. Akers. 2007. A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education* 33: 1-42.

Blankespoor, E., B. P. Miller, and H. D. White. 2014. Initial evidence on the market impact of the XBRL mandate. *Review of Accounting Studies* 19 (4): 1469-1503.

Bloomfield, R. J. 2002. The ''Incomplete Revelation Hypothesis'' and financial reporting. *Accounting Horizons* 16 (3): 233–243.

Bloomfield, R. J. 2008. Discussion of "Annual report readability, current earnings, and earnings persistence". *Journal of Accounting and Economics* (45) (2-3): 248-252.

Bodnaruk, A., T. Loughran, and B. McDonald. 2015. Using 10-K text to gauge financial constraints. *Journal of Financial and Quantitative Analysis* 50 (4): 623-646.

Bonsall, S. B., A. J. Leone, and B. P. Miller. 2017. A plain English measure of financial reporting readability. *Journal of Accounting and Economics* 63 (2-3): 329-357.

Boritz, J. E. 1991. Going…going…gone! *CA Magazine* 124 (4): 22-29.

Boubaker, S., D. Gounopoulos, and H. Rjiba. 2018. Annual report readability and stock liquidity. Working paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3141045.

Bouchaud, J. P., J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand, in Hens, T. and Schenk-Hoppe, K. R. 2009. Handbook of financial markets: Dynamics and evolution. Amsterdam, Holland.

Breen, W., L. R. Glosten, and R. Jagannathan. 1989. Economic significance of predictable variations in stock index returns. *The Journal of Finance* 44 (5): 1177-1189.

Bushee, B. J., J. E. Core, W. Guay, and S. J. W. Hamm. 2010. The role of the business press as an information intermediary. *Journal of Accounting Research* 48 (1): 1-19.

Bushman, R., and A. Smith. 2001. Financial accounting research and corporate governance. *Journal of Accounting & Economics* 31: 237-333.

Campbell, J. L., H. Chen, D. S. Dhaliwal, H. Lu, and L. B. Steele. 2014. The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies* 19 (1): 396-455.

Campbell, J. Y. 1987. Stock returns and the term structure. *Journal of Financial Economics* 18 (2): 373-399.

Campbell, J. Y., and R. J. Shiller. 1988. Valuation ratios and the long-run stock market outlook. *Journal of Portfolio Management* 24 (2): 11-26.

Cao, Z., and G. S. Narayanamoorthy. 2014. Accounting and litigation risk: Evidence from directors' and officers' insurance pricing. *Review of Accounting Studies* 19 (1): 1-42.

Cazier, R. A., and R. J. Pfeiffer. 2016. Why are 10-K filings so long? *Accounting Horizons* 30 (1): 1-21.

Cecchini, M., H. Aytug, G. J. Koehler, and P. Pathak. 2010. Making words work: Using financial text as a predictor of financial events. *Decision Support Systems* 50: 164-175.

Chava, S., and R. A. Jarrow. 2004. Bankruptcy prediction with industry effects. *Review of Finance* 8: 537-569.

Chouliaras, A. 2015. The Pessimism Factor: SEC EDGAR Form 10-K textual analysis and stock returns. Working Paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2627037.

Clatworthy, M., and M. J. Jones. 2001. The effect of the thematic structure on the variability of annual report readability. *Accounting, Auditing & Accountability Journal* 14 (3): 311-326.

Claus, J., and J. Thomas. 2001. Equity premia as low as three percent? Evidence from Analysts' earnings forecasts for domestic and international stock markets. *The Journal of Finance* 56 (5): 1629-1666.

Core, J. 2001. A review of the empirical disclosure literature: Discussion. *Journal of Accounting & Economics* 31: 441-456.

Correia, M., S. Richardson, and I. Tuna. 2012. Value investing in credit markets. *Review of Accounting Studies* 17 (3): 572-609.

Czerney, K., L. L. Lisic., B. Wu, and I. Zhang. 2017. Big 4 auditors, litigation risk, and disclosure tone. Working Paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3351230.

Czyzewski, A. B., and K. Wilkinson. 2014. Does the management discussion and analysis section of the 10-K provide information for retirement decisions? *The Journal of Leadership, Accountability and Ethics* 11 (3): 97-107.

Das, S. R. 2014. Text and context: Language analytics in finance. Boston, United States of America.

Devos, E., and S. Sarkar. 2015. Read the footnotes! Auditor quality, earnings persistence, and the number of footnotes in 10Ks. Working Paper.

Diamond, D., and R. Verrecchia. 1991. Disclosure, liquidity, and the cost of capital. *The Journal of Finance* 66, 1325-1355.

Ditter, D. 2015. The benefits of standardized internet financial reporting. Nuremberg, Germany.

Easley, D., and M. O'Hara. 2004. Information and the cost of capital. *The Journal of Finance* 59: 1553- 1583.

Easley, D., S. Hvidkjaer, and M. O'Hara. 2002. Is information risk a determinant of asset returns? *The Journal of Finance* 57: 2185-2222.

Easton, P. D. 2004. PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. The Accounting Review 79 (1): 73-95.

Elrod, G. B. 2009. Is there predictive value in the words managers use? A key word analysis of the annual report's management discussion and analysis. Texas, United States of America.

Engelberg, J., and S. Sankaraguruswamy. 2007. How to gather data using a web crawler: An application using SAS to search EDGDAR. Working Paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1015021.

Ertugrul, M., J. Lei, J. Qiu, and C. Wan. 2017. Annual report readability, tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis* 52 (2): 811-836.

Fama, E. F. 1965. The behavior of stock-market prices. *Journal of Business* 38 (1): 34-105.

Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25 (2): 383-417.

Fama, E. F. 1991. Efficient capital markets: II. *The Journal of Finance* 46 (5) : 1575-1617.

Fama, E. F., and K. R. French. 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96 (2): 246-273.

Fama, E. F., and K. R. French. 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25 (1): 23-49.

Fama, E. F., and K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.

Fama, E. F., L. Fisher, M. Jensen, and R. Roll. 1969. The adjustment of stock prices to new information. *International Economic Review* 10 (1): 1-21.

Feldman, R., J. Livnat, and B. Segal. 2008. Shorting companies that restate previously issued financial statements. *Journal of Investing* 17 (3): 6-15.

Feldman, R., S. Govindaraj, J. Livnat, and B. Segal. 2008. The incremental information content of tone change in management discussion and analysis. Working Paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1126962.

Ferson, W. E., and C. R. Harvey. 1993. The risk and predictability of international equity returns. *Review of Financial Studies* 6 (3): 527-566.

Filer Manual. 2017. Filer manual - Volume II EDGAR filing. Available online on URL: https://www.sec.gov/info/edgar/edgarfm-vol2-v43.pdf.

French, K. R. 1980. Stock returns and the weekend effect. *Journal of Financial Economics* 8 (1): 55-69.

Frydman, H., E. I. Altman, and D. Kao. 1985. Introducing recursive partitioning for financial classification: The case of financial distress. *The Journal of Finance* 40 (1): 269-291.

Gan, Q., and B. Qiu. 2018. The information content of 10-K file size change. Working paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3004403.

Gandhi, P., T. Loughran, and B. McDonald. 2017. Using annual report sentiment as a proxy for financial distress in U.S. banks. Working paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2905225.

Gandhi, P., T. Loughran, and B. McDonald. 2019. Using annual report sentiment as a proxy for financial distress in U.S. banks. *Journal of Behavioral Finance* 20 (4): 424-436.

Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan. 2001. Toward an implied cost of capital. *Journal of Accounting Research* 39 (1): 135-176.

Gepp, A., K. Kumar, and S. Bhattacharya. 2010. Business failure prediction using decision trees. *Journal of Forecasting* 29: 536-555.

Gerdes Jr, J. 2003. EDGAR-Analyzer: Automating the analysis of corporate data contained in the SEC's EDGAR database. *Decision Support Systems* 35 (1): 7-29.

Godbillon-Camus, B., and C. Godlewski. 2005. Credit risk management in banks: Hard information, soft information and manipulation. Working Paper. Available online on URL: file:///C:/Users/J%C3%B6rg%20Hering/Desktop/SSRN-id882027.pdf.

Gode, D. K., and S. Sunder. 1997. What makes markets allocationally efficient? *The Quarterly Journal of Economics* 112 (2): 603-630.

Goel, S., and J. Gangolly. 2012. Beyond the numbers: Mining the annual reports for hidden cues indicative of financial statement fraud. *Intelligent Systems in Accounting, Finance and Management* 19 (2): 75-89.

Goel, S., S. R. Faerman, and O. Uzuner. 2010. Can linguistic predictors detect fraudulent financial filings? *Journal of Emerging Technologies in Accounting* 7 (1): 25-46.

Grant, G. H., and S. J. Conlon. 2006. EDGAR extraction system: An automated approach to analyze employee stock option disclosures. *Journal of Information Systems* 20 (2): 119-142.

Grossman, S. J, and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *The American Economic Review* 70 (3): 393–408.

Guo, L., F. Shi, and J. Tu. 2016. Textual analysis and machine learning: Crack unstructured data in finance and accounting. *The Journal of Data Science* 2016 (2): 153-170.

Haugen, R. A. and J. Lakonishok. 1988. The incredible january effect. Illinois, United States of America.

Healy, P., and K. Palepu. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical literature. *Journal of Accounting & Economics* 31, 405-441.

Heidari, M., and C. Felden. 2015. Financial footnote analysis: Developing a text mining approach. *Proceedings of the 11th International Conference on Data Mining (DMIN 2015)*. Las Vegas, United States of America. Available online on URL: https://pdfs.semanticscholar.org/4dc9/c82621a6aac8388afe84f264b3adbc87e89e.pdf.

Henry, E., and A. J. Leone. 2016. Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review* 91 (1): 153-178.

Henselmann, K. 2016. Abschlussanalyse 4.0. in: Informationstechnologien als Wegbereiter für den steuerberatenden Berufsstand, edited by C. Bär., A. Fischer, and H. Gulden. Berlin, Germany, Springer-Verlag.

Henselmann, K., and E. Scherr. 2012. Content analysis of XBRL filings as an efficient supplement of bankruptcy prediction? Empirical evidence based on US GAAP annual reports. Working Paper. Available online on URL: http://hdl.handle.net/10419/58246.

Henselmann, K., and E. Scherr. 2013. Content analysis of XBRL filings as an efficient supplement of bankruptcy prediction? Empirical evidence based on US GAAP annual reports, in Scherr, E. 2013. Improving financial reporting using XBRL. Nuremberg, Germany.

Henselmann, K., and J. Hering. 2017. Section sentiment: Form 10-K textual analysis and future stock returns, in: Hering, J. 2018. Textual analysis in accounting and finance. Nuremberg, Germany.

Hering, J. 2016. The annual report algorithm: Retrieval of financial statements and extraction of textual information, in: Hering, J. 2018. Textual analysis in accounting and finance. Nuremberg, Germany.

Hering, J. 2018. Textual analysis in accounting and finance. Nuremberg, Germany.

Hong, H., and J. C. Stein. 1999. A unified theory of underreactions, momentum trading and overreaction in asset markets. *The Journal of Finance* 54 (6): 2143-2184.

Hope, O., D. Hu, and H. Lu. 2016. The benefits of specific risk-factor disclosures. *Review of Accounting Studies* 21 (1): 1005-1045.

Huang, T., S. H. Teoh, and Y. Zhang. 2014. Tone management. *The Accounting Review* 89 (3): 1083-1113.

Humpherys, S. L., K. C. Moffitt, M. B. Burns, J. K. Burgoon, and W. F. Felix. 2011. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems* 50 (3): 585-594.

Jegadeesh, N., and D. Wu. 2013. Word power: A new approach for content analysis. *Journal of Financial Economics* 110 (3): 712-729.

Jiang, W., K. Li, and W. Wang. 2012. Hedge funds and chapter 11. *The Journal of Finance* 67 (2): 513-560.

Kahneman, D., and A. Tversky. 1973. On the psychology of prediction. *Psychological Review* 80 (4):237-251.

Kandel, S., and R. F. Stambaugh. 1996. On the predictability of stock returns: An asset-allocation perspective. *The Journal of Finance* 51 (2): 385-424.

Kearney, C., and S. Liu. 2014. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis* 33 (1): 171-185.

Keim, D. B. 1983. Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics* 12 (1): 13–32.

Kendall, M. G. 1953. The analysis of economic time-series, Part I: Prices. *Journal of the Royal Statistical Society* 116 (1): 11-34.

Koch, P. D. 2011. Investment theory and applications. Working Paper. Available online on URL: https://finpko.ku.edu/myssi/FIN410/inv.04.EG17.pdf.

Kothari, S. P., X. Li, and J. E. Short. 2009. The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review* 84 (5): 1639-1670.

Kravet, T., and V. Muslu. 2013. Textual risk disclosures and investor's risk perceptions. *Review of Accounting Studies* 18 (4): 1088-1122.

Kumar, P. R., and V. Ravi. 2007. Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research* 180: 1-28.

Lakonishok, J., and S. Smidt. 1988. Are seasonal anomalies real? A ninety-year perspective. *Review of Financial Studies* 1 (4): 403-425.

Lambert, R., C. Leuz, and R. Verrecchia. 2007. Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research* 45: 385-420.

Lawrence, A. 2013. Individual investors and financial disclosure. *Journal of Accounting and Economics* 56 (1): 130-147.

Leder, M. 2003. Financial fine print: Undercovering a company's true value. John Wiley & Sons, Hoboken, New Jersey. United States of America.

Lee, Y. 2012. The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research* 29 (4): 1137-1170.

Lehavy, R., F. Li, and K. J. Merkley. 2011. The effect of annual report readability on analyst following and the properties of their earnings forecast. *The Accounting Review* 86 (3): 1087-1115.

Lehmann, B. 2003. Is it worth the while? The relevance of qualitative information in credit rating. Working Paper: Available online on URL: file:///C:/Users/J%C3%B6rg%20He-ring/Desktop/SSRN-id410186.pdf.

Li, F. 2006. Do stock market investors understand the risk sentiment of corporate annual reports? Working Paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=898181.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2/3): 221-247.

Li, F. 2010a. Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature* 29 (1): 143-165.

Li, F. 2010b. The information content of forward-looking statements in corporate filings - A naïve bayesian machine learning approach. *The Journal of Accounting Research* 48 (5): 1049-1102.

Lo, A. W. 2007. Efficient markets hypothesis. Working paper. Available online on URL: https://alo.mit.edu/wp-content/uploads/2015/08/emh2007.pdf.

Lo, A. W., and C. A. MacKinlay. 1999. A non-random walk down wall street. Princeton University Press. United States of America.

Lo, A. W., H. Mamaysky, and J. Wang. 2000. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The Journal of Finance* 55 (4): 1705–1765.

Lo, K., F. Ramos, and R. Rogo. 2017. Earnings management and annual report readability. *Journal of Accounting and Economics* 63 (1): 1-25.

Lopatta, K., M. A. Gloger, and R. Jaeschke. 2017. Can language predict bankruptcy? The explanatory power of tone in 10-K filings. *Accounting Perspectives* 16 (4): 315-343.

Loughran, T., and B. McDonald. 2009. Plain English, readability, and 10-K filings. Working Paper. Available online on URL: https://www3.nd.edu/~tloughra/Plain_English.pdf.

Loughran, T., and B. McDonald. 2011a. When is a liability not a liability? Textual analysis, dictionaries, and 10Ks. *The Journal of Finance* 66 (1): 35-65.

Loughran, T., and B. McDonald. 2011b. Internet appendix for "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks". Available online on URL: http://www.afajof.org/SpringboardWebApp/userfiles/afa/file/Supplements%20and%20Data%20Sets/Internet%20Appendix%20for%20When%20Is%20a%20Liability%20Not%20a%20Liability%20Textual%20Analysis,%20Dictionaries,%20and%2010-Ks%206989-IA-Feb-2011.pdf.

Loughran, T., and B. McDonald. 2011c. Barron's red flags: Do they actually work? *Journal of Behavioral Finance* 12 (2): 90-97.

Loughran, T., and B. McDonald. 2014. Measuring readability in financial disclosures. *The Journal of Finance* 69 (4): 1643-1671.

Loughran, T., and B. McDonald. 2015. The use of word lists in textual analysis. *The Journal of Behavioral Finance* 16 (1): 1-11.

Loughran, T., and B. McDonald. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54 (4): 1187-1230.

Magnusson, C., A. Arppe, T. Eklund, B. Back, H. Vanharanta, and A. Visa. 2005. The language of quarterly reports as an indicator of change in the company's financial status. *Information & Management* 42: 561-573.

Malkiel, B. G. 1992. Efficient market hypothesis, in Eatwell, P. and Eatwell J. 1922. New palgrave dictionary of money and finance. London, United Kingdom.

Malkiel, B. G. 2003. The efficient market hypothesis and its critics. *Journal of Economic Perspectives* 17 (1): 59-82.

Mandelbrot, M. 1966. Forecasts of future prices, unbiased markets, and martingale models. *Journal of Business* 39 (1): 242-255.

Mayew, W. J., M. Sethuraman, and M. Venkatachalam. 2015. MD&A disclosure and the firm's ability to continue as a going concern. *The Accounting Review* 90 (4): 1621-1651.

Merkl-Davies, D. M., and N. M. Brennan. 2007. Discretionary disclosure strategies in corporate narratives: Incremental information or impression management? *Journal of Accounting Literature* 27 (1): 116-196.

Milgrom, P., and N. Stokey. 1982. Information, trade and common knowledge. *Journal of Economic Theory* 26 (1): 17-27.

Miller, B. P. 2010. The effects of reporting complexity on small and large investor trading. *The Accounting Review* 85 (6): 2107-2143.

Miwa, K. 2020. Levels versus changes: Information contents of qualitative verbal information. Working Paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3481618.

Morse, Dale C. Price and trading volume reaction surrounding earnings announcements: A closer examination. *Journal of Accounting Research* 19 (2): 374-383.

Mullainathan, S., and A. Shleifer. 2005. Persuasion in finance. Working paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=864686.

Niederhoffer, V., and M. F. M. Osborne. 1966. Market making and reversal on the stock exchange. *Journal of the American Statistical Association* 61 (316): 879-916.

Nisbett, R., and L. Ross. 1980. Human inference: Strategies and shortcomings of social judgment. New Jersey. United States of America.

Odom, M. D., and R. Sharda. 1990. A neural network model for bankruptcy prediction. *Proceedings of the IJCNN International Joint Conference on neural networks*: 163-168.

Ohlson, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18 (1): 109-131.

Osborne, M. F. M. 1959. Brownian motion in the stock market. *Operations Research* 7 (2): 145-273.

Pendharkar, P. C., and J. A. Rodger. 2004. An empirical study of impact of crossover operators on the performance of non-binary genetic algorithm based neural approaches for classification. *Computers & Operations Research* 31: 481-498.

Pesaran, M. H. 2005. Market efficiency today. Working paper. Available online on URL: https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.researchgate.net/profile/Hashem_Pesaran/publication/24129669_Market_Efficiency_Today/links/09e4150ed4fdadf35d000000/Market-Efficiency-Today.pdf&ved=2ahUKEwj_lcjNyaDvAhVL4eA-KHfqhDCcQFjABegQIBRAD&usg=AOvVaw3FsS3Ss2wWZGtkC_OkkHpW.

Pesaran, M. H., and A. Timmermann. 1994. Forecasting stock returns: An examination of stock market trading in the presence of transaction costs. *Journal of Forecasting* 13 (4): 335-367.

Poterba, J. M., and L. Summers. 1988. Mean reversion in stock returns: Evidence and implications. *Journal of Financial Economics* 22 (1): 27-59.

Purda, L., and D. Skillicorn. 2015. Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. Contemporary Accounting Research 32 (3): 1193-1223.

Pusch, C. 2019. The current reporting behavior of firms facing bankruptcy: Empirical evidence from the U.S. market, in: Pusch, C. 2019. The impact of firm events on capital markets. Nuremberg, Germany.

Rennekamp, K. 2012. Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research* 50 (1): 1319-1354.

Roberts, H. V. 1959. Stock market patterns and financial analysis: Methodological suggestions. *The Journal of Finance* 14 (1): 1-10.

Rogers, J. L., A. V. Buskirk, and S. L. C. Zechman. 2011. Disclosure tone and shareholder litigation. *The Accounting Review* 86 (6): 2155-2183.

Rogers, R. K., and J. Grant. 1997. Content analysis of information cited in reports of sell-side financial analysts. *The Journal of Financial Statement Analysis* 3 (1): 17-30.

Roll, R. and R. J. Shiller. 1992. Volatility in U.S. and Japanese stock markets: Selections from the first annual symposium on global financial markets. *Journal of Applied Corporate Finance* 5 (1): 4-35.

Samuelson, P. A. 1965. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review* 6 (2): 41-49.

Scholes, M. 1969. A test of the competitive hypothesis: The market for new issues and secondary offerings. Chicago, United States of America.

SEC Regulation S-K. 2018. Standard instructions for filing forms under Securities Act of 1933, Securities Exchange Act of 1934 and Energy Policy and Conservation Act of 1975-Regulation S-K. Available online on URL: http://www.ecfr.gov/cgi-bin/textidx?SID=8e0ed509ccc65e983f9eca72ceb26753&node=17:3.0.1.1.11&rgn=div5#se17.3.229_1101.

SEC Release 33-8350. 2003. Interpretation: Commission guidance regarding management's discussion and analysis of financial condition and results of operations. Available online on URL: https://www.sec.gov/rules/interp/33-8350.htm.

Securities and Exchange Commission (SEC). 1934. Securities Exchange Act of 1934. Available online on URL: https://www.sec.gov/about/laws/sea34.pdf.

Securities and Exchange Commission (SEC). 2003a. Improving corporate disclosure – Improving shareholder value, Available online on URL: https://www.sec.gov/news/speech/spch041003cag.htm.

Securities and Exchange Commission (SEC). 2003b. SEC releases guidance regarding MD&A. Available online on URL: https://www.sec.gov/news/press/2003-179.htm.

Securities and Exchange Commission (SEC). 2007. Beginners' guide to financial statements. Available online on URL: https://www.sec.gov/reportspubs/investor-publications/investorpubsbegfinstmtguidehtm.html.

Securities and Exchange Commission (SEC). 2007a. Feedback from individual investors on disclosure. Available online on URL: https://www.sec.gov/news/speech/2007/spch011907ljs.htm.

Securities and Exchange Commission (SEC). 2007b. Closing remarks to the second annual corporate governance summit. Available online on URL: https://www.sec.gov/news/speech/2007/spch032307cc.htm.

Securities and Exchange Commission (SEC). 2009. Form 10-K - Fast answers. Available online on URL: https://www.sec.gov/fast-answers/answers-form10khtm.html.

Securities and Exchange Commission (SEC). 2013. What we do. Available online on URL: https://www.sec.gov/about/whatwedo.shtml.

Seebeck, A., J. Rudolph, and K. Henselmann. 2018. Firms' strategic use of XBRL extensions - Empirical evidence from a sentiment analysis of XBRL tag names, in: Seebeck, A. 2018. The digitalization of financial reporting - How changes in technology are affecting the dissemination, presentation, and analysis of disclosure. Nuremberg, Germany.

Shin, K. S., T. S. Lee, and H. J. Kim. 2005. An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications* 28 (1): 127-135.

Shirata, C. Y., and M. Sakagami. 2008. An analysis of the "going concern assumption": Text mining from Japanese financial reports. *Journal of Emerging Technologies in Accounting* 8: 1-16.

Shirata, C. Y., H. Takeuchi, and H. Watanabe. 2011. Extracting key phrases as predictors of corporate bankruptcy: Empirical analysis of annual reports by text mining. *Journal of Emerging Technologies in Accounting* 8: 31-44.

Shumway, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business* 74 (1): 101-124.

Skinner, D. 1994. Why firms voluntarily disclose bad news. *Journal of Accounting Research* 32, 38-60.

Sutton, S., V. Arnold., J. C. Bedard, and J. Philips. 2009. Where do investors prefer to find nonfinancial information? Working paper. Available online on URL: https://www.journalofaccountancy.com/news/2010/aug/20102682.html.

Tavcar, L. R. 1998. Make the MD&A more readable. *The CPA Journal* 68 (1): 10.

Tennyson, M. B., R. W. Ingram, and M. T. Dugan. 1990. Assessing the information content of narrative disclosures in explaining bankruptcy. *Journal of Business Finance & Accounting* 17 (3): 391-409.

Tetlock, P. C. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62 (3): 1139-1168.

Tetlock, P. C., M. Saar-Tsechansky, S. Macskassy. 2008. More than words: Quantifying language to measure firm's fundamentals. *The Journal of Finance* 63 (3): 1437-1467.

Twedt, B., and L. Rees. 2011. Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports. Working paper. Available online on URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1764604.

Uhr, P., J. Zenkert, and M. Fathi. 2014. Sentiment analysis in financial markets: A framework to utilize the human ability of word association for analyzing stock market news reports.

Working paper. Available online on URL: https://www.eti.uni-siegen.de/ws/publikationen/repository/sentiment_analysis_in_financial_markets_-_a_framework_to_utilize_the_human_ability_of_word_association_for_analyzing_stock_market_news_reports.pdf.

Verrecchia, R. 2001. Essays on disclosure. *Journal of Accounting and Economics* 32: 97-180.

Waud, R. N. 1970. Public interpretation of federal reserve discount rate changes: Evidence on the announcement effect. *Econometrica* 38 (2): 231-250.

Westgaard, S., and N. Wijst. 2001. Default probabilities in a corporate bank portfolio: A logistic model approach. *European Journal of Operational Research* 135: 338-349.

Wilson, R. L., and R. Sharda. 1994. Bankruptcy prediction using neural networks. *Decision Support Systems* 11: 545-557.

Working, H. 1934. A random difference series for use in the analysis of time series. *Journal of the American Statistical Association* 29 (185): 11-24.

Yang, F., B. Dolar, and L. Mo. 2018. Textual analysis of corporate annual disclosures: A comparison between bankrupt and non-bankrupt companies. *Journal of Emerging Technologies in Accounting* 15 (1): 45-55.

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

Zhang, G., M. Y. Hu, E. B. Patuwo, and D. C. Indro. 1999. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research* 116: 16-32.

Zmijewski, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* 22: 59-82.

Zolotoy, L., and P. S. Kalev. Cash-flow news, discount-rate news and the co-skewness risk. Working Paper. Available online on URL: https://acfr.aut.ac.nz/__data/assets/pdf_file/0015/30057/AFM_2012_155.pdf.

General appendix

The general appendix shows where certain chapters of this thesis were presented or were accepted for presentation.

Chapter 2: The effect of disclosure tone on market liquidity and market risk

Accepted for presentation at:

- 42nd European Accounting Association Annual Congress 2019, Paphos, Cyprus
- 51st British Accounting & Finance Association Annual Conference 2019, Birmingham, United Kingdom
- 52nd British Accounting & Finance Association Annual Conference 2020, Southampton, United Kingdom

Chapter 3: Accounting narrative obfuscation in financial statements

Accepted for presentation at:

- 1st Corporate Finance and Asset Pricing Conference 2019, Manchester, United Kingdom
- 6th International Conference on Data Mining and Applications 2020, Zurich, Switzerland

Chapter 4: Form 10-K textual analysis and corporate bankruptcy

Accepted for presentation at:

- 54th British Accounting & Finance Association Annual Conference 2022, Nottingham, United Kingdom

Chapter 5: The effect of annual report readability on financial markets

Presented at:

- 53rd British Accounting & Finance Association Annual Conference 2021, Virtual Conference

Accepted for presentation at:

- 44th European Accounting Association Annual Congress 2022, Bergen, Norway

Appendix for Chapter 2

Table A.1. Number of annual reports filed with the SEC EDGAR system.

Year	Month 1	2	3	4	5	6	7	8	9	10	11	12	Filings (Num.)
2016	169	2,110	3,183	558	203	244	143	198	233	138	194	216	7,589
2015	182	1,777	3,712	635	217	258	167	201	251	137	226	222	7,985
2014	193	1,787	3,573	732	223	264	165	204	287	187	223	246	8,084
2013	195	1,653	3,079	1,247	243	261	227	224	293	181	226	276	8,105
2012	197	1,872	3,508	799	241	292	193	214	300	223	245	309	8,393
2011	210	1,575	4,026	848	254	311	183	241	382	212	259	339	8,840
2010	254	1,476	4,206	956	264	319	202	223	383	246	279	357	9,165
2009	230	1,370	4,734	998	289	373	276	257	366	276	319	351	9,839
2008	103	1,670	4,557	509	192	264	162	176	325	205	185	398	8,746
2007	102	1,003	4,629	1,302	184	269	111	147	261	116	167	283	8,574
2006	97	509	6,123	438	133	345	98	259	317	105	121	307	8,852
2005	127	348	6,322	537	105	287	130	194	356	99	180	332	9,017
2004	176	425	5,924	460	123	313	128	112	343	111	103	349	8,567
2003	168	306	5,643	674	193	314	124	122	350	129	85	360	8,468
2002	259	286	4,412	1,987	234	283	186	134	381	157	239	369	8,927
2001	199	274	4,423	2,305	260	352	176	142	423	165	128	401	9,248
2000	247	299	6,292	761	240	421	180	143	508	189	133	456	9,869
1999	253	248	6,308	1,019	167	437	180	150	524	189	138	509	10,122
1998	251	270	6,398	824	298	430	240	147	554	205	150	520	10,287
1997	246	304	6,015	717	306	435	187	186	574	225	137	567	9,899
1996	157	188	2,484	1,126	179	244	255	161	543	197	167	558	6,259
1995	61	102	1,808	170	77	130	79	79	279	113	79	259	3,236
1994	29	66	1,231	110	58	46	31	36	110	41	56	109	1,923
1993	-	-	-	-	-	-	-	-	-	-	1	3	4
Filings (Num.)	4,105	19,918	102,590	19,712	4,683	6,892	3,823	3,950	8,343	3,846	4,040	8,096	189,998

Notes: The table shows the number of annual reports on Form 10-K and Form 10-K405 filed in each year since 1993.

Table A.2. Variables definitions.

Variable name	Expression
$\mathrm{ATV}_{\mathrm{Ave}}$	The abnormal trading volume: calculated as the cumulative daily total proportion of outstanding shares traded during the event period (trading days -1 to +3), adjusted for the firm-specific average level of outstanding shares traded during the non-event period (trading days -6 to -156).
$\mathrm{ATV}_{\mathrm{Med}}$	The abnormal trading volume: calculated as the cumulative daily total proportion of outstanding shares traded during the event period (trading days -1 to +3), adjusted for the firm-specific median level of outstanding shares traded during the non-event period (trading days -6 to -156).
$\mathrm{ASpread}_{\mathrm{Ave}}$	The abnormal bid-ask spread; calculated as the cumulative daily total proportion of outstanding shares traded during the event period (trading days -1 to +3), adjusted for the average daily bid-ask spread during the non-event period (trading days -6 to -156).
$\mathrm{ASpread}_{\mathrm{Med}}$	The abnormal bid-ask spread; calculated as the cumulative daily total proportion of outstanding shares traded during the event period (trading days -1 to +3), adjusted for the median daily bid-ask spread during the non-event period (trading days -6 to -156).
$\mathrm{ARV}_{\mathrm{MAM}}$	The abnormal stock return volatility; calculated as the sum of the absolute daily abnormal stock returns during the event period (trading days -1 to +3); the abnormal stock returns are computed using the market-adjusted model (MAM), where for the market return, the S&P 500 stock index is subtracted from a firm's stock return (trading days -1 to +3).
$\mathrm{ARV}_{\mathrm{MM}}$	The abnormal stock return volatility; calculated as the sum of the absolute daily abnormal stock returns during the event period (trading days -1 to +3); the abnormal stock returns are computed using the market model (MM), where for the market return, the S&P 500 stock index is included in the regression (trading days -21 to -250).
$\mathrm{ARV}_{\mathrm{4\ Factor}}$	The abnormal stock return volatility; calculated as the sum of the absolute daily abnormal stock returns during the event period (trading days -1 to +3); the abnormal stock returns are computed using the 4-factor model (4 Factor), were factor loadings are estimated by including the excess market returns, a size factor, a value factor, and a momentum factor in the regression (trading days -21 to -250).
$\mathrm{Tone}_{\mathrm{Neg;Full\ Report}}$	The negative textual sentiment in the entire annual report on Form 10-K (submission including exhibits) as measured by the number of negative words according to the Loughran and McDonald business dictionary and its word lists.
Ln MVE	The natural logarithm of the market value of equity; calculated as the share price one day before the Form 10-K filing multiplied by the number of ordinary shares in issue.
PtB	The price-to-book ratio; calculated as the share price one day before the submission divided by the book value per share.
ROA	The return on assets; calculated as a firm's earnings before interest divided by the average of total assets.
Turnover	The average daily share turnover during the prior fiscal year; calculated using the daily number of traded shares divided by the number of outstanding shares.
Volatility	The daily stock return volatility; calculated as the standard deviation of the daily stock returns during the prior fiscal year.
Market Risk	The market risk; calculated as the covariance between the daily returns of the specific stock and the S&P 500 stock index during the prior fiscal year.

Notes: The table presents an overview of the variables used in the study.

Information extraction from financial statements

Annual reports on Form 10-K filed with the SEC have to be formatted in HyperText Markup Language (HTML 3.2/4.0) to be considered as *"official"* SEC EDGAR submissions (Filer Manual 2017, Section 5-1, Section 5-26).[A.1] Annual reports on Form 10-K filed with the SEC are publicly distributed by the EDGAR system (Electronic Data Gathering, Analysis, and Retrieval system). The text version of an annual report (*"Complete Submission Text File"*) provided by the SEC's electronic disclosure system aggregates all information made within a particular submission. Among other document file types, the *"Complete Submission Text File"* (file extension *.txt) contains the core Form 10-K document. In addition, the text version also contains the various exhibits disclosed in a particular filing (formatted in HTML) (Bodnaruk et al. 2015, 643; Loughran and McDonald 2011b, 1; Hering 2016, 11).

Based on the *"Complete Submission Text File"*, the study applies an information extraction algorithm to extract textual information from regulatory filings made with the SEC.[A.2] Using only regular expressions, the applied information extraction algorithm first extracts all documents embedded in the *"Complete Submission Text File"*. Table A.4 presents the regular expression applied to extract all documents disclosed in the *"Complete Submission Text File"*.

Table A.4. Regular expression contained in the information extraction algorithm.

Description	Regular expression
Extraction of documents (Form 10-K, exhibits, pdf, etc.) from the *"Complete Submission Text File"*	(?s)<DOCUMENT>.*?</DOCUMENT>

Notes: The table presents the regular expression contained in the information extraction algorithm for extracting all document file types embedded in the *"Complete Submission Text File"*.

After extracting all information (documents) from the text version of the annual report on Form 10-K the information extraction algorithm deletes all additional (non-relevant) document file types disclosed in the submission.[A.3] Table A.5 presents the regular expressions applied to delete all non-relevant document file types.

[A.1] Annual reports on Form 10-K can also be formatted in American Standard Code for Information Interchange (ASCII) to be considered *"official"* SEC EDGAR submissions. Annual reports on Form 10-K formatted in Portable Document Format (PDF) or XBRL are *"unofficial"* SEC EDGAR submissions (Filer Manual 2017, Section 5-1, Section 5-26).

[A.2] Please note that the information extraction algorithm applied in this study is based on the *"Annual Report Algorithm"* and the *"Items Algorithm"* designed by Hering (2016). Please further note that the parsing procedures are identical to those used in Henselmann and Hering (2017).

[A.3] The information extraction algorithm deletes the cover letter of the SEC submission (RegEx: (?s)<TYPE>COVER.*?</TEXT>) as well as any correspondence between the SEC and a particular EDGAR filer (RegEx: (?s)<TYPE>CORRESP.*?</TEXT>). Please note that the information extraction algorithm also deletes all XBRL documents within the submission.

Table A.5. Regular expressions contained in the information extraction algorithm.

Description	Regular expression
Removal of graphic files	(?s)<TYPE>GRAPHIC.*?</TEXT>
Removal of MS Excel files	(?s)<TYPE>EXCEL.*?</TEXT>
Removal of PDF files	(?s)<TYPE>PDF.*?</TEXT>
Removal of ZIP files	(?s)<TYPE>ZIP.*?</TEXT>

Notes: The table presents the regular expressions contained in the information extraction algorithm for deleting all additional non-relevant document file types embedded in the *"Complete Submission Text File"*.

Next, the information extraction algorithm deletes all metadata and tables as well as their corresponding content before extracting the core Form 10-K report.[A.4] Table A.6 presents the regular expressions implemented in the algorithm to delete non-relevant document content and to extract the core Form 10-K document from the *"Complete Submission Text File"*.

Table A.6. Regular expressions contained in the information extraction algorithm.

Description	Regular expression
Removal of document type information	<TYPE>.*
Removal of sequence information	<SEQUENCE>.*
Removal of filename	<FILENAME>.*
Removal of description	<DESCRIPTION>.*
Removal of head section (including document title)	(?s)<HEAD>.*?</HEAD>
Removal of table content	(?s)(?i)<Table.*?</Table>
Extraction of 10-K section	(?s)<TYPE>10-K.*?</TEXT>

Notes: The table presents the regular expressions contained in the information extraction algorithm for deleting non-relevant document metadata and table content as well as for extracting the core Form 10-K document embedded in the *"Complete Submission Text File"*.

Finally, the information extraction algorithm identifies all items within the core Form 10-K document and extracts all subsections contained in the annual report before deleting all remaining HTML tags and their corresponding attributes. Table A.7 presents the regular expressions implemented in the information extraction algorithm to identify all Form 10-K items and to extract specific subsections from the filing.

Table A.7. Regular expressions contained in the information extraction algorithm.

Description	Regular expression
Identification and renaming of headings (">°Item")	(?s)(?i)(?m)> +Item|>Item|^Item|^ +Item
Removal of multiple empty spaces	(?s) +
Extraction of *"Item 1 - Business"*	(?s)(?i)°Item 1[^AB012345].*?°Item
Extraction of *"Item 1A - Risk Factors"*	(?s)(?i)°Item 1A.*?°Item
Extraction of *"Item 3 - Legal Proceedings"*	(?s)(?i)°Item 3.*?°Item
Extraction of *"Item 5 - Shareholder Matters"*	(?s)(?i)°Item 5.*?°Item
Extraction of *"Item 7 - MD&A"*	(?s)(?i)°Item 7[^A].*?°Item
Extraction of *"Item 8 - Footnotes"*	(?s)(?i)°Item 8.*?°Item
Removal of HTML tags and attributes	(?s)<[^>]*>

Notes: The table presents the regular expressions contained in the information extraction algorithm for identifying all annual report subsections (*"Items"*) and for extracting certain important subsections from the core Form 10-K document embedded in the *"Complete Submission Text File"*.

[A.4] Please note that all reserved character sets (ASCII, ANSI, ISO 8859-1, etc.) are decoded.

Appendix for Chapter 3

Table A.1. Number of annual reports filed with the SEC EDGAR system.

Year	Month												Filings (Num.)
	1	2	3	4	5	6	7	8	9	10	11	12	
2016	169	2,110	3,183	558	203	244	143	198	233	138	194	216	7,589
2015	182	1,777	3,712	635	217	258	167	201	251	137	226	222	7,985
2014	193	1,787	3,573	732	223	264	165	204	287	187	223	246	8,084
2013	195	1,653	3,079	1,247	243	261	227	224	293	181	226	276	8,105
2012	197	1,872	3,508	799	241	292	193	214	300	223	245	309	8,393
2011	210	1,575	4,026	848	254	311	183	241	382	212	259	339	8,840
2010	254	1,476	4,206	956	264	319	202	223	383	246	279	357	9,165
2009	230	1,370	4,734	998	289	373	276	257	366	276	319	351	9,839
2008	103	1,670	4,557	509	192	264	162	176	325	205	185	398	8,746
2007	102	1,003	4,629	1,302	184	269	111	147	261	116	167	283	8,574
2006	97	509	6,123	438	133	345	98	259	317	105	121	307	8,852
2005	127	348	6,322	537	105	287	130	194	356	99	180	332	9,017
2004	176	425	5,924	460	123	313	128	112	343	111	103	349	8,567
2003	168	306	5,643	674	193	314	124	122	350	129	85	360	8,468
2002	259	286	4,412	1,987	234	283	186	134	381	157	239	369	8,927
2001	199	274	4,423	2,305	260	352	176	142	423	165	128	401	9,248
2000	247	299	6,292	761	240	421	180	143	508	189	133	456	9,869
1999	253	248	6,308	1,019	167	437	180	150	524	189	138	509	10,122
1998	251	270	6,398	824	298	430	240	147	554	205	150	520	10,287
1997	246	304	6,015	717	306	435	187	186	574	225	137	567	9,899
1996	157	188	2,484	1,126	179	244	255	161	543	197	167	558	6,259
1995	61	102	1,808	170	77	130	79	79	279	113	79	259	3,236
1994	29	66	1,231	110	58	46	31	36	110	41	56	109	1,923
1993	-	-	-	-	-	-	-	-	-	-	1	3	4
Filings (Num.)	4,105	19,918	102,590	19,712	4,683	6,892	3,823	3,950	8,343	3,846	4,040	8,096	189,998

Notes: The table shows the number of annual reports on Form 10-K and Form 10-K405 filed in each year since 1993.

Table A.2. Variables definitions.

Variable name	Expression
CAR_{MAM}	The cumulative abnormal stock return (marked-adjusted model); calculated as the sum of the absolute daily abnormal stock returns during the event period (trading days -1 to +2); the abnormal stock returns are computed using the market-adjusted model (MAM), where for the market return, the S&P 500 stock index is subtracted from a firm's stock return (trading days -1 to +2).
CAR_{MM}	The cumulative abnormal stock return (market model); calculated using an ordinary least squares (OLS) regression of the daily returns of a company on the daily returns of the market index (S&P 500) during the non-event period (trading days -21 to -250).
$CAR_{4\ Factor}$	The cumulative abnormal stock return (4-factor model); calculated using an ordinary least squares (OLS) regression of the daily returns of a company on the excess market returns, a size factor, a value factor, and a momentum factor during the non-event period (trading days -21 to -250).
ROA_{t+1}	The future return on assets; calculated as the earnings before interest in the next fiscal period divided by the firm's average of total assets in the next year.
$\Delta\ ROA_{t+1}$	The future change in return on assets; calculated as the return on assets in the next year minus the return on assets in the current fiscal period.
$\Delta\ Tone_{Net;Full\ Report}$	The change in the overall disclosure tone for the entire Form 10-K filing; calculated as the net disclosure tone (negative minus positive words) divided by the text length of the annual report minus the overall or net disclosure tone of the previous annual report.
$\Delta\ Tone_{Net;Footnotes}$	The change in the overall disclosure tone for the footnotes section; calculated as the net disclosure tone (negative minus positive words) divided by the text length of the footnotes section minus the overall or net disclosure tone of the footnotes section from the previous year.
Low ROA Dummy	The profitability dummy variable; that takes the value of 1 if a firm year observation is in the lowest quarter of all return on assets observations in the same industry, and 0 otherwise.
Ln MVE	The natural logarithm of the market value of equity; calculated as the share price one day before the Form 10-K filing multiplied by the number of ordinary shares in issue.
PtB	The price-to-book ratio; calculated as the share price one day before the submission divided by the book value per share.
ROA	The return on assets; calculated as a firm's earnings before interest divided by the average of total assets.
Δ ROA	The change in return on assets; calculated as a firm's return on assets minus the return on assets in the previous year.
Leverage	The leverage or dept ratio; calculated as a firm's total liabilities divided by total assets.

Notes: The table presents an overview of the variables used in the study.

Appendix for Chapter 4

Table A.1. Number of annual reports filed with the SEC EDGAR system.

Year	Month												Filings (Num.)
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
2016	169	2,110	3,183	558	203	244	143	198	233	138	194	216	7,589
2015	182	1,777	3,712	635	217	258	167	201	251	137	226	222	7,985
2014	193	1,787	3,573	732	223	264	165	204	287	187	223	246	8,084
2013	195	1,653	3,079	1,247	243	261	227	224	293	181	226	276	8,105
2012	197	1,872	3,508	799	241	292	193	214	300	223	245	309	8,393
2011	210	1,575	4,026	848	254	311	183	241	382	212	259	339	8,840
2010	254	1,476	4,206	956	264	319	202	223	383	246	279	357	9,165
2009	230	1,370	4,734	998	289	373	276	257	366	276	319	351	9,839
2008	103	1,670	4,557	509	192	264	162	176	325	205	185	398	8,746
2007	102	1,003	4,629	1,302	184	269	111	147	261	116	167	283	8,574
2006	97	509	6,123	438	133	345	98	259	317	105	121	307	8,852
2005	127	348	6,322	537	105	287	130	194	356	99	180	332	9,017
2004	176	425	5,924	460	123	313	128	112	343	111	103	349	8,567
2003	168	306	5,643	674	193	314	124	122	350	129	85	360	8,468
2002	259	286	4,412	1,987	234	283	186	134	381	157	239	369	8,927
2001	199	274	4,423	2,305	260	352	176	142	423	165	128	401	9,248
2000	247	299	6,292	761	240	421	180	143	508	189	133	456	9,869
1999	253	248	6,308	1,019	167	437	180	150	524	189	138	509	10,122
1998	251	270	6,398	824	298	430	240	147	554	205	150	520	10,287
1997	246	304	6,015	717	306	435	187	186	574	225	137	567	9,899
1996	157	188	2,484	1,126	179	244	255	161	543	197	167	558	6,259
1995	61	102	1,808	170	77	130	79	79	279	113	79	259	3,236
1994	29	66	1,231	110	58	46	31	36	110	41	56	109	1,923
1993	-	-	-	-	-	-	-	-	-	-	1	3	4
Filings (Num.)	4,105	19,918	102,590	19,712	4,683	6,892	3,823	3,950	8,343	3,846	4,040	8,096	189,998

Notes: The table shows the number of annual reports on Form 10-K and Form 10-K405 filed in each year since 1993.

Table A.2. Variables definitions.

Variable name	Expression
Y_{t+1}	Dummy variable that takes on a value of one if a company files for bankruptcy or zero otherwise.
$Z - Score_{t+1}$	The Altman Z-Score of a firm in the next fiscal year calculated as in Atlman 1968.
$Tone_{Type;Full\ Report}$	The textual sentiment in the entire annual report on Form 10-K (submission including exhibits) as measured by the number of pre-defined words according to the Loughran and McDonald business dictionary and its word lists.
$\Delta\ Tone_{Type;Full\ Report}$	The change in textual sentiment in the entire annual report on Form 10-K (submission including exhibits) as measured by the difference in the disclosure tone between the current and the previous filing.
$Tone_{Type;Section\ X}$	The textual sentiment in a specific subsection of the annual report on Form 10-K as measured by the number of pre-defined words according to the Loughran and McDonald business dictionary and its word lists.
Ln MVE	The natural logarithm of the market value of equity; calculated as the share price one day before the Form 10-K filing multiplied by the number of ordinary shares in issue.
PtB	The price-to-book ratio; calculated as a firm's share price one day before the submission divided by the book value per share.
ROA	The return on assets; calculated as a firm's earnings before interest divided by the average of total assets.
Liquidity	The current ratio; calculated as a firm's total current assets divided by the total current liabilities.
Cash	The cash ratio; calculated as a firm's cash plus short-term investments divided by the average of total assets.
Z-Score	The Altman Z-Score of a firm in the present fiscal year calculated as in Atlman 1968.

Notes: The table presents an overview of the variables used in the study.

Appendix for Chapter 5

Table A.1. Number of annual reports filed with the SEC EDGAR system.

Year	Month 1	2	3	4	5	6	7	8	9	10	11	12	Filings (Num.)
2016	169	2,110	3,183	558	203	244	143	198	233	138	194	216	7,589
2015	182	1,777	3,712	635	217	258	167	201	251	137	226	222	7,985
2014	193	1,787	3,573	732	223	264	165	204	287	187	223	246	8,084
2013	195	1,653	3,079	1,247	243	261	227	224	293	181	226	276	8,105
2012	197	1,872	3,508	799	241	292	193	214	300	223	245	309	8,393
2011	210	1,575	4,026	848	254	311	183	241	382	212	259	339	8,840
2010	254	1,476	4,206	956	264	319	202	223	383	246	279	357	9,165
2009	230	1,370	4,734	998	289	373	276	257	366	276	319	351	9,839
2008	103	1,670	4,557	509	192	264	162	176	325	205	185	398	8,746
2007	102	1,003	4,629	1,302	184	269	111	147	261	116	167	283	8,574
2006	97	509	6,123	438	133	345	98	259	317	105	121	307	8,852
2005	127	348	6,322	537	105	287	130	194	356	99	180	332	9,017
2004	176	425	5,924	460	123	313	128	112	343	111	103	349	8,567
2003	168	306	5,643	674	193	314	124	122	350	129	85	360	8,468
2002	259	286	4,412	1,987	234	283	186	134	381	157	239	369	8,927
2001	199	274	4,423	2,305	260	352	176	142	423	165	128	401	9,248
2000	247	299	6,292	761	240	421	180	143	508	189	133	456	9,869
1999	253	248	6,308	1,019	167	437	180	150	524	189	138	509	10,122
1998	251	270	6,398	824	298	430	240	147	554	205	150	520	10,287
1997	246	304	6,015	717	306	435	187	186	574	225	137	567	9,899
1996	157	188	2,484	1,126	179	244	255	161	543	197	167	558	6,259
1995	61	102	1,808	170	77	130	79	79	279	113	79	259	3,236
1994	29	66	1,231	110	58	46	31	36	110	41	56	109	1,923
1993	-	-	-	-	-	-	-	-	-	-	1	3	4
Filings (Num.)	4,105	19,918	102,590	19,712	4,683	6,892	3,823	3,950	8,343	3,846	4,040	8,096	189,998

Notes: The table shows the number of annual reports on Form 10-K and Form 10-K405 filed in each year since 1993.

Table A.2. Variables definitions.

Panel A: Dependent variables	
Variable name	Expression
$\text{File Size}_{\text{Gross}}$	The gross file size; calculated as the total number of characters in the original filing ("complete submission text file") including ASCII-encoded characters and characters attributable to HTML, XBRL and/or XML encodings.
$\text{File Size}_{\text{Net}}$	The net file size; calculated as the total number of characters in the original filing ("complete submission text file") excluding ASCII-encoded characters and characters attributable to HTML, XBRL and/or XML encodings.
CAR_{MAM}	The cumulative abnormal stock return (marked-adjusted model); calculated as the sum of the absolute daily abnormal stock returns during the event period (trading days -1 to +2); the abnormal stock returns are computed using the market-adjusted model (MAM), where for the market return, the S&P 500 stock index is subtracted from a firm's stock return (trading days -1 to +2).
CAR_{MM}	The cumulative abnormal stock return (market model); calculated using an ordinary least squares (OLS) regression of the daily returns of a company on the daily returns of the market index (S&P 500) during the non-event period (trading days -21 to -250).
$\text{CAR}_{\text{4 Factor}}$	The cumulative abnormal stock return (4-factor model); calculated using an ordinary least squares (OLS) regression of the daily returns of a company on the excess market returns, a size factor, a value factor, and a momentum factor during the non-event period (trading days -21 to -250).
ATV_{Ave}	The abnormal trading volume: calculated as the cumulative daily total proportion of outstanding shares traded during the event period (trading days -1 to +3), adjusted for the firm-specific average level of outstanding shares traded during the non-event period (trading days -6 to -156).
ATV_{Med}	The abnormal trading volume: calculated as the cumulative daily total proportion of outstanding shares traded during the event period (trading days -1 to +3), adjusted for the firm-specific median level of outstanding shares traded during the non-event period (trading days -6 to -156).
$\text{ASpread}_{\text{Ave}}$	The abnormal bid-ask spread; calculated as the cumulative daily total proportion of outstanding shares traded during the event period (trading days -1 to +3), adjusted for the average daily bid-ask spread during the non-event period (trading days -6 to -156).
$\text{ASpread}_{\text{Med}}$	The abnormal bid-ask spread; calculated as the cumulative daily total proportion of outstanding shares traded during the event period (trading days -1 to +3), adjusted for the median daily bid-ask spread during the non-event period (trading days -6 to -156).
ARV_{MAM}	The abnormal stock return volatility; calculated as the sum of the absolute daily abnormal stock returns during the event period (trading days -1 to +3); the abnormal stock returns are computed using the market-adjusted model (MAM), where for the market return, the S&P 500 stock index is subtracted from a firm's stock return (trading days -1 to +3).
ARV_{MM}	The abnormal stock return volatility; calculated as the sum of the absolute daily abnormal stock returns during the event period (trading days -1 to +3); the abnormal stock returns are computed using the market model (MM), where for the market return, the S&P 500 stock index is included in the regression (trading days -21 to -250).
$\text{ARV}_{\text{4 Factor}}$	The abnormal stock return volatility; calculated as the sum of the absolute daily abnormal stock returns during the event period (trading days -1 to +3); the abnormal stock returns are computed using the 4-factor model (4 Factor), were factor loadings are estimated by including the excess market returns, a size factor, a value factor, and a momentum factor in the regression (trading days -21 to -250).
$\text{ROA}_{\text{t+1}}$	The future return on assets; calculated as a firm's future earnings before interest divided by the future average of total assets.
$\text{DY}_{\text{t+1}}$	The future dividend yield; calculated as a firm's dividends per share divided by the future price per share (fiscal year end).
$\text{PR}_{\text{t+1}}$	The future payout ratio; calculated as the future dividends per share divided by the future earnings per share.

Panel B: Independent variables	
Variable name	Expression
$\text{Tone}_{\text{Neg;Full Report}}$	The negative textual sentiment in the entire annual report on Form 10-K (submission including exhibits) as measured by the number of negative words according to the Loughran and McDonald business dictionary and its word lists.
Ln MVE	The natural logarithm of the market value of equity; calculated as the share price one day before the Form 10-K filing multiplied by the number of ordinary shares in issue.
PtB	The price-to-book ratio; calculated as the share price one day before the submission divided by the book value per share.
ROA	The return on assets; calculated as a firm's earnings before interest divided by the average of total assets.
Δ ROA	The change in return on assets; calculated as a firm's return on assets minus the return on assets in the previous year.
Leverage	The leverage or dept ratio; calculated as a firm's total liabilities divided by total assets.
Turnover	The average daily share turnover during the prior fiscal year; calculated using the daily number of traded shares divided by the number of outstanding shares.
Volatility	The daily stock return volatility; calculated as the standard deviation of the daily stock returns during the prior fiscal year.
Market Risk	The market risk; calculated as the covariance between the daily returns of the specific stock and the S&P 500 stock index during the prior fiscal year.

Notes: The table presents an overview of the variables used in the study.