What the CEO says:

The Incremental Information Content in the Language of CEO letters

Johan Sjölund[♣] Mark Söderberg[♦]

Master's thesis in Accounting and Financial Management

Stockholm School of Economics

Spring 2015

Abstract

We investigate if the language of CEO letters on the Stockholm Stock Exchange contains information about the future performance of the firm. We conduct a quantitative content analysis to classify 1,500 CEO letters and a total of 1.8 million words into scores based on the sentiment of each letter. Our main findings are that the sentiment in CEO letters contains information about future operating performance and future firm valuation. These results suggest that the CEO uses language in the CEO letters to provide incremental information beyond quantitative information in the annual report. This paper contributes to the literature on narrative disclosures by demonstrating empirical evidence of information content of CEO letters on the Stockholm Stock Exchange. Furthermore, we add to the limited literature on quantitative sentiment analysis of corporate disclosures in a non-US setting.

Keywords: CEO letter, narrative disclosure, incremental information, content analysis, DICTION

Tutor: Florian Eugster
Thesis Seminar: May 26, 2015

Acknowledgements

We thank our tutor Florian Eugster for valuable guidance. We also give our gratitude to Kenneth French, Tim Loughran and Bill McDonald for contributing to an open research environment. Finally, we thank friends and family for advice and support.

♣ 22352@student.hhs.se
♦ 22017@student.hhs.se

TABLE OF CONTENTS

1 INTRODUCTION

The role of accounting is to a large extent about reducing the information asymmetry between the managers on the inside of the firm and investors on the outside. A general understanding is that both the firm and the market benefit from a low information asymmetry environment (Verrechia, 2001). To aid in reducing the information gap, corporate disclosures exist (Healy and Palepu, 2001). A significant source of information is disclosed through the annual report which to a large extent consists of written narrative disclosures (Hutton, 2004). One of the most well-read written part of the annual report is the CEO letter (Courtis, 2004).

The CEO letter provides a direct link from the top management on the inside of the firm to investors on the outside (Abrahamson and Amir, 1996; Patelli and Pedrini, 2014). By nature, the narrative disclosures provide opportunities for managers to reveal their expectations about the future through the use of language (Davis et al., 2012). Furthermore, the use of language has a substantial influence on how information is processed, perceived and understood (Morris, 2005). Theoretically, the CEO letter could thus contain such information.

However, concerns of the information content in narrative disclosures, such as the CEO letter, have been raised and its usefulness has been questioned. Over time, a substantial increase in the length of disclosures has been reported, especially for the voluntary and narrative sections, and they have developed into exhibiting many public relations and promotional aspects (Beattie et al., 2004). Moreover, narrative disclosures are reported to remain similar between years despite the occurrence of significant firm events (Brown and Tucker, 2011). Finally, preparation costs and uncertain benefits of disclosing additional information might result in a limitation of the information content in the disclosures ((Epstein and Pava, 1995; Healy and Palupe, 2001).

Previous research has been focused on assessing the information content of narrative disclosures and has found it to contain information incremental beyond quantitative financial information. Moreover, development of quantitative sentiment analysis methods has allowed for a systematic approach of information content research on financial texts. However, a less amount of research using this method has been directed towards either the CEO letters or to a non-US setting. In a Swedish context, the study of the information of CEO letters is limited. Although exhibiting similar characteristics, the information content of narrative disclosures is contingent on their context (Patelli and Pedrini, 2014). Collectively, the information contents of narrative disclosures such as CEO letters remain an open empirical question in Sweden.

This thesis aims to add to the literature on the information content of CEO letter by answering the questions: Does the language of CEO letters in Sweden contain information about future firm performance? Do stock prices incorporate this potential information about the future or could investors use it to generate abnormal returns?

To answer this, two hypotheses are developed. First, we test for a positive association between the sentiment in the language of CEO letters and future firm performance. The sentiment is measured through quantitative content analysis and is tested against the future firm performance. Future firm performance is measured by operating measures of ROA, operating cash flow and net profit margin and future valuation measure Tobin's Q. The results support an association between the sentiment of the CEO letter and future firm performance. After testing for firm fixed effects and applying clustered standard errors, the results remain. These results suggest that the CEO letter contains incremental information beyond quantitative information in the annual report.

The second hypothesis tests for an association between the sentiment in CEO letters and future abnormal stock returns. The relationship is evaluated by constructing a trading strategy based on the sentiments of the CEO letters. The results indicate low support for an association with abnormal returns, which suggests that the incremental information found in the CEO letters are already incorporated in stock prices.

We conclude that the language in the CEO letters contains incremental information beyond quantitative information in the annual report. These are non-trivial results suggesting that the CEO uses language in the CEO letters to provide incremental information about the future. This interpretation is in line with the *expectations-adjustment hypothesis* in which managers provide information to adjust the investors' expectations about the future with their own expectations.

This thesis contributes to the existing literature in two ways. First, the results demonstrate empirical evidence of incremental information content of Swedish CEO letters. Second, we contribute by adding to the limited literature on quantitative sentiment analysis of corporate disclosures in a non-US setting.

The disposition of this paper is the following. In Section 2, the previous research and the development of our hypotheses are presented. Section 3 includes the method of extracting the sentiment in CEO letters and the data sample. Section 4 includes the methodology of the tests. Section 5 includes the results. In Section 6, the robustness and limitation of the results are discussed. Section 7 concludes.

2 PREVIOUS RESEARCH

2.1 Theoretical background

A starting point for understanding the potential information content of the CEO letter[1] is the information asymmetry between the managers inside the firm and investors on the outside (Healy and Palupe, 2001). This information asymmetry allows the firm's insiders to possess private information about the future prospects of the firm, based on theory of Akerlof (1970). A consensus view among financial economists is that a low information asymmetry environment is beneficial for all market actors (Verrechia, 2001).

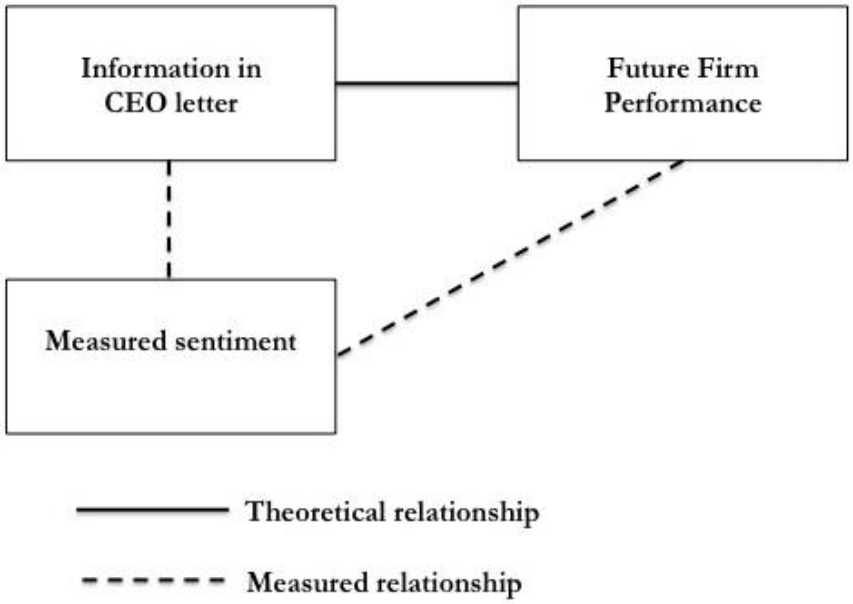
Ideally, the market would thus drive disclosures towards reducing the information asymmetry. This aim to reduce information asymmetry is also one of the main drivers for disclosure (Healy and Palupe, 2001). Further, a prerequisite for disclosures to actually contain incremental information is incentives for the preparers to disclose (Merkl, 2007). This topic is closely related to voluntary disclosure research (as comprehensibly reviewed by Healy and Palupe, 2001).

From an incremental information perspective, preparers have incentives to reduce information asymmetry by disclosing additional information (Merkl, 2007). These incentives exist both on firm-level, such as a decrease of the cost of capital (Kothari et al., 2009), and on individual level, such as enhancing managers' reputation and credibility (Merkl, 2007; Davis et al., 2012). These perceived benefits of a reduced information asymmetry motivates the *expectations-adjustment hypothesis,* under which managers have incentives to provide information as a way of aligning investors' expectations about future performance with their own expectations. (King et al., 1990).

When insiders prepare narrative disclosures, private information and expectations about the future could be reflected in the text. This collection of personally expressed information could be understood as the sentiment of the text (Li, 2010b). Measuring the sentiment would thus reflect private information of i.e. current conditions and future outlooks for firms (Li, 2010b; Kearney, 2014). The CEO letter is one type of narrative disclosure with possible sentimental content. To further understand this relationship, we present a conceptual relationship in figure 1.

[1] The CEO letter also goes by names such as: *CEO letter, A word From the CEO, Presidents letter, CEO comment, Review by the CEO*. In this thesis, these sections are all referred to as CEO letter.

Figure 1.
Conceptual relationship



The theoretical relationship between the information in CEO letters and future firm performance could conceptually be captured by the sentiment in the language of the text. By measuring the sentiment in relation to future firm performance conclusions of the information content of the CEO letter can be drawn. In order to further understand the links within this relationship, we present previous research on the information content of narrative disclosures and sentiment analysis.

2.2 Narrative disclosures research

Several studies evaluate the information content of narrative disclosures, such as the Management Discussion and Analysis (MD&A), officer's comment in earnings press releases and CEO letters (Kearney, 2014). An early study is Hoskin et al. (1986) which examine the incremental information in additional firm disclosures to earnings. The study finds that the officer's comment in earnings press releases provides incremental information in explaining future earnings. Further, Smith and Taffler (1996) examine whether the content in the discretionary narrative disclosure of the Chairman's statement is associated with the risk of bankruptcy. They find that the information content in the Chairman's statement provides information for predicting bankruptcy.

Another important study on the information content of narrative disclosures is Bryan (1997). In this study the association between the MD&A and future stock returns is evaluated by conducting a manual content analysis by classifying the MD&A into smaller parts and relating to future metrics

and returns. The results indicate that narrative disclosures provide incremental information in assessing short-term changes in key financial variables and in stock returns. Thus, qualitative information from management narrative disclosure can act as a complement to financial information.

A few but significant papers have studied the information content of CEO letter (or related topic). Abrahamson and Amir (1996) examine the information content of the President's letters in the US. In the study, they include performance measures such as change in net sales, ROE, and change in DPS. Further, by applying a quantitative content analysis, they classify the letters based on negative words. They find a negative association between the percentage of negative words in the texts and future firm performance and future stock returns. They argue that the president's letter is less regulated than the MD&A statement, which allows management to disclose potentially more useful information.

In a more recent study, Patelli and Pedrini (2014) study the CEO letter is manipulated to mislead interpretations and expectations of future performance or if it contains information useful to predict performance. The study applies a quantitative sentiment analysis and finds a positive association between the optimistic sentiment in the CEO letter and past and future firm performance. The results suggest that incentives to disclose untruthfully are limited under tougher macroeconomic conditions and the CEOs use the letters as a dialogue with shareholders.

In summary, the previous research has shown that qualitative information can provide incremental information to quantitative financial disclosures. Studies on narrative disclosures such as the MD&A and CEO letter have exhibited similar results. Second, prior research on this area has mainly been conducted in a US-setting.

2.3 Sentiment analysis research

Sentiment analysis is the method of understanding and identifying sentiment in texts. Originally, several studies apply a manual content analysis method in order to classify the sentiment in texts as neutral, positive or negative. For example, Hoskin et al. (1986) relied on manual classification by MBA students. However, the manual coding method limits the sample size, the possibilities for replication and includes significant subjective judgments (Davis et al., 2012). Subsequent developments of quantitative content analysis have taken place. Of the quantitative content analysis, two main approaches exist. First, statistical approaches classify sentiment in texts through, for example, machine learning methods (Antweiler and Frank, 2002 ; Li, 2010a). Second, dictionary-based approaches categorize words based on pre-defined rules from dictionaries

(Tetlock, 2007, 2008; Davis et al., 2012). In general, both approaches have been applied with similar results (Kearney, 2014).

Several studies have successfully been able to capture sentiment in financial texts using quantitative content analysis. Prominent studies include Antweiler and Frank (2002) and Tetlock (2007, 2008). Both these studies find a relationship between the sentiment in financial texts and a stock market reaction. This suggests the sentiment as being incremental to current information.

A popular topic of sentiment analysis in subsequent research has been narrative disclosures. Feldman et al. (2010) study MD&A and earnings press releases by applying a dictionary-based approach. They find that changes in the sentiment between the reporting periods are correlated with short-term returns. The results suggest that the MD&A have incremental information content.

In Li (2010a) an attempt to determine the determinants of sentiment is undertaken. The study statistically classifies the sentiment in MD&A into positive or negative and finds these to be a function of current performance, accruals, size, market-to-book, volatility of returns and age. The sentiment is furthermore associated with future firm performance. The paper concludes that the MD&A include forward-looking information in explaining future stock returns and that this is measured by the sentiment.

Davis et al. (2012) study the information value of the officer's in press releases and the market's ability to incorporate this information. The sentiment is measured as the net optimism of the disclosure, instead of only positive or negative classifications, intended to capture the overall writer sentiment. The findings indicate that the levels of the language in the press releases have a predictive value for future firm performance.

In summary, there are significant developments in the sentiment analysis research area. Several approaches for measuring sentiment have been developed and evaluated, although most methods have exhibited successful attempts to capture the sentiment in text.

2.4 Development of hypotheses

In relation to previous research, we position this paper in a Swedish context as limited research on narrative disclosures has been undertaken in this context. Additionally, we place this study as a contemporary attribution to the current questioning of the overall usefulness of narrative disclosures, specifically the CEO letter. To answer our research questions of assessing the incremental information in the language of the CEO letters in Sweden, we develop two hypotheses. In our first hypothesis, we investigate if the sentiment in the CEO letters is associated with future

firm performance. Second, we investigate if the sentiment in the CEO letter is associated with abnormal returns.

Hypothesis 1

A common understanding is that the CEO possess private information about a firm's future prospects, due to the information asymmetry between firm insiders and firm outsiders. The information asymmetry between the preparers and the users of financial information drives the need of disclosures, where the disclosures ideally have the role of reducing the information gap (Healy and Palupe, 2001). Under the *expectations-adjustment hypothesis,* managers have incentives to adjust the outsider's expectations about the future to that of their own knowledge (King et al., 1990; Davis et al., 2012).

As the CEO letter exhibits some special attributes of being expressed directly from the top and offers a direct link of communication to shareholders, the CEO letter is a potential medium for providing incremental information about the future.[2] Based on prior research, this information could be measured as the sentiment in the text. Thus expectations about the future could plausibly be measured through the sentiment of the text. Our first hypothesis is that the sentiment in CEO letters is positively associated with future firm performance.

Hypothesis 2

If a positive association is found in H1, this indicates that the CEO letters contain information about the future firm performance. As such information is acknowledged to be incorporated in stock prices, a second step is to assess if this information about the future firm performance in the CEO letters are already priced into the today's price or in future valuations. If the market incorporates the information captured in CEO letters immediately and accurately today, investors should not be able to earn an abnormal return from sentiment in CEO letters. However, if the information is incorporated to some extent in future valuations investors could earn abnormal returns if realizing the link between the sentiment in CEO letters and future firm performance. Consequently, whether sentiment in CEO letters is associated with abnormal returns remain an empirical question. Our second hypothesis is that the sentiment in CEO letters is associated with future abnormal returns.

[2] We are aware that the CEO might not write the letters themselves, but it can be assumed that the CEO take an active role in the editing and approval of the final text and is held responsible for its content (Jonäll, 2012).

3 CONTENT ANALYSIS & DATA

3.1 Content analysis

The methodology for extracting the sentiment in CEO letters is divided into the parts of collecting CEO letters from annual reports and performing a quantitative content analysis for measuring the overall sentiment in the texts.

As a first step, we collect annual reports from the Finansinspektionen (Swedish Financial Supervisory Authority) database for financial information, Thomson Analytics One and from the respective corporate website. Next, the part containing the CEO letters are localized and extracted for each annual report. Further, we collect information regarding CEO change between years and if the CEO signs the letter. We employ a systematic methodology regarding what to include from the letters, in order to allow for the correct semantic context and not to incorrectly include parts of letter not pertaining to the actual letter. For example, it is only the actual textual content that is analyzed, not pictures, tables or figures. A summary of the methodology for extracting the sentiment in the CEO letters is presented in appendix 9.1.

To measure the sentiments in the CEO letters, we use the established, textual-analysis computer software DICTION 7.0. This software allows for the analysis of a large sample of data in a systematic manner. Further, the systematic approach increases comparability to both prior and future research within the area of sentiment analysis. Next, we classify the words in the CEO letters into different categories and calculate scores based on the pre-defined dictionaries in DICTION. These dictionaries are based on the frequencies of words in contemporary American public discourse (Hart 2000).

More specifically, we measure the sentiment in the CEO letters by applying the DICTION variable *Optimism*.[3] The definition of this variable in our thesis is Net Optimism. Net Optimism is defined as *language endorsing some person, group, concept or event or highlighting their optimistic entailments*. The measure is based on counting the amount of optimistic and pessimistic words according to pre-determined dictionaries. The optimistic word count origins from the dictionaries *praise, satisfaction and inspiration,* and the pessimistic word count comes from the *blame, hardship and denial* dictionaries. A sample of words for each dictionary is presented in Table 1.

[3] In total, DICTION classifies words into 37 different categories and from these it calculates five master variables. Other master variables include Certainty, Activity, Realism, Commonality used by i.e. Patelli (2014).

Table 1.
Sample of words from DICTION Word Lists

	Optimistic dictionaries			Pessimistic dictionaries		
Dictionary	**Praise**	**Satisfaction**	**Inspiration**	**Blame**	**Hardship**	**Denial**
	best	thrilled	commitment	bad	disappointing	nonsense
	better	favorable	quality	undependable	unfortunately	weren't
	important	amazing	honesty	vulnerable	problem	nowhere
	reliable	prospering	excellence	mediocre	failure	without
	positive	encourage	faith	costly	obstacle	nor
	profitable	confident	trust	pessimistic	contempt	notwithstanding
	favorite	pleased	optimism	unstable	conflict	never
	superior	attracting	improvement	harmful	weakness	nothing
Tot words	224	330	151	331	518	39

From DICTION, we extract optimistic and pessimistic scores based on the percentage of words for each category in relation to the total words in the text. We analyze a total of 1,506 CEO letters containing 1,765,000 words. In detail, for each of the optimistic and pessimistic dictionary a raw score is calculated. The included dictionaries are those exhibited in Table 1. Next, we transform the measures to percentage of words out of total words per observation. As a final step, we calculate a Net Optimism score as the difference between the optimistic and pessimistic words.[4] We measure the sentiment as the Net Optimism in order to capture the net sentiment of the text as communicated by the managers (Davis et al., 2012).

The dictionaries are of a crucial role in capturing the sentiment in texts. Prior research suggests that different dictionaries can provide different results (Li, 2010b). Further, Loughran and McDonald (2011) and Kearney (2014) questions whether standard dictionaries are well applicable for analyzing financial texts. However, as the CEO letters are in an unstructured narrative format and allow for more open discourse, generalized dictionaries could be beneficial in capturing sentiment. This as the program has the ability to incorporate the semantic differences between homographs and the negation for words in the dictionaries (Hart, 2000). To test for the selection of dictionaries, we perform robustness tests by using the financial dictionaries, presented in Section 6.

3.2 Sample selection

We cover a total sample of 1,506 firm years listed on the Stockholm Stock Exchange (Nasdaq OMX) between 2004 and 2013. The sample includes companies listed on the Large Cap, Mid Cap and Small Cap on Nasdaq OMX. A raw sample of companies is derived from the Nasdaq OMX listings for each beginning of April. This is performed in order to incorporate companies listed at

[4] This thesis has for example, though we feel optimistic about our results, an optimistic score of 1.18%, a pessimistic score of 1.45% resulting in a Net Optimistic score of -0.27%..

each point in time in order to avoid a survivorship bias, due to the exclusion of delisted companies. The sample is presented in Table 2.

Table 2.
Sample selection

This table shows the sample selection per year in and per industry. The raw sample is selected from the OMX statistics from beginning of April for each year in sample. Companies listed with more than one share or for which no financial information was found are excluded. The 'No info' column on industries represents part of raw sample not linked to an industry as later excluded. AR stands for Annual Report. 'No AR' represents observations excluded when annual reports have not been able to be collected. 'No English Letter' represents observations excluded when collected annual report was not in English or if no CEO letter was included in annual report. Final sample reported as percentage of preliminary sample and per yearly listing on Large Cap, Mid Cap and Small Cap. The data set contains 82 observations that have broken financial years included in regressions, not in trading model.

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
Prel. Sample	270	275	279	265	265	256	251	253	256	252	2,622
No AR	90	94	87	51	48	38	36	29	22	22	517
No English AR	73	66	66	70	65	60	50	48	51	50	599
Final sample	**107 (40%)**	**115 (42%)**	**126 (45%)**	**144 (55%)**	**152 (59%)**	**158 (62%)**	**165 (65%)**	**176 (70%)**	**183 (71%)**	**180 (71%)**	**1,506 (58%)**
Large Cap	51 (77%)	54 (82%)	56 (84%)	61 (87%)	59 (89%)	53 (93%)	51 (91%)	51 (91%)	54 (92%)	56 (92%)	546 (88%)
Mid Cap	31 (41%)	32 (52%)	38 (53%)	44 (56%)	53 (65%)	56 (73%)	57 (77%)	61 (78%)	61 (80%)	54 (81%)	487 (68%)
Small Cap	25 (20%)	29 (20%)	32 (23%)	39 (33%)	40 (34%)	49 (40%)	57 (47%)	64 (54%)	68 (56%)	70 (56%)	473 (39%)

Industry*	Forestry	Cons-truction	Finance	Manu-facturing	Mining	Retail	Services	Trans-portation	Whole-sale	No info	Total
Prel. Sample	31	53	422	1,011	31	119	625	168	61	101	2,622
No AR	0	4	107	112	5	25	137	27	16	84	517
No English AR	14	0	110	214	1	21	169	36	17	17	599
Final sample	**13**	**49**	**209**	**685**	**25**	**73**	**319**	**105**	**28**	**0**	**1,506**
Large Cap	6	33	123	276	12	27	20	49	0	0	546
Mid Cap	4	7	64	175	9	33	150	17	28	0	487
Small Cap	3	9	22	234	4	13	149	39	0	0	473

*Long names on industries: Forestry=Agriculture, Forestry, Fishing; Finance=Finance, Insurance, Real Estate; Transportation=Transportation & Public Utilities; Wholesale=Wholesale trade.

From the raw sample, a preliminary sample is extracted by excluding observations that have more than one listed share. Further, two main exclusions are performed. First, if no annual report is available to be collected it is excluded from the sample. Second, if an annual report or CEO letter is not available in English it is excluded from the sample. For the assignment of Large Cap, Mid Cap or Small Cap for years before 2006, the firms are assigned to the listings of 2007 as not applicable lists before. Companies are assigned an industry classification based on the first two digits in SIC-codes. The SIC-classifications are taken from the SIC-website (www.siccode.com) and company specific SIC-codes are collected from Thomson DataStream, which results in a total of 9 industries in the sample. In the final sample, 82 observations have a broken financial year.

These are included in regressions on sentiment and future firm performance but excluded in the trading model.

From the data set, presented in Table 2, two conclusions can be drawn. First, the sample is skewed over time, affected by the decreased possibility to collect annual reports during earlier years. In addition to this, a higher amount of companies tend to report in English in the later time periods. We collect annual reports in English, which results in a data loss compared to if annual reports in Swedish were to be collected. However, English annual reports are necessary in order to employ a dictionary based content analysis approach. Second, the sample is not proportionally distributed among the Small Cap, Mid Cap and Large Cap lists. This is mainly an effect of companies on Small Cap and Mid Cap is less likely to disclose a CEO Letter in English. Additionally, there is a substantial larger data loss due to missing annual reports for Mid- and Small Cap companies compared to Large Cap.

Compared to studies conducting a quantitative sentiment analysis, the data set is substantially smaller than several other quantitative content analysis studies (i.e. Li, 2010a; Feldman et al., 2010; Davis et al., 2012). Compared to prior research on CEO letters, the data set is similar in size (Abrahamson and Amir, 1996; Patelli, 2014). Finally, the data set in this thesis differs from these studies by being based on a non-US data. The impact and effect on the inference is further discussed in Section 6.

3.3 Financial data

Financial data for the variables included in the regression models and the trading model are collected from DataStream. Firm age is extracted from the Orbis database. Risk factors used for the trading model are obtained from the Kenneth French's webpage[5]. Calculations and definitions of financial variables are defined continuously. The financial variables are collected at the 1st of April each year.

[5] Data obtained with courtesy from Kenneth French website: mba.tuck.dartmouth.edu/pages/faculty/ken.french

4 METHODOLOGY

In order to test for the information content in the measured sentiment of the CEO letters, we adopt a two-stage methodological approach. In the first stage, we test for the information content in the sentiment in CEO letters through two steps. In stage 1a, we investigate the determinants of the sentiment. In stage 1b, we test the hypothesis of the association between the sentiment in CEO letters and future firm performance. In stage 2, we test the second hypothesis of association with future abnormal returns. For this, we construct a trading strategy based on the Net Optimism score.

4.1 Stage 1a: Determinants of sentiment

We study the determinants of the Net Optimism measure with the purpose of understanding the determinants of sentiment and see to what extent this is captured by economical and contextual determinants. Influenced by methods on corporate governance research by Core et al. (1999) and Bowen et al. (2008) we study whether the Net Optimism is to a large extent explained by determinants already included in the sample. In this stage we expect to disentangle the future information in Net Optimism from current economical and contextual determinants. If the explanatory power is low, this would be an indication that the CEO letter could contain other information, such as forward-looking information. This step does not answer a stand-alone hypothesis but is a basis for further analysis. The regression results are presented in Table 8.

We run an OLS regression with Huber-White standard errors. The dependent variable is the Net Optimism measure. We include a baseline regression including independent variables based on Li (2010a). Further, this is regression is augmented by contextual variables for the CEO letters. All variables are measured with the timing of the available information at the point in time of the release of the financial report and market-to-book calculated from the 1st of April each year. Industry and year dummies are included to capture variations in industry and over time. Independent variables are the following:

Current firm performance – measured as the profitability measure return on assets (ROA) intended to capture current firm performance. ROA is defined as earnings before interest expense and taxes divided by the opening balance of assets. The current firm performance is expected to influence the Net Optimism positively.

Firm risk – defined as the standard deviation of ROA for three years prior to the publication of the annual report. This is included to explain changes in sentiment depending on volatility and risk in business. The expected sign is negative.

Size – measures by sales. The size of a firm is hypothesized to capture several aspects of business operations. The expected sign is unclear.

Accruals – defined as the amount of the net income that is not covered by cash flow from operations, scaled by assets. Accruals have been subject of research as a proxy of managers' possible manipulation or of a firm's economic condition (Sloan, 1996; Li, 2010a). In addition, a high amount of accruals today indicates a future decline in earnings, since accruals have to reverse over time. A negative relation between accruals and Net Optimism is expected.

Market-to-Book – there are differences between firms having a high or low market-to-book ratio (MTB) closely related to the growth of a company. A low MTB could indicate that the firm is a value firm. The expected sign is unclear.

Firm age – measured as the age of the firm since its foundation. Managers of young companies are likely to be more cautious when discussing the future, since young firms face more contingencies (Li, 2010a). A negative relation between firm age and Net Optimism is expected.

CEO change – dummy variable defined as 1 if a different CEO than the last writes the current CEO letter. A new CEO could write more optimistically or pessimistically than the last. The expected sign is unclear.

CEO signature – dummy variable defined as 1 if the CEO has signed the CEO letter or not. Provides an indication of the text plausibly being more sanctioned or actually written by the CEO. Nonetheless, the expected sign is unclear.

4.2 Stage 1b: Association with future firm performance

In the next stage, we test for the first hypothesis of the association between Net Optimism in CEO letters and future firm performance. We use four accounting metrics to test for the future firm performance, influenced by Core et al. (1999), Gompers et al. (2003) and Bowen et al. (2008). The first three metrics ROA, operating cash flow and net profit margin are measures of operating performance. The last metric of Tobin's Q is a measure of future firm valuation.

We specify four different regressions with two different horizons. For the operating metrics, we specify a baseline model following Bowen et al. (2008) including the independent variables of the previous year measure, risk measured as standard deviation of three years before and size measured

as sales. For Tobin's Q we follow Gompers et al. (2003) and include control variables of age and sales.

The horizon for the variables is set to one year into the future and average of one and two years into the future following (Core et al., 1999; Li, 2010a).[6] The first horizon is expected to capture for short-term effects and the second horizon is expected to be more stable.

We include several measures in order to capture different aspects of future firm performance. This is in line with research by Bryan (1997), Gompers et al. (2003) and Bowen et al. (2008) but differs from previous sentiment analysis research (Li, 2010a; Davis et al., 2012), which use a sole metric. The metrics included are the following:

ROA – ROA is included as basic measure of future operating performance, as used in Core et al., (1999) and Bowen et al., (2008). The measure is definition as in stage 1a. A choice of ROA would be to instead use ROE, as used in Gompers et al. (2003). However, these measures are found to be closely correlated with each other. ROA is kept as better access to data.

Future operating cash flow – we include the future operating cash flow as a measure of firm performance. This is defined as cash flow from operations scaled by sales. According to Bowen et al. (2008), operating cash flows have the advantage of avoiding any relations between current accruals and future earnings because of the reversing of accruals. However, they also have the disadvantage that the investment in positive NPV projects could result in negative cash flows.

Net profit margin – this measure is defined as net income divided by sales. Net profit margin is included in Gompers et al. (2003) and intended to capture future margins of the firm. It is based on net income, which is often used by financial analysts and exhibits a less correlation with ROA, allowing it to possibly be capturing other aspects of future performance.

Tobin's Q – is intended to capture the effects of the market valuation of the company. This measure is included as in Gompers et al., (2003). Tobin's Q is defined as the market value of assets divided by the book value of assets. The market value of assets is measured as the market value of equity plus the book value of liabilities excluding deferred taxes, divided by book value of assets. A higher value of Tobin's Q indicates a higher stock market valuation.

In our statistical tests, we start with a baseline OLS regression including dummy variables for industry and year effects as controls. This regression specification is in line with Bowen et al. (2008). In all regressions, we originally use of Huber-White standard errors. The usage of this specification

[6] Testing for a longer horizon not possible due to the recent data set.

of the standard errors is in line with our benchmark models of Bowen et al. (2008) and Gompers et al. (2003). These results are presented in Table 9 and 10.

As pointed out by Healy and Palepu (2001), many studies within the area of disclosure suffer from endogeneity problems, which can result in biased estimates. A potential cause of the problem is suffering from omitted variables. Two main problems in the finance literature are that the residuals might be correlated across firms for a given year (time effect) and/or across years for a given firm (firm effect) (Petersen, 2009). In our regressions, a firm effect could include constant unobserved firm characteristics such as the CEO writing style and the culture of the firm. A potential time effect could be the overall market conditions for a given year. In an attempt to address the issues, we test for fixed effects followed by clustered standard errors. As a first step, we include firm fixed effects, in our performance regressions, to address potential individual firm effects.[7] Second, we run the performance regressions with year dummies, to address the time effect, and clustered standard errors by firm to address a possible firm effect (Petersen, 2009). The results of these tests are presented in Table 11 and 12.

4.3 Stage 2: Association with future abnormal returns

In order to evaluate the association between sentiment in the CEO letters and abnormal returns, we form trading strategies based on the measured Net Optimism value for each financial year. We follow methodology employed in Gompers et al. (2003) and Eugster and Wagner (2013).

The stocks are divided into quintiles portfolios based on their Net Optimism score. For the portfolio containing the highest quintile of stocks, the *Top Portfolio* P8000, long positions are taken in the corresponding stocks. For the portfolio containing the lowest quintile of stocks, the *Bottom Portfolio* P0020, short positions are taken. Finally, we create a *Net Portfolio* P8020, where long positions are taken for the highest quintile stocks and short positions taken for the lowest quintile stocks. The return of the Net Portfolio, is calculated the difference in return between the long position and the short positions.

We form the trading portfolios yearly at the beginning of April after the financial-year end, when all annual reports are assumed to be available.[8] The created portfolios are held for a year until being reformed. We calculate continuously compounded monthly returns for all individual stocks in the sample based on the dividend and split adjusted return index (RI) from DataStream.

[7] We perform a Hausman-test supporting the use of fixed effects.
[8] For example, at the beginning of April 2014, portfolios are created based on the Net Optimism score for the financial-year of 2013, reported during 2014.

We evaluate the performance of our portfolios by measuring the abnormal return when controlling for risk factors proven in prior research to explain stock returns. There are several models for estimating abnormal returns based on different risk factors. We employ three different models for calculating abnormal returns; i) The CAPM model by Mossin (1966), ii) the three-factor model by Fama and French (1993) and iii) the extended four-factor model introduced by Carhart (1997). The full four-factor model being defined as:

$$R_{it} - R_{ft} = \alpha + \beta_1 RMRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 WML_t + e_t$$

Alpha (α) is the measure of abnormal return that could be earn in excess of a passive investing strategy in the risk factors. A statistically significant alpha of 5% has the interpretation that a monthly excess return of 5% could be earned by investing in the particular trading strategy. Further, Rit – Rft is the excess return of portfolio *i* over the risk-free rate, RMRF equals the premium return of the market over the risk-free rate, SMB equals the premium return of "Small Minus Big", HML equals the premium return of "High Minus Low", WML equals the premium return of "Winners Minus Losers". The CAPM model only consider the RMRF risk factor in explaining stock returns while Fama and French (1993) considers the additional SMB as a risk factor for size and HML as a risk factor for value. Carhart (1997) expands by adding the WML risk factor. For the data set we use European risk factors, calculated on Western European firms including Sweden.[9] On the full data sample, all three different models for calculating abnormal returns are presented. For further elaborations, we test and argue for the model of Carhart (1997) since the WML risk factor have been proven highly significant in explaining stock returns, i.e. past winners have been shown to outperform past losers (Jegadeesh and Titman, 1993).

In total, we use four different sample splits for the trading model. First, we include the full sample, a sample split excluding financial firms and a sample split between large and small firms. The exclusion of financial firms is performed by excluding companies with the "Finance, Insurance, Real Estate" 2-digit SIC code classification, resulting in a decrease of 205 firm years. Second, we construct a sample split between Large and Small firms to assess possible implications due to the data sample characteristics. We split the sample according to market capitalization on a yearly basis performs the sample split between large and small firms. The results for these sample splits are presented in table 14. As a test for robustness we also calculate value-weighted returns (VW).

[9] We considered using country-specific risk factors, since these have been proven better in explaining stock returns (e.g. Farma and French (1998) and Griffin (2002)). As not country-specific risk factors were available for the full holding period, we use European risk factors.

5 RESULTS

5.1 Sentiment in CEO letters

The sentiment in the CEO letters is measured as the Net Optimism between the optimistic and pessimistic words in the analyzed texts. For each observation sentiment scores are estimated by measures of optimism and pessimism. In Table 3, we present descriptive statistics for the full sample. The mean values show that there are on average 2.48% net optimistic words, 3.26% optimistic words and 0.78% pessimistic words in the sample. The average number of words per CEO letters in the sample is 1172 words.

Table 3.
Net Optimism descriptive statistics

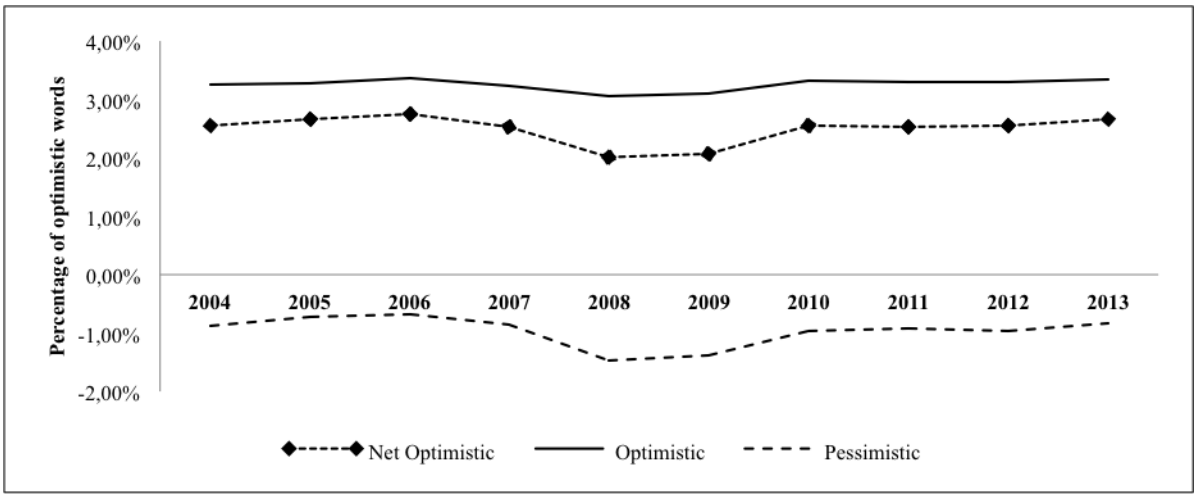
This table shows the total descriptive statistics of sentiment scores for the whole sample using DICTION. NETOPT=Net Optimism calculated as difference between OPT and NEG and where OPT=percentage of optimistic words and PES=percentage of pessimistic words. Reported number of observations, mean, standard deviation, median and quarters.

Measure	mean	sd	median	Q1	Q3
NETOPT	2.48%	1.13%	2.46%	1.78%	3.23%
OPT	3.26%	0.95%	3.19%	2.60%	3.86%
PES	0.78%	0.50%	0.69%	0.43%	1.01%
Number of words	1172	452	1118	876	1405

Figure 2 shows the mean percentage values for the Net Optimism measure, optimistic and pessimistic words for each year. We observe a decline in the Net Optimism for the financial years of 2008 and 2009, which is expected to be co-occurring with the financial crisis and a general decrease in the sentiment.

Figure 2.
Net Optimism per year

This figure represents the Net Optimism mean percentage scores for each year during the period 2004-2013. The y-axis indicates the percentage of optimistic words in texts. Pessimistic words are thus negative. The Net Optimism measure is calculated as the difference between Optimistic and Pessimistic words. Descriptive statistics per year in Table A.2.

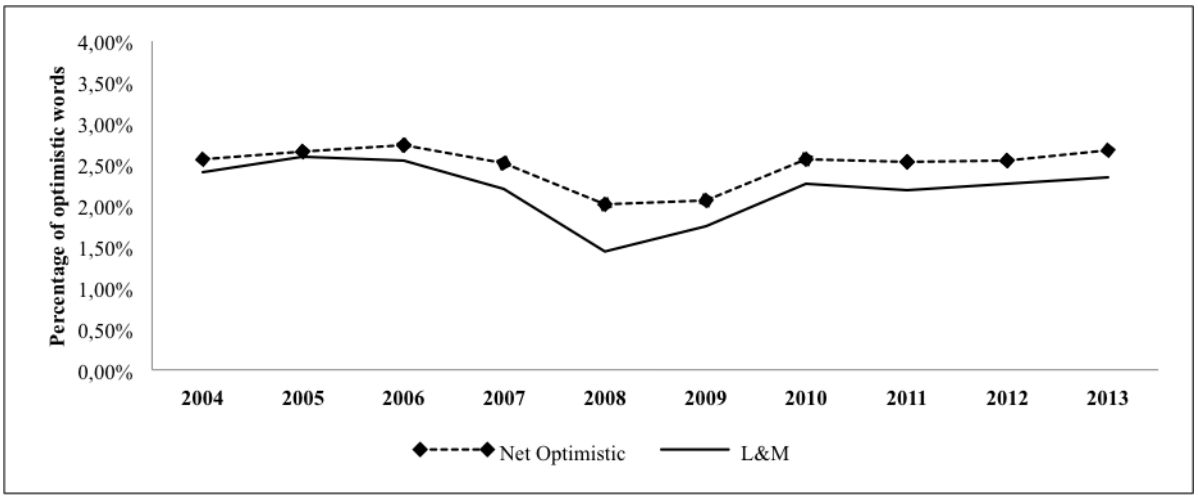


Furthermore, we see that the optimism is relatively stable, while the pessimism is decreasing during the financial crisis. This is in line with the methods and findings of (Li, 2010a; Tetlock, 2007; Feldman et al., 2010) that counting pessimistic words is a measure of the sentiment. Although Net Optimism, which is used in this thesis in line with (Davis et al., 2012), would, by this logic, aggregately capture the same sense as the sole measure of pessimistic words, as it is a direct function of the two.

In Figure 3, we demonstrate the two measures of Net Optimism, DICTION and Loughran and McDonald in relation to each other. We observe that the measures seems to follow each other relatively closely over time, although the Loughran and McDonald measure is consistently lower. The descriptive results for both measures are reported in Table A.2. An illustrative table of sentiment score per company and year is found in Table 3. Sentiment scores per company and year is displayed in Table A.4.

Figure 3.
Measures of Net Optimism over time

This figure represents the Net Optimism mean scores for each year during the period 2004-2013 for both measures of Net Optimism. The y-axis indicates the percentage of optimistic words in texts. The Net Optimism measure is calculated as the difference between Optimistic and Pessimistic Words. Descriptive statistics per year in Table A.2.



In Table 5, the correlation statistics between the measures are displayed. The correlation between the Net Optimism of the two measures is 0.69, suggesting that a similarity between two different measures exists. The strongest correlation of 0.90 is between the DICTION measure of Net Optimism and Optimism.

Table 5.

Correlation Statistics of Sentiment

This table shows the Pearson correlation matrix of textual variables from DICTION (_D) and Loughran and McDonald (_LM). NETOPT = Net Optimism, OPT = Optimistic, PES = Pessimistic words.

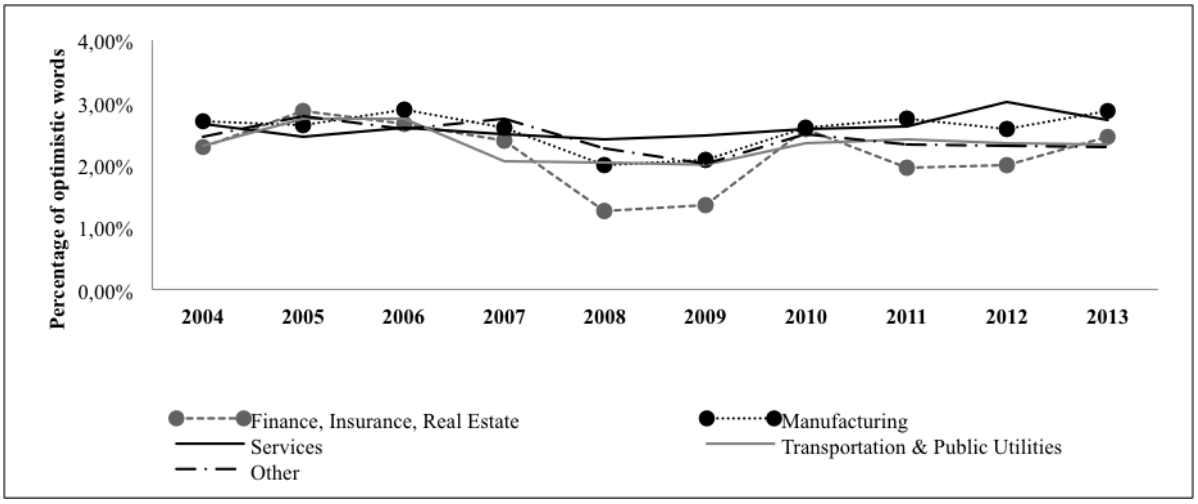
	1	2	3	4	5	6
NETOPT_D	1.00					
NETOPT_LM	0.69	1.00				
OPT_D	0.90	0.59	1.00			
OPT_LM	0.56	0.85	0.59	1.00		
PES_D	-0.56	-0.45	-0.15	-0.16	1.00	
PES_LM	-0.45	-0.61	-0.21	-0.11	0.62	1.00

Figure 4 demonstrates the Net Optimism scores for the five largest industries in our sample. Descriptive statistics for all industries for both measures are reported in Table A.3. From the figure, we observe that the Finance, Insurance and Real Estate industry experiences a significant decline in Net Optimism during the years of 2008 and 2009. For the other industries in the figure, we observe changes in the order during the years indicating changes in sentiment over time between the industries. Although, the relative values of the three largest industries excluding financials show levels between 10 and 15, which is below the mean/median value of 12.4/12.3.

Figure 4.

Net Optimism per industry and year

This figure represents the Net Optimism mean scores as measured by DICTION for the five largest industries in the sample (where the number of firm year observations are more than 100) during the period 2004-2013. The y-axis indicates the percentage of optimistic words.



In summary, the results from the sentiment extraction direct us in three directions. First, they show that the measure of sentiment is relatively similar between the two dictionaries providing support for the accuracy of the measurement. Second, there seems to be variation in Net Optimism over time, possibly coinciding with the overall sentiment of the market. Finally, there seems to be differences in Net Optimism between the industries in the sample, especially for the financial industry.

5.2 Descriptive statistics

We present descriptive statistics for the variables used in the regressions in Section 3. For the regressions on the determinants of sentiment, financial variables are included as well as other explanatory variables. For the regressions on future firm performance, dependent future firm performance variables are used in addition to a number of independent financial variables. The regression variables are displayed in Table 6.

Table 6.
Summary statistics of regression variables

This table shows the descriptive statistics of variables included in the regressions. Financial variables: Return on Assets (*ROA*), scaled by log, profit margin *(PM*), cash flow from operations (CFO), standard deviation of ROA, PM, CFO (*sdROA, sdPM, sdCFO*), accruals (*ACC*), sales (*REV*), market-to-book (*MTB*). Other explanatory variables are: *AGE* is the age of the company for each observation point in time scaled by log, *CHANGE* is 1 if first time CEO writes the letter and 0 otherwise *SIGN* is 1 if CEO has signed the report and 0 otherwise. Measures of future firm performance: future ROA, CFO, FPM indicated by <F> and horizon. Dummy variables for year and industry not displayed in this table. All variables except dummy variables in this table are winsorized to the 1st and 99th percentile. Reported mean, standard deviation, median and quarters. Table with supplementary descriptive statistics is found in table A.5.

Variable	mean	sd	median	Q1	Q3
Financial variables					
ROA	0.0744	0.178	0.0820	0.0268	0.147
sdROA	0.0717	0.106	0.0401	0.0191	0.0819
PM	-0.105	2.127	0.0589	0.0168	0.124
sdPM	1.050	19.10	0.0313	0.0125	0.108
CFO	0.0657	0.147	0.0776	0.0286	0.127
sdCFO	0.0337	0.0612	0.0197	0.00752	0.0397
ACC	-0.0361	0.113	-0.0280	-0.0663	0.00439
REV	14.94	2.114	14.82	13.67	16.39
MTB	0.586	1.269	0.457	-0.155	1.207
Other explanatory variables					
AGE	3.323	1.097	3.178	2.565	4.248
CHANGE	0.129	0.335	0	0	0
SIGN	0.593	0.491	1	0	1
Measures of future firm performance					
FROA1	0.0732	0.172	0.0813	0.0256	0.144
FROA2	0.0727	0.154	0.0757	0.0260	0.140
FCFO1	0.0744	0.152	0.0809	0.0282	0.136
FCFO2	0.0790	0.145	0.0811	0.0319	0.140
FPM1	-0.0747	2.275	0.0595	0.0166	0.123
FPM2	-0.0664	1.618	0.0568	0.0146	0.119
FTQ1	2.821	7.084	1.281	0.977	2.060
FTQ2	2.801	6.145	1.303	0.986	2.086

In Table 6, the variables included in the regressions are presented. The financial variables are winsorized to the 1st and 99th percentage level in order to accommodate extreme values. We present extended descriptive statistics including more percentiles and mean and max values in Table A.5.

Table 7.

Correlation statistics

This table shows the Pearson correlation matrix of variables of Net Optimism, financial variables and other explanatory variables. NETOPT_D is score by DICTION, NETOPT_LM is score by L&M. In Panel A, the correlations for the stage 1 regressions are displayed. In Panel B, the correlations for the stage 2 regressions are displayed. Other variables as defined before.

Panel A: Determinants of Sentiment

	1	2	3	4	5	6	7	8	9	10
NETOPT_D	1.00									
NETOPT_LM	0.69	1.00								
ROA	0.22	0.17	1.00							
sdROA	-0.09	-0.11	-0.18	1.00						
REV	0.09	0.07	0.35	-0.38	1.00					
ACC	0.03	0.03	0.37	-0.03	0.07	1.00				
MTB	0.08	0.06	0.05	0.03	-0.11	-0.12	1.00			
AGE	0.03	0.00	0.06	-0.26	0.42	0.06	-0.13	1.00		
CHANGE	0.05	0.06	-0.08	0.01	-0.01	-0.07	0.03	-0.04	1.00	
SIGN	0.06	0.02	0.07	-0.09	0.27	0.04	-0.10	0.14	-0.04	1.00

Panel B: Variables of Future Firm Performance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
FROA1	1.00															
FROA2	0.90	1.00														
FCFO1	0.81	0.80	1.00													
FCFO2	0.79	0.86	0.93	1.00												
FPM1	0.45	0.39	0.37	0.36	1.00											
FPM2	0.40	0.43	0.37	0.38	0.82	1.00										
FTQ1	0.00	-0.01	0.06	0.05	-0.04	-0.04	1.00									
FTQ2	0.00	-0.01	0.06	0.05	-0.05	-0.05	0.89	1.00								
ROA	0.64	0.66	0.64	0.66	0.28	0.30	0.05	0.03	1.00							
sdROA	-0.17	-0.20	-0.21	-0.21	-0.05	-0.07	0.09	0.11	-0.18	1.00						
CFO	0.67	0.71	0.74	0.75	0.36	0.39	0.06	0.06	0.75	-0.23	1.00					
sdCFO	-0.19	-0.20	-0.22	-0.23	-0.14	-0.16	0.08	0.10	-0.24	0.30	-0.16	1.00				
PM	0.31	0.32	0.37	0.39	0.39	0.42	-0.03	-0.04	0.63	-0.19	0.49	-0.23	1.00			
sdPM	-0.20	-0.22	-0.29	-0.28	-0.31	-0.38	0.01	0.01	-0.28	0.47	-0.36	0.18	-0.47	1.00		
REV	0.27	0.28	0.23	0.22	0.20	0.21	-0.11	-0.12	0.32	-0.38	0.32	-0.37	0.29	-0.28	1.00	
AGE	0.04	0.04	0.00	-0.01	0.05	0.04	-0.12	-0.14	0.06	-0.24	0.03	-0.15	0.08	-0.18	0.41	1.00

In Panel A, the correlations between the variables included in the stage 1a regressions are displayed. In Panel B, the correlation table for the variables included in stage 1b regressions are presented. For the explanatory variables in the regression, ROA correlates with sales and accruals to 0.35-0.37, which is deemed acceptable for inclusion in the regression without too much risk of serial correlation. Age and sales are correlated to 0.41, and remain in the regressions. The variables of change of CEO, CEO signature, and market-to-book exhibit the lowest correlations with the rest of the variables. The dependent variables in the future operating firm performance are correlated with each other. Between future ROA and future operating cash flows, the correlation is high, while lower for net profit margin. However, the measure of Tobin's Q is the least correlated to the other measures being close to zero.

5.3 Stage 1a: Determinants of sentiment

In this stage, we test for the determinants of sentiment, measured by Net Optimism. In Table 8, we present the regression results for the dependent variable Net Optimism. In regression (1), the included variables are from Li (2010a). In regression (2), we augment the baseline regression by adding contextual variables as dummy variables for CEO change and CEO signature.[10] Industry and year dummy variables are not reported in the table.

Table 8.
Determinants of Net Optimism

$$NETOPT_{it} = \beta_0 + \beta_1 ROA_{it} + \beta_2 \delta ROA_{it} + \beta_3 REV_{it} + \beta_4 ACC_{it} + \beta_5 MTB_{it} + \beta_6 AGE_{it} + \beta_7 CHANGE_{it} + \beta_8 SIGN_{it} + \sum_j \beta_9 IND_{ijt} + \sum_k \beta_{10} YEAR_{ikt} + \varepsilon_{it}$$

This table shows the regressions results of OLS regressions on dependent variable Net Optimism (NETOPT) as measured by DICTION. Regressions (1) follows a baseline regression and (2) an augmented regression by adding dummies for change in CEO (CHANGE) and included signature (SIGN). Other independent variables are Return on Assets (ROA), standard deviation in ROA last three years (sdROA), last reported sales scaled by log (REV), Market-to-Book scaled by log (MTB), age of the firm scaled by log (AGE). Dummy variables for industry and year not reported. Robust two-tailed t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (one-tailed when sign of coefficient is predicted; two-tailed otherwise)

		(1)	(2)
Variables	Pred. Sign	*NETOPT*	*NETOPT*
ROA	+	0.0147***	0.0151***
		(6.227)	(6.323)
sdROA	-	-0.00552	-0.00557
		(-1.439)	(-1.467)
REV	?	0.000329*	0.000255
		(1.817)	(1.389)
ACC	-	-0.00770**	-0.00740**
		(-2.070)	(-1.991)
MTB	?	0.000158	0.000167
		(0.628)	(0.666)
AGE	-	-0.000424	-0.000392
		(-1.378)	(-1.282)
CHANGE	?		0.00177*
			(1.889)
SIGN	?		0.00106*
			(1.676)
Constant		0.00536	0.00612
		(1.341)	(1.521)
Observations		1,431	1,431
R-squared		0.135	0.140

In regression (1), we observe that Net Optimism is positively explained by ROA and sales while negatively explained by accruals. As the correlation between ROA and several of the other measures are strong, this is interpreted to capture a lot of the variation of the sample. The signs of standard deviation of ROA and age have signs as predicted, although the variables are not significant. In regression (2), the significance of (1) holds except for all variables except for sales. ROA is significant at the 1% level for both the baseline regression and the extended model. The

[10] Including more contextual variables would be beneficial for future tests.

results is interpreted that when the firm is performing well, the CEO letter tends to have a higher Net Optimism. Unlike previous research performed by Li (2010a), we do not find significance for Market-to-Book. The change of CEO and the inclusion of CEO signature are found to be positively significant with the change in Net Optimism. This could be interpreted that the new CEO express herself more positively. The significance of the signature indicates that it is not coincidental in the annual report and sign of endorsement of the message. Next we find that the firm risk is negative significant with the optimism, as expected.

The reported R-squared is 0.135 for the baseline regression. The R-squared for the (1) regression is in line with the similar regression specifications performed by Li (2010a). The Net Optimism can only to a limited extent be explained by current information. The extended regression (2) increase the R-squared of 0.05 whereby CEO change and CEO sign does not explain much of the variation in Net Optimism. Hence, a potential explanation is that there is additional information in the Net Optimism of the CEO letters than explained by the economic and contextual determinants of CEO change and CEO signature. In a further test of the R-squared, we increase the number of potential variables in the categories of current firm performance and firm risk by including the dependent variables used in future operating firm performance regressions by running an unreported step-wise regression. This has a limited effect on R-squared.

Comparing our results in regression (1), to the benchmark study of Li (2010a) we obtain similar results. The results show that the Net Optimism is positively explained by current firm performance, company risk, the amount of accruals and the firm age. Finally, we find that the Net Optimism is explained to 0.14, which is also in line with Li (2010a).[11]

Overall, the results suggest two things. First, that the Net Optimism is related to other economical determinants available at the time of the publication of the annual report. Second, the limited R-squared gives room for belief that the Net Optimism could be explained, and thus containing, information not explained by the economical determinants. In line with this reasoning, we find positive association with the contextual variables of CEO change and CEO signature. The combination gives room for testing the relation between Net Optimism and future firm performance and thus testing our first hypothesis.

[11] Li (2010a) reports R-squared of 0.14 for the baseline regression and 0.21 for the extended regression.

5.4 Stage 1b: Association with future firm performance

To answer the first hypothesis, we test for the association between Net Optimism and future firm performance. We regress on future operating performance measured as future ROA, presented in column (1) and (2), future operating cash flow, in column (3) and (4), and future net profit margin in (5) and (6). Each regression is based on a horizon of one year and the average of one and two years. These results are presented respectively in Table 9. Industry and year dummies are not reported.

Table 9.

Net Optimism and Future Operating Performance

$$FFP_{it} = \beta_0 + \beta_1 NETOPT_{it} + \beta_2 FFP_{it-1} + \beta_3 \delta FFP_{it} + \beta_4 REV_{it} + \sum_j \beta_5 IND_{ijt} + \sum_k \beta_6 YEAR_{ikt} + \varepsilon_{it}$$

This table shows the regressions results of OLS regressions future operating performance (FFP) measured as future ROA (FROA), future Cash Flow from Operations (FCFO) and future Net Profit Margin (FPM). The horizons are next year (t+1) and average of next year and two years forward (t+2). Explanatory variables include Net Optimism as measured by DICTION (NETOPT), current years value (FFP_{t-1}), standard deviation of FFP (sdFFP) and sales (REV) as defined before. Dummy variables for industry and year not reported. Robust two-tailed t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (one-tailed when sign of coefficient is predicted; two-tailed otherwise)

		(1)	(2)	(3)	(4)	(5)	(6)
		FROA		**FCFO**		**FPM**	
Horizon		t+1	t+2	t+1	t+2	t+1	t+2
Variables	Pred. Sign						
NETOPT	+	0.917***	0.600***	0.312**	0.190*	1.672***	1.580***
		(4.018)	(2.579)	(2.114)	(1.291)	(3.585)	(3.364)
FFP_{t-1}	+	0.414***	0.406***	0.444***	0.424***	0.0808***	0.0592***
		(16.00)	(15.71)	(21.06)	(19.62)	(4.208)	(3.219)
sdFFP	-	-0.0382	-0.0440	-0.0725	-0.145**	-0.00943	-0.0234**
		(-0.664)	(-0.742)	(-1.074)	(-1.991)	(-0.815)	(-2.024)
REV	?	0.000693	0.00249	-0.000538	-0.00195**	0.0125***	0.0166***
		(0.497)	(1.592)	(-0.545)	(-1.987)	(3.853)	(5.020)
Constant		0.0101	-0.0106	0.0377*	0.0791***	-0.284***	-0.336***
		(0.326)	(-0.329)	(1.923)	(4.100)	(-4.023)	(-5.479)
Observations		1,403	1,212	1,484	1,306	1,441	1,253
R-squared		0.453	0.460	0.488	0.499	0.285	0.306

We observe that Net Optimism is positively associated with all measures of future operating performance. For the regression on future ROA, we find that Net Optimism is significant on a 1%-level for one-year and two-year horizon. For the future cash flow from operations, we find significance on a 5%-level for one-year horizon and on a 10%-level for a two-year horizon. For future net profit margin, both one year and two-year horizon are significant on a 1% basis. These results suggest an association between Net Optimism and future operating performance.

The results indicates that one percentage increase in Net Optimism would mean an increase in ROA by 0.9% respective 0.6%. Further, one percentage increase would represent an increase of 0.3% operating future cash flow divided by sales and around 1.6% higher net profit margin. A one-percentage point increase would be about the increase of one standard deviation and a 40% increase

from the mean value of Net Optimism. It is thus a quite large increase in Net Optimism. The effect of this magnitude on the performance metrics is deemed to be limited, but not irrelevant.

Furthermore, we see that, as expected, the current years value of the measures are all significant on one percentage level and positively associated with the measures. This is interpreted as being a valid control variable in an attempt of isolating the effect of the sentiment. Further, the standard deviation is negatively associated as expected. These results are in line with Core et al. (1999) and Bowen et al. (2008). Revenue is positive for ROA and net profit margin, while negative for operating cash flows.

The R-squared for the regressions on ROA and CFO is between 0.4-0.5, which is in line with previous research of similar regressions by Bowen et al. (2008). For the profit margin, the R-squared it is lower, but also in line with measure when included in Gompers et al. (2003). We also observe that the R-squared increases for the two-year horizon, as it is more stable as an average, in line with Core et al. (1999) and Li (2010a).

In order to allow for comparison with previous studies, the control variables are defined as in Bowen et al. (2008). To test for the effects of adding contextual independent variables, we include significant variables of change in CEO and CEO signature from Table 8 in an unreported regression. These have no substantial effect on the results.

For future firm valuation, we regress on dependent variable future Tobin's Q. The results are presented in Table 10. Industry and year variables included but still not reported.

Table 10.

Net Optimism and Future Tobin's Q

$$FTQ_{it} = \beta_0 + \beta_1 NETOPT_{it} + \beta_2 REV_{it} + \beta_3 AGE_{it} + \sum_j \beta_4 IND_{ijt} + \sum_k \beta_5 YEAR_{ikt} + \varepsilon_{it}$$

This table shows the statistics of OLS regressions on Tobin's Q. Tobin's Q calculated as the market value of equity plus the book value of liabilities less the deferred tax liabilities divided by the book value of assets. The horizons are next year (t+1) and average of next year and two years forward (t+2). Explanatory variables include Net Optimism as measured by DICTION (NETOPT), sales (REV) and age (AGE) as defined before. Dummy variables for industry and year not reported. Robust two-tailed t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (one-tailed when sign of coefficient is predicted; two-tailed otherwise)

		(1)	(2)
Horizon		t+1	t+2
Variables	Pred. Sign		
NETOPT	+	3.814***	3.569**
		(2.391)	(2.063)
REV	?	-0.0407***	-0.0491***
		(-4.057)	(-4.649)
AGE	?	-0.0696***	-0.0610***
		(-3.697)	(-3.046)
Constant		0.364**	0.608***
		(2.261)	(3.588)
Observations		1,452	1,263
R-squared		0.211	0.211

For the regression on future Tobin's Q, we find that Net Optimism is significant on a 1%-level for one-year horizon and 5%-level for a two-year horizon. The results indicate that one percentage higher Net Optimism lead to a 3.5-3.8 percentage higher Tobin's Q. As discussed above, an increase of one percentage point of Net Optimism is a substantial increase in the sentiment, as is the corresponding increase in Tobin's Q. The magnitude of the coefficient is, as for the other regressions, limited but not irrelevant. The variables revenue and age are negatively associated with future Tobin's Q indicating a higher valuation for smaller and younger companies. The R-squared for the regressions on ROA and CFO is around 0.2.

As described in section 4, we perform further statistical tests. We present the results for the future firm performance regressions including firm fixed effects and firm clustered standard errors in Table 11. In Panel A, the results for the firm fixed effects regressions are displayed.

Table 11.

Second set of regressions: Net Optimism and Future Operating Performance

This table shows the regression results for regression specified in table 9, using fixed effects on firm level in regressions (1)-(6) and firm clustered standard errors in regressions (7)-(12) Dummy variables for year and industry not reported. *** p<0.01, ** p<0.05, * p<0.1 (one-tailed when sign of coefficient is predicted; two-tailed otherwise)

		FROA		FCFO		FPM	
Horizon		t+1	t+2	t+1	t+2	t+1	t+2
Variables	Pred. Sign						
Panel A: Firm Fixed effects							
		(1)	(2)	(3)	(4)	(5)	(6)
NETOPT	+	0.686***	0.537***	0.214*	0.100	0.949**	0.759*
		(2.877)	(2.457)	(1.335)	(0.672)	(2.029)	(1.619)
FFP_{t-1}	+	0.0723	0.00839	0.125***	0.0934***	-0.00311	-0.0303*
		(1.483)	(0.189)	(3.745)	(2.905)	(-0.213)	(-1.736)
sdFFP	-	0.0726	0.0677	-0.0153	-0.0932*	0.0221**	0.0124
		(1.455)	(1.310)	(-0.260)	(-1.711)	(1.998)	(1.366)
REV	?	-0.0205**	-0.0283***	-0.00964*	-0.00930	0.0119	-0.0138
		(-2.127)	(-3.157)	(-1.933)	(-1.354)	(0.519)	(-0.684)
Constant		0.383***	0.506***	0.219***	0.238**	-0.140	0.230
		(2.715)	(3.743)	(2.992)	(2.344)	(-0.415)	(0.753)
Observations		1,403	1,212	1,484	1,306	1,441	1,253
R-squared		0.098	0.115	0.040	0.055	0.034	0.056

Table 11. Continue

Panel B: Clustered standard errors							
		(7)	(8)	(9)	(10)	(11)	(12)
NETOPT	+	1.062***	0.716***	0.518***	0.410**	1.286*	1.174**
		(4.226)	(2.714)	(3.274)	(2.194)	(1.940)	(1.704)
FFP_{t-1}	+	0.420***	0.415***	0.475***	0.458***	0.103***	0.0791***
		(10.89)	(10.45)	(14.22)	(12.15)	(4.340)	(2.771)
sdFFP	-	-0.0519	-0.0608	-0.0178	-0.100	0.00567	-0.00922
		(-1.453)	(-1.244)	(-0.229)	(-1.257)	(0.484)	(-0.837)
REV	?	0.000135	0.00176	-0.00220	-0.00397***	0.0143***	0.0184***
		(0.0749)	(0.813)	(-1.615)	(-2.629)	(2.737)	(3.105)
Constant		0.0556*	0.0374	0.0829***	0.130***	-0.179**	-0.247**
		(1.784)	(0.988)	(3.689)	(5.283)	(-2.114)	(-2.592)
Observations		1,403	1,212	1,484	1,306	1,441	1,253
R-squared		0.442	0.448	0.453	0.454	0.203	0.218

For the FROA measure, the Net Optimism variable remains significant at the 1%-level for both horizons firm specific fixed effects. However, the positive association for Net Optimism is weaker with compared to the OLS regressions. Thus, an indication is provided of an upward bias in the OLS regressions estimates. For the FCFO measure, the significance for the Net Optimism is reduced on the one-year horizon and the coefficients are weakened indicating an upward bias. Further, the significance is lost for the two-year horizon. For the FPM measure, the significance for the one-year and two-year horizon decreases, but remains significant. In Panel B, we present the results for the OLS regressions with year dummies and firm clustered standard errors. Net Optimism is significant for all performance measures when applying firm clustered standard errors. In total, the additional regressions suggests the presence of a potential bias in our OLS estimates. However, the Net Optimism measure remains statistical significant for all performance measures.

In Table 12, the statistical tests using firm fixed effects and clustered standard errors for the regression on Tobin's Q are presented. The significance for Tobins Q is lost at the first-year horizon for the fixed effects model compared to the OLS regression. Further, the positive coefficient is reduced indicating the cause to be an upward bias in the OLS estimates. In column (3) and (4), when applying firm clustered standard errors the Net Optimism measure is significant at the 5%-level for both time horizons.

Table 12.

Second set of regressions: Net Optimism and Future Tobin's Q

This table shows the regression results for regression specified in table 10, using fixed effects on firm year level in regressions (1) and (2) and firm clustered standard.. In regressions (3) and (4). Dummy variables for year and industry not reported. *** p<0.01, ** p<0.05, * p<0.1 (one-tailed when sign of coefficient is predicted; two-tailed otherwise)

		Firm Fixed effects		**Clustered standard errors**	
		(1)	(2)	(3)	(4)
Horizon		t+1	t+2	t+1	t+2
Variables	Pred. Sign				
NETOPT	+	1.259	1.909**	7.182**	6.957**
		(1.247)	(1.795)	(2.550)	(2.358)
REV	+	0.0819*	0.0458	-0.0589**	-0.0672**
		(1.767)	(0.960)	(-2.329)	(-2.544)
AGE	-	0.0825	0.154	-0.0646	-0.0565
		(0.725)	(1.339)	(-1.408)	(-1.163)
Constant		-1.181*	-0.772	1.205***	1.444***
		(-1.652)	(-1.023)	(3.425)	(3.853)
Observations		1,452	1,263	1,452	1,263
R-squared		0.119	0.076	0.084	0.077

Overall, the indication of biased estimates suggests the presence of some omitted variable in our OLS regressions that are correlated with the specific performance measure and Net Optimism. However, the statistical significance remains for the majority of our future firm performance measures. Thus, the statistical tests to address the potential issue of endogeneity still indicate a positive association between Net Optimism and future firm performance.

In summary, the Net Optimism measure exhibits a significant positive association with future operating performance and valuation. Further, overall the statistical significance remains for Net Optimism when testing for firm fixed effects and firm clustered standard errors. The results suggest that the Net Optimism measure is associated with future firm performance.

5.5 Stage 2: Association with future abnormal returns

For our second hypothesis of the association with future abnormal returns, we construct a trading strategy based on the Net Optimism measure. We present average monthly excess returns on equally-weighted quintile portfolios. The main portfolios of interest are the Bottom Portfolio (P0020), the Top Portfolio (P8000) and the Net Portfolio (LSP8020), which are part of the trading strategy. The results are presented in Table 13.

In Panel A, the alpha for the highest monthly abnormal return is shown for the Net Portfolio of 0.3%, when using the four factor model. The long position Top Portfolio yields a monthly

abnormal return of 0.23% and -0.07% for the short position in the Bottom Portfolio. The rank between the portfolios are as expected. However, none of the abnormal returns are statistically significant.

Table 13.

Full Sample Portfolio Regressions

This table exhibits monthly abnormal returns by portfolio regression type, regression coefficients for the four factor model and portfolio characteristics. The results are for equally-weighted (EW) portfolios based on quintile values for Net Optimism and are reformed on a yearly basis. For example, the top quintile portfolio is P8000 portfolio and the Net Portfolio is LSP8020, which buys the top quintile portfolio (P8000) and sells the bottom quintile portfolio P8020. Panel A displays monthly portfolio abnormal returns in percentage units by portfolio analysis model and the pertaining t-statistics value in parenthesis. The results are based on regressions for the one-factor CAPM model (RMRF), the three factor Fama-French model (RMRF, SMB, HML) and the four-factor Carhart model (RMRF, SMB, HML, WML). Panel B reports the factor coefficients for the four-factor Carhart model and pertaining t-statistics in parenthesis. The sample period is 5/05-5/15.T-statistiscs in parenthesis. ***p<0.01, **p<0.05, *p<0.1

Panel A. Portfolio Alphas						
	(1)	(2)	(3)	(4)	(5)	(6)
Portfolios:	P0020	P2040	P4060	P6080	P8000	LSP8020
CAPM	-0.21	0.36	0.55	0.61	0.23	0.45
	(-0.37)	(0.68)	(1.08)	(1.13)	(0.46)	(1.49)
3-factor model	-0.23	0.31	0.50	0.58	0.20	0.42
	(-0.42)	(0.63)	(1.08)	(1.13)	(0.42)	(1.42)
4-factor model	-0.07	0.46	0.49	0.68	0.23	0.3
	(-0.14)	(0.90)	(1.02)	(1.30)	(0.46)	(0.98)
Panel B. Four factor Regression Coefficients						
RMRF	0.26**	0.22**	0.25**	0.21*	0.26**	0.00
	(2.26)	(2.09)	(2.45)	(1.90)	(2.60)	(0.044)
SMB	1.29***	1.11***	1.22***	1.07***	1.07***	-0.22
	(4.57)	(4.23)	(4.95)	(3.99)	(4.26)	(-1.39)
HML	-0.06	-0.26	-0.15	-0.16	-0.11	-0.06
	(-0.18)	(-0.91)	(-0.55)	(-0.54)	(-0.40)	(-0.33)
WML	-0.18	-0.17	0.01	-0.12	-0.03	0.15*
	(-1.16)	(-1.19)	(0.09)	(-0.80)	(-0.24)	(1.70)
Panel C. Portfolio Characteristics						
Mean excess return	-0.052	0.476	0.660	0.719	0.365	0.416
Portfolio SD	0.065	0.059	0.056	0.059	0.057	0.033
Portfolio skewness	0.167	0.313	-0.932	-0.549	-0.768	-
Median MV	2,833	3,936	3,023	3,809	2,988	-
Median MB	1.242	1.600	1.601	1.543	1.652	-
Mean Obs	29	29	28	28	28	-
Sharpe ratio	-0.008	0.080	0.117	0.122	0.064	-

In Panel B, the coefficients and t-statistics for the four risk factors in the Carhart model are presented. For the risk measures of Market risk premium and Small Minus Big, the coefficients are generally significant for the portfolios. Nonetheless, for the measures of High Minus Low and Winners Minus Losers, the coefficients are not significant. This could lead to biased results if the risk measures are not properly specified. However, as no significance is found using either the CAPM, or Fama-French risk measures, the prospect of changing measures to find significance is deemed limited.

Descriptive statistics for the quintile equally-weighted portfolios are presented in Panel C. The highest mean monthly excess return of over the full sample period is shown for the Net Portfolio of 0.4%. The risk-adjusted excess return, represented by the Sharpe Ratio is also the highest for the Net Portfolio with 0.06%. Further, the Net Portfolio has the highest mean Market Value (MV) of 3,590 and the highest Market-to-Book ratio of 5.42.

We present the results for the sample split in Table 14. The table reports the monthly abnormal returns for the Carhart four-factor model (1997) and the pertaining t-statistics in parentheses. First, we delimited the sample excluding financial firms by arguing that these firms could drive our results. As shown in Figure 4, the drop in Net Optimism during the financial crisis is considerably larger for financial firms. In column (2), our results for the sample excluding financial firms are displayed. We find no indication of any association between sentiment and abnormal returns. A short position in the Bottom Portfolio (P0020) yields a monthly abnormal return of 0.37%. A long position in the Top Portfolio (P8000) results yields an abnormal return of 0.18% and the Net Portfolio 0.55%. However, all these abnormal returns are insignificantly different from zero.

Table 14.
Sample Split Portfolio Regressions

This table displays monthly four factor abnormal returns and the pertaining t-statistics in parentheses of our portfolios, by different samples. The portfolios are based from sorting the Net Optimism score on quintile cutoff values. The Top Portfolio contains the stocks with the highest quintile Net Optimism score (P8000) and the Bottom Portfolio (P0020) contains the stocks with the lowest quintile Net Optimism score. The Net Portfolio (LSP8020) takes a long position in the Top Portfolio (P8000) and a short position in the Bottom Portfolio (P0020). The portfolios are equally-weighted (EW) and reformed on a yearly basis and monthly excess returns are calculated. We run a regression for the mean excess monthly returns on the four Carhart risk factors (RMRF, HML, SMB, WML), where alpha is the constant. The sample period is 5/05-5/15.T-statistiscs in parenthesis., ***p<0.01, **p<0.05, *p<0.1

	Firms		Size	
	(1)	(2)	(3)	(4)
Sample:	Full sample	Excl Financials	Large	Small
P0020	-0.07	-0.37	-0.07	-0.19
	(-0.14)	(-0.65)	(-0.12)	(-0.30)
P8000	0.23	0.18	0.54	0.18
	(0.46)	(0.37)	(1.04)	(0.34)
LSP8020	0.30	0.55	0.60**	0.37
	(0.98)	(1.63)	(2.09)	(0.83)

The results for the sample split between large and small firms are displayed in Column (3) and (4). We find significant monthly abnormal returns in the sample consisting of Large Firms for the Net Portfolio of 0.6%. No other abnormal returns for the sample split are statistically significant. The monthly abnormal return are higher for the Large Firms in the Bottom Portfolio, the Top Portfolio and the Net portfolio compared to small firms.

In summary, we find limited indications of any systematic abnormal returns. We find no significance in the trading portfolios except in the spread portfolio for the sample split Large firms. In addition, the average goodness of fit for the four-factor model is relatively low (20%). Hence, an indication is provided that the model fails to explain the portfolio returns. Moreover, we test for a sample split between firms and still obtain similar results and also observe that the Sharpe Ratios are close to zero for our constructed portfolios which points in the same direction of our results. Overall the results provide indications that the formed trading portfolios on Net Optimism cannot help generate abnormal returns.

6 ANALYSIS AND DISCUSSION

6.1 Summary of results

The aim of this paper is to understand the information content of CEO letters in Sweden. We state two hypotheses with subsequent tests. In our first hypothesis, we find support for our hypothesis that the information content in the CEO letter is associated with future firm performance. In our second hypothesis, we find low support for our hypothesis that the information could be used to generate abnormal stock returns. We elaborate on these hypotheses below.

Evidence on Hypothesis 1

Theoretically, the CEOs could transfer their private information and expectations about the future to the CEO letters and measuring the sentiment of the letter could capture this information. This is in line with previous research by Abrahamson and Amir (1996) and Patelli and Pedrini (2014). We perform a number of regressions on different accounting metrics in order to measure the association between the sentiment in CEO letters and future firm performance, following prior research on information content (Core et al., 1999; Gompers et al., (2003); Bowen et al., 2008). In addition, we address the issues of serial correlation and endogeneity by using a firm fixed effects model and clustered standard errors. For the OLS regressions, we find significance for all measures. In further statistical elaborations on firm fixed effects and clustered standard errors, we find the significance for most or all of the metrics. For the Tobin's Q, we find significance for Net Optimism for both time periods in the OLS and clustered standard errors and for two-year average horizon for the fixed effects and clustered effects. In total, we find strong support for the association between the sentiment in the CEO letter and future firm performance.

We interpret the results as finding strong support for our first hypothesis. The results are interpreted as that the sentiment in CEO letters contains incremental information about future firm performance in relation to the annual report. Hence, the CEO letter would add information content to the annual report. These results are in line with findings of Abrahamson and Amir (1996) and Patelli and Pedrini (2014) studying the information content of CEO letters. Further, the positive association is in line with other studies on management narrative disclosures such as MD&A (Li, 2010a) and earnings press releases (Davis et al., 2012).

Evidence on Hypothesis 2

As the sentiment in CEO letters contain information about the future, the sentiment could potentially be used to generate abnormal returns. We evaluate this by constructing a trading strategy based on the calculated Net Optimism measure. For the results based on the full sample we find insignificant abnormal returns indicating that the information content of the sentiment is already incorporated into market. Overall, we obtain insignificant abnormal returns when excluding financial firms and dividing the sample into large and small firms. Compared to previous research, these results are conflicting. For example, both Feldman et al. (2010) and Tetlock (2008) find significant abnormal returns. Still, these studies differ on several reasons. First, the studies are conducted on a different source of management disclosure (MD&A) and based in a US-setting. Second, our data set is based on substantially fewer observations. Thus, their larger number of observations could support the results of these studies.

We find low support for our hypothesis of the association between the information in CEO letters and future abnormal stock returns. It is difficult to draw inferences from a non-rejected hypothesis. However, given the incremental information content of the CEO letters in relation to the annual report, this information is plausibly related to future abnormal returns. As we find low indications of abnormal returns, this suggests that the information already is incorporated in the stock prices either through the information in the CEO letters or through other sources of information, such as interim reports and conference calls. In a Swedish and European context, this latter explanation would be in line with findings that users of financial reports to a limited extent use the annual reports as a source of information, in comparison to e.g. the interim reports (Hjelström et al., 2014).

6.2 Sensitivity analysis

In order to test the robustness of the results, a series of sensitivity tests is performed. In conclusion, we find that the results are robust for hypothesis 1 and no change in hypothesis 2.

Table 15
Summary of Hypotheses and Robustness Tests

This table summarizes a set of sensitivity tests performed. On the first line, the main findings are found. Subsequent rows include the tests and a short summary of their effect on the main findings. The results are, when applicable, directed to the place in the appendix.

Test	Hypothesis 1: Future firm performance	Hypothesis 2: Future abnormal returns
Main results	H0 Rejected	H0 Not Rejected
Alternative Net Optimism measure Appendix 9.4	**Robust** Significance for all but FCFO1 & FCFO2 Decreased constant for all values	**No change in results**
Excluding financials Appendix 9.5	**Robust** Significance for all measures Increased constant for all measures	Included in main results
Value-weighted portfolios Appendix 9.6	-	**No change in results**

Alternative Net Optimism Measure

The Net Optimism measure is of central importance in this thesis. If this measure would be inadequately specified, the results would lose inference value. This is a valid concern for the robustness of the results. As a robustness check of the measurement, we test a second measure of Net Optimism, based on the Loughran and McDonald (2011) dictionary. To test for this, we include finance-specific word lists by Loughran and McDonald (2011) financial specific dictionaries for optimism increasing and decreasing words.[12] We run the OLS regressions and base the trading strategy on this measure. The results are presented in Appendix 9.4.

For the OLS regressions, we find that for a one year horizon the results remain significant for all measures except the future operating cash flow measure. Moving on to a two-year horizon, the future ROA metric loses significance. We find that the results to be less significant when switching but as a majority of the operating performance and Tobin's Q measures remain robust this test exhibits similar strength of results. We interpret our main results to be robust. For the trading

[12] Dictionaries obtained with courtesy from Bill McDonald website: www3.nd.edu/~mcdonald/Word_Lists.html

model, the monthly abnormal return remains statistically insignificant for all portfolios and the coefficients are similar to the first measure of Net Optimism.

Excluding financials

Several studies exclude financial companies in their data set as these companies exhibit different properties from other industries. In addition to this, during the financial crisis in 2008-2009 the financial companies exhibited a lower Net Optimism, as interpreted in Figure 4. The industry financials reach 13% of the data set and to explore if this industry could have an impact on the results, this group is excluded. We run the OLS regressions for hypothesis on this delimited sample. For the trading model, this test is included in the main results. Results are presented in Appendix 9.5. We find that the significance and coefficients are strengthened and we interpret the results to be robust.

In an additional data related test, the regressions are also run without winsorizing the financial variables and measures of future firm performance. In the study we winsorize the data as we believe it to exhibit some extreme and outlying values that could cause biased results. In order to give an indication for the possible effects of this, we run OLS regressions without winsorizing the data. Table A.5. exhibits descriptive statistics of the min/max and 1/99 percentage level. Regression results are unreported. The results for the regressions are shown to be having a limited effect on the coefficient and the variables remain significant. This is not applicable for the trading model as based on a non-winsorized Net Optimism measure. We find that the significance and coefficients are strengthened and we interpret the results to be robust.

Value-weighted returns

We test our results using VW returns, although we argue for the use of EW returns. Under the VW approach stocks with high capitalization can have a large impact on the portfolio returns. Results are presented in Appendix 9.6. Our results using VW returns are similar to the results using EW-returns, that all alphas remain statistically insignificant for the four-factor model.

6.3 Discussion and limitations

The aim of this thesis has been to investigate the information content in CEO letters. To understand the limitations of our results, we discuss the validity, reliability and generalizability of our study. As a starting point, we use the three conceptual relationships in Figure 1.

First, a prerequisite for the study is that the hypothesized incremental information content in the CEO letter holds. The information content is dependent on the incentives and the ability of the preparers to disclose information. We hypothesize that the CEOs are incentivized to disclose incremental information. However, the competing theory of impression management stipulates that managers can exploit the information asymmetry relationship in self-serving purposes (Merkl, 2007). This would lead to the disclosures not being truthful. The results in the study suggest that CEOs at least on an average level disclose incremental information. However, as the incentives could differ on an individual level, the generalizability is on an aggregate level.

Second, an important aspect to discuss is the ability to measure the sentiment of the analyzed CEO letters. If not measured properly, there is a risk of a decrease of validity. There are ongoing discussions regarding measuring sentiments with dictionaries in accounting and finance research. The dictionary-based approach, applied through DICTION, could have limitating effects on the validity. As it is based on pre-defined dictionaries, the risk arise that the words are out of context (Li, 2010a). Furthermore, as DICTION is based on British and American English (Hart, 2000), this method of measuring sentiment is not ultimate for a non-Anglophone setting. Yet, based on two Net Optimism measures with similar results, we feel confident of the applied method ability to capture the sentiment on at least a rudimentary level. Moreover, the application of a quantitative content analysis method, such as a rule-based dictionary-based approach, allows for an objective methodological approach to measure sentiment. This permits for comparability to previous and future research and increases the reliability of our study.

A third discussion is whether the sentiments in the CEO letters are associated with future firm performance. The employed methods for measuring the link are important for the validity of the relationship and the results. A fundamental problem with disclosure studies in general is the risk of endogeneity (Healy & Palupe, 2001). A central reason of this is the risk of omitted variables causing the regression estimates to be biased (Woolridge, 2009). In this study, we attempt to address this issue of endogeneity by including firm fixed effects models and clustered standard errors. Nonetheless, this is likely to remain a problem and has an impact on the validity of our results as it reduces the possibility for causal inference. The use of instrumental variables or prediction of the

future firm performance based on logit models are suggestions that can help to increase the validity further in future research.

On a more specific note, concerns exist that regarding the actual writer of the CEO Letter. As pointed out by Li (2010a), much of corporate disclosure is written through collaborative efforts by for example managers and public relations staff. In the Swedish context, similar concerns are present and potential "ghost writers" authorizing the texts can exist (Jonäll, 2007). As previously mentioned, the CEO can still be assumed to be responsible and take an active role in the editing and approval of the final text (Jonäll, 2007).

Finally, a discussion on the data sample is warranted. The data set has two main characteristics that could have a limiting effect on the results; it is skewed over time towards the end of the time period and towards large companies. This might limit the inference of the results. A possibility would be that the quality of disclosure is positively related to size, whereby a skewness towards large companies could result in biased results. Further, the sample only contains companies disclosing an annual report in English. A possible selection bias could arise due to the unavailability of English financial reports. As the data sample is containing a substantial amount of the listed companies and covers ten years, we generalize the results to the whole Stockholm Stock Exchange. Overall, the generalization of the results to a larger population is directly limited to CEO letters on the Stockholm Stock Exchange. However, as several mechanisms of the CEO letter follows analogous narrative disclosures these findings add to the research area in a broader context.

7 CONCLUSION

In this thesis we aim to increase the understanding of the information content of the CEO letters on the Stockholm Stock Exchange. We measure the sentiment in the language and test for association with future firm performance and future abnormal returns. Our main findings are that the sentiment in CEO letters in Sweden contains incremental information about future operating performance and future firm valuation.

For our first hypothesis, we test for the information content in the CEO letters in association to future firm performance. We find support for our hypothesis that the CEO letters contain information about future firm performance measured as future operating performance and future valuation. These results suggest that the CEO letter contains incremental information beyond the rest of the annual report. The results are interpreted that CEOs use the language in the CEO letters to provide incremental information to the market in order to reduce information asymmetry.

However, we find low support for the hypothesis that the sentiment in CEO letters is associated with future abnormal returns. These results suggest that the incremental information in the CEO letters already is incorporated in stock prices.

This thesis contributes to the existing literature in two ways. First, empirical evidence of incremental information content of Swedish CEO letters are found. The results add to both the discussion of the usefulness of CEO letters in Sweden and to the new but growing field of sentiment research on corporate narrative disclosures.

Second, the results from this thesis show results of a quantitative sentiment measurement of corporate disclosures in a non-English speaking area. Limited sentiment analysis research has been performed in a non-English speaking region and this shows support for the new but promising research area. Correspondingly, this provides an interesting starting-point for additional research on contextual differences of disclosures measured by their sentiment.

8 REFERENCES

Akerlof, G.A. (1970), "The market for" lemons": Quality uncertainty and the market mechanism". *The quarterly journal of economics,* pp. 488-500.

Abrahamson, E. & Amir, E. (1996), "The information content of the president's letter to shareholders". *Journal of Business Finance & Accounting,* 23 (8), pp. 1157-1182.

Antweiler, W. & Frank, M.Z. (2004), "Is all that talk just noise? The information content of internet stock message boards". *The Journal of Finance,* 59 (3), pp. 1259-1294.

Beattie, V., McInnes, B. & Fearnley, S. (2004), "A methodology for analysing and evaluating narratives in annual reports: a comprehensive descriptive profile and metrics for disclosure quality attributes". *Accounting forum*, 28(3), pp. 205.

Bowen, R.M., Rajgopal, S. & Venkatachalam, M. (2008), "Accounting Discretion, Corporate Governance, and Firm Performance*". *Contemporary Accounting Research,* 25 (2), pp. 351-405.

Brown, S.V. & Tucker, J.W. (2011), "Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications". *Journal of Accounting Research,* 49 (2), pp. 309-346.

Bryan, S.H. (1997), "Incremental information content of required disclosures contained in management discussion and analysis". *Accounting Review*, 72 (2), pp. 285-301.

Carhart, M.M. (1997), "On persistence in mutual fund performance". *The Journal of finance,* 52 (1), pp. 57-82.

Core, J.E., Holthausen, R.W. & Larcker, D.F. (1999), "Corporate governance, chief executive officer compensation, and firm performance". *Journal of Financial Economics,* 51 (3), pp. 371-406.

Courtis, J.K. (2004), "Corporate report obfuscation: artefact or phenomenon?". *The British Accounting Review,* 36 (3), pp. 291-312.

Davis, A.K., Piger, J.M. & Sedor, L.M. (2012), "Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language". *Contemporary Accounting Research,* 29 (3), pp. 845-868.

Epstein, M.J. & Pava, M.L. (1995), "Shareholders' perceptions on the usefulness of MD&As". *Managerial Finance,* 21 (3), pp. 68-83.

Eugster, F. & Wagner, A.F. (2013), "Voluntary Disclosure Quality, Operating Performance, and Stock Market Valuations". *Swiss Finance Institute Research Paper,* (11-25).

Fama, E.F. & French, K.R. (1998), "Value versus growth: The international evidence". *The Journal of Finance,* 53 (6), pp. 1975-1999.

Fama, E.F. & French, K.R. (1993), "Common risk factors in the returns on stocks and bonds". *Journal of Financial Economics,* 33 (1), pp. 3-56.

Feldman, R., Govindaraj, S., Livnat, J. & Segal, B. (2010), "Management's tone change, post earnings announcement drift and accruals". *Review of Accounting Studies,* 15 (4), pp. 915-953.

Gompers, P.A., Ishii, J.L. & Metrick, A. (2003), "Corporate governance and equity prices". *Quarterly Journal of Economics*, 118 (1), pp. 107-156.

Griffin, J.M. (2002), "Are the Fama and French factors global or country specific?". *Review of Financial Studies,* 15 (3), pp. 783-803.

Hart, R.P. (2000), *"DICTION 5.0"*. Austin, TX: Digitex

Healy, P.M. & 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), pp. 405-440.

Hjelström, A., T. Hjelström & E. Sjögren. (2014), *"Decision usefulness explored - An investigation of capital market actors´ use of financial reports"*. Stockholm: Confederation of Swedish Enterprise/Svenskt näringsliv.

Hoskin, R.E., Hughes, J.S. & Ricks, W.E. (1986), "Evidence on the incremental information content of additional firm disclosures made concurrently with earnings". *Journal of Accounting Research,* 24 (1), pp. 1-32.

Hutton, A. (2004), "Beyond financial reporting an integrated approach to disclosure". *Journal of Applied Corporate Finance,* 16 (4), pp. 8-16.

Jegadeesh, N. & Wu, D. (2013), "Word power: A new approach for content analysis". *Journal of Financial Economics,* 110 (3), pp. 712-729.

Jonäll, K. (2009) *"VD:n har ordet: Bilden av det Goda företaget - text och siffror i VD-brev"*. Gothenburg: BAS

Kearney, C. & Liu, S. (2014), "Textual sentiment in finance: A survey of methods and models". *International Review of Financial Analysis,* 33, pp. 171-185.

King, R., Pownall, G. & Waymire, G. (1990), "Expectations adjustment via timely management forecasts: Review, synthesis, and suggestions for future research". *Journal of accounting Literature,* 9 (1), pp. 113-144.

Kothari, S., Li, X. & Short, J.E. (2009), "The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis". *The Accounting Review,* 84 (5), pp. 1639-1670.

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), "Survey of the Literature". *Journal of accounting literature*, 29, pp. 143-165.

Loughran, T. & McDonald, B. (2011), "When is a liability not a liability? Textual analysis, dictionaries, and 10- Ks". *The Journal of Finance*, vol. 66, no. 1, pp. 35-65.

Merkl-Davies, D.M. & Brennan, N.M. (2007), "Discretionary disclosure strategies in corporate narratives: incremental information or impression management?". Journal of accounting literature, 27, pp. 116-196.

Morris, M.W., Sheldon, O.J., Ames, D.R. & Young, M.J. (2007), "Metaphors and the market: Consequences and preconditions of agent and object metaphors in stock market commentary". *Organizational behavior and human decision processes,* 102 (2), pp. 174-192.

Mossin, J. (1966), "Equilibrium in a capital asset market". *Econometrica: Journal of the econometric society,* 34 (4), pp. 768-783.

Patelli, L. & Pedrini, M. (2014), "Is the optimism in CEO's letters to shareholders sincere? Impression management versus communicative action during the economic crisis". *Journal of Business Ethics*, 124 (1), pp. 19-34.

Petersen, M.A. (2009), "Estimating standard errors in finance panel data sets: Comparing approaches". *Review of Financial Studies*, 22 (1), pp. 435-480.

Sloan, R. (1996), "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings? (Digest Summary)". *Accounting review*, 71 (3), pp. 289-315.

Smith, M. & Taffler, R. (1995), "The incremental effect of narrative accounting information in corporate annual reports". *Journal of Business Finance & Accounting*, 22 (8), pp. 1195-1210.

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. & Macskassy, S. (2008), "More than words: Quantifying language to measure firms' fundamentals". *The Journal of Finance*, 63 (3), pp. 1437-1467.

Verrecchia, R.E. (2001), "Essays on disclosure". *Journal of Accounting and Economics*, 32 (1), pp. 97-180.

Wooldridge, J.M. (2009) *"Introductory Econometrics: A Modern Approach"*. 4th ed., Ohio: South-Western Cengage Learning

9 APPENDIX

9.1 Sentiment extraction methodology

In the CEO letter collection process, the following steps have been followed. In a pre-processing section, the annual report has first been downloaded for each company. For each annual report, the CEO letter is then identified. This letter is copied from the annual report file manually and transformed into a text-file. These files are assigned an identification number per firm year, before the quantitative analysis through DICTION.

Certain sections and items of each CEO letter have been excluded and are summarized in Table A.1. Finally, the text files are prepared before the statistical analysis by DICTION, as explained in Section 3.

Table A.1.
CEO letter collection methodology

This table summarizes the inclusion and exclusion of information in the CEO letters.

General aspects	**Included**
• If letter is jointly written by i.e. the CEO and the Chairman it is included in the sample. • If two letters, one by the CEO and one by the Chairman, the letter from the CEO is included in the sample. • If two CEO letters, the letter from the longest serving is included in the sample. • If interview with CEO and no other letter, it is included in the sample.	• Headlines/title/greeting text • Introductory text • The name of the CEO • Bullet points in text
	Not included
	• Tables • Graphs • Quotes

9.2 Supplementary sentiment tables

By comparing the difference between the measures, we see that the optimistic scores for DICTION show a higher mean, median and quarters. However, the L&M measure has a higher average amount of pessimistic words than DICTION. The resulting mean is thus lower for L&M from both the optimistic and pessimistic side. The standard deviations are higher for L&M showing a higher volatility of the measures.

Table A.2.
Descriptive statistics of Net Optimism per year

This table shows the descriptive statistics of Net Optimism variables measured by DICTION and L&M. Shown by year during period 2004-2013. Reported mean numbers of words in CEO letters per year number of observations, mean, min, quarters, median, max and standard deviation.

Year	Measure	mean of words	N	mean	min	p25	p50	p75	max	sd
2004	DICTION	1152	107	2.56%	-0.76%	1.99%	2.55%	3.20%	5.13%	1.11%
	L&M		107	2.40%	-0.15%	1.74%	2.39%	3.13%	5.48%	1.08%
2005	DICTION	1179	115	2.66%	0.39%	2.07%	2.58%	3.30%	5.50%	1.00%
	L&M		115	2.60%	-0.37%	1.91%	2.59%	3.32%	5.39%	1.14%
2006	DICTION	1256	126	2.74%	0.56%	2.00%	2.65%	3.43%	6.30%	1.04%
	L&M		126	2.54%	0.00%	1.94%	2.49%	3.00%	6.92%	1.06%
2007	DICTION	1200	144	2.52%	0.08%	1.77%	2.39%	3.25%	5.16%	1.12%
	L&M		144	2.20%	-0.40%	1.50%	2.22%	2.87%	5.19%	1.08%
2008	DICTION	1198	152	2.01%	-1.86%	1.20%	1.97%	2.71%	7.11%	1.23%
	L&M		152	1.44%	-2.42%	0.39%	1.58%	2.35%	5.14%	1.38%
2009	DICTION	1203	158	2.06%	-0.77%	1.33%	2.14%	2.78%	4.62%	1.09%
	L&M		158	1.74%	-0.92%	0.83%	1.64%	2.52%	4.92%	1.20%
2010	DICTION	1193	165	2.56%	-0.17%	1.95%	2.59%	3.26%	5.16%	1.02%
	L&M		165	2.26%	-0.60%	1.52%	2.29%	2.94%	5.77%	1.00%
2011	DICTION	1158	176	2.53%	-1.65%	1.82%	2.51%	3.28%	5.60%	1.21%
	L&M		176	2.18%	-1.50%	1.39%	2.26%	2.93%	5.77%	1.26%
2012	DICTION	1107	183	2.55%	-0.33%	1.80%	2.53%	3.29%	5.80%	1.08%
	L&M		183	2.27%	-0.27%	1.48%	2.26%	2.98%	5.67%	1.11%
2013	DICTION	1113	180	2.67%	-1.04%	1.95%	2.59%	3.39%	7.03%	1.14%
	L&M		180	2.35%	-1.45%	1.55%	2.39%	3.20%	5.18%	1.14%

Table A.2. shows the number of words and descriptive statistics of Net Optimism scores per year. The mean number of words per CEO letter varies over time without a coherent trend. The scores for Net Optimism decrease during 2008 and 2009, otherwise on a more stable level. In Table A3, we see the results per industry. Finally, in Table A.4, we illustrate our sample by including the top and bottom three observations per year as measured by Net Optimism.

Table A.3.

Descriptive statistics over Net Optimism per industry

This table shows the descriptive statistics of Net Optimism variables measured by DICTION and L&M. Shown per industry using 2-digits SIC-codes. Reported number of observations, mean, min, quarters, median, max and standard deviation.

Industry	Measure	N	mean	min	p25	p50	p75	max	sd
Agriculture, Forestry, Fishing	DICTION	13	0.87%	-1.04%	0.20%	0.73%	1.97%	2.56%	1.08%
	L&M	13	0.98%	-1.45%	0.11%	0.70%	2.50%	3.10%	1.48%
Construction	DICTION	49	2.31%	0.36%	1.85%	2.19%	2.85%	4.34%	0.86%
	L&M	49	2.07%	0.26%	1.40%	2.03%	2.59%	4.87%	0.98%
Finance, Insurance,	DICTION	209	2.16%	-1.86%	1.22%	2.05%	3.06%	6.30%	1.37%
Real Estate	L&M	209	1.91%	-2.42%	1.03%	1.94%	2.74%	6.92%	1.41%
Manufacturing	DICTION	685	2.56%	-1.65%	1.89%	2.54%	3.26%	7.03%	1.12%
	L&M	685	2.22%	-1.81%	1.46%	2.26%	3.00%	5.77%	1.17%
Mining	DICTION	25	2.26%	0.07%	1.76%	2.35%	2.75%	5.07%	1.01%
	L&M	25	1.26%	-0.38%	0.84%	1.43%	1.63%	2.57%	0.73%
Retail Trade	DICTION	73	2.62%	0.12%	1.99%	2.76%	3.38%	4.26%	1.04%
	L&M	73	2.33%	-0.92%	1.66%	2.49%	3.20%	4.40%	1.09%
Services	DICTION	319	2.61%	0.10%	1.97%	2.57%	3.35%	5.18%	0.98%
	L&M	319	2.38%	-0.94%	1.75%	2.38%	3.07%	5.45%	1.06%
Transportation &	DICTION	105	2.32%	-0.33%	1.60%	2.33%	3.03%	5.12%	1.06%
Public Utilities	L&M	105	1.96%	-0.94%	1.18%	1.94%	2.75%	5.77%	1.21%
Wholesale Trade	DICTION	28	2.85%	1.27%	2.22%	2.78%	3.10%	7.11%	1.12%
	L&M	28	2.76%	0.63%	1.92%	2.59%	3.37%	5.23%	1.22%

Table A.4.

Top & Bottom Net Optimism

This table summarizes the top and bottom firms by Net Optimism per year.

2004	2005	2006	2007	2008
		Most Optimistic		
Sandvik	Kaupthing Bank	Swedbank	VBG	Addtech
Ericsson B	Net Insight	Munters	ÅF	Björn Borg
Lundin Petroleum	Investor	SEB	SEB	Oriflame
		Least Optimistic		
Fingerprint Cards	Biotage	Traction	Vostok Nafta	Traction
CTT Systems	Traction	Rottneros	Ledstiernan	Castellum
Wallenstam	Active Biotech	Ratos	Stora Enso	Ratos

2009	2010	2011	2012	2013
		Most Optimistic		
Global Health Partner	Atrium Ljungberg	NOTE	Elekta	Meda
Oriflame	Fabege	Wallenstam	Oriflame	Axis
Björn Borg	ABB	Net Entertainment	DORO	Castellum
		Least Optimistic		
Nordnet	Traction	New Wave Group	Concordia Maritime	Trigon Agri
Vostok Nafta	Sensys Traffic	Traction	NAXS	Black Earth Farming
Castellum	Black Earth Farming	Black Earth Farming	SSAB	Active Biotech

9.3 Supplementary descriptive statistics

Table A.5.

Descriptive statistics of variables

This table shows the descriptive statistics of variables included in the regressions. Financial variables: Return on Assets (*ROA*), scaled by log, profit margin *(PM*), cash flow from operations (CFO), standard deviation of ROA, PM, CFO (*sdROA, sdPM, sdCFO*), accruals (*ACC*), sales (*REV*), market-to-book (*MTB*). Other explanatory variables are: *AGE* is the age of the company for each observation point in time scaled by log, *CHANGE* is 1 if first time CEO writes the letter and 0 otherwise *SIGN* is 1 if CEO has signed the report and 0 otherwise. Measures of future firm performance: future ROA, CFO, FPM indicated by <F> and horizon. Dummy variables for year and industry not displayed in this table. Reported number of observations, mean, median, quarters, 1st and 99th percentile, minimum and maximum and standard deviation.

Variable	N	mean	min	p1	p25	p50	p75	p99	max	sd
Financial variables										
ROA	1,481	0.0744	-1.183	-0.614	0.0268	0.0820	0.147	0.483	1.580	0.178
sdROA	1,450	0.0717	0.000295	0.00184	0.0191	0.0401	0.0819	0.543	1.517	0.106
PM	1,488	-0.105	-30.67	-6.124	0.0168	0.0589	0.124	1.833	20.06	2.127
sdPM	1,471	1.050	0.000307	0.00130	0.0125	0.0313	0.108	10.14	525.2	19.10
CFO	1,494	0.0657	-1.644	-0.559	0.0286	0.0776	0.127	0.354	0.856	0.147
sdCFO	1,494	0.0337	0.00104	0.000387	0.00752	0.0197	0.0397	0.236	1.353	0.0612
ACC	1,494	-0.0361	-1.373	-0.443	-0.0663	-0.0280	0.00439	0.240	0.722	0.113
REV	1,484	14.94	7.907	9.399	13.67	14.82	16.39	19.16	19.74	2.114
MTB	1,484	0.586	-3.646	-2.054	-0.155	0.457	1.207	4.369	6.096	1.269
Other explanatory variables										
AGE	1,498	3.323	0	0.693	2.565	3.178	4.248	5.883	6.001	1.097
CHANGE	1,506	0.129	0	0	0	0	0	1	1	0.335
SIGN	1,506	0.593	0	0	0	1	1	1	1	0.491
Measures of future firm performance										
FROA1	1,448	0.0732	-1.017	-0.598	0.0256	0.0813	0.144	0.477	1.580	0.172
FROA2	1,253	0.0727	-0.808	-0.548	0.0260	0.0757	0.140	0.439	1.029	0.154
FCFO1	1,494	0.0744	-1.150	-0.558	0.0282	0.0809	0.136	0.434	1.199	0.152
FCFO2	1,315	0.0790	-0.891	-0.527	0.0319	0.0811	0.140	0.440	0.896	0.145
FPM1	1,466	-0.0747	-30.67	-5.051	0.0166	0.0595	0.123	1.858	22.54	2.275
FPM2	1,275	-0.0664	-21.68	-6.419	0.0146	0.0568	0.119	1.789	11.42	1.618
FTQ1	1,469	2.821	0.190	0.435	0.977	1.281	2.060	29.60	146.7	7.084
FTQ2	1,279	2.801	0.255	0.483	0.986	1.303	2.086	32.24	77.66	6.145

Table 8.3. is a more comprehensive type of Table 6. The data set exhibits some extreme values as displayed in the min and max column in comparison to the 1th and 99th percentage level.

9.4 Sensitivity analysis: Alternative Net Optimism measure

Table A.6.

Net Optimism and Future Operating Firm Performance

$$FFP_{it} = \beta_0 + \beta_1 NETOPT_{it} + \beta_2 FFP_{it-1} + \beta_3 FFP_{it} + \beta_4 REV_{it} + \sum_j \beta_5 IND_{ijt} + \sum_k \beta_6 YEAR_{ikt} + \varepsilon_{it}$$

This table shows the regression results for the sensitivity test using the financial dictionary of Loughran &McDonald. The table is specified as in table 9. Dummy variables for industry and year are not reported. Robust two-tailed t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (one-tailed when sign of coefficient is predicted; two-tailed otherwise)

		(1)	(2)	(3)	(4)	(5)	(6)
		FROA		FCFO		FPM	
Horizon		t+1	t+2	t+1	t+2	t+1	t+2
Variables	Pred. Sign						
NETOPT	+	0.555***	0.330*	0.113	0.00405	0.935**	0.940**
		(2.513)	(1.486)	(0.673)	(0.0255)	(2.056)	(2.011)
FFP_{t-1}	+	0.422***	0.411***	0.448***	0.427***	0.0828***	0.0611***
		(16.34)	(15.95)	(21.30)	(19.81)	(4.312)	(3.317)
sdFFP	-	-0.0396	-0.0446	-0.0701	-0.142*	-0.00999	-0.0240**
		(-0.683)	(-0.753)	(-1.030)	(-1.938)	(-0.865)	(-2.067)
REV	?	0.000769	0.00257	-0.000465	-0.00186*	0.0127***	0.0168***
		(0.543)	(1.619)	(-0.469)	(-1.899)	(3.873)	(5.002)
Constant		0.0108	-0.00938	0.0377*	0.0799***	-0.283***	-0.333***
		(0.345)	(-0.290)	(1.914)	(4.132)	(-3.988)	(-5.438)
Observations		1,403	1,212	1,484	1,306	1,441	1,253
R-squared		0.447	0.457	0.487	0.498	0.279	0.301

Table A.7.

Net Optimism and Future Tobin's Q

$$FTQ_{it} = \beta_0 + \beta_1 NETOPT_{it} + \beta_2 REV_{it} + \beta_3 AGE_{it} + \sum_j \beta_4 IND_{ijt} + \sum_k \beta_5 YEAR_{ikt} + \varepsilon_{it}$$

This table shows the regression results for the sensitivity test using the financial dictionary of Loughran &McDonald on Tobin's Q. The table is specified as in table 10. Dummy variables for industry and year not reported. Robust two-tailed t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (one-tailed when sign of coefficient is predicted; two-tailed otherwise)

		(1)	(2)
Horizon		t+1	t+2
Variables	Pred. Sign		
NETOPT	+	3.457***	3.199**
		(2.434)	(2.118)
REV	+	-0.0405***	-0.0487***
		(-4.027)	(-4.617)
AGE	-	-0.0686***	-0.0605***
		(-3.647)	(-3.022)
Constant		0.352**	0.599***
		(2.181)	(3.533)
Observations		1,452	1,263
R-squared		0.210	0.210

Table A.8.

Portfolio regressions

This table displays monthly four factor abnormal returns for the sensitivity test using the financial dictionary of Loughran & McDonald. The table is specified as in table 12. The sample period is 5/05-5/15.T-statistiscs in parenthesis. ***p<0.01, **p<0.05, *p<0.1

	(1)	(2)	(3)
Portfolios:	P0020	P8000	LSP8020
Full sample	-1.1	0.188	1.29
	(-1.218)	(0.373)	(1.536)
Excl Financials	-1.591	0.256	1.85*
	(-1.563)	(0.525)	(1.912)
Large	-1.18	0.371	1.55
	(-1.316)	(0.753)	(1.864)
Small	0.0193	0.308	0.288
	(0.0292)	(0.588)	(0.596)

9.5 Sensitivity analysis: Excluding financials

Table A.9.

Net Optimism and Future Operating Performance

$$FFP_{it} = \beta_0 + \beta_1 NETOPT_{it} + \beta_2 FFP_{it-1} + \beta_3 \delta FFP_{it} + \beta_4 REV_{it} + \sum_j \beta_5 IND_{ijt} + \sum_k \beta_6 YEAR_{ikt} + \varepsilon_{it}$$

This table shows the regression results for sensitivity test excluding financials on future operating performance. The table is specified as in table 9. Dummy variables for industry and year are not reported. Robust two-tailed t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (one-tailed when sign of coefficient is predicted; two-tailed otherwise)

		(1)	(2)	(3)	(4)	(5)	(6)
		FROA		FCFO		FPM	
Horizon		t+1	t+2	t+1	t+2	t+1	t+2
Variables	Pred. Sign						
NETOPT	+	1.252***	0.982***	0.437***	0.428***	1.895***	1.670***
		(5.470)	(3.994)	(2.564)	(2.409)	(4.414)	(3.752)
FFP_{t-1}	+	0.452***	0.453***	0.454***	0.422***	0.117***	0.108***
		(19.95)	(19.15)	(21.97)	(19.16)	(6.337)	(5.641)
sdFFP	-	-0.0123	-0.0513	-0.0256	-0.135*	-0.00305	-0.0130
		(-0.236)	(-0.894)	(-0.368)	(-1.755)	(-0.263)	(-1.084)
REV	?	0.00141	0.00332**	0.000213	-0.00101	0.0154***	0.0189***
		(0.970)	(2.035)	(0.185)	(-0.873)	(4.984)	(5.636)
Constant		-0.00941	-0.0290	0.0218	0.0599***	-0.335***	-0.369***
		(-0.290)	(-0.851)	(1.007)	(2.738)	(-4.950)	(-6.058)
Observations		1,209	1,047	1,280	1,127	1,245	1,084
R-squared		0.550	0.568	0.484	0.484	0.343	0.365

Table A.10.

Net Optimism and Future Tobin's Q

$$FTQ_{it} = \beta_0 + \beta_1 NETOPT_{it} + \beta_2 REV_{it} + \beta_3 AGE_{it} + \sum_j \beta_4 IND_{ijt} + \sum_k \beta_5 YEAR_{ikt} + \varepsilon_{it}$$

This table shows the regression results for the sensitivity test excluding financials on Tobin's Q. The table is specified as in table 10. Dummy variables for industry and year not reported. Robust two-tailed t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (one-tailed when sign of coefficient is predicted; two-tailed otherwise)

		(1)	(2)
Horizon		t+1	t+2
Variables	Pred. Sign		
NETOPT	+	3.880**	3.295*
		(1.973)	(1.533)
REV	+	-0.0474***	-0.0570***
		(-4.141)	(-4.747)
AGE	-	-0.0800***	-0.0717***
		(-3.600)	(-3.024)
Constant		0.473***	0.756***
		(2.613)	(3.966)
Observations		1,253	1,092
R-squared		0.144	0.138

9.6 Sensitivity analysis: Value-weighted returns

Table A.11.

Value-Weighted + Original Net Optimism measure

This table displays monthly four factor abnormal returns for the sensitivity test using value-weighted returns. The table is specified as in table 12. The sample period is 5/05-5/15.T-statistiscs in parenthesis. ***p<0.01, **p<0.05, *p<0.1

	(1)	(2)	(3)
Portfolios:	P0020	P8000	LSP8020
Full sample	0.33	0.41	0.07
	(0.62)	(0.83)	(0.26)
Excl Financials	0.24	0.24	0.24
	(0.43)	(0.75)	(0.44)
Large	0.57	0.57	0.57
	(1.03)	(1.11)	(-0.15)
Small	0.47	0.24	-0.23
	(0.78)	(0.44)	(-0.55)

Table A.12.

Value-Weighted + Other Net Optimism measure

This table displays monthly four factor abnormal returns for sensitivity test using value-weighted return and the alternative net optimism measure using the financial dictionary from Loughran & McDonald. The table is specified as in table 12. The sample period is 5/05-5/15.T-statistiscs in parenthesis. p-values denoted, ***p<0.01, **p<0.05, *p<0.1

	(1)	(2)	(3)
Portfolios:	P0020	P8000	LSP8020
Full sample	-1.58	-0.0603	1.52
	(-1.302)	(-0.134)	(1.315)
Excl Financials	-2.17	-0.127	2.04
	(-1.641)	(-0.268)	(1.593)
Large	-0.832	-0.113	0.719
	(-0.741)	(-0.249)	(0.650)
Small	0.415	0.153	-0.262
	(0.656)	(0.274)	(-0.556)