I

Louvain School of Management

The Processing of Information in Financial Disclosures

The use of bullet points

Author(s): Hill-Derive Anouchka
Supervisor(s): Thewissen James
Academic year 2023.-2024.
Master subject and focus:
Master [120] Business Engineering, with a specific orientation in Financial Engineering
Daytime schedule

Declaration Regarding AI Tool Usage in Master's Thesis

We recognize that AI tools might be valuable aids during the master's thesis work, but they are not infallible. Remember that transparency fosters trust, and acknowledging AI's role enhances the credibility of your work.

Therefore, when deciding to use such a tool, you need to adhere to the following principles of responsible use of AI.

1. **Critical Evaluation :**
   - We critically assessed the AI-generated output, ensuring its alignment with our research objectives.
   - Any modifications or corrections were made based on our expertise and domain knowledge.
2. **Transparency :**
   - We acknowledge the use of Chatgpt transparently, emphasizing that it contributed to our work but did not replace human judgment.
   - Our commitment to transparency ensures the integrity of this thesis.
3. **Ethical Considerations :**
   - We actively monitored for biases or unintended consequences introduced by the AI tool.
   - Our ethical responsibility guided our decisions throughout the research process.

Declaration **(This declaration is mandatory and must appear on the first page (after the title page) of the document.**

During the preparation of this master's thesis, the author(s) utilized Chatgpt for the following purpose:

1. Chatgpt was used to help with and correct our code in the program Rstudio

2. The AI was also employed to expand our ideas to test our data and the robustness of our results

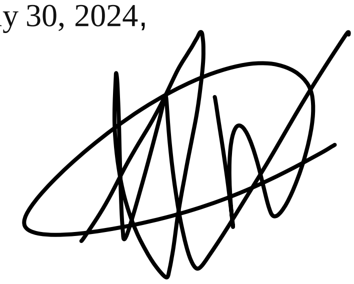
3. The last reason we made us of Chatgpt was to explain in simpler terms some concepts needed for this paper when their explanation was not clear enough in scientific papers

After using Chatgpt, the author(s) diligently reviewed and edited the content produced by the tool. We take full responsibility for the final content presented in this thesis.

By signing this declaration, we affirm that the content of this master's thesis reflects our original work, augmented by the responsible use of AI.

Feel free to customize the text by replacing "[NAME TOOL / SERVICE]" with the actual AI tool you used and providing a concise explanation of its purpose. Remember to sign and date the declaration appropriately.

Read and approved on May 30, 2024,



These few words to thank Dr. James Thewissen, Professor of Finance at the Louvain School of Management, for his extremely valuable advice throughout my entire thesis development process. Through my exchanges with Dr. Thewissen I got the opportunity to learn so much about financial disclosures and their importance for financial markets. Thanks to his experience and scientific point of view, he helped me to bring my thesis to a successful conclusion.

Table of Contents

1 Introduction 3

2 Literature review and hypothesis 7

2.1 Organizing information 7

2.1.1 The bullet point format 7

2.1.2 Organizing information in financial disclosures 9

2.1.3 Processing fluency theory and market efficiency 11

2.2 hypotheses development 12

3 Empirical research method 14

3.1 Data collection 14

3.2 Summary statistics 15

3.3 Research method 17

3.3.1 Part 1: Factors influencing the presence of bullet points in EPR 17

3.3.2 Part 2: Test of Hypothesis 1 18

3.3.3 Part 3: Test of Hypothesis 2 19

3.4 Variables description 19

3.4.1 Dependent Variables 19

3.4.2 Independent Variables 20

3.4.3 Control Variables 21

3.5 Model description 22

3.5.1 Part 1: Factors influencing the presence of bullet points in EPR 22

3.5.2 Part 2: Test of Hypothesis 1 22

3.5.3 Part 2: Test of Hypothesis 2 23

4 Empirical results 24

4.1 Findings 24

4.1.1 Part 1: Factors influencing prevalence of bullet points in EPR 24

4.1.2 Part 2: H1: "The number of bullet points in earnings press releases improves stock market performance until it reaches a curving point." 25

4.1.3 Part 3: H2: "Bullet points participate to reduce information asymmetry of stock markets and therefore improve the prediction accuracy of financial markets" 27

4.2 Robustness testing 28

4.2.1 New dependent variable 28

4.2.2 Cross-validation technique 29

5 Limitations 31

6 Conclusion 32

1 Introduction

This paper aims to put the research about effective communication in financial disclosures a step further. In order to do this, we need to understand a theory that connects information and stock markets which is the Market Efficiency Hypothesis (EMH). This hypothesis states that stock markets use all publicly available information to calculate market prices. However, limitations of this theory are that investors are rational people with cognitive limitations making them unable to process all the information and their perception varies depending on their knowledge, the environment, their personal beliefs and how the information is presented (Hodnett & Hsieh, 2012; Lee, 2012). The latter will be our area of focus for this research.

It has been shown that the way financial information is written influences investors and therefore stock markets (Henry, 2008). This makes the formatting of information a new tool to include in market predictions. Consequently, theories are developed to analyze the best ways to organize information to enhance the fluidity of the processing procedure.

The latter discovery brought scientists to focus on developing processing fluency theories that are based on the principle that people favor information that is quickly processed (Alter & Oppenheimer, 2006). Easily understood disclosures tend to reduce information asymmetry and draw in greater investments (Lawrence, 2013). This leads to a major concept of processing fluency: readability, or as defined by the Cambridge Dictionary, “the quality of being easy and enjoyable to read”. Until now, research has focused on the linguistic aspect centered on the content’s complexity. Therefore, this paper desires to distinguish itself by bringing forward the structural aspect of readability which refers to how information is presented and analyze how the specific format of bullet points impacts stock markets.

This research focuses on bullet points because this format is considered a heuristic tool that organizes the information effectively, makes the understanding faster and holds the reader’s attention (Ho, 2020; Ho, 2023; Kozak & Hartley, 2011; Ledin & Machin, 2015). We hypothesize that it can reduce information asymmetry and thus increase market efficiency because as Alter and Oppenheimer (2006) explain, “When people attempt to understand complicated information, they tend to simplify the task by relying on mental shortcuts or heuristics”. This makes bullet points a perfect candidate to increase the fluency of financial disclosures. Our findings show that indeed, up to four bullet points in an earnings press release (EPR), there is a positive impact on companies' stock prices.

Furthermore, since regulators understood financial disclosures influence market's stability and quality (Goldstein & Yang, 2017), they are looking for new rules to increase the readability of financial disclosures. Indeed, they consider it as a tool that, "levels the playing field" in financial markets, improving market efficiency and liquidity while also potentially lowering enterprises' cost of capital (Goldstein & Yang, 2017). This paper is an attempt to guide them towards the idea that investors would benefit from a regulation requiring firms to include highlights in their EPRs that consist of four bullet points.

Additionally, the existence of sentiments in the qualitative parts of financial disclosures has an impact on investors and "market reaction around the release of the accounting narrative tends to be positively associated with the expressed managerial sentiment" (Boudt & Thewissen, 2018). As Boudt and Thewissen (2018) express in their paper, research from many authors has shown that sentiment analysis is a promising tool for trend identification, market predictions and reduces the information asymmetry between companies and stakeholders (Davis, Piger, and Sedor, 2012; Patelli and Pedrini, 2013 cited by Boudt & Thewissen, 2018). Therefore, we add sentiment metrics to our analysis of bullet points.

Based on the findings that the presence of sentiments and the organization of the information influence processing fluency of business disclosures, we want to provide a first approach on how the presence of bullet points impact investors and, consequently, the financial markets.

The methodology used for this study starts with **a display of literature** to understand the background behind our topic **followed by an empirical research** divided into three main segments:

- First, we make use of linear regressions to understand what determines the prevalence of bullet points in earnings press releases and we discover that with years going by, companies make gradually more use of this format which could be linked to the growing discoveries about processing fluency theories. Moreover, positive sentiments are usually prevalent indicating that it is used by companies doing well or have a desire to look so.

- Second, the research focuses on the impact of bullet points on markets. We use polynomial regression models and determine that the amount of bullet points has, as hypothesized, a concave curvilinear relationship with stock prices.
- The third segment aims at determining how significant bullet points are on information asymmetry and therefore, the accuracy of stock market forecasts. We use the same methodology and apply regression analyses to see the relationship between the number of bullet points and error of predictions. Our findings show that bullet points help increase market efficiency up to a certain point. As for the relation with the stock price, the relationship between analysts' error and the number of bullets is concave help. However, regression models are not the best predictive models in this case which opens up to a need for further analysis.

In conclusion, this research seeks to enhance the theory of processing fluency in business disclosures by highlighting the effectiveness of a specific heuristic tool for improving readability: the bullet point format. Our contribution has three main aspects:

1. While research exists on the linguistic aspect of readability, there is a gap regarding **structural readability**. Research has shown that more complex and longer financial disclosures are associated with companies being more volatile and less successful (Li, 2008 cited by Loughran and McDonald, 2014) and several metrics have been developed to measure the communications content's complexity such as the FOG index or the Plain English initiative created by the U.S Securities and Exchange Commission (SEC). Also, there is proof that improving linguistic processing fluency has a positive impact on stocks, Alter and Oppenheimer (2006) show that stocks with a smoother name, or easy to pronounce have better results. However, no research can be found about how improving the structure of financial disclosures can be helpful as well. Based on Processing Fluency Theories and market efficiency, we empirically tested that organizing the information using up to four bullet points improves the fluency and therefore increases market efficiency
2. Although heuristics have been widely studied, there is a lack of literature on the **specific heuristic of bullet points**. Heuristic cues such as star ratings have been widely proved to be effective (Bigne, Simonetti, Ruiz & Kakaria, 2021 cited by Zhu, Huang & Liu,

2023). Our paper adds bullet points as a new heuristic shortcut that is efficient in financial communications to the literature.

3. As regulators develop new rules to improve the clarity of financial disclosures, our findings offer valuable insights. The SEC's has already developed the Plain English initiative that aims to present complex information more clearly for investors. However, this rules concerns the disclosures’ content and only a few structural criteria exist for effective communication. According to Trautmann and Hamilton (2003), cited by Arslan-Ayaydin, Boudt, & Thewissen (2016), earnings press releases must be "accurate and complete so as not to mislead." Consequently, this paper aims to offer a concrete solution for improving the fluency of business communications to regulators and open a new path to the literature of this theory.

This paper aims to offer a concrete solution for improving the fluency of business communications, and to open a new path to the literature of this theory.

2 Literature review and hypothesis

2.1 ORGANIZING INFORMATION

For clarity and comprehension, the way information is organized is very important. A well-structured piece of writing stimulates the reader to go easily from one idea to the next. When there are complex concepts to comprehend, it simplifies and facilitates the understanding. Information can be better read, remembered, and understood by organizing the information using specific approaches (Johnson & Christensen, 2011). The information can be formatted by dividing every part into digestible sections, arrange it using headings, bullet points, or visual aids.

Furthermore, a well-organized piece of paper makes important topics stand out as they are memorized and understood better by the readers (McKoon, 1977). As said earlier, in this paper, we focus on organizing information in bullet points.

Also, research has shown that the disclosure of information in the financial world has an impact on the market (Engelberg, 2008; Henry, 2006; Lawrence, 2013), managers need to carefully organize the information they want to communicate to their investors in their financial disclosures, as it can be a powerful tool. Indeed, the tone and the format of financial disclosures impact how investors understand and analyze the information (Arslan-Ayaydin, Boudt & Thewissen, 2016). In this section, we outline what the literature says about bullet points, financial disclosures and processing fluency theories.

2.1.1 The bullet point format

2.1.1.1 Definition

How does this format help understand information quicker?

For starters, we define it as a way to present information in a well-organized, logical and efficient manner (Ho, 2020). The use of this kind of communication becomes increasingly popular because it catches the reader's attention and helps him understand quickly the information (Ho, 2023). In the Educational field, prior research shows that the use of PowerPoints and with it, listed points increases student's attention, make the lectures more appealing and help memorize (Clark, 2008; Johnson & Christensen, 2011; Ding & Liu, 2012). The bullet points are a great assistance because thanks to their concise configuration, students can understand the material immediately and stay focused on what the teacher is explaining (Clark, 2008).

2.1.1.2 Bullet points as heuristics

Can bullet points be considered as heuristics ?

As stated by Gilovich, Griffin and Kahnemann (2002) and cited by Zhu, Huang and Liu (2023), heuristic cues are "simple, efficient rules that guide individuals' judgments and influence decision-making processes." Also, when individuals face complex or ambiguous situations, they tend to simplify the latter by using mental shortcuts or heuristics (Alter & Oppenheimer, 2006) as opposed to delving into complicated thoughts (Mahmood, Luffarelli & Mukesh, 2019 cited by Zhu, Huang & Liu, 2023). Accordingly, heuristics are usually used early in the communication process in order to let them act as a guideline and a tool to preserve cognitive activity (Ferran and Watts 2008 cited by Zhu, Huang and Liu, 2023).

Moreover, heuristics, by helping process more fluently, rise positive feelings and familiarity which leads to easier understanding and memorization (Alter & Oppenheimer, 2006). Bullet points also aspire to improve processing fluency by raising positive feelings thanks to its clear structure and therefore enhance understandability and memorization (Ledin & Machin, 2015) which confirms that they can properly be considered as heuristic shortcuts.

2.1.1.3 Psychology behind bullet points

Why is the bullet point format easy on the brain?

The use of heuristics such as bullet points is justified by the fact that according to Zhang, Zhao, Cheung and Lee (2014) "people consider a few informational cues or even a single informational cue and form a judgment based on these cues" (cited by Zhu, Huang and Liu, 2023). The latter is confirmed by Barker (2010) that tried to understand the psychology behind bullet points. The listing of highlights in such a way acts as concise containers, that make the brain emphasise on these specific pieces of information. Each bullet by being succinct is a signal that the item is digestible and independent from the others which aligns with the general human difficulty to process long continuous texts that ask for higher concentration and thus, are more demanding for the brain.

Another reason explaining why bullet points helps the understanding, is that the material is written in a few words with a “visual hierarchy”, expression used by Amic G. Ho (2023) which makes it easier for the reader to break down the main ideas and understand them. As Ledin and Machin (2015) explained, using the same logic as for heuristics, a text, that is organized, creates a sense of structure, which is reassuring for the brain, and brings a positive feeling to the reader, making it more appealing and helping him memorize the information. When the reader is motivated to read, it has a positive impact on the understanding and the memorizing.

Clarity and efficiency are essential keys to communicate, as Brock and Joglekar (2011) showed in their study that a lower density on a slide or in a text, thanks to formats like bullet points, is directly combined with effectiveness in grasping the ideas. Their results show that the right density is 20 words divided in 3 bullet points. It should be noted, however, that since this was evaluated using slide shows for oral presentations, we need to carefully use the results as our focus is on disclosures that are read.

2.1.1.4 Limitations of the bullet point format

Lastly, even though the bullet point format is a very effective and helpful communication tool, it also has its drawbacks. As Buchko, Buchko & Meyer (2012) explain, the risk of making use of bullet points is that it might oversimplify the information and even promote “intellectual simplicity” (Tufte, 2003 cited by Buchko, Buchko & Meyer, 2012). When information is overly condensed, the reader might not understand it entirely or not grasp the whole point that the writer is trying to make which might be considered as a bias. Consequently, this format needs to be used, but carefully.

2.1.2 Organizing information in financial disclosures

2.1.2.1 Financial disclosures

Financial disclosures are written reports that companies prepare to communicate relevant financial information to their stakeholders. Wang (2013) explains that these communications are supposed to provide a thorough and accurate representation of the business's financial condition and performance, in order to increase transparency, accountability, and help investors make well-informed decisions.

It has been widely discovered that these disclosures and how they are written impact the stock markets (Henry, 2008; Lawrence, 2013), as they connect companies with investors. Indeed, investors are people that can be convinced, influenced and sensitive to the information they receive. In his article, Lee (2012) referred to Hirst and Hopkins (1998) and Elliott (2006) to declare that all professionals of stock markets have limited attention and experience "information overload", exactly as a non-professional. Papers hard to read by being long, unorganized or overly complex, according to Li (2008) as referred by Lee (2012) are disliked, and make stocks more volatile and riskier. It is therefore essential for companies to use the most attractive and efficient ways of organizing information. Consequently, bullet points, known to be an efficient and logical way of formatting information (Ho, 2020) can be a tool for companies to reduce investor's overflood of market information and attract them.

However, companies are aware of the impact their disclosures have on investors, and thus, tend to format them in a way that diverts investors' focus from bad earnings (Lee, 2013). It has been known for a long time that businesses alter the numbers they publish to make the company look better in the investor's eyes. Similarly, managers now use information formatting and magnify the tone they use to trick investors (Arslan-Ayaydin, Boudt & Thewissen, 2016), in order to diminish the risk of scaring away their investors, and to "maximize the firm's value by minimizing the potential negative value effect" (Markowitz, 1952 cited by Arslan-Ayaydin, Boudt & Thewissen, 2016) of their information disclosures. Nevertheless, investors are aware of these dishonest practices which makes them less sensitive to overly positive disclosures. Also, financial disclosures are studied thoroughly by regulators to "improve financial market quality and stability" and protect the investors (Goldstein & Yang, 2017). Regulators forbid managers to provide misleading information in their disclosures but whereas false numbers are easily discovered, it is more difficult to regulate qualitative information.

2.1.2.2 Earnings press releases

This paper focuses on one specific financial disclosure: earnings press releases. EPR's are usually written by a business's public relations departments and is published to inform investors and the press of important materials regarding the company. Henry, Thewissen & Torsin (2021) explain that research has shown the significance of this type of financial communications as of “economic importance” to transfer information to investors (Kothari, 2001 cited by Henry, Thewissen & Torsin, 2021), earnings press releases have more impact on the market compared to other financial disclosures mandated by the SEC such as annual reports (Li & Ramesh, 2009). Additionally, information provided in earnings press releases is usually more qualitative and make it therefore, easier for companies to play within regulatory constraints.

2.1.3 Processing fluency theory and market efficiency

A theory that combines information and stock markets is called market efficiency. The Efficient Market Hypothesis (EMH) was first employed by Fama in 1970. We define market efficiency as the extent to which stock prices take into account all relevant information. According to this theory, if the market is efficient, it is impossible for investors to outperform it. Also, literature recognizes three degrees of efficiency:

- First, in a weak form of efficiency, the current stock prices incorporated all the information from the past. This implies that former price movements have no utility to forecast future prices.
- Second, semi-strong efficient markets adapt quickly to new public information making it possible for an investor to have an advantage when he possesses private and unavailable information from that market.
- Third, stock prices, with the strongest degree of efficiency, reflect all the information, public and private. In such a market, all the information has been priced, so no investors, even owners of private information, can find advantages to outperform the market.

Nonetheless, we must notice an important critique. It has been found in decision research and behavioural finance, that investors and market professionals, as opposed to the assumption of the EMH, are irrational people (Statmann, 1999 cited by Hodnett & Hsieh, 2012). They have limited cognitive abilities, which makes it impossible for them to process all the information available (Hodnett & Hsieh, 2012), and that increases information asymmetry and consequently, reduces market efficiency.

A response to this limitation is the **processing fluency theory** based on the principle that individuals favor easily processed information, or “fluent information” (Alter & Oppenheimer, 2006). Accordingly, a person's judgments and decision-making are influenced by how easily they receive information. When information is easy to process, it tends to be perceived as more familiar, truthful, and pleasing (Alter & Oppenheimer, 2006; Reber, Schwarz, & Winkielman, 2004), which can result from various factors such as clarity, simplicity, and familiarity. Readability, defined as the ease with which a reader can understand written text (Flesch, 1948), is an essential concept of processing fluency. When a paper is clear and concise, and has a coherent structure, it has a high readability, which naturally improves decision-making and enhances the reader's capacity to understand and memorize information (Flesch, 1948).

Processing fluency and readability are directly related to financial markets, as they have a big influence on how investors perceive, and thus respond to financial disclosures. Financial reports, that are organized and clear, using bullet points to highlight important information, are easier to read and understand. Song and Schwarz (2008) confirm the latter and explain that the decision-making process of investors becomes faster and more accurate when readability is enhanced, which has a positive impact on investor behavior and as a consequence reduces information asymmetry and improves market efficiency. Lawrence (2013) goes further and found that people invest more in companies that make better disclosures. Particularly, people own more shares in companies that provide clear and easy-to-read financial disclosures.

2.2 HYPOTHESES DEVELOPMENT

From the past exploration of literature, a question can therefore be asked: What informational format is best adapted to earnings press releases?

We know that the use of cognitive writing tools or heuristics improve readability and therefore positively affect investors and stock markets. This paper focuses on the bullet point format. We consider it as a heuristic shortcut and a structural readability tool that should enhance processing fluency of financial disclosures, as it is supposed to boost investors' attention and confidence (Zhu, Huang & Liu, 2023). Consequently, that should reduce information asymmetry between firms and stakeholders and improve market efficiency.

This brings us to our first hypothesis:

H1: "The number of bullet points in earnings press releases improves stock market performance until it reaches a curving point."

Knowing that the analysis of financial disclosures has informational value to predict stock market movements and based on the processing fluency principle, we expand our research and establish a second hypothesis to be tested:

***H2: "Bullet points participate to reduce information asymmetry of stock markets and they improve the prediction accuracy of financial markets"*.**

Our research methods to test these hypotheses will be developed in the next section of this paper.

3 Empirical research method

Based on the processing fluency principle and market efficiency hypotheses, we develop in this section a research method to test the impact of bullet points on financial markets. We start this third chapter of the paper by explaining the dataset we employed for our research, we continue with an analysis of some summary statistics to have a better understanding of our data. The third part of this section describes the empirical methodology that we based on linear and polynomial regression analyses. Furthermore, we explain the dependent, independent and control variables we decided to use in our regressions and we conclude this chapter by diving into the description of our regression models.

3.1 DATA COLLECTION

The dataset is composed of 54 725 earnings press releases containing bullet points, and 62 variables selected as control variables to analyze the relationship between the financial world, bullet points and sentiment analysis. Since this paper is a first step in this research, we included only a set of these variables that were estimated to be most relevant.

For the sake of this analysis, we selected variables that were relevant to our research and the unused variables were removed from the dataset, leaving it with 54 725 EPR and31 variables. The list and definition of each variable used in this paper can be found in Table 1 of the Appendix.
Our data come from the database COMPUSTAT, known to be specific for financial, statistical and market-related applications.

Collecting the dataset was a complex process because the files are in HTML format, which has a distinct syntax, that is typically unknown from the companies having data on Compustat. This is problematic to our data collection because managers writing EPR are not aware that a node in the HTML code is created when they make a bullet point in their file, consequently, if this person deletes the bullet, they will only remove the bullet point itself without eliminating the node behind it in the code creating nodes with actual bullet points and nodes with “fake” bullets. The issue is that our program does not distinguish these genuine and “fake” bullets, leading to a general dataset including all nodes which therefore includes errors.

It has been chosen to automate the procedure in order to expand the number of observations for analysis rather than selecting each EPR with genuine bullet points which should have been done by an individual and would have led to a significantly smaller dataset. Even though this decision led to some errors in our dataset, thanks to the law of large numbers, these errors cancel out each other our, which minimizes their overall impact on our analysis.

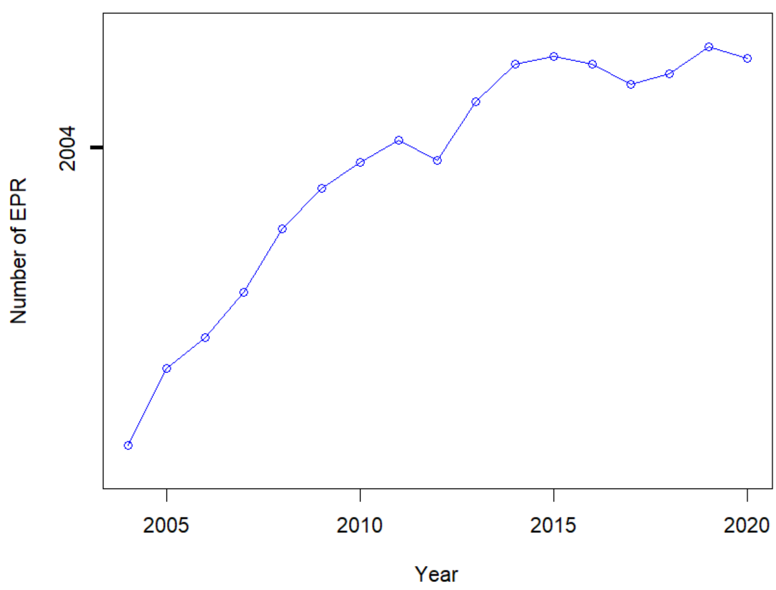
Lastly, all this work has been done by Dr. Thewissen and his colleagues, and the dataset has been generously given to us for the purpose of this paper.

3.2 SUMMARY STATISTICS

By starting with **general statistics** to better understand our data, we are getting some interesting results that can be seen in Table 2. In our sample, businesses use, on average, five bullet points for each press release, which is backed up by the fact that five is also the variable's median. Nonetheless, with a standard deviation of 2.56, the range of bullet points goes, as a consequence, from 2 to 10 bullets and the proportion of EPR containing 2 and 10 bullet points is respectively 15% and 12%.

Also, as shown on Graph 1, we see that the prevalence of bullet points increases with **time**. A deduction that could be made is that along with the discoveries about the significance of organizing information, companies try to improve their writings.

Graph 1: Average Amount of EPR with Bullet Points Per Year



Moreover, the number of bullet points does not appear to be much impacted by the **locations of the company's headquarters**. With three exceptions, France, Argentina, and Israel, using 7 bullet points on average, most of the countries have a mean of four to six bullet points per earnings press release.

To get a sense of **the kind of businesses** that make up our dataset, we look into the accounting and financial metrics. We see that the Return on Assets (ROA) ratios range from --2.52 to 0.67 and have an average value of 0.0023, the stock prices vary from 0.26 to 1 525 000 and the amount of total assets for a semester goes from 3.78 to 3 246 076. These metrics including wide ranges of values, consequently, we deduce that our dataset is composed of various companies of different sizes and strategies. Also, 21% of the press releases of the dataset are associated to a negative ROA, meaning that the companies, who wrote these papers, were making losses on their assets at the publishing . Having a diversified set of data is helpful to grasp a full overview of what influences the use of bullet points in the business world and avoid selection bias. When a sample is not representative of the population being studied, we risk selection bias to arise, which can skew the findings and consequently lead to incorrect conclusions.

Regarding the **textual analysis measures**, businesses composing our data write press releases within the same range of readability since the average FOG index is 22.02 and has a rather small standard deviation of 3.5. According to the generally accepted scale of this index, this is a high value and therefore a low readability. However, the FOG index is known to be unadapted to financial disclosures because it is based on sentence's length and words' complexity. It indicates a low readability when a text contains a high proportion of complex words defined as words with two or more syllables. However, multi-syllable terms are very frequent in business language and earnings press releases even though these are not difficult to understand (Loughran & Mcdonald, 2014). Additionally, the amount of words varies a lot between the disclosures, we see that it varies from 38 words to 30 656.

Companies tend to express predominantly positive **sentiment** in their bullet points. **Sentiment metrics** are measured using Henry's sentiment dictionary and is more widely explained in the section 3.4.2 Independent Variables. Indeed, as we can see in Table 2, the average values for the positive sentiment metric, PositiveHenry is higher compared to its negative sentiment metrics. However, as indicated by the very high standard deviation values there is significant variability in how this sentiment is communicated across different press releases. Consequently, it is difficult to make any assumptions from these results.

3.3 RESEARCH METHOD

3.3.1 Part 1: Factors influencing the presence of bullet points in EPR

The first part of the research tests the **prevalence of bullet points in earnings press releases** and what are the **characteristics of the companies making the most use of this format**. To achieve this goal, we adopt a quantitative approach. The method we chose is the linear regression analysis. The methodology involves constructing several regression models, with the dependent variable being the count of bullet points (*nbrbullets*). We want to determine the impact of independent variables on the presence of bullet points while taking other related factors into account. To improve the continuity of the non-negative variables, we take the natural logarithm +1 of these variables such as nbrbullets.

Regression models are estimated using ordinary least squares (OLS) regressions. The approach of the model is to estimate the parameters of the linear regressions by minimizing the sum of the squared differences between the observed values of the dependent variable and the values predicted by the model. This computes coefficient estimates for each independent variable, indicating how much and the direction of their impact on the dependent variable. The general formula of linear regression models we use is the following:

$$Y = \alpha + \beta_i X_i + \varepsilon$$

- Where **Y** denotes the dependent variable (*nbrbullets*)
- $\boldsymbol{\alpha}$ is the intercept
- $\boldsymbol{\beta_i}$ represents the coefficient of the independent variable $\boldsymbol{X_i}$
- $\boldsymbol{\varepsilon}$ is the standard error of the model

The independent variables are grouped according to their definition and one linear regression model corresponds to the test of the dependent variable with one group.

The first category represents **accounting performance metrics** and includes the variables total assets per semester (*atq*), return on assets (*EARNTA*) and the standard deviation of ROA (*sd*). These are the variables we use to evaluate the influence of the companies' financial health and profitability on their bullet point usage.

Second, **market metrics**, which includes return (*return*), losses (*loss*) and volatility of the stock return (*risk_monthly*) are used to grasp the market dynamics.

We include a **second group of market metrics** counting the cumulative abnormal returns for the short and long term trading day windows (*car1, car5, car3, car180*) to reflect the stock market's response before and after the release of earnings information.

We investigate **temporal trends** by grouping the variables representing the year (*year*) and month (*month*) of the EPR. This helps us evaluate if the use of bullet points change over time.

The following category of variables are the **analysts' consensus estimates**, including mean earnings forecasts (*Consensus_mean*), their standard deviation (*STD_consensus)* and the difference between the actual and forecasted values (*errormean*) to provide insights into the alignment or deviation from market expectations.

Lastly, a group counting **textual analysis metrics** and **sentiment analysis tools**, including the FOG index (*FOG*) and the library Henry, (*PositiveHenry, NegativeHenry*) help us to have insights into the sentiments expressed in the bullet points.

3.3.2 Part 2: Test of Hypothesis 1

The following goal of this analysis is testing our first hypothesis by establishing the **curving point of the number of bullet points at which stock prices stop increasing and start lowering down**. Our approach employs polynomial regressions, a specific type of regression analysis that models the relationship between the dependent and independent variables in the n^{th} degree. While linear regression is suited for analyzing linear relationships between variables, the polynomial regression model is particularly valuable to investigate nonlinear relationships, especially curvilinear patterns (Ostertagová, 2012).

$$Y = \alpha + \beta_i^n X_i^n + \varepsilon$$

- Where the term $\boldsymbol{\beta_i^n}$ represents the coefficient of the i^{th} variable at the nth degree.

We hypothesize that the correlation between stock price and the number of bullet points is curvilinear, making the polynomial model better fitted for our analysis. In this second approach, we have a model without control variables and a second with. We chose the variables that were determined to be significantly impacting the number of bullet points in the first part of the research. By integrating significant independent variables into our analysis, we want to improve the control over their effects and can better evaluate how they impact the relationship between the number of bullet points and stock prices. This approach is supposed to help in several ways, it mitigates bias, enhances accuracy and improves the fit of the model (Frölich, 2008). However, it is important to check for multicollinearity and overfitting by assessing these issues through diagnostic tests (Frölich, 2008).

3.3.3 Part 3: Test of Hypothesis 2

The last part of our empirical research is to **confirm or infirm that the presence of bullet points containing sentiments reduce the asymmetry of information on the financial market and increases the stock market efficiency**.

To accomplish that, we analyze regression models using the *errormean* as dependent variable. We first start with a linear regression having only the independent variable and no control variables. Next, we develop the analysis by adding the same control variables we used with the first hypothesis. We further test for non-linear relationships by applying polynomial transformations on the latter linear models. Here again, we predict a concave relationship between the variables.

3.4 VARIABLES DESCRIPTION

3.4.1 Dependent Variables

For the purpose of our research, **three dependent variables** were needed. Firstly, to establish what the factors influencing the utilization of bullet points are, we need the variable ***nbrbullets***. It represents the number of bullet points contained in the earnings press release concerned.

The second explained variable we use is ***stock_price*** to test our hypothesis of a curvilinear relationship between the amount of bullet points and stock prices. It indicates the value of the stock at the time the company released its earnings press release.

The last variable serves to analyze the informational value of bullet points in the prediction of stock price's movements. The variable we choose for this purpose is ***errormean*** which represents the average difference between the analysts forecast and the actual value of earnings per share at the publishing moment of the EPR

3.4.2 Independent Variables

In the first part of our research, we use different independent variables to determine what impacts the presence of bullet points in earnings press releases.

The variables *year* and *month* are utilized to test whether **time** impacts the number of bullet points. The former represents the year the earnings news release was published, while the latter shows the month the disclosure was made.

We also looked at the relationship between the **firms' accounting** and the bullet points they included in their disclosures using the variables *atq, ibq, EARNTA, sd* and *loss*. They stand for, respectively, the company's assets per semester, income before taxes and depreciation, return on assets (ROA), the ROA ratio's standard deviation over the last five years representing companies' profitability, and a binary variable that equals to 1 when the ROA ratio's value is negative or 0 otherwise.

Financial-related variables we use are *return*, which shows the stock price return of the previous 12 months before the EPR was published, and *risk_monthly*, which is the standard deviation of the monthly stock price return over the previous 12 months and is a measure of the company's volatility and consequently, a risk measure.

Furthermore, we use variables associated to **financial analysts metrics**. The average value of the analysts' predictions for the value of earnings per share is represented by the variable *consensus_mean*. *Actual* is a variable that represents profits per share as they actually occur, and *STD_consensus* is the latter variable's standard deviation. By displaying the gap between the actual value of the profits per share and the analysts' forecast*, errormean* demonstrates the error of the analysts' average and median consensus.

We also test the **cumulative abnormal return (CAR)** over different event windows surrounding the earnings press release. The latter is displayed by the variables *car1, car3, car5,* and *car180*. *Car1* records the CAR between the day before the EPR and the day after the publication, *car3* covers the window between the three days prior to and the three days following the EPR, *car5* covers the period between the five days prior to and the five days following the disclosure and, *car180* records the CAR between the day before the EPR and the 180 days following the release. These variables are meant to reflect the short- and long-term responses of investors to the market.

The variables related to **textual analysis**. *FOG* and *NumberOfWords* are measures of readability. The former represents the FOG index defined by Loughran and Mcdonald (2014) as a “linear combination of average sentence length and the proportion of complex words (words with more than 2 syllables)” and the latter indicates the amount of words contained in the bullet points.

Finally, the variables *PositiveHenry* and *NegativeHenry* indicate the amount of words representing **positive or negative words** according to the word lists developed by Henry. We preferred the Henry dictionary to the DICTION and Loughran&McDonald libraries because it was specifically designed for financial text analysis and tailored to grasp the nuances of financial vocabulary and market sentiments (Kearney & Liu, 2014; Todd, Bowen & Moshfeghi, 2024). The DICTION library uses predefined dictionaries for sentiments and other categories to assess tone across several domains making it less adapted to our needs and the Loughran & McDonald's while being also adapted to financial text analysis, its sentiment library includes categories such as litigations and uncertainty which we determined to be unnecessary (Kearney & Liu, 2014; Todd, Bowen & Moshfeghi, 2024).

The second and third part of our research only uses **one independent variable** of interest, *Nbrbullets* that indicates **the number of bullets in the earnings press release**. However, our regression models used in these sections are implemented with control variables that we develop in the following part of this paper.

3.4.3 Control Variables

We employ **control variables** to complement our regression models to avoid obtaining biased estimates of coefficients. They encounter for other factors that could have an impact on the dependent variables and help isolate the effect of the independent variable of interest.

All variables composing our dataset can be considered as control variables. In order to avoid unnecessary complex model and overfitting, we make a selection in these variables. We decide to use the results of the first part of our analysis, or the variables that have a significant relation with bullet points. Variables are considered significant if their p-value is lower than 0.5.

The **variables we determined to be significant** with the linear regressions with the number of bullets are: *atq, EARNTA, sd, loss, risk_monthly, year, month, FOG, PositiveHenry,* *NegativeHenry* and *NumberOfWords*. The definition of these variables can been found in the section 3.4.2 Independent Variables or in Table 1 of the appendices.

3.5 MODEL DESCRIPTION

3.5.1 Part 1: Factors influencing the presence of bullet points in EPR

The first regression models we need are **multiple linear regression models** to test the significance of the group of independent variables j on the variable representing the number of bullets:

$$\log(1 + Nbrbullets) = \alpha + (\beta_i Independent\ Var_i)_j + \varepsilon_j$$

As explained in the section 3.3 Research method we classified the variables of our dataset according to their definition. We transformed the non-negative variables (dependent and independent) using the natural logarithm + 1 transformation to make the models more continuous. Our expectations are that the variables of time, accounting and financial metrics will be most significantly related to the number of bullets.

3.5.2 Part 2: Test of Hypothesis 1

The second part of our analysis encounters **linear and polynomial regression models** with and without control variables and aims at determining if there is a relationship between the number of bullet points and the stock prices, a market performance metric for companies. The first model is a linear regression without control variables:

$$\log(1 + stockprice) = \mathrm{\alpha} + \mathrm{\beta}\log(1 + \mathrm{Nbrbullets}) + \mathrm{\varepsilon}$$

Here again we use the natural logarithm transformation. We believe the coefficient will show a negative significant relationship between the dependent and independent variables but the model accuracy will not be convincing as we expect a curvilinear relation. Consequently, we attempt to increase to model's fit by adding control variables:

$$\log(1 + stockprice) = \alpha + \mathrm{\beta}\log(1 + \mathrm{Nbrbullets}) + \gamma Control\ Var + \varepsilon_i$$

Where the control variables are chosen as explained in part 3.4.3 Control Variables and attempt to account for confounding factors and improve the accuracy of the model. We anticipate that as the complexity of the model increases, the proportion of the data that is explained will increase as well however, we risk to obtain an overfitting model. Also, again, as we expect a curvilinear relationship, this model will not be convincing enough and leads us to the third regression model, the polynomial regression without control variables:

$$\log(1 + stockprice) = \alpha + \beta_1 \log(1 + \text{Nbrbullets}) + \beta_2 \log(1 + Nbrbullets)^2 + \varepsilon$$

The expected non-linear relationship is captured by polynomial terms. We stop at a regression of second degree because we hypothesize a curvilinear relationship. If the polynomial degree is too high, the fitted curve risks to be too flexible and become oddly shaped. In order to increase the accuracy of the model, again, we add control variable and obtain our fourth model:

$$\begin{aligned}\log(1 + stockprice) \\ &= \alpha + \beta_1 \log(1 + \text{Nbrbullets}) + \beta_2 \log(1 + Nbrbullets)^2 + \gamma_1 Control\,Var \\ &+ \gamma_2 Control\,Var^2 + \varepsilon\end{aligned}$$

The same approach is used regarding the control variables as in the linear regression. Both coefficients of the independent variable are expected to be negative in order to have a concave curvilinear relationship.

3.5.3 Part 2: Test of Hypothesis 2

The third part of the analysis uses the same approach as the latter with linear and polynomial regressions but differs on the dependent variable. The chosen one to represent the impact of bullet points on the asymmetry of information and therefore on forecastings' accuracy is the *errormean* which we expect to have a curvilinear relationship with the number of bullets too.

4 Empirical results

4.1 FINDINGS

4.1.1 Part 1: Factors influencing prevalence of bullet points in EPR

The preliminary results were centered on analyzing what variables have a significative impact on the use of bullet points to find out what determines their presence. Our linear regression models show that the variables from the **market metrics** group, **assets per semester, Return on Assets (ROA), the standard deviation of ROA and loss are significantly impacting the number of bullet points**. Whereas, companies with higher assets per semester tend to use more bullet points companies than companies with a lower capital, businesses's with higher volatility and lower or negative ROA ratios make less use of it.

Regarding **market metrics**, only the **volatility of the stocks return are significant** with a negative coefficient meaning that companies, having higher volatility, or in other words, being riskier tend to use the classical continuous text rather than a shorter format.

Also, **regarding the sentiment analysis metrics**, we see that bullet points are **more likely used to present positive information**. Whereas the variable containing the positive words has a positive coefficient estimate, the variable corresponding to negative words in the Henry library has a negative coefficient.

Lastly, as was seen with summary statistics, the **use of bullet points increases with years,** but are unexpectedly **more prevalent at the start of the calendar year**.

Furthermore, as we see in Table 3, the coefficients of **cumulative abnormal return variables are not significant**. Since we hypothesized a relationship where bullet points are supposed to reduce information asymmetry, we took our analysis a step further. We tested whether polynomial transformations would improve the fitting of these regressions, and if more variables would become significant. However, only the variables *car1* and *car180* become relevant, indicating non-linear relationships between the number of bullet points and cumulative abnormal return with very short and very long windows, or market response impacts bullet points only on short and long periods. The polynomial transformation's results should be taken carefully because even though the adjusted R-squared values increase with the degrees, the residual standard errors also rise meaning that whereas the model is better fitted to the data, the model's predictions have bigger errors on average. This could be a consequence of overfitting where the model catches noise in the training data rather than only the relationship.

To conclude this first batch of results, the companies that are more likely to use bullet points, **possess more assets** and **present positive information** or use a **confident tone**. However, as was explained in the literature, companies having less performing results tend to prefer continuous texts and lower their readability in order to drown their performance in complicated writings. Indeed, our empirical results definitely show that the information companies decide to highlight using bullet points is meant to be in their favour.

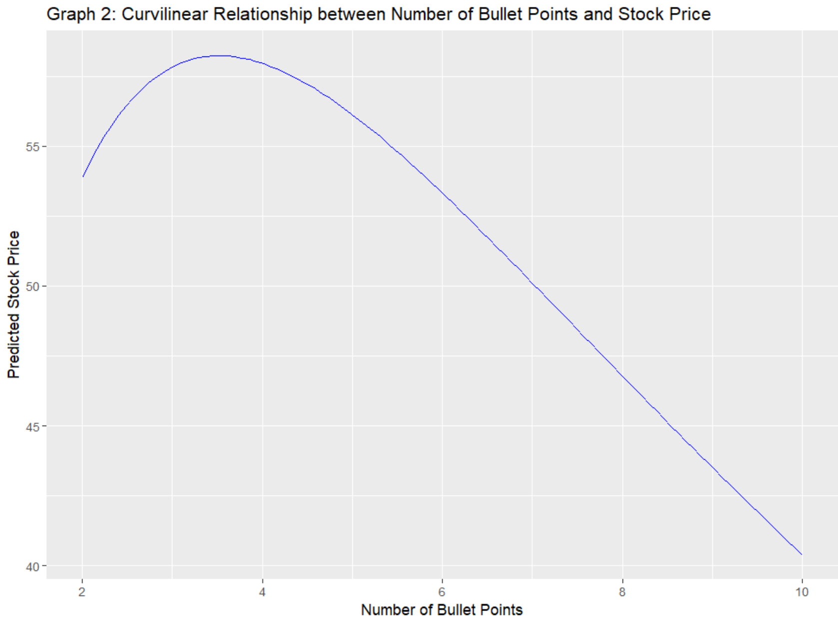
4.1.2 Part 2: H1: "The number of bullet points in earnings press releases improves stock market performance until it reaches a curving point."

The second part of this analysis is dedicated to determining the nature of the relationship between the stock prices and the number of bullet points. As hypothesized, the results in Table 4a and 4b show a significant and nonlinear correlation.

We started by modelling a polynomial regression with the natural logarithm of the stock price variable as dependent variable and the natural logarithm of the number of bullet points as only independent variable. **Our results provide an initial confirmation of our theory, demonstrating a negative and concave relationship between the variables.**

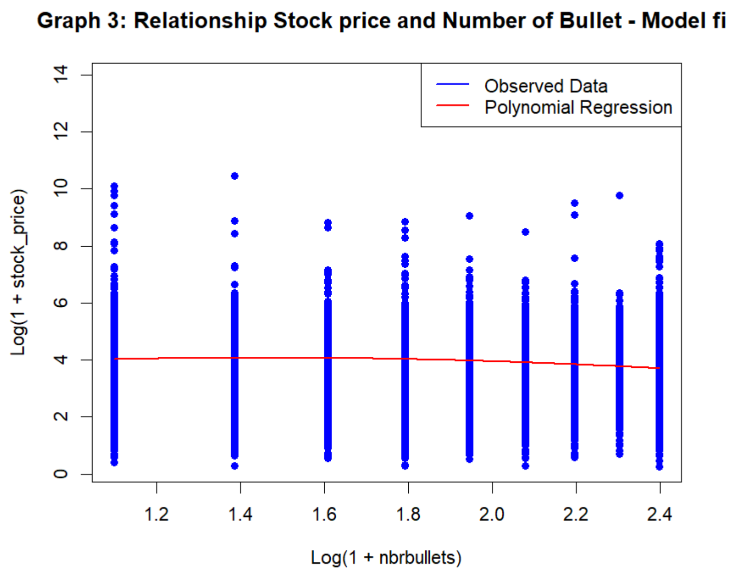
Further evidence **that the relationship between stock price and the number of bullet points is curvilinear**, is provided by the observation that raising the degree of the polynomial regression does not really increase the value of the adjusted R squared or the residual standard error and the estimated coefficients associated with higher degrees are not significant.

Also, as we can see on Graph 2, when adding control variables we have an additional confirmation that the relationship is definitely concave and curvilinear.



These results support the findings of Brock and Joglekar (2011), who suggested that three bullet points is the ideal number to elicit the optimum response from readers. Graph 2 demonstrates distinctly that the curve's optimal point is located around 3.5. The inference that can be drawn from this, is that businesses should carefully make use of the bullet point format. While it is initially a more advantageous format, employing it excessively or entirely to structure financial disclosures can have negative effects.

Nonetheless, an important observation is that the values of the adjusted R squared stay very low meaning that as can be seen on Graph 3, our model only partially explains the variance of the dependent variable.



Alternative modeling approaches should be explored to improve the prediction performance of the model. These will be discussed in section 5. Limitations.

4.1.3 Part 3: H2: "Bullet points participate to reduce information asymmetry of stock markets and therefore improve the prediction accuracy of financial markets"

This section of the study examines if bullet points are a useful tool for reducing information asymmetry and, improving market efficiency. The predicted variable we employ in this section is *errormean*, which aims to determine if bullet points help improve financial analysts' predictions. The results in Table 5a and 5b are surprising because they show that the **estimated coefficient for the number of bullet points in the linear regression model is only statistically significant at 0.1 making the results unconvincing since the p-value is higher than 0.05.** Also, the model's residual standard error is very high, and the adjusted R-squared is very low, implying a model that poorly fits the data.

When we include control variables in the linear regression, the **significant relationship between the number of bullet points and forecasting errors is again not confirmed**. Even though, the value of the adjusted R squared improves compared to the simple regression model, it is still low, and the residual standard error decreases only slightly, indicating potential overfitting.

Similarly, polynomial regressions support an **unsignificant relationship between forecasting errors and bullet points**, but the model evaluation metrics show again a poor fitting of the data. Therefore, while the number of bullet points appears to have no informational value in predicting stock market movements, further research is required to develop more accurate models that might better analyze the nature of this relationship.

To conclude, only the first hypothesis is confirmed, bullet points impact companies' stock price but extensive research has to be done to determine if this heuristic shortcut improves information asymmetry. Also, even if the first relationship is concave, managers need to use the bullet point format carefully since, as shown on Graph 2 abusing of this format is actually detrimental.

4.2 ROBUSTNESS TESTING

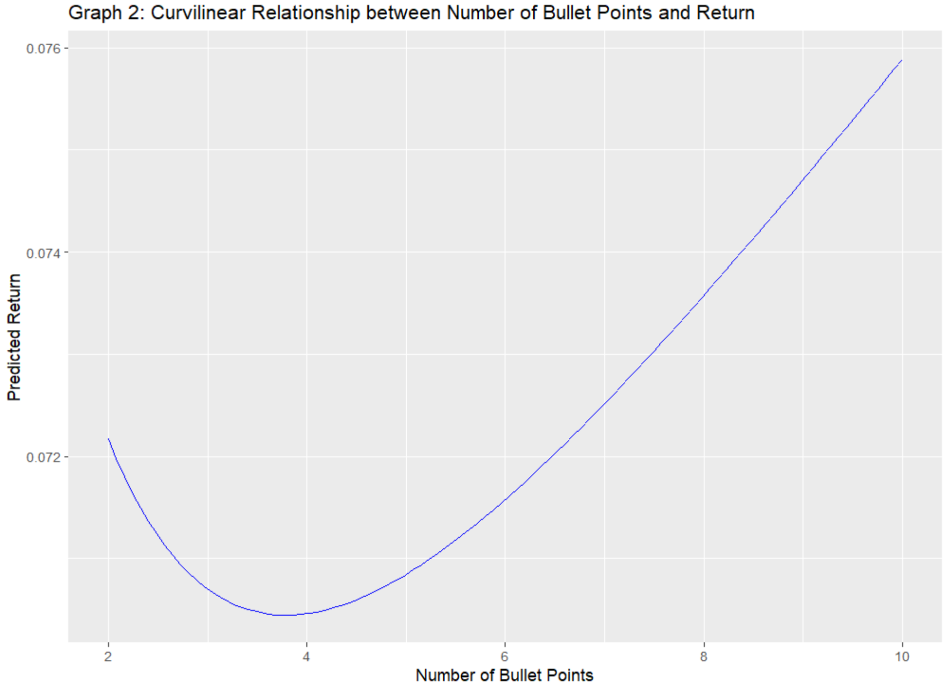
To verify the validity and dependability of our results, we run two robustness tests in order to confirm that our findings are not affected by particular model parameters, presumptions, or sample changes. This aims to confirm the consistency and generalizability of our findings and to strengthen our confidence in our conclusions. As the results of the second hypothesis are not really convincing, we only test the robustness of the first hypothesis.

4.2.1 New dependent variable

To see whether our findings are stable and reliable we first test them using another dependent variable.

With the empirical results, we confirmed that there was indeed a concave curvilinear relationship between the number of bullet points and the stock price. In order to see if this relationship is stable and reliable, we do a new regression analysis using another variable assessing market performance. We use the variable *return* as dependent variable. Furthermore, we keep *nbrbullets* as independent variable and the same control variables. Surprisingly, the results obtained are a convex curvilinear relationship as we can see in Table 6a and 6b as with Graph 4:

Graph 2: Curvilinear Relationship between Number of Bullet Points and Return



The difference of relationship between the regressions is probably because stock price and return respond differently to specific information or variables in earnings press releases. There might be variations in the curvilinear shapes of the stock price which is a metric of the short-term market reaction, and the return, the longer-term performance measure. Whereas The market's valuation of a company at a given moment can be directly determined by looking at the stock price that is affected by market conditions, investor behavior, or corporate performance, the return is a measure of the stock's success over a given time. It takes into consideration price fluctuations and dividends, and displays the total profit or loss that investors have incurred. Past performance, expectations, and general market trends all have an impact on returns. We conclude thus that the variable *return* is not adapted to execute or robustness test.

Therefore, we change our approach and test the curvilinear relationship by evaluating the model on different subgroups of the data to make sure that the concavity holds throughout various environments. The subgroups are based on the category of variables. We use the same subgroups as in the testing of the variables influencing the prevalence of bullet points in earnings press releases. The results in Table 7 of this second test do confirm the concave curvilinear relationship between the stock price and the number of bullet points. However, the value of the coefficients varies in-between the different groups leaving an opportunity for further analysis to determine what is exactly the behavior of stock prices when they are impacted by bullet points.

4.2.2 Cross-validation technique

In order to assess if there is a concave curvilinear relationship between stock price and bullet points, we tested our models using the cross-validation method. This is a statistical technique used to evaluate the performance of a model by dividing the data into a number of folds, then training the model on a subset of the data, and finally validating it on the remaining data. This process is repeated many times to make sure that the model's performance is robust and does not depend on any particular subset of the data. For our analysis, we employ a 10 fold cross-validation which divides the data in ten subsets, trains the model on nine, and consequently validates it on the tenth subset. The results of both cross-validations can be found in Table 8.

We started by cross-validating the model without control variables showed poor results. The residual mean squared error (RMSE) has a value of 0.9569 and a the value of the adjusted R-squared of 0.0050 is also very low which shows that the model does not explain the data well as has high prediction error. Nonetheless, the model including control variables performed significantly better. This can be seen by the lower value of RMSE of 0.8324 and a much higher R-squared of 0.2477. These results imply that the relationship between bullet points and stock price is better explained when we take additional relevant variables into account. The results of the cross-validation therefore confirms the robustness of the curvilinear relationship. The decrease of the RMSE and increase of the adjusted R squared indicate that the more complex model is more accurate and provides a better fit to the data.

5 Limitations

First, whereas analyzing the impact of bullet points in financial disclosures as a new way to organize the information brings a brand new perspective to the literature, finding existing literature to explore this relationship was challenging since the bullet point format has not been investigated that much even in areas such as psychological or educational sciences.

Second, collecting the dataset was extremely challenging due to the HTML format, as explained earlier the latter is unfamiliar to companies using Compustat. Managers are unaware that creating and deleting bullet points can leave "fake" nodes behind in the HTML code. Our program cannot differentiate between real and fake bullet points, leading to dataset errors.

Third, regarding part 2 and 3 of our results, the **model's prediction accuracy of the data is low**, as can be seen by the R-squared values and residual standard errors. Therefore, it suggests that other factors may significantly influence stock prices and forecasting errors. This creates an opportunity for further research.

Fourth, we did **not test for potential multicollinearity among variables**, which could affect the stability of the coefficients.

Lastly, the use of a polynomial regression model, even though it captures non-linear relationships, may lead to **overfitting or unnecessary complex models**, as we can see with the residual standard errors of some of our models. Therefore, **other approaches** could be interesting to investigate. Some machine learning techniques such as the **Random Forest or Support Vector Machines (SVM)** could be adapted. Random Forest is a learning method that builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting. It's particularly adapted to capture complicated relationships as in our case. SVM is a supervised learning algorithm that we use to classify data and to do regression analyses. It classifies data by finding an optimal line or hyperplane that maximizes the distance between each class in an N-dimensional space. Because it can handle both linear and nonlinear classification tasks, it is commonly employed in machine learning.

6 Conclusion

In conclusion, this research contributes to the theory of processing fluency in business disclosures by analyzing the effectiveness of a specific heuristic shortcut which improves structural readability: the bullet point format. Our paper aims to bring a contribution that is constituted of three main aspects:

1. **Structural Readability**: Whereas existing research focuses on the linguistic aspect of readability, there is a gap in understanding the structural aspect. We analyze the impact of bullet points on the processing fluency of financial disclosures, and therefore bring forward the importance of presenting and organizing information to improve comprehension.
2. **Heuristic Analysis**: Even though lots of research has been done on heuristics such as the original location of the company or the brand popularity, the literature on the specific heuristic of bullet points is still very small. Our study fills this gap by evaluating how bullet points is an efficient way of presenting complex financial information and, therefore make the investors' decision-making processes easier.
3. **Regulatory Implications**: As regulators desire to improve the fluency of financial disclosures by creating new regulations, our findings can offer valuable ideas. We demonstrated the positive impact of bullet points on market efficiency and investor decision-making, which could lead to the standardizing use of bullet points in earnings press releases.

Finally, our paper highlights how essential it is to take into account the language and the structural readability of financial disclosures on the basis of processing fluency theories. Our empirical results show that as was predicted by the literature, companies tend to use bullet points to present themselves positively. We also demonstrate that bullet points have informational value to predict financial markets since there is a significant concave relationship between the number of bullets and stock prices. Wherefore, this paper furthers the research of how the presentation of information influences investor behavior and the market responses by showing that bullet points are an effective heuristic tool.

Bibliography

Alter, A. L., & Oppenheimer, D. M. (2006). Predicting short-term stock fluctuations by using processing fluency. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(24), 9369–9372. https://doi.org/10.1073/pnas.0601071103

Arslan-Ayaydin, Ö., Boudt, K., & Thewissen, J. (2016). Managers set the tone: Equity incentives and the tone of earnings press releases. *Journal of Banking & Finance*, *72*, S132–S147. https://doi.org/10.1016/j.jbankfin.2015.10.007

Barker, D. (2010, January 3). *The psychology of the bullet point*. Online https://deanebarker.net/tech/blog/psychology-of-the-bullet-point/, consulted on 14 April 2024

Boudt, K., & Thewissen, J. (2018). Jockeying for position in CEO letters: impression management and sentiment Analytics. *Financial Management*, *48*(1), 77–115. https://doi.org/10.1111/fima.12219

Brock, S. E., & Joglekar, Y. (2011). Empowering PowerPoint: Slides and teaching Effectiveness. *Interdisciplinary Journal of Information, Knowledge, and Management*, *6*, 085–094. https://doi.org/10.28945/1366

Buchko, A. A., Buchko, K. J., & Meyer, J. M. (2012). Is there power in PowerPoint? A field test of the efficacy of PowerPoint on memory and recall of religious sermons. *Computers in Human Behavior*, *28*(2), 688–695. https://doi.org/10.1016/j.chb.2011.11.016

Clark, J. (2008). Powerpoint and Pedagogy: Maintaining student interest in university lectures. *College Teaching*, *56*(1), 39–44. https://doi.org/10.3200/ctch.56.1.39-46

Ding, X., & Liu, J. (2012). Research on Five Stakeholders & Five Relationships of Higher Engineering Education in China. *International Journal of Education and Management Engineering*, *2*(9), 61–65. https://doi.org/10.5815/ijeme.2012.09.10

Engelberg, J. (2008). Costly Information Processing: Evidence from Earnings Announcements. *AFA 2009 San Fransisco Meetings Paper*. https://doi.org/10.2139/ssrn.1107998

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221–233. https://doi.org/10.1037/h0057532

Frölich, M. (2008). Parametric and nonparametric regression in the presence of endogenous control variables. *International Statistical Review*, *76*(2), 214–227. https://doi.org/10.1111/j.1751-5823.2008.00045.x

Goldstein, I., & Yang, L. (2017). Information disclosure in financial markets. *Annual Review of Financial Economics*, *9*(1), 101–125. https://doi.org/10.1146/annurev-financial-110716-032355

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *the Journal of Business Communication/Journal of Business Communication*, *45*(4), 363–407. https://doi.org/10.1177/0021943608319388

Henry, E., Thewissen, J., & Torsin, W. (2021). International earnings announcements: tone, forward-looking statements, and informativeness. *European Accounting Review*, *32*(2), 275–309. https://doi.org/10.1080/09638180.2021.1963795

Ho, A. (2023). Types and role of the bullet point system in reading experience. *AHFE International*. https://doi.org/10.54941/ahfe1003704

Ho, A. G. (2020). Practical application of bullet point systems for constructing an information hierarchy. In *Advances in intelligent systems and computing* (pp. 255–261). https://doi.org/10.1007/978-3-030-51626-0_32

Hodnett, K., & Hsieh, H. (2012). Capital Market Theories: Market efficiency versus Investor Prospects. *International Business & Economic Research Journal (IBER)*, *11*(8), 849. https://doi.org/10.19030/iber.v11i8.7163

Johnson, D. A., & Christensen, J. (2011). A comparison of Simplified-Visually Rich and Traditional presentation styles. *Teaching of Psychology*, *38*(4), 293–297. https://doi.org/10.1177/0098628311421333

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, *33*, 171–185. https://doi.org/10.1016/j.irfa.2014.02.006

Kozak, M., & Hartley, J. (2011). Writing the conclusions: how do bullet-points help? *Journal of Information Science*, *37*(2), 221–224. https://doi.org/10.1177/0165551511399333

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting & Economics*, *56*(1), 130–147. https://doi.org/10.1016/j.jacceco.2013.05.001

Ledin, P., & Machin, D. (2015). HOW LISTS, BULLET POINTS AND TABLES RECONTEXTUALIZE SOCIAL PRACTICE. *Critical Discourse Studies*, *12*(4), 463–481. https://doi.org/10.1080/17405904.2015.1039556

Lee, Y. (2012). The effect of quarterly report readability on information efficiency of stock prices*. *Contemporary Accounting Research*, *29*(4), 1137–1170. https://doi.org/10.1111/j.1911-3846.2011.01152.x

Li, E. X., & Ramesh, K. (2009). Market reaction surrounding the filing of periodic SEC reports. *The Accounting Review*, *84*(4), 1171–1208. https://doi.org/10.2308/accr.2009.84.4.1171

Loughran, T., & Mcdonald, B. (2014). Measuring readability in financial disclosures. *the Journal of Finance*, *69*(4), 1643–1671. https://doi.org/10.1111/jofi.12162

McKoon, G. (1977). Organization of information in text memory. *Journal of Verbal Learning and Verbal Behavior*, *16*(2), 247–260. https://doi.org/10.1016/s0022-5371(77)80050-9

Ostertagová, E. (2012). Modelling using Polynomial Regression. *Procedia Engineering*, *48*, 500–506. https://doi.org/10.1016/j.proeng.2012.09.545

readability. (2024). Online https://dictionary.cambridge.org/dictionary/english/readability, consulted on 29 April 2024

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*(4), 364–382. https://doi.org/10.1207/s15327957pspr0804_3

Song, H., & Schwarz, N. (2008). If It's Hard to Read, It's Hard to Do: Processing Fluency Affects Effort Prediction and Motivation. *Psychological Science, 19*(10), 986-988. https://doi.org/10.1111/j.1467-9280.2008.02189.x

Todd, A., Bowden, J., & Moshfeghi, Y. (2024). Text-based sentiment analysis in finance: Synthesising the existing literature and exploring future directions. *International Journal of Intelligent Systems in Accounting, Finance & Management*, *31*(1). https://doi.org/10.1002/isaf.1549

Tufte, E. & Genevieve Liang. (2003). PowerPoint is evil: power corrupts. PowerPoint corrupts absolutely. In *Wired News*. https://bioinfo.uib.es/~joemiro/TecBasAula/PPEvil.pdf

Wang, S. A. (2013). Financial disclosure and communication. In *Palgrave Macmillan US eBooks* (pp. 139–162). https://doi.org/10.1057/9781137351876_8

Zhu, Z., Huang, Q., & Liu, H. (2023). How heuristic cues impact crowdfunding performance: The moderating role of platform competition intensity and platform demand potential. *Journal of Business Research*, *160*, 113797. https://doi.org/10.1016/j.jbusres.2023.113797

Appendices

Table 1: Variable description

Variable	Variable description
ciknumeric	Central Index Key of the companies in the dataset without the triple 0 infront
date	Date at which the earnings press release (EPR) has been published
nbrbullets	The amount of bullet present in the EPR
year	Year in which the EPR was published
month	Month in which the EPR was published
atq	Total assets in the company per semester
ibq	Income before taxes and depreciation
EARNTA	Ratio Return On Assets (ROA)
sd	Standard deviation of ROA on the 5 past years
loss	Indicates whether the value of ROA is negative. Loss equals 1 if ROA has a negative value and 0 if it has a positive value
return	Return of the stock price on the past 12 months
risk_monthly	Indicates the standard deviation of the monthly stock return on the past 12 months.
stock_price	Represents the stock price when the EPR was published
loc	Location of the headquartes of the company
Consensus_mean	Indicates the average of analysts' consensus forecast of Earnings-Per-Share (EPS)
Actual	Actual value of the EPS realized by the company
STD_consensus	Standard deviation of the variable consensus_mean
errormean	Indicates the mean of the errors of the analysts' forecasts. If the actual value of the EPS is lower than the forecast the errormean has a negative value.
indfmt	Indicates the industry in which is the company
car1	Cumulative Abnormal Return for the company for the [day-1 ; day+1] trading day window
car5	Cumulative Abnormal Return for the company for the [day-5 ; day+5] trading day window
car3	Cumulative Abnormal Return for the company for the [day-3 ; day+5] trading day window
car180	Cumulative Abnormal Return for the company for the [day-1 ; day+180] trading day window
FOG	Indicates the FOG index of the EPR. Measure of readability
PositiveHenry	Amount of words indicating a positive sentiment according the Henry library
NegativeHenry	Amount of words indicating a negative sentiment according the Henry library
NumberOfWords	Indicates the amount of words of the EPR

Table 2: Summary Statistics

Table 2: Summary Statistics

	Variable	mean	sd	min	med	max
1	nbrbullets	5.350	2.630	2	5	10
2	atq	142,574.400	449,370.900	4.920	5,236.460	3,246,076
3	ibq	319.330	1,310.840	-31,764	33.410	22,236
4	EARNTA	0.010	0.030	-1.370	0.010	0.670
5	sd	0.010	0.020	0	0	1.400
6	return	0.090	0.060	0.010	0.080	1.140
7	risk_monthly	0.080	0.390	-0.980	0.050	14.230
8	stock_price	133.280	8,350.340	0.300	30.150	1,060,000
9	Consensus_mean	-6.350	769.720	-97,650	0.420	175.760
10	Actual	-6.470	794.310	-100,800	0.450	178
11	STD_consensus	0.600	52.250	0	0.030	6,627.490
12	errormean		18.800	-1,386.410	0	2,000
13	car1	0	0.070	-0.700	0	1.370
14	car3	0	0.080	-0.870	0	1.210
15	car5	0	0.090	-0.930	0	1.070
16	car180	-0.020	0.330	-3.140	-0.020	3.880
17	FOG	21.730	3.750	1.650	21.340	122.490
18	PositiveHenry	60.220	50.420	0	46	924
19	NegativeHenry	28.730	27.830	0	20	389
20	NumberOfWords	2,938.840	1,765.580	60	2,504.500	30,656
21	bulletsquare	35.560	32.190	4	25	100

Table 2: Average Number of Bullets per Year

Year	Nbrbullets
2004	161
2005	636
2006	831
2007	1,110
2008	1,508
2009	1,760
2010	1,919
2011	2,056
2012	1,929
2013	2,299
2014	2,529
2015	2,576
2016	2,532
2017	2,403
2018	2,469
2019	2,639
2020	2,564

Table 2: Average Number of Bullets by Country

Country	Nbrbullets
ARG	10
BMU	5.92
CAN	5.13
CHE	5.99
CHN	5.25
CYM	3.64
FRA	8.62
GBR	3.62
IRL	3.85
ISR	6.50
LUX	5.03
NLD	5.88
SGP	4.60
SWE	5.40
USA	5.35
ZAF	5

Table 3: Linear regressions with *Nbrbullets* as dependent variable

	Dependent variable: nbrbullets) (1)	(2)	(3)
atq)	0.022***		
	(0.001)		
ibq	0.00000		
	(0.00000)		
EARNTA	-0.174***		
	(0.043)		
sd)	-0.198***		
	(0.070)		
loss	0.012**		
	(0.005)		
return)		-0.050*	
		(0.030)	
risk_monthly		-0.009**	
		(0.004)	
car1			-0.123*
			(0.066)
car3			0.014
			(0.082)
car5			0.109*
			(0.061)
car180			0.001
			(0.007)
Constant	1.573***	1.759***	1.760***
	(0.008)	(0.003)	(0.002)
Observations	54,682	54,725	32,323
R2	0.015	0.0002	0.0002
Adjusted R2	0.015	0.0001	0.0001
Residual Std. Error	0.412 (df = 54676)	0.416 (df = 54722)	0.425 (df = 32318)
F Statistic	166.116*** (df = 5; 54676)	4.146** (df = 2; 54722)	1.993* (df = 4; 32318)
Note:			*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: nbrbullets) (1)	(2)	(3)
FOG)	-0.054***		
	(0.018)		
PositiveHenry)	0.060***		
	(0.005)		
NegativeHenry)	-0.011**		
	(0.005)		
NumberOfWords)	0.130***		
	(0.008)		
year)		29.083***	
		(1.110)	
month)		-0.018***	
		(0.004)	
Consensus_mean			0.0005
			(0.0003)
Actual			-0.0005
			(0.0003)
STD_consensus			-0.0003
			(0.001)
errormean			0.0003
			(0.0004)
Constant	0.710***	-219.477***	1.760***
	(0.061)	(8.447)	(0.002)
Observations	32,316	32,316	32,311
R2	0.051	0.022	0.0002
Adjusted R2	0.051	0.022	0.0001
Residual Std. Error	0.414 (df = 32311)	0.421 (df = 32313)	0.425 (df = 32306)
F Statistic	434.733*** (df = 4; 32311)	358.285*** (df = 2; 32313)	1.600 (df = 4; 32306)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4a: Linear regressions with *stock_price* as dependent variable (with and without control variables

	Dependent variable:	
	log(1 + stock_price)	
	(1)	(2)
log(1 + nbrbullets)	-0.124***	-0.258***
	(0.013)	(0.011)
log(1 + atq)		0.168***
		(0.002)
EARNTA		2.396***
		(0.157)
log(1 + sd)		1.256***
		(0.243)
loss		-0.123***
		(0.016)
log(1 + risk_monthly)		0.480***
		(0.014)
log(1 + year)		81.136***
		(2.433)
log(1 + month)		0.040***
		(0.008)
log(1 + FOG)		0.303***
		(0.039)
log(1 + PositiveHenry)		0.082***
		(0.011)
log(1 + NegativeHenry)		-0.114***
		(0.010)
log(1 + NumberOfWords)		-0.151***
		(0.017)
Constant	3.702***	-614.673***
	(0.023)	(18.470)
Observations	32,316	32,316
R2	0.003	0.225
Adjusted R2	0.003	0.225
Residual Std. Error	0.958 (df = 32314)	0.845 (df = 32303)
F Statistic	97.941*** (df = 1; 32314)	781.215*** (df = 12; 32303)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 4b: Polynomial regressions with *stock_price* as dependent variable (with and without control variable

	Dependent variable: log(1 + stock_price) (1)	(2)	(3)	(4)
poly(log(1 + nbrbullets), degree = 2)1	-9.480*** (0.957)			-18.706*** (0.868)
poly(log(1 + nbrbullets), degree = 2)2	-7.524*** (0.957)			-13.330*** (0.857)
poly(log(1 + nbrbullets), degree = 3)1		-9.480*** (0.957)		
poly(log(1 + nbrbullets), degree = 3)2		-7.524*** (0.957)		
poly(log(1 + nbrbullets), degree = 3)3		-1.094 (0.957)		
poly(log(1 + nbrbullets), degree = 4)1			-9.480*** (0.957)	
poly(log(1 + nbrbullets), degree = 4)2			-7.524*** (0.957)	
poly(log(1 + nbrbullets), degree = 4)3			-1.094 (0.957)	
poly(log(1 + nbrbullets), degree = 4)4			3.069*** (0.957)	
poly(log(1 + atq), degree = 2)1				75.446*** (1.088)
poly(log(1 + atq), degree = 2)2				-8.003*** (0.956)
poly(EARNTA, degree = 2)1				12.752*** (0.988)
poly(EARNTA, degree = 2)2				14.042*** (0.931)
poly(log(1 + sd), degree = 2)1				0.589 (0.980)
poly(log(1 + sd), degree = 2)2				-4.014*** (0.880)
poly(log(1 + year), degree = 2)1				27.671*** (0.925)
poly(log(1 + year), degree = 2)2				14.026*** (0.849)
poly(log(1 + risk_monthly), degree = 2)1				32.125*** (0.885)
poly(log(1 + risk_monthly), degree = 2)2				-3.673*** (0.854)
loss				-0.055*** (0.016)
poly(log(1 + month), degree = 2)1				4.103*** (0.846)
poly(log(1 + month), degree = 2)2				-2.644*** (0.841)
poly(log(1 + FOG), degree = 2)1				6.588*** (0.948)
poly(log(1 + FOG), degree = 2)2				-3.411*** (0.897)
poly(log(1 + PositiveHenry), degree = 2)1				11.990*** (1.422)
poly(log(1 + PositiveHenry), degree = 2)2				4.076*** (1.132)
poly(log(1 + NegativeHenry), degree = 2)1				-15.845*** (1.434)
poly(log(1 + NegativeHenry), degree = 2)2				-9.266*** (1.060)
poly(log(1 + NumberOfWords), degree = 2)1				-14.928*** (1.673)
poly(log(1 + NumberOfWords), degree = 2)2				-3.566*** (1.151)
Constant	3.484*** (0.005)	3.484*** (0.005)	3.484*** (0.005)	3.492*** (0.005)
Observations	32,316	32,316	32,316	32,316
R2	0.005	0.005	0.005	0.250
Adjusted R2	0.005	0.005	0.005	0.249
Residual Std. Error	0.957 (df = 32313)	0.957 (df = 32312)	0.957 (df = 32311)	0.831 (df = 32292)
F Statistic	79.968*** (df = 2; 32313)	53.748*** (df = 3; 32312)	42.894*** (df = 4; 32311)	466.776*** (df = 23; 32292)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5a: Linear Regressions with *errormean* as dependent variable (with and without control variable

	Dependent variable: errormean (1)	(2)
log(1 + nbrbullets)	-0.425*	-0.297
	(0.246)	(0.255)
log(1 + atq)		0.100*
		(0.054)
EARNTA		14.880***
		(3.491)
log(1 + sd)		25.205***
		(5.400)
loss		0.534
		(0.346)
log(1 + risk_monthly)		2.174***
		(0.310)
log(1 + year)		-127.453**
		(54.087)
log(1 + month)		0.127
		(0.187)
log(1 + FOG)		-0.063
		(0.867)
log(1 + PositiveHenry)		-0.779***
		(0.253)
log(1 + NegativeHenry)		0.444**
		(0.216)
log(1 + NumberOfWords)		0.502
		(0.388)
Constant	0.749*	966.584**
	(0.445)	(410.526)
Observations	32,311	32,311
R2	0.0001	0.003
Adjusted R2	0.0001	0.003
Residual Std. Error	18.799 (df = 32309)	18.774 (df = 32298)
F Statistic	2.990* (df = 1; 32309)	8.331*** (df = 12; 32298)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 5b: Polynomial Regressions with *errormean* as dependent variable (with and without control variable

	Dependent variable: errormean (1)	(2)
poly(log(1 + nbrbullets), degree = 2)1	-32.517*	-17.271
	(18.801)	(19.444)
poly(log(1 + nbrbullets), degree = 2)2	25.982	16.926
	(18.801)	(19.192)
poly(log(1 + atq), degree = 2)1		-1.546
		(24.365)
poly(log(1 + atq), degree = 2)2		-49.122**
		(21.409)
poly(EARNTA, degree = 2)1		69.925***
		(22.135)
poly(EARNTA, degree = 2)2		-28.604
		(20.843)
poly(log(1 + sd), degree = 2)1		170.001***
		(21.949)
poly(log(1 + sd), degree = 2)2		31.456
		(19.696)
loss		0.917**
		(0.358)
poly(log(1 + risk_monthly), degree = 2)1		163.560***
		(19.823)
poly(log(1 + risk_monthly), degree = 2)2		-436.209***
		(19.114)
poly(log(1 + year), degree = 2)1		-64.567***
		(20.709)
poly(log(1 + year), degree = 2)2		42.244**
		(19.009)
poly(log(1 + month), degree = 2)1		4.982
		(18.943)
poly(log(1 + month), degree = 2)2		9.105
		(18.827)
poly(log(1 + FOG), degree = 2)1		-5.832
		(21.226)
poly(log(1 + FOG), degree = 2)2		8.946
		(20.079)
poly(log(1 + PositiveHenry), degree = 2)1		-80.728**
		(31.849)
poly(log(1 + PositiveHenry), degree = 2)2		80.568***
		(25.341)
poly(log(1 + NegativeHenry), degree = 2)1		95.308***
		(32.106)
poly(log(1 + NegativeHenry), degree = 2)2		39.941*
		(23.735)
poly(log(1 + NumberOfWords), degree = 2)1		19.948
		(37.464)
poly(log(1 + NumberOfWords), degree = 2)2		-72.130***
		(25.782)
Constant	0.001	-0.136
	(0.105)	(0.117)
Observations	32,311	32,316
R2	0.0002	0.020
Adjusted R2	0.0001	0.019
Residual Std. Error	18.798 (df = 32308)	18.616 (df = 32292)
F Statistic	2.450* (df = 2; 32308)	28.510*** (df = 23; 32292)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 6a: Robustness test - Linear Regressions with *return* as dependent variable

	Dependent variable: return (1)	(2)
log(1 + nbrbullets)	-0.0001	0.004***
	(0.001)	(0.001)
log(1 + atq)		-0.004***
		(0.0002)
EARNTA		-0.043***
		(0.011)
log(1 + sd)		0.446***
		(0.017)
log(1 + risk_monthly)		-0.023***
		(0.001)
loss		0.035***
		(0.001)
log(1 + year)		-2.159***
		(0.170)
log(1 + month)		0.002***
		(0.001)
log(1 + FOG)		0.001
		(0.003)
log(1 + PositiveHenry)		-0.005***
		(0.001)
log(1 + NegativeHenry)		0.017***
		(0.001)
log(1 + NumberOfWords)		-0.009***
		(0.001)
Constant	0.094***	16.570***
	(0.002)	(1.287)
Observations	32,316	32,316
R2	0.00000	0.165
Adjusted R2	-0.00003	0.165
Residual Std. Error	0.064 (df = 32314)	0.059 (df = 32303)
F Statistic	0.030 (df = 1; 32314)	533.650*** (df = 12; 32303)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 6b: Robustness test – Polynomial Regressions with *return* as dependent variable

	Dependent variable: return (1)	(2)
poly(log(1 + nbrbullets), degree = 2)1	-0.011 (0.064)	0.217*** (0.055)
poly(log(1 + nbrbullets), degree = 2)2	0.206*** (0.064)	0.244*** (0.054)
poly(log(1 + atq), degree = 2)1		-1.590*** (0.069)
poly(log(1 + atq), degree = 2)2		1.599*** (0.061)
poly(EARNTA, degree = 2)1		0.113* (0.063)
poly(EARNTA, degree = 2)2		-0.404*** (0.059)
poly(log(1 + sd), degree = 2)1		1.036*** (0.062)
poly(log(1 + sd), degree = 2)2		-0.559*** (0.056)
loss		0.025*** (0.001)
poly(log(1 + risk_monthly), degree = 2)1		-1.481*** (0.056)
poly(log(1 + risk_monthly), degree = 2)2		4.178*** (0.054)
poly(log(1 + year), degree = 2)1		-0.418*** (0.059)
poly(log(1 + year), degree = 2)2		0.931*** (0.054)
poly(log(1 + month), degree = 2)1		0.368*** (0.054)
poly(log(1 + month), degree = 2)2		0.228*** (0.053)
poly(log(1 + FOG), degree = 2)1		0.219*** (0.060)
poly(log(1 + FOG), degree = 2)2		0.064 (0.057)
poly(log(1 + PositiveHenry), degree = 2)1		-1.080*** (0.090)
poly(log(1 + PositiveHenry), degree = 2)2		-0.548*** (0.072)
poly(log(1 + NegativeHenry), degree = 2)1		2.346*** (0.091)
poly(log(1 + NegativeHenry), degree = 2)2		0.677*** (0.067)
poly(log(1 + NumberOfWords), degree = 2)1		-0.584*** (0.106)
poly(log(1 + NumberOfWords), degree = 2)2		-0.122* (0.073)
Constant	0.094*** (0.0004)	0.090*** (0.0003)
Observations	32,316	32,316
R2	0.0003	0.333
Adjusted R2	0.0003	0.333
Residual Std. Error	0.064 (df = 32313)	0.053 (df = 32292)
F Statistic	5.146*** (df = 2; 32313)	700.963*** (df = 23; 32292)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 7: Robustness test – Regressions with *stock*_price as dependent variable and subgroups as independent variables

	Dependent variable:			
	log(1 + stock_price)			
	(1)	(2)	(3)	(4)
poly(log(1 + nbrbullets), degree = 2)1	-17.580*** (0.886)	-13.623*** (0.952)	-0.000 (0.000)	-13.750*** (0.977)
poly(log(1 + nbrbullets), degree = 2)2	-16.246*** (0.890)	-6.703*** (0.949)	0.000*** (0.000)	-9.753*** (0.965)
poly(log(1 + atq), degree = 2)1	58.943*** (1.082)			
poly(log(1 + atq), degree = 2)2	-19.951*** (0.993)			
poly(ibq, degree = 2)1	5.341*** (1.017)			
poly(ibq, degree = 2)2	4.667*** (0.964)			
poly(EARNTA, degree = 2)1	18.851*** (0.944)			
poly(EARNTA, degree = 2)2	16.339*** (0.956)			
poly(log(1 + sd), degree = 2)1	-2.096** (0.997)			
poly(log(1 + sd), degree = 2)2	-3.543*** (0.916)			
poly(log(1 + year), degree = 2)1		27.924*** (0.955)		
poly(log(1 + year), degree = 2)2		12.816*** (0.947)		
poly(log(1 + month), degree = 2)1		-4.194*** (0.943)		
poly(log(1 + month), degree = 2)2		2.683*** (0.944)		
poly(log(1 + return), degree = 2)1			-0.000*** (0.000)	
poly(log(1 + return), degree = 2)2			-0.000 (0.000)	
poly(log(1 + stock_price), degree = 2)1			172.448*** (0.000)	
poly(log(1 + stock_price), degree = 2)2			-0.000 (0.000)	
poly(log(1 + PositiveHenry), degree = 2)1				25.150*** (1.481)
poly(log(1 + PositiveHenry), degree = 2)2				3.819*** (1.204)
poly(log(1 + NegativeHenry), degree = 2)1				-3.697** (1.511)
poly(log(1 + NegativeHenry), degree = 2)2				-7.855*** (1.153)
poly(log(1 + NumberOfWords), degree = 2)1				-0.116 (1.739)
poly(log(1 + NumberOfWords), degree = 2)2				3.177** (1.295)
Constant	3.484*** (0.005)	3.484*** (0.005)	3.484*** (0.000)	3.484*** (0.005)
Observations	32,316	32,316	32,316	32,316
R2	0.171	0.038	1.000	0.022
Adjusted R2	0.170	0.037	1.000	0.022
Residual Std. Error	0.874 (df = 32305)	0.941 (df = 32309)	0.000 (df = 32309)	0.949 (df = 32307)
F Statistic	664.639*** (df = 10; 32305)	210.210*** (df = 6; 32309)	4,300,750,667,656,057,335,240,806,864,202.000*** (df = 6; 32309)	90.095*** (df = 8; 32307)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 8: Cross-validation test with 10 k-folds of the regressions using *stock_price* as dependent variable (with and without control variables)

```
Linear Regression

32316 samples
    1 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 29085, 29083, 29085, 29085, 29084, 29083, ...
Resampling results:

  RMSE       Rsquared     MAE
  0.9569013  0.005104927  0.7268009

Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
Linear Regression

32316 samples
   12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 29085, 29085, 29085, 29083, 29084, 29084, ...
Resampling results:

  RMSE       Rsquared   MAE
  0.8339153  0.2457568  0.6135783

Tuning parameter 'intercept' was held constant at a value of TRUE
```

*note: There is an error in the title of the results, these are not the cross-validations of the linear regressions with *stock_price* as dependent variable but these are the results of the cross-validation of the polynomial regressions.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve
Boulevard Emile Devreux 6, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique
www.uclouvain.be/lsm