Bachelor Thesis in Accounting and Financial Management
Stockholm School of Economics
Department of Accounting
Spring 2023

When Good News are Bad News

Finding Value Relevant Information in the Sentiment of Annual Reports

Abstract

The qualitative aspects of corporate disclosures are rarely considered by stakeholders, meaning that a vast amount of potentially value relevant information goes to "waste". We study the sentiment contained in corporate disclosures, investigating the value relevance of its qualitative aspects to determine the potential value for stakeholders in considering this information. This is done by calculating the net positive sentiment in Swedish large-cap firms' annual reports during the period 2008 to 2020 and relating this to future firm performance. Our results show that net positive sentiment in the annual reports is significantly negatively related to future firm performance. This indicates that the qualitative information in Swedish large-cap firms' annual reports is value relevant and could be used by stakeholders to further reduce information asymmetry to managers.

Keywords: sentiment, tone, information asymmetry, earnings management, finance-specific word lists, term weighting, firm performance, dictionary

Authors: Emil Erne (25045) and Elias Mårtensson (25070)
Counselor: Liwei Zhu

Table of contents

1. Introduction

The median word count of Swedish large-cap annual reports increased by 20 percent between 2008 and 2020 (see figure A2) and increases in length can also be seen in other types of corporate disclosure (Davis et al. 2012). The vast amount of information that companies publish through numerous corporate publications to convey information to its stakeholders is only getting longer. Apart from meeting regulatory requirements, these publications serve to reduce the information asymmetry between managers and stakeholders. The information contained in corporate disclosures can be divided into two general categories, quantitative accounting information and qualitative subjective narrative information. While the quantitative data is communicated widely and considered by most stakeholders, the less objective and transparent nature of the qualitative data makes it more difficult to use (Healy, Palepu 2001). As the qualitative aspects of corporate disclosures are not widely considered (Bartlett, Chandler 1997), vast amounts of qualitative data goes to "waste" despite its potential to include value relevant information for stakeholders. Moreover, research has shown that companies may use subjective narrative objectives to manage earnings, create a favorable impression of the company, and manipulate the perception of their performance (Dechow, Skinner 2000). As a result, the consideration of the qualitative aspects of corporate disclosure may help complement the quantitative data and further reduce information asymmetry between stakeholders and managers.

Information asymmetry is an important issue in financial reporting and decision making, and sentiment analysis could play an important role in addressing this. Ettredge et al. (2022) found that private firms with greater information asymmetry tend to be more aggressive in their tax planning, potentially leading to less accurate financial statements. Similarly, DeGeorge et al. (2016) note that differences in accounting standards and practices across countries can create information asymmetry and impede the comparability of financial statements. Subjective narrative in corporate disclosures is provided through explanatory statements, future outlooks and other clarifications. The sentiment information contained within these texts could possibly provide stakeholders with managements' intended and unintended outlook on the firm's situation. In order to minimize information asymmetry between managers and stakeholders, this study attempts to find value relevant information within the qualitative information conveyed in corporate disclosures by Swedish companies.

Sentiment analysis of corporate disclosures has become one of the most prominent methods in producing quantifiable measures of qualitative information. When conducting sentiment analysis on a body of text, usage of a word list based or machine learning based approach has to be decided upon. Both of these approaches are used in prior literature (see Li 2010a and Davis et al. 2012) and no clear standard exists as to a preferred approach, with both having their relative strengths and weaknesses. A word list based approach contributes to increasing replicability and avoidance of potential issues with inter-rater reliability (Henry, Leone 2016) and, as a result, this study employs the word list based approach.

When using a word list based approach to sentiment analysis, there are three key considerations to make. Firstly, the choice of word list is central in the capturing of text sentiment. While there exists a multitude of different word lists, there is yet to be an established standard choice used for analysis within the financial and accounting literature. However, the findings of Loughran and McDonald (2011) show the advantages of using a finance-specific word list rather than a general sentiment word list. Therefore, this paper mainly employs the finance-specific word list developed in that paper. Secondly is the application of a weighting scheme to the extracted sentiment. In line with Loughran and McDonald (2011), the weighting scheme applied in this study is a term frequency inverse document frequency weighting. As opposed to an equal weighting scheme, where each positive and negative term bears equal weighting for the calculation of a document's overall sentiment, the inverse document frequency takes other factors into account. These factors are the diminishing value of the sentiment carried by a word when its frequency increases within both a specific document as well as within the sample as a whole (Loughran, McDonald 2011). Lastly, different literature aims to capture different types of sentiment contained in text, such as risk (Li 2006) or positivity/negativity, with the latter being most common. To measure the positivity and/or negativity of texts, the measurements constructed are typically net positive sentiment such as Davis et al. (2012), or solely negative sentiment, such as Loughran and McDonald (2011).

Sentiment analysis in the financial context can be conducted on several forms of corporate disclosures. Prior research on sentiment within corporate disclosures and association with firm performance or stock market reaction have mainly focused on quarterly earnings press releases and the management discussion and analysis (MD&A) section of the 10-K[1]. Davis et al. (2012) establish two main results: (1) sentiment within quarterly earnings press releases is found to be predictive of future firm performance, and (2) unexpected levels of sentiment within quarterly earnings are found to be associated with positive abnormal returns. MD&A section sentiment analysis has further provided evidence of the relationship of its contents and future firm performance (Davis, Tama‑Sweet 2012), as well as excess stock market returns (Feldman et al. 2010). Loughran and Mcdonald (2011), using their own financial word lists, found relationships between sentiment in 10-Ks (and the MD&A section within it) and excess market returns and standardized unexpected earnings. We have instead chosen to conduct our sentiment analysis using another type of corporate disclosure, companies' annual reports. The annual report is a form of corporate disclosure that summarizes an entire business year to the companies' stakeholders. As such, it is considerably longer than other forms of corporate disclosures, meaning that it contains both more content (the sample of words within an observation is greater) as well as a larger variety of information to analyze. Therefore, we argue that annual reports could contain more uncensored and therefore potentially more value relevant sentiment information. This makes annual reports a type of corporate disclosure which forms an interesting base for conducting sentiment analysis.

Most commonly, previous research investigates the relationship between corporate communication sentiment and its association with two different areas, future firm performance and market reaction. Due to the nature of the annual report, with limited new quantitative information for the stock market to react on, we have chosen to focus on firm performance. Firm performance has previously been measured using different key metrics such as standard unexpected earnings (Loughran, McDonald 2011) and return on assets (Davis et al. 2012). We follow the general method of (Davis et al. 2012) and employ future return on assets as a proxy for future firm performance, as also used in other prior literature such as (Li 2010a, Price et al. 2012).

Limited prior research investigating value relevant information within corporate disclosures has been conducted in the Swedish domain and using annual reports. Due to domain differences as well as differences regarding the structure of different types of corporate disclosure, we argue that prior literature is not generalizable. As such, this paper intends to

[1] The 10-K is a comprehensive account of a company's financial performance required by the SEC.

investigate whether there is value relevant information contained in Swedish companies' corporate disclosures. This research question is investigated by analyzing the relationship between the net positive sentiment within Swedish companies' annual reports and future firm performance measured as return on assets in the fiscal year following the annual report. We examine the qualitative aspects of 331 annual reports from 33 companies listed on the Swedish OMXS30 index during 13 years between 2008 and 2020. The qualitative aspects are quantified through a textual analysis program we created which extracts and counts words with positive or negative sentiment based on the finance-specific word list created by Loughran and McDonald (2011). Extracted positive and negative sentiments are combined to create a net positive sentiment variable quantifying language usage in the annual reports. To investigate this relationship, this paper employs a similar model to Davis et al. (2012) for investigating the association between future firm performance and earnings press release sentiment.

Our results establish a significant relationship at the 5% level between annual report sentiment and the firm performance in the following year. This indicates that, in line with prior literature conducted on other types of corporate disclosure, annual reports also contain valuable relevant information. Moreover, the negative direction of the significant relationship indicates that there is a relationship between having positive sentiment in the annual report and a more negative future firm performance. These results are in line with prior sentiment analysis conducted on 10-Ks, a type of corporate disclosure with many similarities to the annual report (Loughran, McDonald 2011). However, compared to other literature considering other types of corporate disclosures, the directional relationship is opposite as their results indicate a positive relationship between corporate disclosure net positive sentiment and future firm performance (Davis et al. 2012). We consider how the differences in the structure of different types of corporate disclosures can explain these results.

Our paper contributes to the literature in two main ways. Firstly, we contribute to the limited sentiment analysis conducted on financial publications in the Swedish domain. The establishment of a significant relationship between net positive sentiment in the annual reports and firm performance indicates that there is sentiment value in the corporate disclosures of Swedish companies, which paves the way for future research within the area. Secondly, we contribute to the limited prior literature basing its sentiment analysis on the annual report and suggest more extensive use of this corporate disclosure in the future.

The remainder of this paper is organized as follows: Section 2 presents previous related literature. Section 3 outlines the methodology. Section 4 presents the findings, and section 5 discusses these findings. Section 6 provides the conclusions, contributions, limitations of this study as well as suggestions for further research.

2. Literature review

Research on sentiment analysis can be traced back to the 1940's with the political endeavor of measuring public opinion during World War II (Stagner 1940, Knutson 1945). The number of published papers on sentiment analysis was relatively low until the 2010's when the subject gained popularity, and topics moved beyond mostly product reviews to a wide range of topics including financial contexts (Mäntylä et al. 2018). Conducting sentiment analysis within financial contexts has evolved from labor intensive studies where words were manually classified according to sentiment (Abrahamson, Amir 1996) to data-driven methods with specific financial word lists created for the financial context. Apart from reducing manual labor, the computational comparison to specific word lists has made the transformation more objective. This has helped in increasing the accuracy of transforming the qualitative aspects of text to quantitative measures (Li 2010b). These aspects have likely contributed to the increased popularity of the area of research.

Sentiment association with both future firm performance and market reaction has been used to investigate if corporate communication has value relevant information content. Prior research has analyzed sentiment in corporate communication and found it to contain incremental information in annual reports, 10-Ks and narrative disclosures within these (Li 2006, Loughran, McDonald 2011, Feldman et al. 2010, Davis, Tama‐Sweet 2012), conference calls (Davis et al. 2015, Price et al. 2012), earnings press releases (Davis et al. 2012, Henry 2008) and news articles (Tetlock 2007, Tetlock et al. 2008).

One of the most influential papers within the domain of sentiment analysis on financial text was published by Loughran and McDonald (2011). Their paper laid the foundation for improved capturing of financial sentiment by developing their own finance-specific word lists. Sentiment studies on financial text prior to the development of finance-specific word lists often expressed limitations due to their use of general purpose word lists. Li (2010a) partially motivates the use of the machine learning approach rather than the word list based approach with the possibly low prediction power of a general word list on financial texts. Loughran and McDonald's (2011) use their finance-specific sentiment word list to examine the negative sentiment in relation to market reactions, standardized unexpected earnings (SUE), and sentiment differences between the 10-K as a whole and the MD&A section. Within a four day period [0;3], they find that the negative general purpose Harvard word list is not significantly related to excess returns, while their finance-specific negative word list has a significant negative relationship. When investigating SUE's relation to both the Harvard negative word list and their finance-specific word list, they find that both word lists are significantly positively related to SUE. They argue that the latter result and its relationship between an increase in negative word usage and a greater earnings surprise might be explained by managers' attempt to manage market expectations.

Davis et al. (2012) investigate net positive sentiment (positive less negative sentiment proportions) in quarterly earning press releases during a five year period between 1998 and 2003, and its association with future firm performance (return on assets in the four subsequent quarters). They find that net positive sentiment is positively related to future firm performance. In addition, the study acknowledges the possibility of managers utilizing opportunistic language to mislead stakeholders. Regardless, their results suggest that language is employed to communicate value relevant information to investors and stakeholders. This means that stakeholders are suggested to find conveyed sentiment in earnings press releases credible to some degree. Furthermore, Davis and Tama-Sweet (2012) investigate pessimistic language in the MD&A section of the 10-K (while controlling for sentiment in earnings press releases) in relation to future firm performance measured as the average return on assets for the four subsequent quarters. Their sample period extends from 1998 to 2003. They find that the degree of pessimistic sentiment can predict firm performance in the subsequent quarters indicating incremental information content within the MD&A section.

The relationship between corporate disclosure sentiment and subsequent stock market performance has been studied in different corporate communication contexts. These studies further indicate that the qualitative aspects of corporate disclosures contain valuable information. Davis et al. (2012) examined the unexpected net positive language within earnings press releases and the market reaction by investigating change in sentiment compared to the previous quarter. They find that increased unexpected sentiment values are associated with positive abnormal returns around a three day period [-1;1] where zero is the day the earnings press release is issued.

Further contributions to the literature on value relevance of information content within corporate communication relating to market return has been examined on both the MD&A section and conference calls. Feldman et al. (2010) examine whether sentiment within the MD&A section in quarterly reports is associated with excess market return by investigating change in sentiment with sample data between 1993 and 2007. Sentiment was found to be significantly associated with returns around the SEC filing date [-1;1] as well as drifts in excess returns, suggesting that market participants act due to the MD&A section containing incremental non-financial information. Price et al. (2012) investigate the positive scaled by negative sentiment within quarterly earnings conference calls and subsequent stock market reaction. The subsequent stock market reaction is measured in cumulative abnormal return (CAR) across a three day period [-1;1] where day 0 is the date of the conference call. Additionally, they consider a longer period to measure any relation between sentiment and post-earnings announcement drift [2;60]. They find that relative positivity in conference calls is significantly related to both the initial CAR period and the post-earnings announcement drift CAR period.

Considering risk sentiment is another way to investigate sentiment in corporate communications. Li (2006) examines sentiment related to risk or uncertainty in annual reports and its association with future earnings and stock returns with sample data between 1994 and 2005. The study indicates that an increase in risk sentiment in annual reports is linked to decreased future earnings. Furthermore, substantial negative stock market returns are related to a larger increase in annual report risk sentiment in comparison to firms only experiencing smaller increases in risk sentiment.

Further literature expands the scope of research beyond that based on sentiment within companies' own corporate communication. Tetlock (2007) investigates the relationship between pessimistic sentiment in the *Wall Street Journal* (*WSJ*) section "Abreast of the Market" and the stock market returns with sample data between 1984 and 1999. Higher media pessimism is found to predict downwards pressure on the stock market and high market volume. However, given the fluctuations in market prices during the trading day, they could not conclude that pessimistic media sentiment efficiently reflects negative information not previously incorporated in stock market prices. Tetlock et al. (2008) extend this study to include all content regarding companies in both *WSJ* and *Dow Jones News Service between* 1980 and 2004. They find that the market incorporates qualitative information in news stories into stock prices with a slight delay, indicating a relatively efficient stock market and providing more evidence for the findings of Tetlock (2007). Beyond stock market returns, they also investigate qualitative information in relation to individual firms' performance through earnings, where low firm earnings are found to be related to news stories with pessimistic sentiment. These findings further provide evidence that sentiment can contain value.

As evident in the research discussed so far, different research considers different types of sentiment. While some research investigates other types of sentiment, such as risk (Li 2006), most research is concerned with the positivity and/or negativity of texts. The most common measurements used for positivity and negativity in texts are net positive sentiment as used by Davis et al. (2012), and negative sentiment as used by Loughran and McDonald (2011).

2.1 Word list based approach versus machine learning approach

The two leading methods to assess sentiment within qualitative information is using a word list based approach or machine learning classifier algorithms. A word list based approach, often called "word frequency counts" or "bag of words", counts words of interest in a body of text based on a word list. The methodology behind this will be explained further in section 3 of this paper. The alternative machine learning approach usually comprises of the following general two steps: (1) Classifying a training dataset into different sentiment categories using a method which usually involves manual coding of the dataset, and (2) utilizing the manually classified training set as input to an algorithm which classifies the rest of the dataset into the categories. An example of a study utilizing machine learning is Li (2010a) conducted on the subsection forward-looking statement in 10-K MD&As. Here, the training set includes 30000 sentences that were manually classified as positive, negative, neutral or uncertain. The machine learning algorithm then classifies 13 million sentences into the sentiment categories. The author finds a significant positive relationship between sentiment and future earnings.

Henry and Leone (2016) describe two main disadvantages of using machine learning algorithms in comparison to a word frequency based methodology. Firstly, they argue that the manual classification of the training dataset poses problems regarding inter-rater reliability within and across studies. As an example of this, Das and Chen (2007) showcase the limited consensus between the people classifying sentences for the training dataset. This paper indicates that there was only 72 percent consensus between two persons in the classification of sentiment in sentences. Secondly, research replication is problematized since the machine learning classification produces varying classifications depending on the given training dataset. Additionally, the machine learning research design includes many classification choices made by the authors (or the people classifying sentences) and may therefore differ across studies and reduce objectivity. In contrast, a word frequency approach does not have the inter-rater reliability problem and, given that identical word lists, weighting methods and models are used, two different studies will obtain equal measures of sentiment on the same dataset. This drastically increases replicability.

2.2 Sentiment word lists

Research involving sentiment analysis often employs textual analysis software and their built in sentiment word lists in order to extract sentiment (see Davis et al. 2012 or Davis, Tama-Sweet 2012). The Diction and General Inquirer textual analysis softwares have been widely used in prior financial sentiment analysis despite not being developed for the financial context. The word lists in Diction were developed within the political communications context and has been used in a variety of contexts from politics (Ghazal Aswad 2019) to education (Kondracki et al. 2002) in addition to finance. General Inquirers' word lists was developed within the social psychology context and contains sentiment word lists from the Harvard IV dictionary and the Lasswell word list (Stone et al. 1966). Similarly to Diction, the General Inquirer word list has been applied in many different research contexts (Alpert et al. 2000). Both the Diction and General Inquirer word lists are used in many different settings and these word lists are considered general, not context specific. A general word list such as General Inquirer is most commonly used within sentiment analysis (Loughran, McDonald 2011).

Consensus regarding what specific word list to use has not been established within the topic of sentiment analysis in financial texts. However, context specific word lists have been shown to best capture actual sentiment within financial text. Loughran and McDonald (2011) developed financial word lists to better capture sentiment within financial text since previous literature analyzing financial text often used word lists which misclassified financial language. For example, they found that General Inquirer's negative word list used in Tetlock (2007) and Tetlock et al. (2008) misclassified almost 74 percent of negative words within a financial context. Misclassification means that a word is classified as negative in the general word list while not providing actual negative sentiment in a financial context. Examples of this include common financial words such as "cost", "tax", "liability" and "foreign", as well as industry-specific words such as "crude" in the oil industry, "cancer" in the medical field, and "tire" in the automobile industry. The noise created by the misclassification of words contributes to general word lists lacking the predictive power within the financial setting that finance-specific word lists have (Henry, Leone 2016). Further support for using a finance-specific word list when analyzing financial text was found in examining conference call sentiment (Price et al. 2012).

Despite this, papers written after the introduction of finance-specific word lists still often use general word lists in their research. The general word lists in these papers are primarily used to complement the use of a financial word list, with the financial word lists consistently providing greater predictive power (for example Davis, Tama-Sweet 2012).

2.3 Weighting schemes

Another distinction between prior sentiment research is in regards to what weighting scheme to use (Loughran, McDonald 2011, Davis et al. 2012). In order to analyze sentiment within a body of text, an important choice regarding the weighting scheme of extracted words must be made. There are mainly two weighting schemes, equal weighting and inverse document frequency (idf) weighting. The idea behind idf weighting, according to Loughran and McDonald (2011), is to acknowledge that simply counting words or calculating the proportion of the number of positive or negative words in a document does not accurately capture the sentiment value of each word. To capture this value more accurately, term weighting, contrary to the unweighted proportion, considers two additional factors. Firstly, the importance of a specific word within a document is addressed by accentuating the sentiment value of words that are common in the specific document. Secondly, the importance of a word within the entire sample of documents is considered, and reduces the contribution towards the overall sentiment value of a document word on what proportion of documents in the entire sample contains it. Loughran and McDonald (2011) suggest the usage of idf weighting in calculation of sentiment within financial disclosures. Furthermore, they argue that the raw count of a certain word, or equal weighting, is not the best measure of its information content, as some less frequently used words may carry greater individual sentiment weight than a very frequently occurring word.

Henry and Leone (2016) criticize the usage of idf weighting in calculation of sentiment where the relevance of each document is equal. Since term-frequency idf weighting is determined by the commonality of a word, the sentiment in a document is always dependent on the content of the other documents in the sample if idf weighting is applied. They argue that this renders the use of idf weighting questionable in sentiment analysis. Furthermore, they state that idf weightings impede the possibility of replication given the sample dependency. They instead argue for the use of equal weighting since it enhances transparency and replicability. Equal weighting counts each sentimental word equally towards the overall sentiment, regardless of number of occurrences and other aspects.

2.4 Research problem and hypothesis development

Based on the review of previous sentiment research, this study attempts to find value relevant information within the qualitative information conveyed in Swedish large cap companies corporate disclosure. We use annual reports for the type of corporate disclosure, despite prior research preferring other corporate disclosures, due to two reasons. Firstly, given the length of a typical annual report (see figure A2) and its purpose of summarizing an entire year of business as well as providing an outlook on the firm’s future, it includes a wide range of content that could potentially include value relevant information. Secondly, historical annual reports are publicly available whereas access to historical press releases, earnings calls and other company publications frequently requires access to paid platforms.

In order to quantify the value relevance of the qualitative aspects of the annual reports, this study employs a net positive sentiment measurement as used by Davis et al. (2012). Moreover, in order to investigate whether this information is value relevant, its relationship to future firm performance (as used by Loughran and McDonald (2011)) is considered. The research question is formalized by the following hypothesis:

Hypothesis: *There exists a relationship between net positive sentiment used in Swedish firms' annual reports and their future performance.*

Testing this hypothesis aims to reveal whether there is information content in the qualitative aspects of corporate disclosures. This helps investors and other stakeholders in determining whether or not the qualitative aspects of the corporate disclosures could be used to help make different kinds of decisions.

Furthermore, the hypothesis helps ensure our research contributes to the prior research conducted on this topic in two ways. Firstly, prior research regarding value relevant information of qualitative aspects of corporate disclosures within the Swedish domain is limited. Due to geographical differences, we argue that they would not be directly transferable to the Swedish domain without further research. The most prominent geographical differences that would affect the transferability include differences in culture and language use, as well as regulatory requirements. Secondly, there are limited prior studies conducted on value relevance of qualitative information specifically in annual reports. Moreover, we argue that results obtained by studies looking at the sentiment of other types of corporate disclosures are not necessarily transferable to our context since different types of corporate disclosure have considerable structural differences.

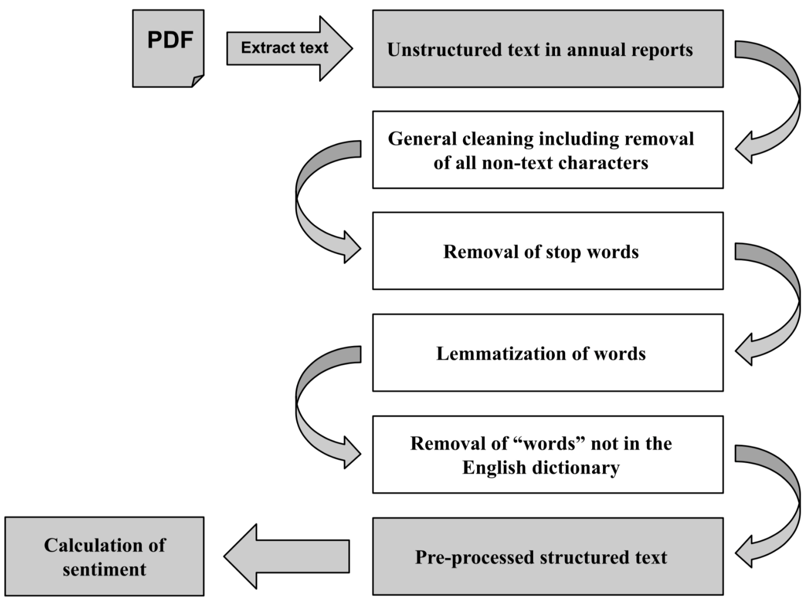
3. Method

3.1 Variables

The variables used to investigate the relationships between annual report sentiment and firm performance are based on prior research and literature on the topic. More specifically, we employ the same model specifications (with some deviations) as used by Davis et al. (2012) to find association with future firm performance. The variables used in constructing the model are described below:

The sentiment data is central to our research problem and included as the variable *SENTNP* in our model. This variable represents the net positive sentiment in the annual reports downloaded from the companies' web pages. In order to calculate the sentiment found in the annual reports, we first extract the text from the files to construct a complete list of the words contained in a file. For this we employ the R-package Tabulizer due to its ability to capture text in pdf-files of more complex designs, such as multiple columns. Next, the extracted text needs to be cleaned, a process outlined in Figure 1. Apart from general cleaning, such as removing non-alphabetic characters and punctuation, one of the key steps in this process is the removal of stopwords, such as "do" and "the", from the list of words. The stopwords to be removed are obtained from the R-package TidyText's list "stopwords". Another key step in the process is the lemmatization of words, which transforms all words into their standard form. As an example, the word "achieving" is replaced by the word "achieve". For the lemmatization, we use the TextStem package's "lemmatize_strings" function. The final major step in the cleaning process is ensuring that all the words in the list are actual words in the English language. This is done by matching all the words in the list to a master dictionary (a word list of all words in the English language) and removing any words that are not successfully matched. The master dictionary used by us is the master dictionary employed by Loughran and McDonald (2011) for the same purpose. It consists of 86531 words and is based on the 12of12inf dictionary with additions made for words occurring frequently in financial texts. This process results in a list of all pre-processed words for every sample annual report.

FIGURE 1
Text cleaning process



After conducting the text cleaning, the sentiment is calculated by matching the words in each report to a sentiment word list. As a base for our sentiment analysis, the Loughran & McDonald sentiment word lists, which classifies a total of 4150 words into six different sentiments: “positive”, "negative”, “litigious”, “constraining”, “superfluous” and “uncertainty”, is used. As we are concerned about the net positive sentiment in each report, we solely consider the “positive” (354 words) and “negative” (2355 words) sentiment categories and disregard the remaining sentiments. Next, a term frequency idf term weighting scheme is applied to all positively and negatively classified words using equation 1 as employed by Loughran and McDonald (2011). The first term of equation 1 weakens the impact of high frequency words with a log transformation. The second term of equation 1 modifies a word’s impact based on its commonality in the sample on the document level. Applying term weighting differentiates our models from Davis et al. (2011) but is consistent with other literature such as Loughran and Mcdonald (2011). As argued by Jurafsky and Martin (2009, p.12), term weighting “has an enormous impact on the effectiveness of a retrieval system”. The method for applying this term weighting equation to the unweighted sentiment proportions is briefly outlined below:

1. All the individual lists of positive and negative words in each annual report are combined and a table is created showing:
    a. All unique sentiment words in the entire sample.
    b. The number of occurrences of the specific word (in all the reports combined) in the sample.
    c. The number of reports in the sample in which the word occurs.
2. Based on the data in the table above, a fourth column is created and each word is assigned a weighting according to the weighting formula described (see equation 1).
3. The contribution of a word towards a document’s total positive and negative sentiments is calculated by multiplying the number of occurrences of the word in a specific document by its weighting factor.
4. The proportion of positive and negative sentiment in each annual report is calculated by summing the positive and negative contributions of all words respectively.
5. The proportion of negative sentiment is subtracted from the proportion of positive sentiment resulting in the variable *SENTNP*.

$$w_{i,j} = \begin{cases} \frac{(1+\log(tf_{i,j}))}{(1+\log(a_j))} \log \frac{N}{df_i} & if\ tf_{i,j} \geq 1 \\ 0 & otherwise \end{cases}$$

(1)

In equation 1, N is the total number of annual reports in the sample, df_i the number of annual reports containing at least one occurrence of the i^{th} word, $tf_{i,j}$ the raw count of the i^{th} word in the j^{th} annual report and a_j the word count in the j^{th} annual report.

The regressand in our model is future return on assets (*FUTROA*) which intends to measure future firm performance. We calculate future return on assets by dividing the net income of the following year by the average total assets in the following year. This is a slight deviation from Davis et al. (2012) resulting from the difference in the type of corporate disclosure considered. As they use quarterly press releases, future firm performance is calculated as the average return on assets in the following four quarters. While we could have considered the average of four future periods as well, we argue that 4 years is too long of a period for most potential valuable information in the annual report to be relevant. Instead, we choose to consider future performance in one future period (1 year) in the calculation of future firm performance, which is consistent with the period (4 quarters) that Davis et al. (2012) use.

In addition to *SENTNP* and *FUTROA*, we also include several control variables intended to control for the financial results of the firm and the market situation. The general argument for employing these variables is that the quantitative information relating to a firm's performance in the current period will likely have an effect on the firm's performance in the following period. The control variables are described below.

Our model includes the current or "unlagged" return on assets as *ROA* to account for current firm performance. Moreover, the standard deviation of the future return on assets (*FUTROASTDEV)* is included to account for the volatility in firm performance in the following year. This is a necessary deviation from Davis et al. (2012) due to the differences in calculating *FUTROA* resulting from the different types of corporate disclosures used. Davis et al. (2012) account for the volatility in an individual firm's future firm performance for each year (the performance in the 4 subsequent quarters). This is not possible for us due to the single component of future firm performance. Due to the significance in Davis et al. (2012)

measure of future firm performance volatility, we construct a different variable for this. Rather than accounting for intra-firm volatility, we account for inter-firm volatility by considering the standard deviation of the firm performance of all firms in a given year. This constitutes the *FUTROA* variable in our model.

We include two variables to account for earnings surprises and their effects on the firm performance. The first variable is the *SURP* variable which measures the actual surprise (actual net income less analyst consensus of net income) for the full year scaled by the market capitalization of the firm at the end of the period. The second variable considering earnings surprises is *BEAT*, a dummy variable equal to 1 if the firm beat the full year analyst consensus of net profit for the period (if SURP > 0) and 0 otherwise (if SURP ≤ 0).

Next, we control for any dividend increases during the year using *DIVINC*, where the variable is equal to 1 if there is a positive change in dividend and 0 otherwise. Furthermore, we control for any non-recurring income with the variable *NREPOS*, which is 1 if the net effect of the non-recurring item on period earnings is positive and 0 otherwise. Similarly, we control for non-recurring expenses with the variable *NRENEG*, which is 1 if the net effect of the non-recurring item on period earnings is negative and 0 otherwise. We also include the dummy variable *LOSS* which equals 1 if net income is negative in the period and 0 otherwise.

Another variable that we construct is *LOGREV*, which takes the logarithm of the company's revenue for a specific year. This measure accounts for differences in firm performance caused by the size of a given firm in a given year. As the revenue distribution of the sample observations is highly positively skewed (see Figure A3), we use the logarithm of revenue in the analysis.

Our model also includes a number of other measures relating to a firm's financial performance. This includes the individual components of return on assets: *PM* is profit margin (net income divided by revenue) and *AT* is asset turnover (revenue divided by average assets). In addition, *LEV* is firm leverage (debt divided by assets) and *B2M* is book-to-market value (assets divided by market cap).

We also include fixed effects for every year (*YEAR*) and individual firm (*FIRM*). The year fixed effects are typical in prior research (including Davis et al. (2012)) and account for changed levels of firm performance in the sample throughout the period. The firm fixed effect is an adaptation of the more frequently used industry fixed effects. Given the static nature of a large cap index, the total number of firms in our sample is relatively small. As a result, more than half of the 21 industries in our sample (classified using the first two digits of the SIC-code) only consist of a single firm. In effect, this means that the majority of our observations would use firm fixed effects even if we employed industry fixed effects. To avoid the inequality of some firms having firm fixed effects and some having industry fixed effects, and due to the reasonable number of variables that including firm fixed effects requires for our sample, we choose to diverge from Davis et al. (2012) in this regard.

Another divergence from the model used by Davis et al. (2012) is the exclusion of the *DET_FS* variable. The purpose of *DET_FS* is to account for the inclusion of detailed financial statements in the document. However, given that a detailed financial statement is always included in the annual report, this variable is not relevant for our model.

The final model consists of 56 independent variables, of which 42 variables account for firm and year fixed effects.

All accounting data is Institutional Brokers Estimate System (I/B/E/S) data extracted from Refinitiv Eikon.

3.2 Research design

Given our research problem and its similarity to Davis et al. (2012), the model we construct is a similar regression model based on the variables specified above. The model seeks to investigate if there is a relationship between the *SENTNP* and the *FUTROA* variables. It further controls for the other variables of interest as outlined in the previous section. The resulting model used to explain *FUTROA* is presented in equation 2.

$$FUTROA_i = \beta_0 + \beta_1 SENTNP_i + \beta_2 ROA_i + \beta_3 FUTROASTDEV_i + \beta_4 LOGREV_i + \beta_5 SURP_i + \beta_6 BEAT_i + \beta_7 LOSS_i + \beta_8 NREPOS_i + \beta_9 NRENEG_i + \beta_{10} DIVINC_i + \beta_{11} PM_i + \beta_{12} AT_i + \beta_{13} LEV_i + \beta_{14} B2M_i + \sum_i \beta_{15i} FIRM_{ij} + \sum_j \beta_{16j} YEAR_{ij} + \varepsilon_i \tag{2}$$

FUTROASTDEV is the standard deviation of *ROA* for all firms in a given year, $FIRM_{ij}$ is an indicator variable controlling company *i* in the j^{th} year, and $YEAR_{ij}$ is an indicator variable controlling for company *i*'s annual report released during the j^{th} year. Both *FIRM* and YEAR are therefore included to capture firm and year fixed effects. *ROA* accounts for previous firm performance while *FUTROASTDEV* and *LOGREV* control for the effects of size and risk on future firm performance. *SURP*, *BEAT*, *LOSS*, *DIVINC*, *NREPOS*, *NRENEG*, *PM*, *AT*, *LEV*, and *B2M* control for other quantitative information provided in the annual report. *SENTNP* is the sentiment measure which will indicate if the sentiment in the document is associated with future firm performance.

3.3 Sample selection

The domain included in our model is represented by the firms in the Swedish large cap index OMXS30[2]. The choice to limit our sample to Swedish large cap firms was made for a number of reasons. Given the regulatory requirements on large cap firms, these firms have both accounting data and historical annual reports readily available for download. This, together with the established nature of large-cap firms, ensures more comparability of the data as all firms included in the analysis must produce somewhat similar materials and report in accordance with similar regulatory requirements. The data on the composition of the OMXS30 index was received by Nasdaq upon request.

As for the time horizon, we use the period 2008-2020 due to three reasons. Firstly, the records received from Nasdaq on the composition of the index was only in an easily accessible electronic format starting from 2008. Secondly, the availability of digital historical annual reports diminishes prior to 2010. Combined, these two factors mean that any years prior to 2008 would consist of largely incomplete data. Thirdly, when the data analysis was conducted (March 2023), all companies had not yet released their financial results for 2022 which, due to the lagged nature of our dependent variable, would have been necessary to include 2021 in our sample.

Given the choice of OMXS30 firms and the time horizon of 2008-2020, the total number of firms included in our analysis is 33. Although the size of the index is 30 firms, some of the constituents have been replaced during the time period (we include firms that are included in

[2] OMXS30 is an index consisting of the 30 most traded Swedish stocks listed on the Nasdaq stock exchange.

the OMX30 index in the first half of the specific year). The reasons for a total number of firms close to 30 despite the prior point is that some firms are excluded due to multiple share classes, in which case we use the B-share and data missing for some firms.

Altogether, the resulting sample consists of 331 observations (firm-years) on which the regressions are conducted.

TABLE 1

Sample composition

Preliminary sample	n = 390
Exclusions due to multiple share classes	-13
Exclusions due to incomplete data	-46
Final sample	n = 331

Notes: *This table presents the preliminary sample, the reasons for exclusion from the preliminary sample, and the resulting final sample. One observation refers to a companys' annual report and corresponding financial and accounting information in one year.*

4. Results

4.1 Description of data

Table 2 presents descriptive statistics for all variables included in the model except the firm fixed effect variables *FIRM* and *YEAR*. In the sample, reported earnings met or exceeded analysts' earnings expectations 59,9 percent of times, indicating a slight conservatism in the analyst expectations. 4,2 percent of our observations had negative earnings. In 59 percent of observations, the dividends increased. The average profit margin was 22,4 percent, average asset turnover was 71,6 percent, average leverage 65,4 percent and average return on assets 5,8 percent. The average revenue of the observations was 10bn SEK, and in 91,3 percent of observations the net of non-recurring income and expenses was negative, whereas in 18,7 percent it was positive (meaning that all of our observations had some non-recurring income or expenses).

TABLE 2

Descriptive statistics

	Mean	Median	Maximum	Minimum	Std. dev.
SENTNP	-0,0003	-0,0003	0,0002	-0,0010	0,0002
ROA	0,0607	0,0581	0,4151	-0,2310	0,0600
LOGREV	10,795	10,878	11,734	8,5809	0,4718
SURP	-0,0337	0,0002	0,1231	-11,773	0,6463
BEAT	0,5994	1,0000	1,0000	0,0000	0,4908
LOSS	0,0422	0,0000	1,0000	0,0000	0,2013
NREPOS	0,1867	0,0000	1,0000	0,0000	0,3903
NRENEG	0,7982	1,0000	1,0000	0,0000	0,4020
DIVINC	0,5904	1,0000	1,0000	0,0000	0,4925
PM	0,2248	0,1019	82,3596	-88,0528	6,7803
AT	0,7145	0,7815	1,9798	0,0025	0,4721
LEV	0,6537	0,6277	1,5145	0,0172	0,2160
B2M	0,6120	0,4098	5,4118	-0,0998	0,6785

***Notes**: This table presents descriptive statistics for all financial and sentiment variables. SENTNP is the net positive sentiment measured as positive sentiment less negative sentiment. ROA is net income scaled by the average asset level during the year. FUTROA is the ROA in the subsequent year. FUTROASTDEV is the standard deviation of FUTROA for a given year. SURP is the difference between actual earnings and I/B/E/S analyst earnings consensus scaled with market capitalization at the end of the fiscal year. BEAT is equal to 1 if SURP ≥ 0 and 0 otherwise. LOSS is equal to 1 if earnings are negative and 0 otherwise. SENTNP is the net positive sentiment, calculated as the difference of weighted positive sentiment and weighted negative sentiment scaled by total words. LOGREV is the logarithm of revenue in a given year. PM is the profit margin measured as net profit scaled by revenue. NREPOS is equal to 1 if the net effect of non-recurring items on earnings is positive, and 0 otherwise. NRENEG is equal to 1 if the net effect of non-recurring items on earnings is negative, and 0 otherwise. AT is the asset turnover calculated as sales in a given year scaled by average total assets in a given year. LEV is firm leverage measured by debt scaled by assets (both measured at the end of the current fiscal year). B2M is book value of equity scaled by market value of equity (both measured at the end of the current fiscal year). DIVINC controls for dividend increases and is equal to 1 if there is a dividend increase, and 0 otherwise.*

Table 3 presents a correlation matrix for the accounting and sentiment variables. Many of the variables are shown to be significantly correlated, making a multivariate analysis for the hypothesis appropriate.

TABLE 3

Correlation statistics

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 *SENTNP*	1,00													
2 *FUTROA*	0,06	1,00												
3 *ROA*	0,07	0,64***	1,00											
4 *LOGREV*	-0,22***	-0,09	0,02	1,00										
5 *SURP*	-0,01	-0,12*	0,27***	0,25***	1,00									
6 *BEAT*	-0,04	0,01	0,08	0,13*	0,07	1,00								
7 *LOSS*	-0,06	-0,11*	-0,44***	-0,18**	-0,26***	-0,13*	1,00							
8 *NREPOS*	0,00	-0,01	-0,01	-0,17**	0,03	-0,07	0,01	1,00						
9 *NRENEG*	0,01	-0,02	-0,02	0,18**	-0,03	0,08	-0,01	-0,95***	1,00					
10 *DIVINC*	0,00	0,15**	0,19***	0,05	0,06	-0,04	-0,25***	0,05	-0,05	1,00				
11 *PM*	0,00	-0,02	0,35***	-0,05	0,72***	0,02	-0,21***	0,13*	-0,13*	-0,00	1,00			
12 *AT*	0,08	0,26***	0,33***	0,36***	0,08	0,09	-0,11*	-0,20***	0,18***	-0,04	-0,03	1,00		
13 *LEV*	-0,06	-0,12*	-0,08	0,16**	0,12*	0,00	-0,04	-0,12*	0,11*	-0,06	-0,04	-0,05	1,00	
14 *B2M*	-0,13*	-0,25***	-0,27***	-0,16**	-0,06	0,03	0,07	0,11*	-0,17**	-0,01	0,01	-0,26***	-0,24***	1,00

Notes**: This table presents pearson coefficients in a correlation matrix for the financial and sentiment variables. Variable definitions can be found in Table 2.".", "*", "**", and "" denote statistical significance at the 10 percent, 5 percent, 1 percent and 0,1 percent levels respectively.*

4.2 Hypothesis testing

Our hypothesis investigates the relationship between net positive annual report sentiment (*SENTNP*) and future firm performance (*FUTROA*). As a result, we are interested in whether *SENTNP* is significantly associated with *FUTROA*.

The multivariate regression result is presented in Table 4 (with firm fixed effects omitted for presentation purposes). As evident, there exists a negative and significant relationship at the 5 percent level between the *SENTNP* variable and the dependent *FUTROA* variable. As a result, our hypothesis is accepted. This suggests that there is a relationship between a higher net positive sentiment in the report, and lower firm performance in the following year (and vice-versa).

In addition, a number of the control variables included in our model are significantly related to the dependent *FUTROA*.

The variables with a significant and positive relationship to *FUTROA* are: *BEAT* is positively significant at the 5% level, indicating that there is a positive relationship between a firm beating the analyst consensus of its earnings and its future performance. *PM* is positively significant at the 10 percent level, indicating that there is a positive relationship between a firm's profit margin and its performance in the following year. The variables that are significantly and negatively related to future firm performance are: *SURP* is negatively significant at the 1 percent level, indicating that there is a negative relationship between the scaled earnings surprise compared to analyst consensus and future firm performance. *FUTROASTDEV* is negatively significant at the 5 percent level, indicating that there is a negative relationship between more volatility in future return on assets in a given year and a specific firm's future return on assets. *NRENEG* is negatively significant at the 10 percent level, indicating that there is a negative relationship between a firm's future performance and whether the effect of the non-recurring item was negative on period earnings. *B2M* is significant at the 10 percent level, indicating that there is a negative relationship between the book-to-market value of a firm and its future performance. The adjusted R^2 value of the model is 0.6591.

TABLE 4

Test of the association between net positive sentiment in annual reports and future firm performance

Variable	Coefficient	*t*-stat
SENTNP	-43,8263*	-2,31
ROA	0,0417	0,20
FUTROASTDEV	-14,3680*	-2,36
LOGREV	0,0507	1,27
SURP	-0,0351**	-3,31
BEAT	0,0119*	2,03
LOSS	0,0075	0,36
NREPOS	-0,0352	-1,52
NRENEG	-0,0476.	-1,92
DIVINC	0,0063	1,11
PM	0,0017.	1,75
AT	0,0094	0,33
LEV	0,0485	1,10
B2M	-0,0161.	-1,87
Intercept	0,6833	1,02
YEAR	INCLUDED	
FIRM	INCLUDED	
Adjusted R2	0,6591	
Sample size	331	

Notes**: This table presents regression results for equation 2, which is used to test the hypotheses on whether there is an association between future firm performance measured by return on assets in the following year and net positive sentiment. The regression includes FIRM and YEAR, which are firm and year fixed effect variables. Variable definitions can be found in Table 2. t-statistics are calculated using White (1980) heteroskedasticity robust standard errors. "." , "*", "**", and "" denote statistical significance at the 10 percent, 5 percent, 1 percent and 0,1 percent levels respectively.*

4.3 Summary of results

The results obtained means that we can establish a significant relationship between annual report sentiment and firm performance in the following year and that our hypothesis is accepted. The significant relationship is negative, signifying that there is a relationship between more net positive sentiment in the annual report and more negative future firm performance and vice versa. Moreover, the adjusted R^2 of our model is 0,6591.

5. Discussion of results

The negative coefficient of the net positive sentiment variable, *SENTNP,* is relatively high at approximately -43,8. However, given the low values of the variable (median value of -0,0003 as shown in table 1), this is considered reasonable. In addition, the adjusted R^2 of our model indicates a strong explanatory power of our model at 65,91% of the variance. This is slightly higher than the adjusted R^2 of 64,0% obtained by Davis et al. (2012) employing a similar model.

Our results indicate consistency with the results of Loughran and McDonald (2011). This is the most comparable prior research due to the similarity of the corporate disclosures considered (10-Ks and annual reports), the sentiment word lists used and the regressand used. They find a significant positive relationship between negative sentiment in 10-Ks and future firm performance, while our results indicate a significant negative relationship between the net positive sentiment and future firm performance in annual reports. This indicates directional consistency.

A possible explanation for the direction of our results is that managers' might attempt to manage expectations and over-compensate for bad financial performance in the qualitative aspects of its corporate disclosures by framing the numbers and the outlook more positively (Dechow, Skinner 2000). Loughran and McDonald (2011) argue that this might be done through a lower proportion of negative words, which helps explain their and our results where increased negative (in our case decreased net positive) sentiment is associated with better future firm performance. This perspective would also suggest that companies that present sound financial results do not necessarily concern themselves with bragging or overstating these. The reasons for this could be that they simply do not need to convince its stakeholders about their performance. However, if a firm is struggling and presents disappointing results, they may feel the need to express themselves more positively and/or less negatively. The argument of compensation as a defense mechanism in general is also supported by the works of influential psychologist Sigmund Freud[3].

Moreover, in line with Davis et al. (2012), we find a significant relationship between the net positive sentiment and future firm performance. However, their results indicate a positive relationship rather than a negative relationship as found in our study. Similarly, Davis and Tama-Sweet (2012) find a significant negative relationship between sentiment (pessimistic), and future firm performance. The significant relationship between sentiment and firm performance we establish is consistent with these findings in that managers use language to communicate future firm performance expectations in corporate disclosures. However, we differ from these studies in that our identified relationship is inverted. To explain this, it is important to consider the most differentiating factor between our research and prior research, the type of corporate disclosure considered.

The annual report differs from other corporate publications in its purpose and content. While an earnings press release often is a short text intended to convey a clear message, an annual report serves the purpose of summarizing an entire year of business for the stakeholders. Given the different corporate publications' purposes, the content also varies. Here, the annual report is interesting as it includes a large variety of content - everything from the CEO letter to the footnotes of the income statement is included.

Another key difference is the length between annual reports and earnings press releases. In our sample of annual reports, the median word count was approximately 44000 in 2020, while the median earnings press release length in Davis et al. (2012) was approximately 1700 in 2003. Despite the gap in years between the observations, this gives an indication of the differences in length. As earnings press releases are shorter, one explanation for the discrepancy in the direction of the relationships is that those texts may be more carefully

[3] Freud S. 1983. The Nuero-Psychosis of Defense.

crafted and looked over to avoid any corporate missteps. While the press release may be carefully overlooked to ensure that the message communicated is directly in line with the intended message, the annual reports, despite some parts surely experiencing the same treatment, may be less “censored”. In addition, the length of the annual report means that there is a much larger amount of information contained in each observation.

In combination, the systematic differences between different types of corporate disclosure may indicate why our results do not match those obtained by studies looking at other forms of corporate disclosure.

In conclusion, as prior research has established (see Davis et al. 2012 and Loughran, McDonald 2011), the significance of our results establishes that the qualitative information content within corporate disclosures contains value relevant information. The relationship found between language usage and future firm performance indicates managers’ communication of expectations in corporate disclosures. This suggests that the qualitative aspects of corporate disclosures could be used to reduce information asymmetry between companies and stakeholders. As such, stakeholders may benefit from complementing their analysis of companies with consideration of the language used in corporate disclosures.

5.1 Alternative methods

In line with previous literature, we also test our hypothesis using sentiment based on a general word list rather than a word list adapted to financial texts, such as the Loughran and McDonald word list used in our model. We have previously outlined the reasons for choosing to employ a financial word list in our main analysis. This includes that the adaptations allow for the sentiment captured to be more representative of the intended sentiment in financial texts (with “tax”, for example, not being considered a negative word). This is shown by prior literature finding stronger relationships between corporate disclosure sentiment and dependent variables such as future firm performance when the sentiment is calculated using a financial word list (see Loughran, McDonald 2011 and Price et al. 2012).

Despite the stronger predictive power of financial sentiment word lists found by previous studies, most prior literature uses general sentiment word lists (see Davis et al. 2012 or Tetlock 2007). As a result, in combination with the differences in domains, we have also conducted a secondary analysis using sentiment calculated based on a general word list. The main argument for the general word list potentially working better than the financial word list in the Swedish domain is the difference in official language. Although Sweden's English proficiency is considered to be very high[4], the language used in Swedish companies' annual reports could be less advanced and finance-specific than the language used in annual reports of companies based in countries with English as their official language.

We use the general positive and negative word lists in General Inquirer (employed by Tetlock (2007) and Tetlock et al. (2008)) which were developed in the social psychology setting and are based on the Harvard IV-4 word list (used as Loughran and McDonald's general word list). The results of running our model with a net positive sentiment variable based on the General Inquirer word list are presented in table A5. As the sign of the variable is consistent with our main results, these results are qualitatively consistent with the findings of our main analysis. However, the variable is not significant in the model. This is inconsistent with previous literature where general word lists also establish significant relationships between the corporate disclosure sentiment and future firm performance (see Loughran, McDonald 2011 and Price et al. 2012).

The potential explanations for this discrepancy include reasons similar to those given for the inverted direction between our results and results in prior studies based on other types of corporate disclosure. In addition, the inclusion of detailed financial data (including financial statements) in the annual report, in combination with the general word list's issues in classifying words that occur frequently in the presentation of financial data (such as "tax" and "liability"), may lead to significantly more noise being produced. As an example of the noise produced when using the general dictionary compared to the financial dictionary, the word "tax" appears an average of 214 times in our sample reports, increasing the count of negative words by 214 despite no actual sentiment value.

As for the model and its other variables, the results are very similar to our main results. It is worth noting that the explanatory power of this model is slightly reduced to 65,45%.

[4] EF English Proficiency Index 2022

Concludingly, we can establish that the argument of a general word list potentially working better than a financial word list in our domain due to language reasons, does not hold. The reason for substantial lack of significance can likely be attributed to the general word list producing more noise due to misclassifications of words. This means that the result from this model supports the claim in Henry and Leone (2016, p. 157) that "general word lists likely lack predictive power in a capital market setting". However, the general word list does provide results that are qualitatively consistent with the results provided by the finance-specific word list.

Another alternate consideration made was for the construction of a model considering solely the negative sentiment (as Loughran, McDonald 2011) rather than the net positive sentiment. However, due to an extremely high correlation of -0,96 (see table A6) between the negative and the net positive sentiments extracted from our sample, the construction of this model was considered redundant. This is likely due to the difference in length between the Loughran and McDonald (2011) positive (354) and negative (2355) word lists.

6. Conclusions

This paper has built on prior literature in the field of sentiment analysis in the financial context to investigate the value relevance of qualitative information in corporate disclosures in the Swedish domain. This research question has been investigated through the relationship between Swedish large-cap firms' net positive annual report sentiment and future firm performance as measured by the firms' return on assets in the following year. We examine the net positive sentiment of 331 annual reports from 33 companies listed on the Swedish OMXS30 index during 13 years between 2008 and 2020. By using Loughran and McDonald's (2011) finance-specific word lists and applying a term weighting scheme to account for the frequency of specific words, we have established a significant negative relationship between net positive annual report sentiment and future firm performance. This indicates an association between a company expressing itself more net positively in its annual report and lower future firm performance (and vice-versa). Moreover, this suggests that managers might adapt their usage of language depending on expectations regarding future firm performance. As a result, the sentiment in annual reports might help reduce information asymmetry between companies and stakeholders by indicating managers' expectations about the future. Finally, as similar previous studies have indicated within international domains, our results indicate that there is value relevant information in the qualitative aspects of Swedish annual reports.

6.1 Contributions

Establishing a significant relationship between the net positive sentiment in Swedish annual reports and future firm performance means that we make contributions to the literature.

There has been limited established sentiment analysis research conducted on the value relevance of the qualitative aspects of corporate publications in the Swedish domain. As a result, establishing a significant relationship between sentiment of Swedish firms' annual reports and future performance of the firms means that the tools used for identification of value relevant information in other domains are, at least to some extent, transferable to the Swedish domain.

Moreover, there has to our knowledge been a limited amount of established research on the value relevance of qualitative information in annual reports specifically. It is difficult to say which type of corporate disclosure contains the most value relevant qualitative information given the lack of comparable studies in the same domain but on a different type of corporate disclosure. However, our research highlights the potential to expand the scope of current literature by using a different type of corporate disclosure than previously used as the basis for sentiment analysis in the financial context.

6.2 Limitations

There are a few potential limitations to our research of which we are aware. These will now be addressed.

Firstly, while we have conducted "sanity checks" on the aggregated results as well as random checks on some individual reports, we rely extensively on the Tabulizer package in R for extracting the text from the annual reports correctly. As a result of the checks that we have conducted, we feel confident that most content is correctly extracted but acknowledge the fact that some annual reports use complex HTML-coding in certain parts that mean that text in some figures, for example, may be excluded.

Secondly, the method for extracting the sentiment data has been developed by us and has therefore not been tested in other settings. The reason for writing this code ourselves is that most prior literature uses software that is not free to access. Moreover, and likely as a result of the latter, very few papers include detailed instructions on how to actually extract the sentiment data. Again, we feel confident that our model works and have iterated the development carefully to arrive at our final version as well as conducted several random tests, but the fact that it is a self-developed model is worth noting.

Thirdly, our sample size is fairly limited as a result of the specific domain chosen (OMXS30), the limited data electronically kept by and received from Nasdaq, as well as the limited availability of electronic historical annual reports published prior to 2008. In addition, the manual labor required to extract annual reports due to the lack of a central database (such as EDGAR for SEC filing) was a limiting factor.

Lastly, there are some data gaps due to missing accounting data in the Eikon database. When a single piece of accounting information required for the model does not exist, the data point has to be omitted. This creates an “incomplete” sample in that some firms are not included for all the years that they should have been.

6.3 Future research

Based on our findings, there is a variety of future research that we would suggest to further develop our understanding of the value relevance of Swedish firms’ corporate publications’ qualitative aspects. By replacing the annual reports with other forms of corporate disclosures, including quarterly reports, press releases and investor calls would align the research more with prior literature in other domains and offer valuable insights into the Swedish domain as compared to the international domain.

Another alternative would be expanding the sample to extend further than only the OMX30 index would also be valuable in contrasting the different domains. This is due to prior literature generally not only considering firms in the specific country’s large cap index, but also looking at smaller firms. As previously mentioned in the discussion, the annual reports of these firms might be of a different nature with a resulting different type of sentiment.

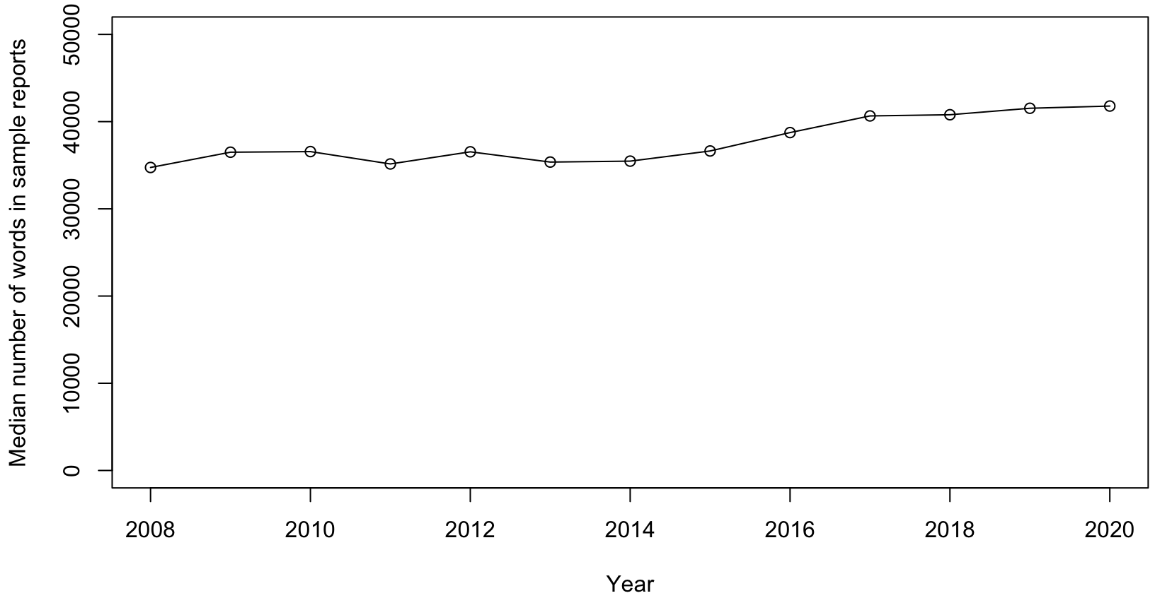
Although slightly more complex, it would be interesting to exchange the language of the annual reports in the sample from English to Swedish. This would enable investigation of the difference in sentiment due to the language barrier and might result in very different relationships being identified. The main challenge in conducting this research would be the development of a Swedish sentiment word list. In addition, as suggested by our research as well as previous research, this word list should preferably be finance-specific. While it would be impossible to completely replicate a specific English word list such as the Loughran and McDonald word list due to the different sentiments and connotations of directly translated words, this research could prove very interesting.

Furthermore, one could complement our quantitative results by conducting qualitative studies on the production of the annual report. This could offer insights into many of the issues and phenomena discussed in this paper, including the possible explanations given for the negativity of the relationship between net positive annual report sentiment and future firm performance. Some aspects that would be very interesting to investigate would be the process of constructing the annual report and, more specifically, who writes it, with what input, and based on what, and also to contrast this with stakeholder usage of the annual report.

In conclusion, our research has helped pave the way for future research within sentiment analysis in Sweden and we look forward to reading about future findings on the topic.

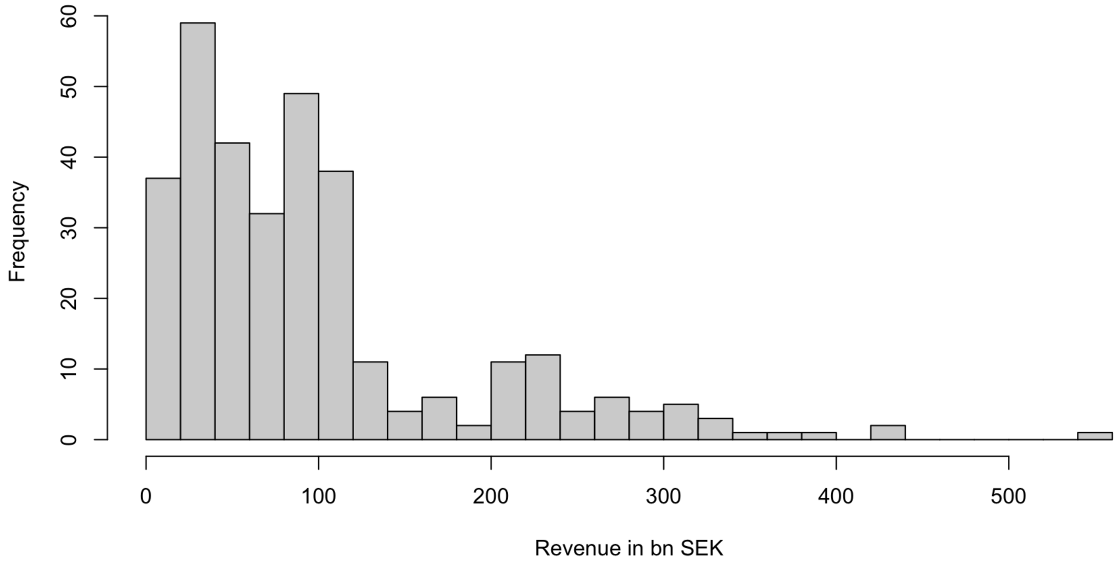
Appendix

FIGURE A2
Time series of median nr of words in sample reports by year



Notes: *This figure presents the median number of words in the annual reports in the sample for each year between 2008 and 2020.*

FIGURE A3
Distribution of revenue in observations



Notes: *This figure presents the frequency distribution of Revenue in bn SEK for each company year during the sample period between 2008 and 2020.*

TABLE A5

Regression results for net positive sentiment using General Inquirer word list for calculation of SENT variable. Regressand: FUTROA

Variable	Coefficient	*t*-stat
SENTNP_GI	-23,4265	-0,60
ROA	0,0499	0,24
FUTROASTDEV	-14,8120*	-2,50
LOGREV	0,0488.	1,22
SURP	-0,0353**	-3,31
BEAT	0,0118*	2,08
LOSS	0,0104	0,50
NREPOS	-0,0335.	-1,31
NRENEG	-0,0435	-1,64
DIVINC	0,0071	1,17
PM	0,0017.	1,76
AT	0,0139	0,49
LEV	0,0461	1,04
B2M	-0,0164*	-1,99
Intercept	0,7569	1,14
YEAR	INCLUDED	
FIRM	INCLUDED	
Adjusted R2	0,6545	
Sample size	331	

Notes: *This table presents regression results based on equation 2 with the deviation of a new sentiment variable. SENTNP_GI is the net positive sentiment measured as positive sentiment less negative sentiment using the General Inquirer word lists. ROA is net income scaled by the average asset level during the year. FUTROA is the ROA in the subsequent year. FUTROASTDEV is the standard deviation of FUTROA for a given year. SURP is the difference between actual earnings and I/B/E/S analyst earnings consensus scaled with market capitalization at the end of the fiscal year. BEAT is equal to 1 if SURP ≥ 0 and 0 otherwise. LOSS is equal to 1 if earnings are negative and 0 otherwise. SENTNP is the net positive sentiment, calculated as the difference of weighted positive sentiment and weighted negative sentiment scaled by total words. LOGREV is the logarithm of revenue in a given year. PM is the profit margin measured as net profit scaled by revenue. NREPOS is equal to 1 if the net effect of non-recurring items on earnings is positive, and 0 otherwise. NRENEG is equal to 1 if the net* *effect of non-recurring items on earnings is negative, and 0 otherwise. AT is the asset turnover calculated as sales in a given year scaled by average total assets in a given year. LEV is firm leverage measured by debt scaled by assets (both measured at the end of the current fiscal year). B2M is book value of equity scaled by market value of equity (both measured at the end of the current fiscal year). DIVINC controls for dividend increases and is equal to 1 if there is a dividend increase, and 0 otherwise. t-statistics are calculated using White (1980) heteroskedasticity robust standard errors. "." , "*", "**", and "***" denote statistical significance at the 10 percent, 5 percent, 1 percent and 0,1 percent levels respectively.*

TABLE A6
Pearson correlation statistics between net positive and negative sentiment variables

	1	2
1 *SENTNP*	1,00	
2 *SENTNEG*	-0,96***	1,00

Notes**: This table presents a correlation matrix for the net positive sentiment variable (calculated as positive less negative sentiment) SENTNP and the negative sentiment variable SENTNEG. ".", "*", "**", and "" denote statistical significance at the 10 percent, 5 percent and 1 percent levels respectively.*

References

ABRAHAMSON, E. and AMIR, E., 1996. The information content of the president's letter to shareholders. *Journal of Business Finance and Accounting,* **23**, pp. 1157-1182.

ALPERT, M., ROSENBERG, S.D., POUGET, E.R. and SHAW, R.J., 2000. Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry research,* **97**(2-3), pp. 107-118.

BARTLETT, S.A. and CHANDLER, R.A., 1997. The corporate report and the private shareholder: Lee and Tweedie twenty years on. *The British Accounting Review,* **29**(3), pp. 245-261.

DAS, S.R. and CHEN, M.Y., 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management science,* **53**(9), pp. 1375-1388.

DAVIS, A.K., GE, W., MATSUMOTO, D. and ZHANG, J.L., 2015. The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies,* **20**, pp. 639-673.

DAVIS, A.K., PIGER, J.M. and SEDOR, L.M., 2012. Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research,* **29**(3), pp. 845-868.

DAVIS, A.K. and TAMA‐SWEET, I., 2012. Managers’ use of language across alternative disclosure outlets: earnings press releases versus MD&A. *Contemporary Accounting Research,* **29**(3), pp. 804-837.

DE GEORGE, E.T., LI, X. and SHIVAKUMAR, L., 2016. A review of the IFRS adoption literature. *Review of Accounting Studies,* **21**, pp. 898-1004.

DECHOW, P.M. and SKINNER, D.J., 2000. Earnings management: Reconciling the views of accounting academics, practitioners, and regulators. *Accounting horizons,* **14**(2), pp. 235-250.

ETTREDGE, M., MAO, J. and STONE, M.S., 2022. Small Audit Firm De-registrations from the PCAOB-Regulated Audit Market: Strategic Considerations and Consequences. *Available at SSRN 3572291,* .

FELDMAN, R., GOVINDARAJ, S., LIVNAT, J. and SEGAL, B., 2010. Management’s tone change, post earnings announcement drift and accruals. *Review of Accounting Studies,* **15**, pp. 915-953.

GHAZAL ASWAD, N., 2019. Exploring charismatic leadership: A comparative analysis of the rhetoric of Hillary Clinton and Donald Trump in the 2016 presidential election. *Presidential studies quarterly,* **49**(1), pp. 56-74.

HEALY, P.M. and PALEPU, K.G., 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics,* **31**(1-3), pp. 405-440.
HENRY, E., 2008. Are investors influenced by how earnings press releases are written? *The Journal of Business Communication (1973),* **45**(4), pp. 363-407.
HENRY, E. and LEONE, A.J., 2016. Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review,* **91**(1), pp. 153-178.
JURAFSKY, D. and MARTIN J.H., 2009, Speech and Language Processing (Prentice Hall, Upper Saddle River, NJ).
KNUTSON, A.L., 1945. Japanese opinion surveys: the special need and the special difficulties. *Public opinion quarterly,* **9**(3), pp. 313-319.
KONDRACKI, N.L., WELLMAN, N.S. and AMUNDSON, D.R., 2002. Content analysis: Review of methods and their applications in nutrition education. *Journal of nutrition education and behavior,* **34**(4), pp. 224-230.
LI, F., 2010a. The information content of forward‐looking statements in corporate filings—A naïve Bayesian machine learning approach. *Journal of Accounting Research,* **48**(5), pp. 1049-1102.
LI, F., 2010b. Textual analysis of corporate disclosures: A survey of the literature. *Journal of accounting literature,* **29**(1), pp. 143-165.
LI, F., 2006. Do stock market investors understand the risk sentiment of corporate annual reports? *Available at SSRN 898181,* .
LOUGHRAN, T. and MCDONALD, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10‐Ks. *The Journal of finance,* **66**(1), pp. 35-65.
MÄNTYLÄ, M.V., GRAZIOTIN, D. and KUUTILA, M., 2018. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review,* **27**, pp. 16-32.
PRICE, S.M., DORAN, J.S., PETERSON, D.R. and BLISS, B.A., 2012. Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance,* **36**(4), pp. 992-1011.
STAGNER, R., 1940. The cross-out technique as a method in public opinion analysis. *The Journal of social psychology,* **11**(1), pp. 79-90.
STONE, P.J., DUNPHY, D.C. and SMITH, M.S., 1966. The general inquirer: A computer approach to content analysis.

TETLOCK, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance,* **62**(3), pp. 1139-1168.

TETLOCK, P.C., SAAR‐TSECHANSKY, M. and MACSKASSY, S., 2008. More than words: Quantifying language to measure firms' fundamentals. *The journal of finance,* **63**(3), pp. 1437-1467.

WHITE, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society,* , pp. 817-838.