Norwegian School of Economics

Bergen, Spring, 2019

NHH

Investor Sentiment and Stock Returns

Using Newspaper and Google Search Sentiments to Predict Returns on Oslo Stock Exchange

Bård Øymar and Maria Myrseth

Supervisor: Aksel Mjøs

Master thesis, MSc in Economics and Business Administration, Finance

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

We hypothesize that current sentiments can predict future stock returns, and we construct sentiment indexes based on Norwegian newspapers and Norwegian Google searches respectively. The indexes measure changes in the occurrence of economic terms with a negative sentiment, like *refinancing*, *recession* and *fraud*. We investigate if the indexes predict future returns on Oslo Stock Exchange.

Our first finding is that an increase in a weekly newspaper index predicts negative return two weeks later. A one standard deviation increase in the index is associated with 0.4% lower return for the 10% largest stocks. The effect is only apparent for large stocks. Our finding suggests that the index explains 0.6% of returns in week two.

The second finding is that an increase in a monthly Google search index predicts positive return the next month. A one standard deviation increase in the index is associated with a 1.2% higher return the subsequent month. The effect is strongest for large stocks. Our finding suggests that the index explains 5.9% of next month's return.

Contents

ABSTRACT **2**

ACNOWLEDGEMENTS **6**

1. INTRODUCTION **7**

1.1 HYPOTHESIS, MOTIVATION AND VALUE OF THE THESIS 10

2. LITTERATURE REVIEW **12**

3. EMPIRICAL METHOD **14**

3.1 DATA ON STOCK RETURNS 14

3.2 CONSTRUCTION OF SENTIMENT INDEXES 19

3.2.1 Newspaper corpus *20*

3.2.2 Google Trends *20*

3.2.3 Term selection *23*

3.2.4 Processing of each term *24*

3.2.5 Index composition *25*

3.2.6 Index definition *28*

3.2.7 Sentiment Index statistics *29*

3.3 OTHER DATA AND VARIABLES 31

3.3.1 Fund Flow *32*

3.3.2 Corporate Confidence *33*

3.3.3 Consumer Confidence *33*

3.3.4 VIX *34*

3.3.5 Spread *35*

3.3.6 Brent *35*

3.3.7 Volatility *36*

3.3.8 Summary statistics for control variables *36*

3.4 EMPIRICAL STRATEGY 38

3.4.1 Average returns 38

3.4.2 Cross-sectional analysis 40

3.5 ECONOMETRIC CONCERNS 41

3.6 VALIDITY AND RELIABILITY 42

4. RESULTS AND DISCUSSIONS 44

4.1 WEEKLY NEWSPAPER INDEX 44

4.1.1 Average return 44

4.1.2 Cross-sectional analysis 49

4.1.3 Conclusions 52

4.2 MONTHLY NEWSPAPER INDEX 52

4.2.1 Average return 52

4.2.2 Cross-sectional analysis 54

4.2.3 Conclusions 54

4.3 MONTHLY GOOGLE SEARCH INDEX 55

4.3.1 Average return 55

4.3.2 Cross-sectional analysis 61

4.3.3 Conclusions 64

4.4 ROBUSTNESS CHECKS 64

4.5 DISCUSSIONS AND COMPARISONS 65

4.5.1 Comparison to Larsen and Thorsrud (2017) 66

4.5.2 Comparison to Fisher and Statman (2000) 66

4.5.3 Discussion of findings on Newspaper Index 67

4.5.4 Discussion of findings on the Google Search Index 69

5. CONCLUSIONS **70**

6. APPENDIXES **71**

6.1 APPENDIX 1: ALTERNATIVE TABLE, WEEKLY NEFA, AVERAGE RETURN 71

6.2 APPENDIX 2: ALTERNATIVE TABLE, WEEKLY NEFA, CROSS CHECKS 72

6.3 APPENDIX 3: ALTERNATIVE TABLE, WEEKLY NEFA, SIZE PORTFOLIOS 73

6.4 APPENDIX 4: ALTERNATIVE TABLE, WEEKLY NEFA, SIZE PORTFOLIOS 74

6.5 APPENDIX 5: ALTERNATIVE TABLE, MONTHLY FEARS, SIZE PORTFOLIOS 75

REFERENCES **76**

Acnowledgements

Our work has relied on help from several people. We hereby thank the following persons for their contributions.

First and foremost Aksel Mjøs at NHH for his supervision. It has stimulated us first to explore alternative options, and later directed us in the right directions. Gisle Andersen at NHH for helping us with the newspaper corpus and linguistic concerns in the construction of our sentiment indexes. Øivind Anti Nilsen at NHH for helping us with econometric concerns and interpretations. Bernt Arne Ødegaard at NHH and UiS for answering questions about his data. Our gratitude also extends to the fact that he has constructed and published much of the data we rely on. The Norwegian Fund and Asset Management Association with Ida Aamodt-Hansen for providing us with data that is not publicly available.

1. Introduction

At least since Keynes introduced the notion of animal spirits, it has been thought that emotions and attitudes are drivers in the economic outlook (Keynes, 1936). This spirit is a crowd psychology, which implies that not only individual investors, but also crowds of investors might exhibit the same spirit. This has later been formalized in the finance literature as an investor sentiment (Schleifer and Summers, 1990). Baker and Wurgler (2007) defines the investor sentiment as a belief about future cash flows and investment risks that is not justified by the facts at hand. They argue that this sentiment also affects investment decisions for crowd of investors.

The value of stocks should be determined by all available information that concerns its value. Only new information should move prices, and this should move prices in the direction and magnitude that the information implies (Bodie, Kane and Marcus, 2018, pp. 334). This does however depend on the finite abilities of human beings, and it is not obvious that prices are not moved by changes in sentiments or biases, such that the efficient market hypothesis postulates (Fama, 1970).

A behavioural finance tradition focuses on the consequences of these finite abilities, particularly of individual retail investors. Research shows that individual investors suffer from faults like the availability bias, where they are biased towards investing in assets that are easily available and familiar (Barber and Odean, 2008). They suffer from the disposition effect, where they sell winning assets too soon, in order to realize a gain, and keep losers too long, in order to avoid realizing a loss (Odean, 1998). They are also overconfident in the ability to predict the market (Heuer, Merkle and Weber, 2017). An overview of biases and dispositions can be found in Barber and Odean (2013).

Yet the efficient market hypothesis (EMH) postulates that other investor groups counteract any investments based on biases and sentiments, such that the overall market is efficient (Bodie, Kane and Marcus, 2018, pp. 334). EMH states that the value of a stock, or an overall market, already reflects all the available information regarding its value (Fama, 1970). If an asset receives positive attention by some investors who press prices up, other investors will sell the asset such that the price stays at market consensus. This is true for all levels of the market, whether it regards a single stock, a sector or the whole market[1].

In an efficient market, new information should not take long to move prices, but move them as soon as investors are able to trade based on the information (Bodie, Kane and Marcus, 2018, pp. 334). As news are priced in the market immediately, and other investors counteract any mispricing put forth by individuals, neither biases nor an investor sentiment should cause market inefficiencies. This implies that it should be impossible to beat the market by studying this behaviour and acting on it.

Yet much literature find evidence of market inefficiencies. We discuss some of these in chapter 2. One explanation is that there are limitations to arbitrage. Grossman and Stiglitz (1980) shows that there are practical limitations to the efficiency of markets, and any mispricing will only be exploited by others if the reward from doing so exceeds the cost. A perfectly efficient market is thus impossible, but this does not exclude the possibility of highly efficient markets. In the words of Bodie, Kane and Marcus: "*Rather than ask the qualitative question, Are markets efficient? we ought instead to ask a more quantitative question: How efficient are markets?"* (Bodie, Kane and Marcus, 2018, pp. 347).

As markets are not perfectly efficient, it is reasonable to expect that the investor biases and sentiments we have discussed are not fully counteracted by other investors. If this is true, then some measurement of this might be able to predict the movements of subsequent stock prices. This idea prerequisite that other investors do not fully counteract irrational sentiment investments and that the market is not fully efficient.

However, it is not easy to obtain a broad measure for sentiments. We suggest that the variation of linguistic terms that occupies us is a good proxy for this. This approach has proven to be successful in several studies (Herve, Zouaoui and Belvaux, 2019; Da, Engelberg and Gao, 2015; Tetlock, Saar-Tsechansy and Macskassy, 2008; Tetlock, 2007). Whenever some terms occur more often, it suggests that we are relatively more concerned with whatever the term denotes. When for instance the term *recession* occurs more often, it suggests that one might

[1] Yet some make the case that EMH describes the micro level of individual stocks better than the macro level of the market (Jung and Schiller, 2002).

have a worse outlook on the economy. Moreover, this outlook might lead to a subsequent move in asset prices.

We suggest that newspapers and Google searches are both sources where the occurrence of terms fluctuate along with an economic outlook, and that these make good proxies for the information investors are exposed to and what goes on in the mind of investors.

We construct two alternative sentiment indexes from Norwegian newspapers and Google searches respectively. We use the same method counting the occurrence of economic terms with an unambiguous sentiment. The method is suggested by Da, Engelberg and Gao (2015). For newspapers, we construct both a weekly and monthly index, and for Google searches, we construct a monthly index. We investigate if the indexes can predict average returns on Oslo stock exchange and cross-sections of small and large stocks.

We utilize terms like *recession*, *financial crisis* and *refinancing* (we use the equivalent Norwegian terms). We are thus not aiming at measuring specific news regarding the value of stocks, but rather the spirit or attitude of the newspapers and Google users. Our investigation is thus founded in the belief that the spirit of investors affects investment decisions, and that it is possible to obtain a measure for this spirit.

Newspapers communicate information to the public. Some of this information should affect stock prices; it can concern the expectancy of future revenues of a single stock (like a new contract), or the outlook of the whole market (like news on brexit). We call this fundamental news. An efficient market will react to such news and adjust prices accordingly. It is however obvious that most information directly concerning the value of stocks will reach the market before it is read in newspapers. Newspapers also publish news that should not affect stock prices, like an article on the housing market crash of Norway in 1899. We call this noise, in accordance with Kyle (1985) and Black (1986).

As newspapers publish both irrelevant information and relevant information that should immediately be obtained by the market, it will not be possible in an efficient market to predict subsequent stock prices by studying the sentiment of newspapers. In a market with some degree of inefficiencies, this is however not necessarily impossible.

By simply counting the occurrence of terms in newspapers, we are not able to distinguish fundamental news from noise. If newspapers are able to predict changes in stock prices in a subsequent period, it can be either due to fundamental news taking time to settle, or potential noise affecting asset prices[2].

All people and not only investors perform Google searches. It is thus a source to understand what occupies the mind of a representative Google user. Yet when some group is relatively more occupied with a topic or notion, it is quite likely to be reflected in increased search activity on that topic (Da, Engelberg and Gao, 2015). Given the role of investor sentiments, this can potentially either coincide with or precede changes in stock prices.

Our empirical investigation is concerned with the Norwegian equity market. NOU 2018: 5 (2018) and Norges Bank (2018) gives a comprehensive overview over the Norwegian financial system and markets. Norway has well-functioning and internationally integrated markets. A few large companies dominate the Norwegian equity market. By 2016, the five largest companies make up more than 54% of OSE (NOU 2018: 5, 2018, pp. 41). Some studies make the case that small stocks have a larger fraction of private retail investors and are thus more sensitive to changes in an investor sentiment (Kumar and Lee, 2006; Barber, Odean and Zhu 2009). OSE is thus particularly fitting to identify any differences between small and large stocks.

1.1 Hypothesis, motivation and value of the thesis

Our hypothesis is that sentiment in the current can predict stock returns in the future. Whenever the sentiment is either positive or negative, this contributes towards either increasing or decreasing stock prices. To investigate this, we limit the scope to our constructed indexes and Oslo stock exchange (OSE). Secondarily we investigate if the sentiment has different predictions for small and large stocks.

We are not the first ones aiming at answering this hypothesis, yet there are reasons that motivate our investigation. Most related literature investigates American markets, which differ in several ways, most notably in size. Another issue is that a sentiment is a rather abstract entity. It is not straightforward to comprehend what is being measured, and the method applied

[2] Some papers do however argue that noise leads to price-reversal and a return to fundamentals, whereas fundamental news leads to price continuation, due to initial under reaction (French and Roll, 1986; Larsen and Thorsrud, 2017).

is of importance. As far as we know, we are the first ones applying the chosen method to measure long-term sentiments of newspapers and Google searches.

The value of the paper is twofold. The sentiment indexes can potentially contribute towards predicting future stock returns. In addition, the indexes can contribute towards understanding the role of newspapers and Google searches in the pricing of stock markets.

The rest of the thesis is structured as follows. Chapter 2 reviews existing literature. Chapter 3 explains our empirical method, hereunder data samples, index construction and model specifications. Chapter 4 presents and discusses results. Chapter 5 concludes.

2. Litterature review

Several studies construct different sentiment indexes and relate these to asset prices, the most famous one being Lee, Schleifer and Thaler (1991). They argue that the discount on Closed-End Funds (CEFs) serve as a proxy for investor sentiment. CEFs are publicly traded funds that often trade at a discount relative to the funds' assets. They argue that the price differs due to the sentiment of less informed individual investors who invest in CEFs, and not directly in the underlying assets. This is supported by findings that CEF discount correlate with the return on small cap stocks, who are relatively more held by the same group of individual investors.

Warther (1995) finds that stock returns are positively correlated with unexpected monthly fund flows, but not expected flows. Other papers also find evidence that mutual fund flows serve as a proxy for investor sentiment (Ben-Rephael, Kandel and Wohl, 2012).

Da, Engelberg and Gao (2015) creates a sentiment index based on the daily occurrence of search terms in Google. Single search terms that are significantly correlated with the return of broad indexes are combined in a composite indicator named FEARS – Financial and Economic Attitudes Revealed by Search. The search terms have negative, unambiguous economical interpretations, like *recession* and *bankruptcy*. They find that an increase in FEARS predicts a decrease in stock prices the current day, followed by a partly reversal during the two following days. The effect is stronger in stocks that are attractive to individual retail investors and are harder to arbitrage. This latter result supports the assumption that the indicator serves as a proxy for investor sentiment, and that the sentiment matters for stock prices.

In a similar fashion, Joseph, Wintoki and Zhang (2011) uses Google search occurrence for single stock tickers. They find that the current week's search intensity predicts next week's stock returns and trading volume. They also find that the effect is stronger in stocks that are volatile and hard to arbitrage. This paper investigates single stock returns and not average returns, but it supports the notion that Google searches serves as a proxy for investor sentiment.

Herve, Zouaoui and Belvaux (2019) examines investor sentiments and the return of French stocks. They also measure investor sentiment based on Google and find that it predicts stock returns. An alternative index constructed from data on Wikipedia page traffic does not yield the same results.

Fisher and Statman (2000) measures the monthly sentiment of three groups of market participants, namely Wall Street strategists, newsletter writers and individual investors. The sentiment measures are based on surveys of the groups. They find a significant correlation between newsletter writers and individual investors, whereas the strategists do not correlate with the other two groups. Another finding is that the sentiment of both individual investors and newsletter writers partly reflect an expectation that short-term development continues, meaning that short-term positive return creates bullish sentiment. Further, they find that a positive sentiment amongst individual investors predicts a negative return to S&P500 the next month.

Barber and Odean (2008) and Fang and Peress (2009) argues that media coverage of single stocks predicts stock returns. The theory is that investors are biased towards investing in stocks that are easily available, and that this effect moves prices. We are concerned with average returns and not the return of single stocks, but these studies support the notion that newspapers' sentiment matter for stock prices.

Tetlock (2007) measures the sentiment in a popular column in Wall Street Journal. This is done by classifying every word in the column, ranging from positive to negative. He finds that a high pessimistic sentiment predicts negative daily returns on the Dow Jones Industrial Average, followed by a return to fundamentals. Whenever the sentiment is unusually high it predicts unusually high trading volume, independent of the sentiment being positive or negative.

Larsen and Thorsrud (2017) analyzes the content of the Norwegian business newspaper Dagens Næringsliv (DN) by classifying each article into topics and sentiment. They find that the content significantly predicts a daily return that continues, peaking after 14 days. The study applies a significantly different methodology from ours, and it is thus relevant to compare our results to those of Larsen and Thorsrud. This will be done in subchapter 4.5.1.

3. Empirical method

We have hypothesised that the sentiment in the current can predict stock returns in the future. We now proceed to explain how we test this empirically. Secondarily we want to test if the sentiment has different predictions for small and large stocks.

In the following we explain our data, define variables, explain the construction of our sentiment indexes and the strategy of our empirical investigations. As data on average return is needed in the construction of the indexes, this is defined under 3.1. Part 3.2 explains and defines the construction of our indexes. Part 3.3 explains and defines any other variables. In part 3.4 we explain our empirical strategy and specify models. In part 3.5 we highlight some econometric concerns and explain how we deal with these. Part 3.6 is dedicated to reflections on the validity of our investigation.

3.1 Data on stock returns

We use several alternative measures for the return of OSE. They are all provided by Bernt Arne Ødegaard[3]. We use both weekly and monthly calculated returns. The monthly returns are obtained directly from Ødegaard. We calculate the weekly return based on daily returns from Ødegaard. The cumulative weekly return is defined as[4]:

$$((1+r_{monday})\times(1+r_{tuesday})\times(1+r_{wednesday})\times(1+r_{thursday})\times(1+r_{friday}))-1=r_{week} \quad (1)$$

The following figure shows the development of OSE during the relevant sample period from 1998 until 2018.

[3] Data from Ødegaard, Bernt Arne are found at:
http://finance.bi.no/~bernt/financial_data/ose_asset_pricing_data/index.html

[4] This formula corresponds to Ødegaards data, as the published monthly return corresponds to the monthly return we obtain by applying the formula above with the return for each day of the month.

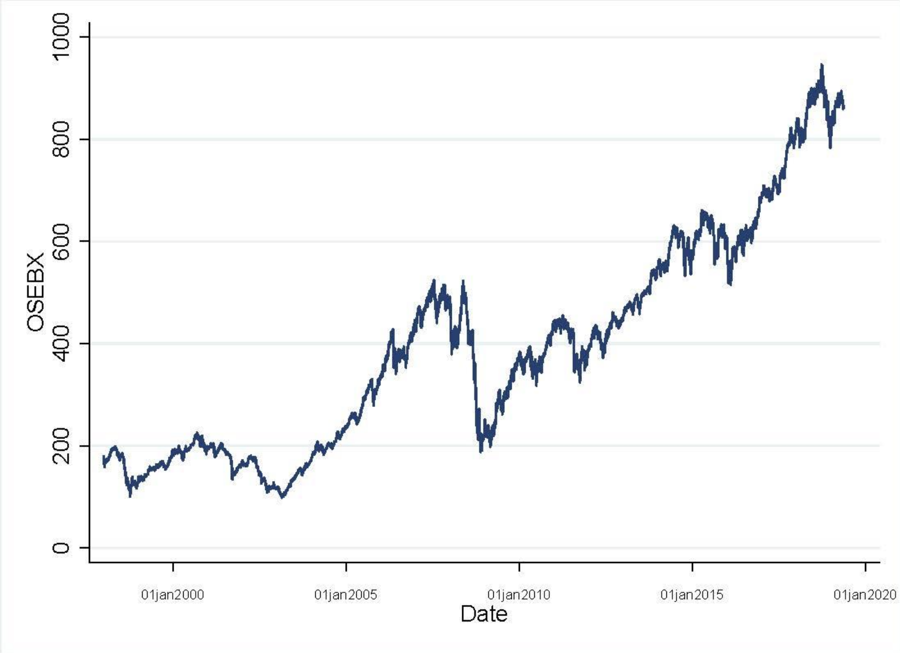

Figure 1: Graphical presentation of the main OSEBX index on OSE from 1998-2018.

Average return

For average return, we use two main measures, *Allshare* and *OBX*. Two alternatives, *VW* and *EW*, are later used to test if our findings are consistent when using alternative measures. We first present the variables verbally. At the end of the subchapter, we present a table with summary statistics.

Allshare is a value-weighted portfolio of all stocks on OSE. The amount of stocks included has varied during the sample period, averaging at approximately 210 stocks (Ødegaard, 2019). *OBX* is a value-weighted portfolio consisting of the 25 most liquid stocks on OSE. Neither *Allshare* nor *OBX* include dividends. We later explore if our sentiment indexes work better at predicting the return of *Allshare* or *OBX*, and thereby answer if the sentiment indexes have different implications for small and large stocks.

VW is a value-weighted portfolio of stocks on OSE, where the return includes dividends[5]. Small and illiquid stocks are filtered out from the portfolio. A specification of the criterions is found in Ødegaard (2019). *EW* is a similarly constructed portfolio, but with an equally

---

[5] We find no evidence that the inclusion of dividends affect our analysis. As *EW* and *VW* are only used to control our results, we do not discuss this in detail.

weighting of the stocks. *EW* thus gives small stocks a relatively greater weight than *VW*. We also explore if our sentiment indexes work better at predicting the return of *EW* or *VW*.

We define each variable as *Allshare Return, OBX Return, EW Return* and *VW Return.* Occasionally we refer to returns in general; we note this simply as *Return*. In this latter case the model we specify or the assertion we make is valid for all four return variables.

Figure 2 shows the weekly return of *Allshare*, as well as the distribution of weekly returns for *Allshare*. Figure 2 manifests large movements around the financial crisis of 2008. We will later apply yearly dummies in our analysis to account for yearly characteristics.

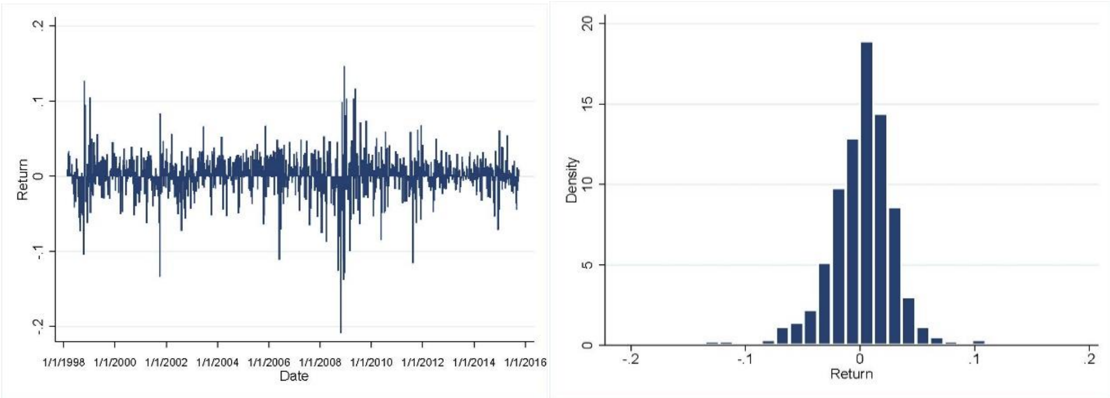

Figure 2: To the left: Graphical presentation of weekly Allshare return for the period 1998-2015. To the right: the distribution of weekly Allshare return.

Table 1 shows summary statistics for weekly returns between January 1998 and September 2015. This is the relevant sample period for our weekly analysis. Note that *EW* and *VW* has a higher mean due to the inclusion of dividends in the calculations. The week with the most negative return is the second week in October 2008. The numbers are in percentages, and are not annualized.

Return	**Mean**	**Std. Dev**	**Min**	**0.25**	**Median**	**0.75**	**Max**
Allshare	0.180	3.023	-20.881	-1.336	0.490	1.910	14.660
OBX	0.167	3.350	-21.948	-1.331	0.454	1.924	18.334
EW	0.467	2.095	-14.875	-0.458	0.673	1.559	12.019
VW	0.430	2.996	-19.798	-1.018	0.623	2.093	15.673

Table 1: Summary statistics for weekly return from January 1998 to September 2015.

Table 2 shows summary statistics for monthly return between January 1998 and December 2018. This is the relevant sample period for our monthly analysis. The numbers are in percentages, and the numbers are not annualized.

Return	Mean	Std. Dev	Min	0.25	Median	0.75	Max
Allshare	0.847	5.755	-23.934	-2.285	1.290	4.305	15.047
OBX	0.782	6.195	-25.352	-2.573	1.135	4.684	17.225
EW	1.238	4.658	-18.327	-1.202	1.530	3.979	12.189
VW	1.610	5.578	-21.041	-1.431	1.748	5.157	16.715

Table 2: Summary statistic for monthly return from January 1998 to December 2018.

Size portfolios

We also use size portfolios constructed by Ødegaard. Illiquid stocks are first filtered out; then stocks are divided into ten portfolios dependent on market size (Ødegaard, 2019). Each portfolio thus entails approximately the same number of stocks. The internal weighting inside each portfolio is equal. As few large stocks dominate OSE, portfolio 10 (with the 10% largest stocks) have a larger market value than the other nine portfolios combined. We are thus able to compare effects on small stocks and large stocks.

Table 3 shows summary statistics for the weekly return of the size portfolios between January 1998 and December 2015. The values are in percentages. Note that the sample period is of importance. We find significantly different statistics for different sample periods.

Return	Mean	Std. Dev	Min	0.25	Median	0.75	Max
Portfolio 1	0.595	1.910	-5.653	-0.564	0.487	1.671	13.263
Portfolio 2	0.644	2.514	-10.243	-0.789	0.436	1.904	14.131
Portfolio 3	0.595	2.440	-12.243	-0.777	0.598	2.000	16.454
Portfolio 4	0.601	2.731	-11.496	-0.863	0.595	1.949	20.438
Portfolio 5	0.509	2.692	-13.720	-0.874	0.443	1.858	17.270
Portfolio 6	0.416	2.813	-17.248	-0.965	0.582	2.047	15.945
Portfolio 7	0.384	2.774	-13.857	-0.976	0.526	1.911	16.358
Portfolio 8	0.356	3.035	-21.592	-1.228	0.395	2.020	20.156
Portfolio 9	0.140	3.568	-17.441	-1.596	0.388	2.107	19.002
Portfolio 10	0.210	3.719	-24.884	-1.437	0.481	1.988	24.759

Table 3: Summary statistics for weekly return series for the different size portfolios from January 1998 to December 2015.

The following table shows a correlation matrix for the ten size portfolios and average returns. Recall that *OBX*, *Allshare* and *VW* are value weighted. It is thus as expected that the portfolios for large stocks correlate more with average returns. To some extent, we see the same pattern for the equally weighted *EW* variable, which is less expected. This suggests that small stocks on average are less related to the market, which also might indicate that they are less related to measures on average sentiments.

Portfolio	**EW**	**VW**	**Allshare**	**OBX**
Portfolio 1	0.34	0.17	0.17	0.16
Portfolio 2	0.44	0.26	0.26	0.25
Portfolio 3	0.46	0.29	0.29	0.27
Portfolio 4	0.53	0.35	0.35	0.33
Portfolio 5	0.65	0.52	0.51	0.49
Portfolio 6	0.71	0.60	0.60	0.57
Portfolio 7	0.81	0.73	0.73	0.71
Portfolio 8	0.79	0.73	0.72	0.70
Portfolio 9	0.83	0.84	0.84	0.83
Portfolio 10	0.82	0.95	0.93	0.95

Table 4: The table shows correlations between the ten size portfolios and average return from January 1998 until December 2018. It is calculated from weekly returns.

This is also visible in the graphical presentation in the following model. We clearly see that portfolio 10 correlates with Allshare, and that portfolio 1 is less correlated and less volatile.

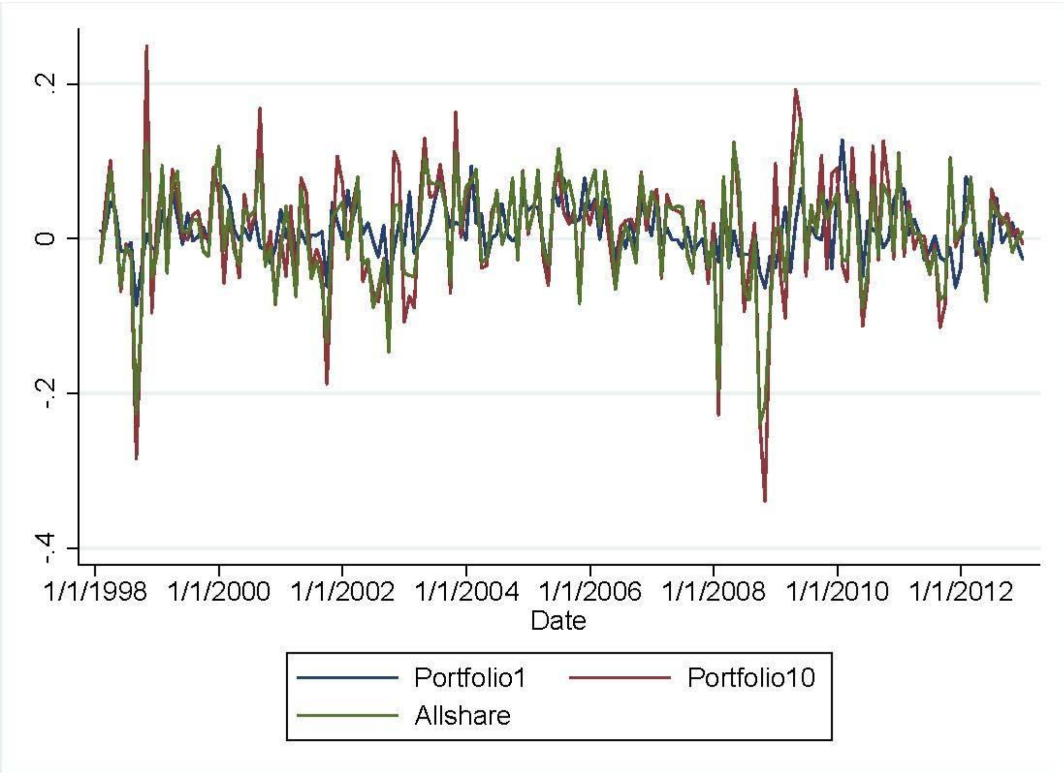

Figure 3: The graph shows the monthly return of Allshare, Portfolio 1 and Portfolio 10 between January 1998 and December 2012.

Table 5 presents summary statistics for monthly return for the size portfolios between January 1998 and December 2018. The values are in percentages.

Return	Mean	Std. Dev	Min	0.25	Median	0.75	Max
Portfolio 1	1.539	3.699	-8.633	-0.949	1.019	3.836	13.456
Portfolio 2	1.399	5.203	-18.356	-1.718	1.294	4.253	21.775
Portfolio 3	1.398	5.173	-17.624	-1.591	1.475	4.414	21.879
Portfolio 4	1.263	5.418	-18.658	-1.500	1.318	3.943	18.772
Portfolio 5	1.539	5.971	-19.159	-19.953	1.224	4.537	22.669
Portfolio 6	1.289	5.561	-19.794	-1.752	1.239	5.030	18.844
Portfolio 7	1.482	5.733	-22.632	-1.811	1.628	4.892	15.658
Portfolio 8	1.286	6.230	-18.828	-2.047	1.368	5.179	27.111
Portfolio 9	0.660	6.828	-24.987	-3.231	1.216	4.873	22.849
Portfolio 10	0.934	6.900	-33.861	-2.591	1.216	4.766	24.909

Table 5: Summary statistics for monthly return for the different size portfolios from January 1998 to December 2018.

3.2 Construction of Sentiment Indexes

In the following we explain the construction of our indexes. The method is based on Da, Engelberg and Gao (2015), from now on denoted DEG. We state any deviation from their construction. The method is the same for both newspapers and Google searches. We therefore explain the construction simultaneously, stating whenever the data or the processing of the data differs.

DEG creates a daily index based on Google searches, named Financial and Economical Attitude Revealed by Search, or FEARS in short. As our monthly index using Google search is similar, we use their name. We name the index constructed from newspapers NEFA, after Newspapers Economic and Financial Attitude. NEFA is constructed both monthly and weekly.

In short, the indexes work as follows. We measure the occurrence of some predetermined terms in the past and present. We then analyse which terms that have historically been able to predict returns. The current occurrence of these terms composes the current index, which is expected to predict the direction of returns in the following periods.

The subchapter is structured as follows. Under 3.2.1 and 3.2.2 we explain the data sources. Under 3.2.3 we explain how we select relevant terms. Under 3.2.4 we explain how the data on each relevant term is processed. Under 3.2.5 we explain how each index is constructed with the use of data on single terms, and under 3.2.6 we define the variables formally. Under 3.2.7 we show summary statistics for the sentiment indexes.

3.2.1 Newspaper corpus

We use the *Norwegian Newspaper Corpus*[6]. It is published by the Clarino project that is a joint language project with the Norwegian research council and several Norwegian universities and institutions, amongst them NHH. The data contains 11 major Norwegian newspapers from January 1998 until September 2015, including DN that specializes in business news. The total number of words in the corpus is 1 509 076 098. The database lets us count daily occurrence of terms. Each term can be traced back to its origin in an article in a newspaper on a specific date. To obtain weekly and monthly series, we simply add up the daily occurrences. We can thus obtain time series on the occurrence of any term in the corpus over the sample period. The time series we obtain are later adjusted for seasonality and trend. A general observation is that there is surprisingly little seasonality and trend in any term obtained.

3.2.2 Google Trends

Similar to DEG, we use Google Trends as the source for internet searches[7]. The service provides a time series for the occurrence of any search term in a chosen time span and a chosen country. As we are interested in Norwegian markets, we limit the search area to Norway and use only Norwegian terms. Figure 4 shows raw time series for the monthly occurrence of the terms *overskudd* and *utbytte* (*profit* and *dividend*) when we limit the area to Norway. Table 6 shows summary statistics for the same terms.

[6] The corpus can be found at: http://clarino.uib.no/korpuskel/page

[7] Google Trends can be found at https://trends.google.com/trends/?geo=US

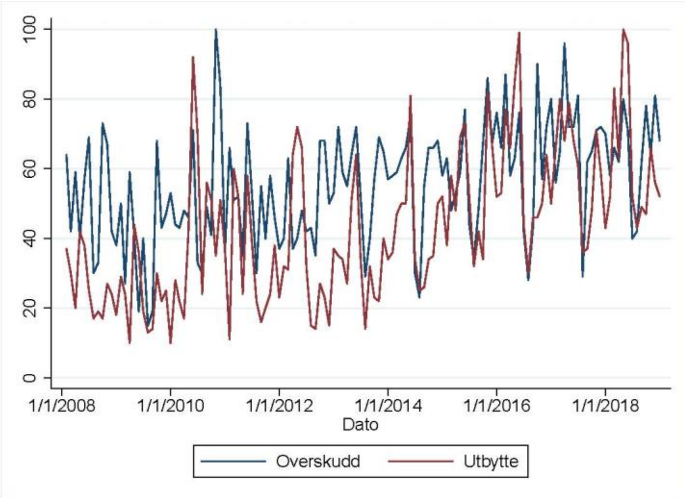

Figure 4: The figure shows the unadjusted series that Google provides for monthly occurrences of the search terms overskudd and utbytte in Norway, 2008-2018. At each period, the occurrence is expressed relative to the period with the highest occurrence. The month with the highest occurrence is given the value 100.

Term	Mean	Std. Dev	Min	0.25	Median	0.75	Max
Overskudd	55.7	17.0	15.0	42.0	57.5	68.0	100.0
Utbytte	43.1	20.9	10.0	25.5	38.5	56.0	100.0

Table 6: Summary statistics for the unadjusted time series Google Trends provide for the occurrence of the search terms overskudd and utbytte in Norway, 2008-2018.

The value of each period is not the absolute occurrence of the term, but the occurrence relative to the period with the highest occurrence, which is given the value 100. We see that for *overskudd* this is October 2010 and for *utbytte* it is April 2018. Every other month are ascribed a value between 0 and 100. If *overskudd* is searched 9500 times in October 2010 and 6000 times in January 2015, then January 2015 is ascribed the value 63 (which is given by 6000/9500 x 100).

This implies that the time series do not express the level of the search terms. *Overskudd* can for instance be much more searched than *utbytte*. We simply have no way to see the level of search activity. This indexing fits our purpose, since we are interested in the relative changes in the search terms, and not the absolute volume. By measuring the change from period to period, we get a measure for changes in search activity. A change from 40 to 44 from one period to the next will thus have the same value as a change from 70 to 77. This makes sense since the two values can hide the same search volumes.

From the statistics in table 6, we see that the two series have different mean, median and standard deviation. These values are all relative to the maximum value of 100.

Figure 4 also exemplifies some other properties of the time series. We see that the two terms follow a similar pattern; the correlation is 0.5. It makes intuitively sense that the two terms correlate. We also see that they have an obvious seasonal component where they tend to peak during the spring months, and reach bottom during late summer. The increased activity during the spring coincides with the publication of preceding years financial statements. The last observation is that the pattern perhaps follows a trend.

Figure 5 shows the raw data for monthly occurrence of the term *underskudd* (*deficit*) for both newspaper and Google search. The correlation between the series is 17.6%. The correlation increases to 20.3% when lagging Google searches one month. This supports the assumption that fluctuations are not arbitrary, but relate to a sentiment.

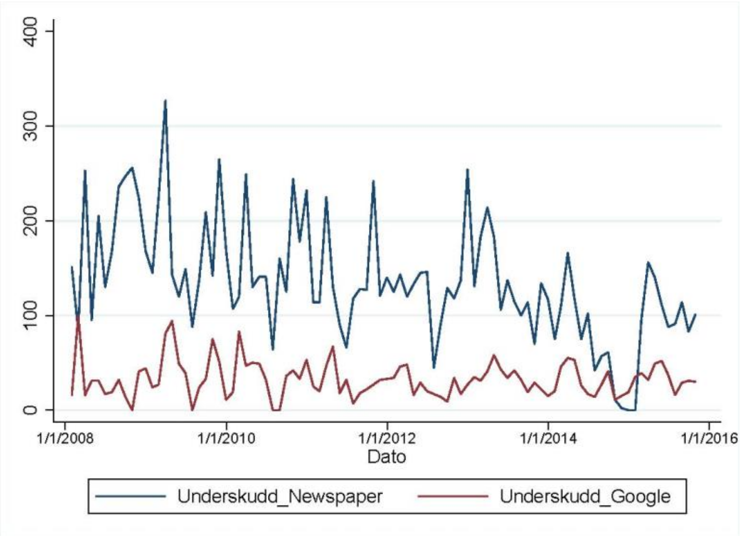

Figure 5: Monthly occurrence of the term underskudd in newspapers and internet search in Norway, 2008-2015. The time series are unadjusted and presented in the units that the data sources provide.

Term	**Mean**	**Std. Dev**	**Min**	**0.25**	**Median**	**0.75**	**Max**
Underskudd Newspaper	136.5	62.8	0.0	101.0	129.5	166.0	327.0
Underskudd Google	32.6	19.2	0.0	19.0	31.0	42.0	100.0

Table 7: Summary statistic for the occurrence of the term underskudd in newspapers and internet search in Norway, 2008-2015. Unadjusted and presented in the units that the data sources provide.

Recall that the level of the scales are not comparable. The newspaper counts absolute occurrence and Google search counts occurrence relative to the period’s maximum occurrence. which is given the value 100.

3.2.3 Term selection

We have previously explained the sources for our data. We are able to obtain data on the occurrences of any term we want from both newspapers and Google searches. As we want to construct sentiment indexes that can predict stock returns, we need some criteria to filter out relevant terms to use in the composition of our indexes.

DEG follows existing literature in using Harvard IV-4 and the Lasswell dictionaries (Tetlock, 2007; Tetlock, Saar-Tsechansky and S Macskassy, 2008)[8]. These dictionaries provide lists of terms that are economic and have either a positive or a negative sentiment. Terms that qualify as both being economic and having an unambiguous sentiment are included in the list that composes the index constructed in DEG. Our challenge is that these dictionaries are limited to English terms. As we are concerned with Norwegian terms and markets, we need to find a substitute. We do however not know of any dictionary that classifies terms in this way. We are therefore left with the option to filter the terms ourselves. To ensure objectivity, we strictly abide to predetermined lists and criterions. The criterions we use are as follows.

1. The term has to be economical. *Depresjon* (*depression*) does for instance not qualify, as it has other and more common uses outside economics.

2. The term needs to have either a clear positive or a negative sentiment. Negative understood in a broad sense. *Rente* and *inflasjon* (*interest rate* and *inflation*) are not negative in a strict sense, but they are negative in the sense that whenever one is relatively more concerned about these notions, it expresses some negative awareness.

These criterions are applied when we systematically go through the following lists. The Norwegian central banks *Ord og uttrykk*[9] – a list of terminology used by the central bank.

[8] Harvard IV-4 and Lasswell dictionaries can be found at http://www.wjh.harvard.edu/~inquirer/homecat.htm

[9] The Norwegian central bank has removed the list from the original webpage norgesbank.no/Ord-og-Uttrykk. It is however possible to visit the page through the following third party that offers access to former webpages. https://web.archive.org/web/20180111194314/http://www.norges-bank.no/Ord-og-uttrykk/

and *Økonomileksikon* – a comprehensive encyclopedia in economics and business administration by Arild Lillebø (Lillebø, 2005). Several terms appear in both sources.

This method to select terms does involve some discretion from our part, which is an obvious weakness. However, by strictly abiding to predetermined lists and criteria we exclude any biases as far as we are able. Some terms, like *BNP*, *aksje* and *jobb* (*GDP*, *stock* and *work*) do not have an obvious sentiment, and their inclusion can be discussed. However, recall criteria number two, that it is sufficient that increased awareness of the notion implies a sentiment. If for instance the awareness of *aksje* increases, this can be interpreted as a positive shift. Under subchapter 4.4 it will also be made clear that the inclusion of irrelevant terms does not bias the results.

The filtering of terms results in an initial list of 98 terms[10]. This is the starting point for the construction of the two alternative sentiment indexes.

As Norway is a small country, Google Trends provides limited data. We thus have to exclude some terms from the Google search index due to insignificant data. This is also the reason why we are unable to create a weekly index based on Google Trends. Some weeks simply have too little activity to provide reliable data.

3.2.4 Processing of each term

We thus have a list of relevant terms. We then obtain one time series for each term from both the newspaper corpus and Google Trends as explained previously. We define the occurrence of each term *j* in a given period *t* as $SVI_{j,t}$ and $NWC_{j,t}$, after Search Volume Indicator and

[10] Mislighold, resesjon, konkurs, svindel, inflasjon, aksje, arbeidsledighet, arbeidsledig, korrupsjon, korrupt, styringsrente, rente, gjeld, børskrakk, finanskrise, underskudd, budsjettunderskudd, bruttonasjonalprodukt, sparken, bestikkelse, kreditor, fattigdom, minstelønn, oppsigelse, nedleggelse, nedlagt, nedbemanning, jobbsøkere, kostnad, nødhjelp, bot, forelegg, obligasjoner, sparing, investering, investere, omfordeling, økonomi, kredittkort, forbrukslån, kreditt, arbeidsuke, arbeidsmengde, avkastning, emisjon, overskudd, arbeidsavklaringspenger, arbeidsmiljøloven, avgift, bedrageri, beslag, avvikling, bistand, bostøtte, boligkrakk, dagpenger, BNP, deflasjon, erstatning, fattig, finans, folketrygden, finansiering, fond, hjemløs, forsikring, inkasso, konjunktur, gjeld, inntekt, jobb, krakk, korreksjon, kursfall, likviditet, lavkonjunktur, lønnsomhet, NAV, minstepensjon, lønn, pengepolitikk, refinansiering, pensjon, purring, privatøkonomi, samfunnsøkonomi, skatt, spekulasjon, subsidier, spare, svindler, sysselsetting, tariffavtale, tvangsauksjon, trygd, utbytte, utviklingshjelp, volatilitet.

Newspaper Wordcount respectively. We now proceed to explaining how the time series for each term is processed.

We first divide each period by the time series average. Each period then expresses the occurrence of a term relative to the average occurrence. We then log-transform each series to normalize and account for extreme values. For each period, we want to know the relative change in the occurrence of the term j; so each period t is given the value delta $SVI_{j,t}$ and delta $NWC_{j,t}$ defined by:

$$\Delta SVI_{j,t} = ln\ \left(SVI_{j,t}\right) - ln(SVI_{j,t-1}) \tag{2}$$

$$\Delta NWC_{j,t} = ln(NWC_{j,t}) - ln(NWC_{j,t-1}) \tag{3}$$

To account for seasonality, we regress each series on monthly dummies and keep the residual. We also detrend any series. In order to make the terms comparable we divide each series by their standard deviation. As a result, a change in the occurrence of a term is weighted equally as the others when included in an index. In addition, a one standard deviation increase in the term *overskudd* is directly comparable to a similar increase in the term *utbytte*.

3.2.5 Index composition

We now have the time series needed to compose the indexes. They have been log transformed, detrended, seasonally adjusted, weighted to have the same standard deviation and express changes relative to the previous period. We first give a short intuitive explanation of the indexes before we more formally explain the further construction. At the end of the subchapter, we explain that we only use terms with a negative sentiment.

The purpose of the indexes is to predict return in a future period. At the start of January 1st, we do for instance want to predict the return in either week one or January. The information available is previous returns *Return* up to December 31st and previous and current values of ΔNWC_j and ΔSVI_j.

By relating $\Delta NWC_{j,t-1}$ to $Return_t$ and $\Delta SVI_{j,t-1}$ to $Return_t$, we can obtain an idea of which terms that have historically predicted *Return*. These terms will constitute the index. The intuition is quite simple; whenever the occurrence of a term increases, it contributes to the index in the direction that the term has historically predicted. If for instance ΔNWC_j for *overskudd* is known to have a positive correlation to next period's return, then a high value of ΔNWC_j for *overskudd* will contribute to a higher (more negative) combined NEFA.

We now explain the construction more formally. In the following paragraph, we use ΔNWC as an example, but the same goes for ΔSVI. At any point in time, the index can only be constructed by previous and current observations. The last relevant pair of observations for the construction of $NEFA_t$ is thus $\Delta NWC_{j,t-1}$ and $Return_t$.

We use a dynamic rolling regression model where the construction each period utilizes every pair of observations up to the last available one. When constructing $NEFA_{week2}$, that is expected to predict returns in week 3, the construction is based on every pair of observation up to $\Delta NWC_{j,week1}$ and $Return_{week2}$. After week 3 passes, we want to construct $NEFA_{week3}$ to predict the returns in week 4. The model is then rolled over such that the last pair of observations utilized is $\Delta NWC_{j,week2}$ and $Return_{week3}$. In this way each $NEFA_t$ are dynamic and constructed out of sample, as it is only based on previous observations. In other words, we never use any observations that are not known at the time of the construction.

By rolling over the index each period, we allow for the possibility that terms play different roles during the sample period. The fact that one term could predict returns ten years ago does not necessarily mean that it can do so today. The alternative is to measure the role of the terms over the whole sample period, but this would involve using future observations in the construction of the index. We are thus rolling over the model for each time in order to avoid using future observations and at the same time we utilize the latest known observations.

We use an OLS regression method to find the historical relation between $\Delta NWC_{j,t-1}$ and $Return_t$. For each period t we perform the following regressions respectively:

$$Return_t = \beta_0 + \beta_1 \, \Delta NWC_{j,t-1} + \mu_t \tag{4}$$

$$Return_t = \beta_0 + \beta_1 \, \Delta SVI_{j,t-1} + \mu_t \tag{5}$$

This is done for each term each period. As a result, we obtain an updated coefficient β_1 and p-value for each term each period.

Table 8 shows the coefficient and p-value for some weekly ΔNWC_j over the whole sample period. This serves as an example. Recall that the significance and coefficient for each term is updated each period. We see that both terms associated with private finance (like *refinansiering*) and macro economical terms (like *lavkonjunktur*) yields results. The idea is to reduce the idiosyncrasy by combining the terms in one index, such that the significance of the index is greater than for each term alone.

Term	**Coefficient**	**P-value**
Refinansiering	-0.0017	0.086
Lavkonjunktur	-0.0016	0.114
NAV	-0.0015	0.126
Bostøtte	-0.0015	0.144
Likviditet	-0.0014	0.171
Bedrageri	-0.0014	0.175
Utbytte	-0.0013	0.191
Inflasjon	-0.0012	0.221
Minstelønn	-0.0011	0.247
Privatøkonomi	-0.0011	0.271

Table 8: The table presents terms with a negative relation to weekly Allshare return when utilizing the whole sample period. The dependent variable is weekly Allshare return one week into the future.

We now need to decide on a criterion for selecting terms in each period. The relevant parameter is the p-value. There are two main alternatives; either choosing a significance cut-off, implying that every term with a p-value below a certain value is included; or choosing a predetermined amount of terms each period. The first alternative will exclude terms that are not significant, but fewer terms implies greater potential idiosyncrasy. The latter alternative will secure a constant diversification of the idiosyncrasy, but it will potentially include insignificant terms in the index. We thus face a trade-off. DEG selects the latter alternative and includes the 30 most significant terms in each period. In our case, we do however find that this involves the inclusion of terms with a high p-value in their own right. We therefore choose to deviate from the methodology of DEG and include a predetermined number of 10 terms.

The drawdown of this choice is that we potentially allow for more idiosyncrasy, which implies that the index at a certain point in time is more likely to be affected by irrelevant factors. This can be exemplified; the term *gjeld* might increase in a period due to an idiosyncratic factor like a song titled *gjeld* being released. The fewer words that composes the index in the period, the greater the idiosyncratic effect of the song will be on the index. Over time this idiosyncrasy should however be diversified. This will be further discussed under chapter 4.4, where we test if alternative constructions leads to different results.

Before moving on to formally define the indexes we first need to clarify two aspects. As the selection of terms to compose NEFA and FEARS depends on historic correlation between

$\Delta NWC_{j,t-1}$ and $\Delta SVI_{j,t-1}$ and $Return_t$ respectively, it matters which return variable we use. By using *Allshare*, the indexes are biased towards predicting *Allshare* and not *OBX*. By alternatively using *OBX* in the construction, the index should be relatively better at predicting returns of large stocks. To fully understand the indexes relation to small and large stock returns, we construct indexes using both variables. In the following main analysis, we use *Allshare* unless anything else is stated. To avoid any confusion we denote the alternative constructions as *NEFA Allshare*, *NEFA OBX*, *FEARS Allshare*, *FEARS OBX*, etc. We continue to refer to NEFA and FEARS in the general cases.

The other aspect that needs to be clarified is that we exclude terms with positive sentiment. We find that almost all terms that have historically predicted returns have a negative sentiment. This is true for both newspapers and Google searches. It is thus seemingly only a variation in negative terms that are associated with returns. By excluding positive terms, the indexes become simpler and more intuitive, which is an advantage. This finding is consistent with DEG and Tetlock (2007), and DEG also exclude positive terms.

3.2.6 Index definition

We now have decided which terms should be included in the index for each period. The index in each period is constructed by previous observations, and the ten terms that have most significantly predicted returns are included.

Recall that we have treated each term such that they are equally weighted. Formally, the value of the index for period t is given by the average of the 10 terms j included:

$$FEARS_t = \sum_{j=1}^{10} \Delta SVI_{j,t} \Big/ 10 \tag{6}$$

$$NEFA_t = \sum_{j=1}^{10} \Delta NWC_{j,t} \Big/ 10 \tag{7}$$

We thus obtain a value for our indexes for each period and can use this as a variable in our analysis. Our goal is to see if these indexes can predict subsequent returns, and if it has different implications for small and large stocks.

For NEFA we have both a weekly and a monthly index. As the construction depends on changes from week to week and month to month respectively, the monthly NEFA is not equal to the sum of weekly NEFAs. Quite contrary, there is no mathematical reason why weekly and monthly NEFA should provide the same results.

3.2.7 Sentiment Index statistics

Weekly NEFA

The following figure presents the distribution and the time series for the weekly NEFA index calculated with Allshare return. We observe no systematic changes in the index during the sample period, which is between 1998 and 2015.

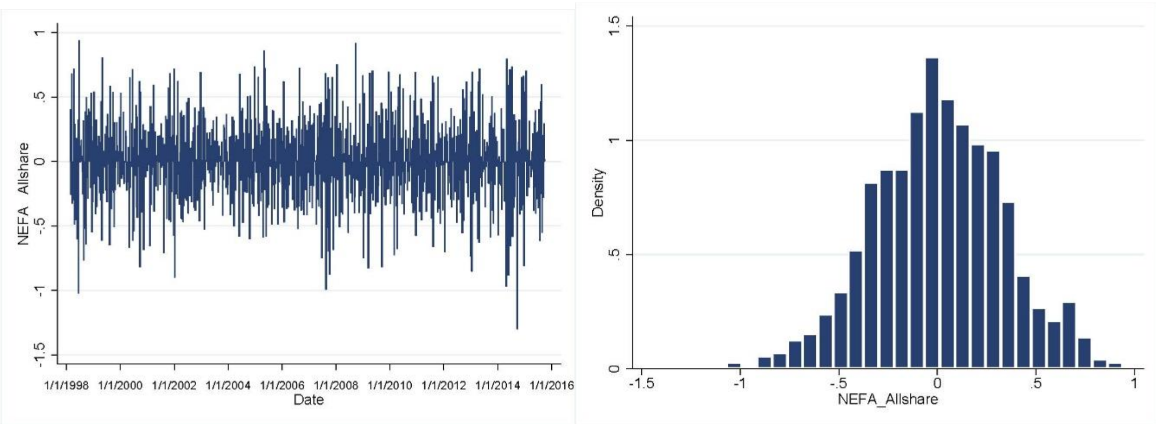

Figure 6: To the left: Graphical presentation of the time series for the weekly NEFA sentiment variable constructed with Allshare return, 1998-2015. To the right: the variable's distribution.

Table 9 presents summary statistics for the weekly NEFA variables. Note that the values are close to identical when using different return variables in the construction.

Variable	Mean	Std. Dev	Min	0.25	Median	0.75	Max
NEFA Allshare	0.005	0.334	-1.302	-0.221	0.004	0.220	0.940
NEFA OBX	0.003	0.348	-1.749	-0.223	0.003	0.230	1.083
NEFA EW	-0.001	0.355	-1.636	-0.235	0.000	0.231	1.088
NEFA VW	-0.002	0.330	-1.505	-0.214	-0.002	0.216	1.037

Table 9: Summary statistics on the weekly NEFA variables, constructed based on data from 1998-2015.

Monthly NEFA

Figure 7 shows the distribution and time series for the monthly NEFA index calculated with Allshare between 1998 and 2015.

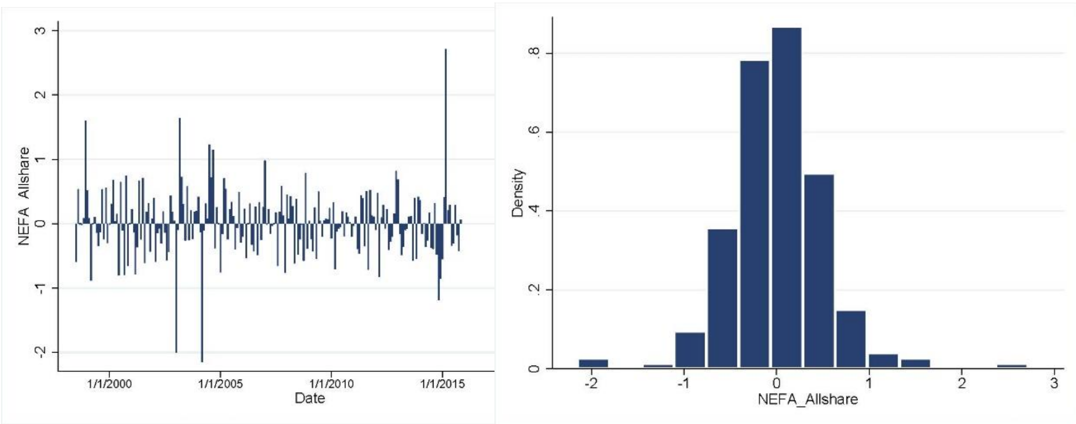

Figure 7: To the left: Graphical presentation of the time series for the monthly NEFA sentiment variable, constructed with Allshare return, 1998-2015. To the right: the variable's distribution.

Sentiment Index	Mean	Std. Dev	Min	0.25	Median	0.75	Max
NEFA Allshare	0.004	0.521	-2.149	-0.301	0.008	0.285	2.716
NEFA OBX	0.000	0.496	-2.004	-0.324	0.007	0.305	2.716
NEFA EW	-0.004	0.515	-2.281	-0.307	-0.019	0.297	2.572
NEFA VW	-0.011	0.510	-1.815	-0.338	-0.015	0.231	2.707

Table 10: Summary statistics for the four monthly NEFA variables, constructed based on data from 1998-2015.

Monthly FEARS

Figure 8 shows the distribution and the time series for the monthly FEARS index calculated with Allshare between the 2008 and 2018.

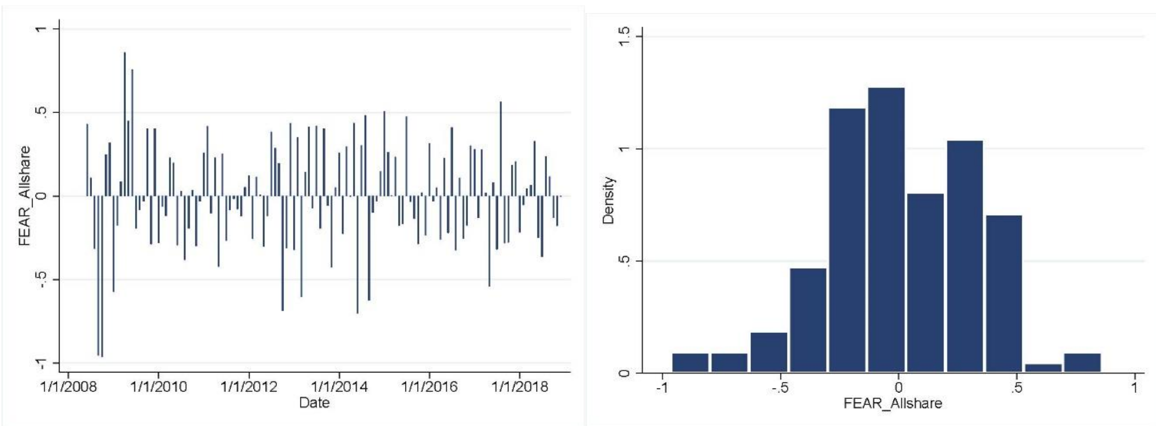

Figure 8: To the left: Graphical presentation of the time series for the monthly FEARS variable, constructed with Allshare return, 2008-2018. To the right: the variable's distribution.

Sentiment Index	Mean	Std. Dev	Min	0.25	Median	0.75	Max
FEARS Allshare	-0.001	0.324	-0.965	-0.226	-0.004	0.252	0.861
FEARS OBX	0.002	0.334	-0.965	-0.225	-0.010	0.234	1.255
FEARS EW	-0.004	0.323	-0.940	-0.185	0.018	0.193	0.727
FEARS VW	-0.008	0.336	-0.969	-0.235	0.005	0.234	1.254

Table 11: Summary statistics for the four monthly FEARS variables, constructed based on data from 2008-2018.

Conclusions

We see that the statistics for each variable is very similar between the four alternative constructions. This suggests that it does not matter much if we use *Allshare*, *OBX*, *EW* or *VW* as the return variable in the construction.

Recall that the indexes measure changes in occurrence between periods. This implies that weekly and monthly NEFA are less connected than one might expect. An increase that is present in the weekly data is not necessarily present in the monthly data. A high occurrence in the first week will matter less for the monthly index if the other three weeks have a normal occurrence. In other words, weekly volatility in the occurrence can be diversified away in the monthly occurrence. In addition, terms that relate to returns on a monthly basis, are not necessarily the same terms that relate to returns on a weekly basis. The monthly and weekly indexes might thus measure different forms of a newspaper sentiment.

The correlation between monthly NEFA and monthly FEARS are 0.04. This suggests that NEFA and FEARS measure different forms of sentiments.

3.3 Other data and variables

We have previously defined our main variables. We now proceed to explain and define other data and variables.

We want to check if our indexes represent new knowledge, or if it is simply existing knowledge in disguise. By including previously known variables we find if our variables significantly predict returns even after controlling for this existing knowledge.

The inclusion of control variables implies a trade-off between making a model less efficient by including redundant variables, and making the results biased due to exclusion of desired variables. By utilizing several variables, we are however able to test alternative models where we vary the amount of control variables.

There are two reasons to include a variable in our case. The first one is if a variable is known to possibly predict stock returns. Næs, Skjeltorp and Ødegaard (2008) analyses which factors that have historically been drivers of the return on OSE. For average returns, they find that changes in oil prices as well as macro-economic factors like yield spread matters, and on this basis we include *Brent* and *Spread*. We also include a variable that measures the volatility on OSE in the preceding month, noted as *Volatility*.

The other reason for inclusion is if a variable possibly express similar properties as our indexes. In that case, our indicators might simply serve as proxies for existing variables. This latter is the rationale behind the inclusion of the control variables *Producer confidence*, *Consumer Confidence*, *VIX* and *Fund flow*.

Several of the variables are log-transformed in order to obtain normalized time series. This is specified in each subchapter.

3.3.1 Fund Flow

Ben-Rephael, Kandel and Wohl (2012) argues that fund flows serve as a proxy for investor sentiment. The theory is that whenever investors are positive (negative) towards future returns it will result in a net inflow (outflow) into mutual funds. Warther (1995) also finds that fund flows predict stock returns. Fund flows into stock funds thus serves as a valid control variable.

The Norwegian Mutual Fund Association (VFF) has supplied data on monthly fund flows. The data distinguishes between types of funds – stock funds, bond funds and balanced funds. Our sample range from the start of 1998 to the end of 2018. We have manually calculated the data from before 2008 from data provided by VFF. Newer data can be obtained from the web page of VFF[11].

Our variable measures monthly net flow into stock funds, where the net flow is in-flow minus out-flow in nominal terms. This implies that we do not adjust the net flow relative to the market value of the funds. We thus do not account for the effect that a change in fund value has on

[11] Data on fund flow after 2007 can be found at: https://vff.no/historisk-statistikk

net fund flow. For the publicly available part of the sample, after 2008, we are however able to estimate relative fund flow. When doing so we find no evidence that any results are altered. We thus do not find any evidence that our absolute variable is inferior to a relative measure of fund flow.

The time series are detrended, seasonally adjusted and divided by 1000. The variable is positively related to a sentiment. Whenever the in-flow to stock funds exceeds the out-flow, the variable is positive.

3.3.2 Corporate Confidence

The Norwegian central bank publishes quarterly reports on the sentiment of the production side of the economy, called the regional network report[12]. It is formed based on a network of 1500 professionals who covers all industrial sectors and geographical regions. We use aggregated historical data on all sectors and regions measuring the anticipated growth in production for the following six months. As this measures the sentiment of industrial production, it serves as a control variable for our indexes. The data is recorded and ranges from October 2002 until February 2019.

In order to normalize the series, we log transform and record the change in each report relative to the previous report. The report has been produced at varying intervals. Each period is ascribed the value from the last report. This implies that some values are ascribed to more periods than others.

The variable is positively related to a sentiment; whenever the outlook of future production betters relative to previous period, the variable increases.

3.3.3 Consumer Confidence

Finance Norway publishes a quarterly report on consumer confidence[13]. It is based on a standardized survey on households' economic outlook and expectancy. The Indicator

[12] Data on the Regional Network are found at: https://www.norges-bank.no/Publisert/Publikasjoner/Regionalt-nettverk/2019/12019/

[13] Data on Finance Norway's Consumer Confidence Indicator are found at: https://www.finansnorge.no/aktuelt/sporreundersokelser/forventningsbarometeret1/forventningsbarometeret-2019/troen-pa-egen-okonomi-er-sterkere-enn-troen-pa-landets-okonomi/

measures the occurrence of optimistic respondents relative to pessimistic respondents and is seasonally adjusted. We log transform the time series and record the change from last quarter. It is recorded from 1992, so it covers all our sample periods. In our case the indicator serves as an alternative and control variable to our sentiment indexes.

The variable is positively related to a sentiment; whenever the consumer confidence increase relative to previous period, the variable increases.

3.3.4 VIX

The Norwegian Volatility Index (NOVIX) measures the volatility implied from 30-day options on OBX. As the demand for put-options increases (decreases) relative to buy-options, the NOVIX increases (decreases). This serves as a measure for the market's fear of a decline in OSE during the following 30 days. NOVIX is recorded only from April 2016[14]. Due to the lack of historical data on NOVIX, we use the original Chicago Board Option Exchange Volatility Index (VIX), calculated after the same principles for the broad American S&P 500 index[15]. We record the seasonally adjusted average monthly and weekly value, log transform the series and then use the period changes.

The variable is negatively related to a sentiment; whenever the market's fear of a decline in S&P 500 increases, the variable increases.

The substitution of a Norwegian volatility index for VIX is obviously not favorable, as we wish to measure the sentiment of the Norwegian market. However, the lack of a historical Norwegian alternative makes it the best option. The legitimacy of using VIX as a control variable depends on there being a significant correlation between VIX and a Norwegian alternative like NOVIX. We analyze the relationship between NOVIX and VIX from April 2016 until April 2019 and find a daily correlation of 0.56. This implies that VIX does correlate with the Norwegian sentiment to some extent, but not perfectly. It also complies with the idea that the Norwegian market is significantly integrated with international markets (here represented by S&P 500), but not perfectly. The substitution is thus a weakness. Note also the

---

[14] Data on NOVIX are found at: https://novix.xyz/

[15] Data on the CBOE VIX are found at: http://www.cboe.com/vix

assumption that the correlation we find after 2016 is representative for the period from 1998 until 2016.

3.3.5 Spread

Historical data on 10 year government bond and 3 month treasury bill is published by the Norwegian Central bank[16]. Based on this we construct a yield spread composed of 10 year bond minus 3 month bill as suggested by Næs, Skjeltorp and Ødegaard (2008).

$$Yield\ spread = Yield_{10\ yr\ bond} - Yield_{\ 3\ m\ bill} \tag{8}$$

The yield spread serves as a macro indicator for the financial market's expectancy of long-term prospects relative to the short-term current outlook. We record the average spread for each week and month. We then log transform the series and record the change from previous period.

Two factors lead to changes in the variable; changes in 10-year bond yield and changes in 3-month bill yield. Increased yield is associated with an increasingly positive economic outlook (Bodie, Kane, Marcus, 2018, pp. 477). An increase in 3-month bill yield will reduce the spread, whereas an increase in 10-year bond yield will increase the spread. Changes in yields can occur due to a variety of reasons. The relation to stock returns is thus complex, dependent on the time-horizon of the relationship and which factors that drive the changes.

3.3.6 Brent

We use historical data on Brent Spot oil price denominated in American dollar[17]. As OSE is significantly exposed to the oil industry, an increased oil price is expected to increase the valuation of future earnings (Næs, Skjeltorp and Ødegaard, 2008). We find both monthly and weekly data, where the price of each period is the average daily price of all days noted. Following Næs, Skjeltorp and Ødegaard (2008), we use dollar prices in order to isolate the effect of oil price changes from changes in the USDNOK exchange rate. We are interested in

[16] Data on government bonds/ bills are found at: https://www.norges-bank.no/Statistikk/Rentestatistikk/Statsobligasjoner-Rente-Manedsgjennomsnitt-av-daglige-noteringer/ and https://www.norges-bank.no/Statistikk/Rentestatistikk/Statsobligasjoner-Rente-Daglige-noteringer/

[17] Data on Brent Spot are found at: https://fred.stlouisfed.org/series/DCOILBRENTEU

the effect of changes in the oil price and not the absolute level. We therefore log transform the series and record the change from the previous period.

The variable is positively associated with stock returns. When oil prices increase, this potentially lead to expectancy of increased cash flows for stocks.

3.3.7 Volatility

One particular concern is that our sentiment indexes are highly correlated with the stock market volatility. Evidence show that investors demand a higher return in periods of greater volatility (French, Schwert and Stambaugh, 1987). In this case, it might be that increased volatility coincides with increased sentiment indexes, and that it is the volatility, and not the sentiment that predicts subsequent stock returns.

To control for this we include a variable that measures the volatility on OSE in the preceding 30 days. We follow the methodology suggested by French, Schwert and Stambaugh (1987), where the standard deviation of the preceding 30 days for stock index m, at day t, is defined as the following.

$$\sigma_{mt}^2 = \sum_{n=1}^{30} return_{i,t}^2 + 2\sum_{n=1}^{29} return_{i,t} return_{i+1,t} \tag{9}$$

Where $return_i$ is daily return for day i. We calculate this with daily returns for each return variable. When we use *Allshare* as dependent variable, we measure the prior 30-day volatility of *Allshare*, and when we use *OBX* as dependent variable, we measure the volatility of *OBX*, and so on. For simplicity, we note it simply as *Volatility* in all cases.

The intuitive interpretation of the variable is that increased volatility predicts increased returns in subsequent periods, as investors demand a premium to hold a relatively more volatile asset (French, Schwert and Stambaugh, 1987).

3.3.8 Summary statistics for control variables

In the following, we present summary statistics for our control variables. Some variables are estimated both weekly and monthly. The following table presents statistics for the variables used in the weekly analysis.

Variable	Mean	Std. Dev	Min	Median	Max	Period	Frequency
VIX	0.000	0.093	-0.389	-0.004	0.476	Jun. 1998 – Sep. 2015	Weekly
Spread	0.001	0.053	-0.353	-0.003	0.442	Jan. 2003 – Sep. 2015	Weekly
Brent	0.008	0.190	-1.043	0.022	0.476	Jan. 1998 – Sep. 2015	Weekly
Fund Flow	0.000	2.703	-15.370	-0.113	9.478	Jan. 1999 – Sep. 2015	Monthly
Consumer Confidence	0.000	0.246	-0.785	0.005	0.832	Apr. 1998 – Sep. 2015	Quarterly
Corporate Confidence	0.000	0.189	-1.053	0.013	0.439	Nov. 2002 – Sep. 2015	Quarterly

Table 12: Summary statistics for the weekly control variables used in the prediction of future return.

Table 13 presents statistics for the variables used in the monthly analysis.

Variable	Mean	Std. Dev	Min	Median	Max	Period	Frequency
VIX	0.000	0.164	-0.361	-0.023	0.719	Jun. 1998 – Des. 2018	Monthly
Spread	0.003	0.090	-0.215	-0.006	0.609	Jan. 2003 – Des. 2018	Monthly
Brent	0.003	0.040	-0.135	0.008	0.087	Jan. 1998 – Des. 2018	Monthly
Fund Flow	0.000	2.758	-15.000	-0.004	9.995	Jan. 1999 – Des. 2018	Monthly
Consumer Confidence	-0.005	0.350	-1.253	-0.012	1.295	Apr. 1998 – Des. 2018	Quarterly
Corporate Confidence	0.015	0.171	-1.040	0.027	0.483	Nov. 2002 – Des. 2018	Quarterly

Table 13: Summary statistics for the monthly control variables used in the prediction of future return.

Note that the values is in reference to the frequency. The monthly standard deviation of VIX is for instance 0.164, and the weekly standard deviation of VIX is 0.093. The standard deviation of Fund Flow is on a monthly basis in both tables. Note also the different sample periods. The sample periods presented in the two tables are the periods relevant for our weekly and monthly analysis respectively. Our weekly NEFA index does for instance stop at the end of September 2015, and we thus cut the sample periods of the weekly control variables at that time.

Table 14 presents a correlation matrix for our control variables. It has been estimated from our weekly data. There is no time lag, implying the values denote the correlation between variables occurring at the same point in time.

	VIX	Spread	Brent	Fund Flow	Consumer Confidence	Corporate Confidence	Allshare Return
VIX	1.00	-0.08	0.00	-0.13	0.00	0.00	-0.47
Spread	-0.08	1.00	-0.03	0.00	-0.04	-0.12	-0.02
Brent	0.00	-0.03	1.00	0.01	-0.01	0.00	0.01
Fund Flow	-0.13	0.00	0.00	1.00	0.30	0.11	0.22
Consumer Confidence	0.00	-0.04	-0.01	0.30	1.00	0.60	0.18
Corporate Confidence	0.00	-0.12	0.00	0.11	0.60	1.00	0.08
Allshare Return	-0.47	-0.02	0.01	0.22	0.18	0.08	1.00

Table 14: Correlation matrix for weekly control variables.

We observe no big surprises. We see a strong negative relation between returns and *VIX*, indicating that decreasing stock prices and an increasing fear of future decline coincide. We also observe a strong positive relation between *Fund Flow* and returns. This is not surprising, as stock funds are composed of stocks, thus a net flow into funds is associated with a demand for stocks. We also see a positive relation between *Consumer Confidence* and returns. This can express either that the expectancy of consumers about the future is rightful, and/ or that this confidence contributes towards increasing the relative demand for stocks.

The strongest correlation between two control variables is that of *Consumer Confidence* and *Production Confidence*. It is no surprise that these are related, as they both express confidence about the future. Note also the positive correlation between *Consumer Confidence* and *Fund Flow*. This suggests that increased confidence coincides with increased demand for risky funds.

The fact that some control variables correlate can affect our results. This is discussed later, where we also explain how we encounter the issue.

3.4 Empirical strategy

We have now defined a hypothesis, our data and variables. Our main hypothesis is that our sentiment indexes can predict stock returns on OSE. Secondarily we want to see if the indexes have different predictive power on small and large stocks.

We now proceed to explain how we test this empirically. In the following, we distinguish between tests on average returns and cross-sectional returns on small and large stocks.

The notation *Sentiment Index* refers to the general case of our sentiment indexes. In the following models, this notation means that we estimate similar models for both NEFA and FEARS respectively. Instead of specifying each model one time for NEFA and another time for FEARS, we simply use the notation Sentiment Index.

3.4.1 Average returns

In the following we specify how we test if our indexes predict average returns on OSE. The questions we wish to answer is in which direction the indexes predict prices to move, which periods are predicted and what the magnitude of the effect is. We use similar models as suggested by DEG.

We are interested in our sentiment indexes' ability to predict the future. We thus control for other variables known in the present. The rationale is that our index only represents new knowledge if the knowledge is not already expressed by existing variables. The control variables include the seven variables presented in subchapter 3.3, as well as lagged values of stock return and yearly dummy-variables.

We estimate the following model for weekly NEFA, monthly NEFA and monthly FEARS separately.

$$Return_{t+k} = \beta_0 + \beta_1\, Sentiment\ Index_t + \sum_{m=1}^{M} \gamma_m\, Control_t^m + \mu_{t+k} \qquad (10)$$

If β_1 is significant we see that the sentiment index at time t predicts $Return_{t+k}$ even after controlling for alternative variables and lagged returns. We estimate the model for several periods ($Return_t$ until $Return_{t+5}$) to see if the indexes predict returns in the current and future periods. Our hypothesis is that the sentiment indexes predict future return. If this is true then $NEFA_t$ and $FEARS_t$ will be significant when using $Return_{t+1}$, $Return_{t+2}$ or another future period as dependent variable.

To fully understand the role of the sentiment indexes we also run regressions with NEFA and FEARS as dependent variables.

$$Sentiment\ Index_{t+k} = \beta_0 + \beta_1\, Return_t + \sum_{m=1}^{M} \gamma_m\, Control_t^m + \mu_{t+k} \qquad (11)$$

This model tests if the sentiment indexes are explained by the control variables or previous and current returns. If the independent variables are not significant, it supports the notion that NEFA and FEARS serve as original and independent indexes.

We also perform an additional test to test the independence of the sentiment indexes. This is done by repeating the first models with $Return_{t+k}$ as dependent variable, only this time we omit the sentiment indexes as an independent variable. We then compare the results of the models with and without the sentiment index, and see if the coefficient of the control variables are altered. If they are not significantly altered, it supports the assertion that NEFA and FEARS serve as independent indexes. The relevant pair of models to compare is in this case model 10 and model 12.

$$Return_{t+k} = \beta_0 + \sum_{m=1}^{M} \gamma_m \, Control_t^m + \mu_{t+k} \tag{12}$$

The models specified above still leave us with many choices. We can alter between using the return of *Allshare, OBX, EW* and *VW,* we can vary the amount of control variables, and lag variables. We run several models and report those that are representative for our findings. Whenever the results depend on choices we make we report alternative tables to support our discussions.

3.4.2 Cross-sectional analysis

Our second objective is to see if our indexes have different implications for small and large stocks. We take the results of our analysis of average return, and perform additional tests for this. If FEARS for instance shows to predict returns for the first subsequent month, and no other future period, then we will proceed with additional analysis for this period.

The first test is to compare models with *OBX Return* and *Allshare Return* as dependent variables. As *OBX* is composed of large stocks, the comparison will contribute towards understanding if the sentiment index has different predictions for large stocks. We estimate the following models.

$$OBX\ Return_{t+k} = \beta_0 + \beta_1 \, Sentiment\ Index\ Allshare_t + \sum_{m=1}^{M} \gamma_m \, Control_t^m + \mu_{t+k} \tag{13}$$

$$Allshare\ Return_{t+k} = \beta_0 + \beta_1 \, Sentiment\ Index\ Allshare_t + \sum_{m=1}^{M} \gamma_m \, Control_t^m + \mu_{t+k} \tag{14}$$

By comparing β_1 in model 13 to β_1 in model 14, we obtain an idea if the sentiment index predicts a greater effect for either *OBX* or *Allshare*.

Recall that the construction of the indexes depend on which return variable we use. When we use *Allshare* in the construction, the index is biased towards predicting the return of *Allshare* and not *OBX*. To control for this effect, we estimate the same two models using the alternative indexes constructed with *OBX*. These should be biased towards predicting the return of *OBX* and not *Allshare*. We therefore also estimate the following models.

$$OBX\ Return_{t+k} = \beta_0 + \beta_1 \, Sentiment\ Index\ OBX_t + \sum_{m=1}^{M} \gamma_m \, Control_t^m + \mu_{t+k} \tag{15}$$

$$Allshare\ Return_{t+k} = \beta_0 + \beta_1 \, Sentiment\ Index\ OBX_t + \sum_{m=1}^{M} \gamma_m \, Control_t^m + \mu_{t+k} \tag{16}$$

By comparing these four models, we obtain an understanding of whether the sentiment indexes predict the return of small and large stocks differently, and if the finding depends on the construction of the sentiment index.

We then proceed to models using the size portfolios as dependent variables. Recall that this is ten portfolios ranked after market size of the stocks included.

Both *Allshare* and *OBX* are value weighted, and weigh large stocks heavily. The sentiment indexes are thus potentially biased towards predicting the return of large stocks when both *Allshare* and *OBX* are used in the construction. To control for this we estimate models using both *Allshare* and *OBX* in the construction, as well as the equally weighted *EW* variable. The sentiment index should not be biased towards predicting large stocks when using *EW* in the construction. We estimate the following models for each size portfolio s.

$$Return_{t+k}^{s} = \beta_0 + \beta_1 \, Sentiment\ Index\ Allshare_t + \textstyle\sum_{m=1}^{M} \gamma_m \, Control_t^m + \mu_{t+k} \tag{17}$$

$$Return_{t+k}^{s} = \beta_0 + \beta_1 \, Sentiment\ Index\ OBX_t + \textstyle\sum_{m=1}^{M} \gamma_m \, Control_t^m + \mu_{t+k} \tag{18}$$

$$Return_{t+k}^{s} = \beta_0 + \beta_1 \, Sentiment\ Index\ EW_t + \textstyle\sum_{m=1}^{M} \gamma_m \, Control_t^m + \mu_{t+k} \tag{19}$$

By comparing the results for the ten size portfolios, we see if our sentiment indexes has different implications for small and large stocks. Secondarily we control if the results depend on the construction of the sentiment indexes or not.

3.5 Econometric concerns

In the following, we highlight some econometric concerns. We adjust all our variables as previously explained, and we have normally distributed variables we are able to work with. Yet we perform several tests to control for any problems.

Working with time series data might reveal problems of autocorrelation and non-stationarity. We test all our variables for non-stationarity and discover no concerns. We find some evidence of autocorrelation in some models. To account for this we use the Cochrane-Orchutt procedure and report altered results.

We find some evidence of heteroscedasticity in our models. To account for this we estimate heteroscedastic robust standard errors in all models.

Under chapter 3.3.8 we saw that some variables are correlated. *Consumer Confidence* and *Producer Confidence* are for instance correlated by 0.6, which might not be surprising. Correlated control variables might lead to less efficient coefficient estimates. To account for this, we estimate alternative models with the exclusion of various control variables that could potentially cause multicollinearity. We report alternative results if omitting variables alter the results.

Another concern is that the results depend on the specification of our control variables. We have applied Ramsey RESET tests to accommodate this. In addition, we have tested alternative adjustments of the control variables, and find that the variables applied in the reported models are specified and adjusted in the most reliable way.

Some control variables do not overlap with the full sample period of NEFA. We thus have to choose between using the full sample period and all control variables. To account for this we report both alternatives for the models that yields the most interesting results.

3.6 Validity and reliability

In the following, we reflect on what our empirical tests are able to answer. Our main hypothesis is that it is possible to predict future returns with a sentiment index. Indexes of this sort can be constructed in indefinite ways, and it is obvious that negative findings will not disprove the hypothesis. However, given a valid and reliable construction of our tests, a positive finding is able to confirm the hypothesis.

The internal validity of our tests rely on the primary data sources of Google Trends and the newspaper corpus. Particularly the newspaper corpus is constructed with different research in mind. We thus face some risk that the corpus is not consistent over the sample period. We have however controlled all included terms for deviations and find no reason to distrust the consistency. One advantage is that the corpus lets us control each data point manually. So whenever one term experiences a high occurrence during a period, we are able to verify the date, newspaper, article and sentence of each occurrence.

A disadvantage with Google Trends is that the occurrence of each term is calculated based on a large sample of Google searches and not complete data on all searches performed in the period. The time series for each term might therefore deviate some from the true historic search activity for the term. This is especially true for terms with a low level of search activity.

Even more, the internal validity relies on the construction of our indexes. One valid objection is that alternative construction methods can yield different results. We identify that the most questionable step is the selection of the terms that are included in the index each period. To accommodate this we perform alternative tests under subchapter 4.4 to check if our findings are sensitive to this selection criterion.

Another objection might be that our method can potentially yield results even with unrelated time series. In each period, we include only the significant terms out of a large pool. This might give associations to data mining. With a large enough pool some might argue that some terms will correlate, even without any systematic relation to returns. In response, we will highlight the fact that terms are selected only on basis of preceding data, such that the index for any one period is constructed out of sample. In addition, we perform a comprehensive test where we repeat the whole process, only with random, nonrelated terms. This is reported in subchapter 4.4.

Another issue is that the consumption pattern of newspapers and Google searches might have changed during the time span. When controlling both single terms and the complete indexes we find little evidence of any systematic changes in the pattern.

The external validity depends on the universality of our tests and findings. As previously noted, a negative finding will have little universal value, as other indexes can still confirm the hypothesis. Yet a positive finding has general implications. Our sentiment indexes are constructed with Norwegian data in the Norwegian language, and the results are thus applicable to the Norwegian market. The findings have general implications for the role of newspapers and Google searches in Norway. As OSE represents a well-functioning and internationally integrated market it is sensible to expect findings that are consistent with other markets. It is however likely that the role of Google searches and newspapers can be partly different between Norway and other countries.

4. Results and discussions

We now proceed to the results of our analysis. We present the results for weekly NEFA, monthly NEFA and monthly FEARS in subchapters 4.1, 4.2 and 4.3 respectively. For each subchapter we first present results for average returns before moving to the cross-sectional analysis of small and large stocks.

In subchapter 4.4 we present robustness tests we have performed to verify our results. In subchapter 4.5, we relate our findings to existing papers and theories.

4.1 Weekly Newspaper Index

The following subchapter presents the results for weekly NEFA. It is based on 872 weekly observations between 1998 and 2015.

4.1.1 Average return

Table 15 presents results for our models with weekly NEFA. We found that an increase in NEFA isolated predicts negative returns in the second subsequent week. The numbers suggest that a one standard deviation increase in NEFA is associated with 0.27% less return. The sign of the coefficient is as expected, as an increase in NEFA implies an increase in the occurrence of negative terms.

The result is significant at the 5% significance level. The results are similar when using *OBX*, *EW* and *VW* in the construction of NEFA. An alternative table with *NEFA OBX* is found in appendix 1. The results are similar when altering the number of control variables included, and the length of the sample period.

Recall that we constructed NEFA to predict the following week and not the second week. It is thus more surprising that NEFA is related to the returns in the second subsequent week. This is however consistent with findings in Larsen and Thorsrud (2017). This is discussed further under subchapter 4.5.

We also observe that the R-squared is declining the longer into the future one goes. The R-squared is below 0.12 for the second week. This relatively low value reflects that short-term returns are initially unpredictable. We see that the R-squared is above 0.35 for the first subsequent week. This is partly due to the explanatory power of VIX. The yearly dummy variables and the lagged return variables also contribute to inflate R-squared. This is discussed further under subchapter 4.3.1.

We also see that NEFA coincides with the return in the current week. The sign is positive, implying that increased NEFA coincides with positive returns. We find no evidence that NEFA forecasts any later periods than the second week. We have also estimated models with weekly NEFA where we use monthly aggregated return as dependent variables. We find no explanatory power for neither the first nor later months.

	(1) Allshare Return(t)	(2) Allshare Return(t+1)	(3) Allshare Return(t+2)	(4) Allshare Return(t+2)	(5) Allshare Return(t+2)	(6) Allshare Return(t+3)	(7) Allshare Return(t+4)
NEFA Allshare	0.00772*	0.00164	-0.00831*	-0.00783*	-0.00752*	0.00423	-0.00169
	(2.10)	(0.51)	(-2.26)	(-2.50)	(-2.43)	(1.16)	(-0.46)
VIX	-0.0862***	-0.157***	-0.0281*	-0.0315**	-0.0285**	-0.0141	-0.0118
	(-8.55)	(-14.96)	(-2.29)	(-2.94)	(-2.66)	(-1.12)	(-0.93)
Spread	-0.00690	-0.0375*	-0.0653**			0.0376	0.00716
	(-0.42)	(-2.10)	(-3.06)			(1.69)	(0.32)
Brent	0.903	1.011	0.505			0.904	1.186
	(1.35)	(1.27)	(0.52)			(0.86)	(1.12)
Consumer Confidence	0.0124*	0.0156*	0.0244**	0.0166*	0.0160***	0.0262**	0.0223*
	(2.18)	(2.30)	(2.93)	(2.56)	(3.61)	(2.91)	(2.47)
Fund Flow	0.000904***	0.00120***	0.00123**	0.00133***	0.00111**	0.000929*	0.000394
	(3.31)	(3.70)	(3.12)	(3.51)	(3.12)	(2.19)	(0.93)
Corporate Confidence	-0.912	-1.023	-0.522			-0.913	-1.200
	(-1.36)	(-1.28)	(-0.53)			(-0.86)	(-1.13)
Volatility	-0.0197	-0.0153	0.0164	0.0405	-0.0176	0.0596	0.0457
	(-0.62)	(-0.41)	(0.36)	(1.05)	(-0.56)	(1.22)	(0.93)
Yearly Dummy Variables	YES	YES	YES	YES	NO	YES	YES
Return(t-3)	0.0780*	-0.0598	0.0137	0.0108	0.0149	-0.0353	-0.0299
	(2.02)	(-1.72)	(0.34)	(0.31)	(0.43)	(-0.86)	(-0.72)
Return(t-2)	-0.233***	0.0205	-0.105**	-0.0660	-0.0595	-0.000560	-0.0349
	(-5.77)	(0.59)	(-2.62)	(-1.91)	(-1.73)	(-0.01)	(-0.85)
Return(t-1)	0.470***	-0.0205	-0.0233	-0.0198	-0.00893	-0.117**	-0.00351
	(12.46)	(-0.60)	(-0.59)	(-0.58)	(-0.26)	(-2.88)	(-0.09)
Return(t)		0.000323	-0.0277	-0.00272	0.0193	-0.0316	-0.114**
		(0.01)	(-0.69)	(-0.08)	(0.57)	(-0.78)	(-2.80)
_cons	-0.0222	-0.0229	-0.00916	0.00180	0.00286	-0.0230	-0.0278
	(-1.17)	(-1.02)	(-0.33)	(0.40)	(1.48)	(-0.77)	(-0.93)
N	662	662	662	872	872	661	660
R^2	0.391	0.362	0.118	0.091	0.065	0.077	0.062
adj. R^2	0.369	0.338	0.085	0.064	0.055	0.043	0.026

*Table 15: The table relates Allshare weekly returns to NEFA Allshare. The dependent variables are contemporaneous returns (model 1) and future Allshare weekly return in the next 4 weeks (model 2-7). The independent variable is the constructed sentiment index NEFA Allshare. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year = 2003 for model 1, 2, 3, 6, 7. Base year =1999 for model 4. *, ** and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

Table 16 shows additional analysis of NEFA. The first two models use the return in the second subsequent week as dependent variable. NEFA is omitted from model 2. If the control variables are not altered between the two models it supports the idea that NEFA has independent explanatory power. We see that none of the control variables are significantly altered, although we see some change in the coefficient for VIX. This suggests that NEFA is independent to a large extent, but that it is related to VIX to some extent.

The same is observed in the third model with NEFA as dependent variable. We see that VIX coincides with NEFA. The fourth model uses next week's NEFA as dependent variable. We see that neither VIX nor other control variables explain next week's NEFA. We find the same result for other future periods.

We see some evidence that the previous week's returns predict NEFA for the subsequent week. The return two weeks prior to week t has a positive sign, whereas returns three weeks prior has a negative sign. This result is however not robust, as the results disappear when using the current NEFA, and NEFA for the second and third subsequent weeks as dependent variables. Both models have an R-squared close to zero, which indicates that the sentiment of newspapers are unpredictable by our other variables.

The comparison of the first two models in table 16 yields another insight. By including NEFA in the model, we see that the R-squared increases by 0.006, from 0.112 to 0.118. This number suggests that the weekly NEFA can explain approximately 0.6% of the return in the second subsequent week[18]. We find approximately the same value when excluding yearly dummy variables and lagged returns in the models. This is not a large number, but it is still a significant contribution towards forecasting something that is initially unpredictable.

[18] The corresponding value when comparing the adjusted R-squared is 0.05. The values depend on the specification of the model and we do not interpret this as exact values, rather as approximations.

	(1)	(2)	(3)	(4)
	Allshare Return(t+2)	Allshare Return(t+2)	NEFA Allshare(t)	NEFA Allshare(t+1)
NEFA_Allshare	-0.00831*			
	(-2.26)			
VIX	-0.0281*	-0.0309*	0.341**	0.137
	(-2.29)	(-2.51)	(2.90)	(1.16)
Spread	-0.0653**	-0.0664**	0.0284	-0.126
	(-3.06)	(-3.11)	(0.15)	(-0.66)
Brent	0.505	0.508	0.566	1.733
	(0.52)	(0.52)	(0.07)	(0.22)
Consumer Confidence	0.0244**	0.0243**	0.0197	0.0188
	(2.93)	(2.91)	(0.29)	(0.28)
Fund Flow	0.00123**	0.00122**	0.000348	-0.00338
	(3.12)	(3.10)	(0.11)	(-1.04)
Corporate Confidence	-0.522	-0.525	-0.611	-1.767
	(-0.53)	(-0.54)	(-0.08)	(-0.22)
Volatility	0.0164	0.0150	0.0579	-0.240
	(0.36)	(0.33)	(0.16)	(-0.64)
Yearly Dummy Variables	YES	YES	YES	YES
Return(t-3)	0.0137	0.00613	0.554	-1.209**
	(0.34)	(0.15)	(1.35)	(-2.93)
Return(t-2)	-0.105**	-0.106**	0.331	1.005*
	(-2.62)	(-2.65)	(0.78)	(2.38)
Return(t-1)	-0.0233	-0.0228	-0.194	0.0972
	(-0.59)	(-0.57)	(-0.46)	(0.23)
Return(t)	-0.0277	-0.0331	0.738	0.269
	(-0.69)	(-0.83)	(1.78)	(0.65)
_cons	-0.00916	-0.00923	-0.0104	-0.0290
	(-0.33)	(-0.33)	(-0.05)	(-0.13)
N	662	662	662	662
R^2	0.118	0.112	0.033	0.030
adj. R^2	0.085	0.080	-0.002	-0.005

*Table 16: The first two models relate Allshare weekly returns to NEFA Allshare. The dependent variables are return two weeks into the future. In the first model, the independent variable is the constructed sentiment index NEFA Allshare. In the second model, we check whether the exclusion of NEFA alters the coefficients of the control variables. The third and fourth models relate NEFA Allshare to other control variables. The dependent variables are Allshare in the current (model 3) and NEFA Allshare in the following week (model 4). The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2003. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

4.1.2 Cross-sectional analysis

We now proceed to the results of our additional analysis on small and large stocks. We limit the analysis to the return of the second subsequent week, as this is the period that NEFA has shown to predict average returns.

We first cross-check the returns of *Allshare* and *OBX*. We use *NEFA OBX* as independent variable, and the return of *Allshare* as dependent variable, and opposite. Table 17 shows the results. We observe two things. The first is that constructing NEFA with *Allshare* yields slightly larger absolute coefficients and higher significance. The second and most interesting observation is larger absolute coefficient and higher significance for *OBX return* as a dependent variable. This is not due to using *OBX* in the construction of NEFA, since we control for using *NEFA Allshare*. Although the difference is not enough to conclude, this suggests that large stocks returns are more sensitive to changes in newspaper sentiment.

We observe a similar and even stronger pattern when cross-checking *EW* and *VW*. The results for these models are found in appendix 2. Recall that *VW* weighs large stocks relatively heavier than *EW*. We observe the same pattern that both the absolute coefficient is larger and the significance is higher for *VW return* as a dependent variable. This also gives support to the idea that large stocks returns are more sensitive to changes in newspaper sentiment.

The findings are similar when altering the composition of control variables and utilizing the full sample period.

	(1) Allshare Return(t+2)	(2) Allshare Return(t+2)	(3) OBX Return(t+2)	(4) OBX Return(t+2)
NEFA Allshare	-0.00831*		-0.00977*	
	(-2.26)		(-2.40)	
NEFA OBX		-0.00737*		-0.00897*
		(-2.08)		(-2.30)
VIX	-0.0281*	-0.0285*	-0.0354**	-0.0359**
	(-2.29)	(-2.32)	(-2.62)	(-2.65)
Spread	-0.0653**	-0.0652**	-0.0776***	-0.0774***
	(-3.06)	(-3.06)	(-3.32)	(-3.31)
Brent	0.505	0.443	0.277	0.204
	(0.52)	(0.45)	(0.26)	(0.19)
Consumer Confidence	0.0244**	0.0244**	0.0284**	0.0284**
	(2.93)	(2.94)	(3.12)	(3.13)
Fund Flow	0.00123**	0.00122**	0.00125**	0.00124**
	(3.12)	(3.09)	(2.93)	(2.91)
Corporate Confidence	-0.522	-0.459	-0.297	-0.224
	(-0.53)	(-0.47)	(-0.28)	(-0.21)
Volatility	0.0164	0.0168	0.0304	0.0309
	(0.36)	(0.37)	(0.68)	(0.69)
Yearly Dummy Variables	YES	YES	YES	YES
Return(t-3)	0.0137	0.0119	0.0214	0.0198
	(0.34)	(0.29)	(0.53)	(0.49)
Return(t-2)	-0.105**	-0.104**	-0.104**	-0.103**
	(-2.62)	(-2.60)	(-2.60)	(-2.59)
Return(t-1)	-0.0233	-0.0216	-0.0176	-0.0158
	(-0.59)	(-0.54)	(-0.44)	(-0.40)
Return(t)	-0.0277	-0.0255	-0.0465	-0.0440
	(-0.69)	(-0.64)	(-1.17)	(-1.11)
_cons	-0.00916	-0.00761	-0.00569	-0.00388
	(-0.33)	(-0.28)	(-0.19)	(-0.13)
N	662	662	662	662
R^2	0.118	0.117	0.116	0.116
adj. R^2	0.085	0.084	0.083	0.083

*Table 17: The table relates Allshare weekly returns and OBX weekly return to NEFA Allshare and NEFA OBX. The dependent variables are Allshare return two weeks into the future (model 1 and 2) and OBX return two weeks into the future (model 3 and 4). The independent variable is the constructed sentiment index, NEFA Allshare or NEFA OBX. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2003. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

We then proceed to the results of our analysis on the size portfolios. Table 18 shows results for NEFA regressed on all ten size portfolios in the second subsequent week. We see that NEFA only predicts the return of portfolios composed of large stocks.

The numbers suggest that a one standard deviation increase in NEFA is associated with 0.41% less return two weeks later for the 10% largest stocks. The finding is significant at a 1% level. This coincides with the previous crosscheck on *OBX* and *Allshare*, but the result is more unambiguous.

	(1) Portfolio 1 Return(t+2)	(2) Portfolio 2 Return(t+2)	(3) Portfolio 3 Return(t+2)	(4) Portfolio 4 Return(t+2)	(5) Portfolio 5 Return(t+2)	(6) Portfolio 6 Return(t+2)	(7) Portfolio 7 Return(t+2)	(8) Portfolio 8 Return(t+2)	(9) Portfolio 9 Return(t+2)	(10) Portfolio 10 Return(t+2)
NEFA Allshare	0.00378	-0.000198	-0.00214	-0.00335	-0.00822**	-0.00220	-0.00341	-0.00465	-0.00800	-0.0124**
	(1.61)	(-0.08)	(-0.80)	(-1.20)	(-2.93)	(-0.67)	(-1.15)	(-1.42)	(-1.92)	(-2.85)
VIX	-0.0200**	-0.0317***	-0.0293***	-0.0382***	-0.0269**	-0.0300**	-0.00950	-0.0210	-0.0400**	-0.0353*
	(-2.60)	(-3.70)	(-3.38)	(-4.48)	(-2.78)	(-2.71)	(-0.73)	(-1.88)	(-2.89)	(-2.44)
Spread	-0.0151	-0.0285	-0.0239	-0.0178	-0.0444**	-0.0610**	-0.0387	-0.0490*	-0.0892***	-0.100***
	(-1.12)	(-1.91)	(-1.60)	(-1.24)	(-2.59)	(-3.14)	(-1.10)	(-2.48)	(-3.72)	(-3.98)
Brent	1.659***	0.677	1.302*	0.915	0.351	0.969	0.814	-0.327	0.220	-0.306
	(3.38)	(1.29)	(2.39)	(1.89)	(0.54)	(1.33)	(1.17)	(-0.45)	(0.25)	(-0.34)
Consumer Confidence	0.00318	0.0216***	0.00636	0.0297***	0.0155*	0.0213**	0.0238*	0.0306***	0.0234*	0.0324**
	(0.64)	(3.83)	(1.15)	(5.49)	(2.31)	(2.80)	(2.38)	(3.79)	(2.58)	(3.30)
Fund Flow	0.000761**	0.000712**	0.000932***	0.00118***	0.00115***	0.000619	0.00118**	0.00116**	0.00121**	0.00124**
	(3.11)	(2.64)	(3.44)	(4.66)	(3.47)	(1.71)	(2.58)	(3.13)	(2.72)	(2.66)
Corporate Confidence	-1.667***	-0.682	-1.311*	-0.924	-0.355	-0.978	-0.828	0.312	-0.241	0.275
	(-3.38)	(-1.29)	(-2.40)	(-1.90)	(-0.55)	(-1.34)	(-1.18)	(0.43)	(-0.28)	(0.30)
Volatility	0.0456	0.0965*	0.0659	0.0895*	0.123*	0.0752	0.0994	0.0970	0.125	0.106
	(1.14)	(2.20)	(1.46)	(2.16)	(2.23)	(1.25)	(1.00)	(1.58)	(1.69)	(1.34)
Yearly Dummy Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Return(t-3)	0.0347	-0.00310	0.0173	0.000734	0.0716	-0.0592	-0.0132	-0.00306	0.0552	0.0562
	(0.90)	(-0.08)	(0.44)	(0.02)	(1.82)	(-1.47)	(-0.25)	(-0.08)	(1.38)	(1.40)
Return(t-2)	-0.0289	-0.0526	-0.00956	-0.0246	-0.0775*	-0.0331	-0.0891	-0.103*	-0.0358	-0.0973*
	(-0.74)	(-1.40)	(-0.24)	(-0.63)	(-1.98)	(-0.83)	(-1.81)	(-2.58)	(-0.90)	(-2.44)
Return(t-1)	-0.0202	0.0294	-0.0154	-0.0145	-0.0131	-0.110**	-0.0370	-0.0278	-0.0860*	0.00656
	(-0.51)	(0.78)	(-0.39)	(-0.37)	(-0.33)	(-2.77)	(-0.67)	(-0.69)	(-2.18)	(0.17)
Return(t)	-0.0643	-0.0761	-0.0812*	-0.0942*	0.0322	-0.0485	-0.0224	0.0419	-0.0404	-0.0223
	(-1.62)	(-1.94)	(-2.06)	(-2.42)	(0.83)	(-1.22)	(-0.43)	(1.05)	(-1.02)	(-0.57)
_cons	-0.0526**	-0.0192	-0.0323	-0.0234	-0.00561	-0.0283	-0.0255	0.0141	-0.00424	0.0105
	(-2.94)	(-0.99)	(-1.64)	(-1.32)	(-0.24)	(-1.06)	(-1.01)	(0.53)	(-0.13)	(0.31)
N	662	662	662	662	662	662	663	662	662	662
adj. R^2	0.072	0.111	0.100	0.195	0.103	0.093	0.063	0.095	0.089	0.094

*Table 18: The table relates weekly returns of 10 size portfolios on OSE to NEFA Allshare. The dependent variables are return two weeks into the future for each size portfolio, respectively. The independent variable is NEFA Allshare. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow, historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2003. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

We find the same results when using *NEFA OBX*, and when changing the composition of control variables and altering the length of the sample period. An alternative table for the latter can be found in appendix 3.

A peculiar result is that portfolio 5, consisting of the 10% stocks that are just below the median size, is significant. This does not fit with the overall picture that only large stocks relate to NEFA. One possible explanation is that the portfolio contains one or more stocks that for some reason (unrelated to size) relates to NEFA.

Both *Allshare* and *OBX* are value weighted. Since a few large stocks dominate OSE, the return of these are given much weight when constructing both *NEFA Allshare* and *NEFA OBX*; these might therefore both be biased towards predicting large stocks. To control for this we also construct an alternative *NEFA EW* and regress the size portfolios with this variable. Recall that *EW* is equally weighted and that *NEFA EW* thus has no bias towards predicting returns of large stocks over small stocks. The results are however similar, and we can conclude that the results are not dependent on which variable we use in the construction of NEFA. This table is found in appendix 4.

4.1.3 Conclusions

We find that our weekly newspaper sentiment index significantly predicts the return of large stocks on OSE in the second subsequent week. NEFA is largely unpredictable by our other variables, but is ambiguously related to prior returns, and it is partly coinciding with VIX. The newspaper sentiment thus only predict the return of large stocks. Recall that the terms are not associated with any specific type of stock, but rather general economic terms. This is discussed further under subchapter 4.5.

4.2 Monthly Newspaper Index

We now proceed to the results of our analysis on the monthly NEFA index. The results are based on the 208 monthly observations between 1998 and 2015.

4.2.1 Average return

Table 19 shows results for our models on monthly NEFA. We see some evidence of a relationship between NEFA and returns two and three months into the future, but these are not present when utilising the full sample period. The results are the same when using *OBX*, *EW* and *VW* in the construction of FEARS. We have also run models using subsequent weekly returns as dependent variables. This does not yield any results either. We thus conclude that the monthly NEFA index does not show any convincing results for average returns.

The table also shows results for models using NEFA as dependent variable. We see that none of the variables forecast NEFA for the subsequent month. The same is the case for later months. In the current month, we see that NEFA coincides with Brent oil price. Increased oil prices seems to coincide with an increasingly negative sentiment in newspapers. This is surprising as an increased oil price is expected to have a positive effect on the Norwegian economy. The relationship is however not very robust to alterations of sample period and control variables.

We also see some evidence that positive return in the previous month predicts a decline in the negative sentiment in the current month. This relationship is however not apparent in the model for next month's return, as current return is not significant when using next month's NEFA as dependent variable.

	(1) Allshare Return(t)	(2) Allshare Return(t+1)	(3) Allshare Return(t+2)	(4) Allshare Return(t+3)	(5) Allshare Return(t+3)	(6) Allshare Return(t+4)	(7) NEFA Allshare(t)	(8) NEFA Allshare(t+1)
NEFA Allshare	-0.0140	0.00261	0.0150	0.0177	0.00456	-0.0117		
	(-1.84)	(0.33)	(1.54)	(1.87)	(0.59)	(-1.22)		
VIX	-0.0492*	-0.209***	-0.0399	0.0475	0.0496	0.0377	0.0963	-0.274
	(-2.06)	(-8.30)	(-1.33)	(1.63)	(1.90)	(1.28)	(0.36)	(-1.32)
Spread	0.0268	-0.0176	0.0174	0.0239		0.0470	0.323	-0.779
	(0.63)	(-0.40)	(0.33)	(0.47)		(0.92)	(0.70)	(-1.92)
Brent	0.111*	0.0312	-0.0411	-0.00924		0.0109	1.265*	-0.565
	(2.15)	(0.58)	(-0.62)	(-0.14)		(0.17)	(2.17)	(-1.22)
Consumer Confidence	0.0489*	0.0291	0.0681*	0.0827**	0.0458	0.0496	0.459	0.0934
	(2.09)	(1.20)	(2.23)	(2.76)	(1.80)	(1.62)	(1.62)	(0.83)
Fund Flow	0.00432***	-0.000882	-0.00183	-0.00136		-0.000156	0.0232	0.00664
	(3.38)	(-0.65)	(-1.06)	(-0.81)		(-0.09)	(1.46)	(0.52)
Corporate Confidence	0.00263	0.0289	0.0608	0.0493		-0.0657*		
	(0.11)	(1.14)	(1.89)	(1.56)		(-2.03)		
Volatility	-0.810***	-0.194	0.0914	0.216	0.163	0.0158	-2.987	-0.760
	(-5.36)	(-1.13)	(0.43)	(1.04)	(0.95)	(0.07)	(-1.53)	(-0.45)
Yearly Dummy Variable	YES	YES	YES	YES	YES	YES	YES	YES
Return(t-3)	-0.143*	-0.0979	-0.190*	-0.156	-0.148*	-0.130	-0.970	-1.109
	(-2.02)	(-1.32)	(-2.15)	(-1.82)	(-2.05)	(-1.50)	(-1.23)	(-1.61)
Return(t-2)	-0.248***	-0.0769	-0.168	-0.264**	-0.201**	-0.174	-1.342	-0.341
	(-3.39)	(-0.97)	(-1.80)	(-2.91)	(-2.75)	(-1.90)	(-1.65)	(-0.50)
Return(t-1)	-0.0715	-0.101	-0.108	-0.193	-0.190*	-0.255*	-2.609**	-0.0439
	(-0.81)	(-1.11)	(-1.01)	(-1.85)	(-2.38)	(-2.41)	(-2.76)	(-0.06)
Return(t)		0.0740	-0.0867	-0.0959	-0.0389	-0.172	-1.545	-1.002
		(0.81)	(-0.79)	(-0.90)	(-0.49)	(-1.59)	(-1.57)	(-1.19)
_cons	0.0805***	0.0475**	0.0387	0.0322	-0.00114	0.0549*	0.414	0.239
	(5.29)	(2.80)	(1.75)	(1.47)	(-0.04)	(2.44)	(1.95)	(1.23)
N	152	152	152	152	208	152	152	189
R^2	0.607	0.596	0.346	0.374	0.301	0.342	0.179	0.112
adj. R^2	0.536	0.520	0.223	0.255	0.205	0.217	0.039	-0.024

*Table 19: The table relates Allshare monthly returns to monthly NEFA Allshare. The dependent variables are contemporaneous returns (model 1), future Allshare mothly return in the next 4 months (model 2-6), and NEFA allshare in the current and in the following month. In the first six models, the independent variable is the NEFA sentiment index. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. In model 6 and 7 we see if there is a relation between the control variables and the independent variable. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2003 for models 1, 2, 3, 4, 6, 7, 8. Base year=1999 for model 5. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

Note that the R-squared is higher in our monthly models then in our weekly models. The main reason is the increased explanatory power of the constant, the lagged return variables and the yearly dummy variables. This issue is discussed further under subchapter 4.3.1.

4.2.2 Cross-sectional analysis

The monthly NEFA has not resulted in any convincing results for average return. Yet there is a possibility that it yields results for either small or large stocks when analysed separately. We now proceed to the results of the cross-sectional analysis.

When cross-checking *OBX* and *Allshare*, we find no differences. This suggests that neither small nor large stocks are particularly sensitive to changes in monthly NEFA. The results are similar when using different compositions of control variables, and altering the sample period.

This corresponds to the analysis of size portfolios. We find no evidence that NEFA predicts neither small nor large stocks. The results are the same when altering the composition of control variables and utilizing the full sample period, and when using *OBX*. *EW* and *VW* in the construction of NEFA respectively.

We thus conclude that the monthly NEFA does not predict small or large stocks when analysed separately.

4.2.3 Conclusions

We find no convincing evidence that the sentiment of newspapers predict subsequent returns on OSE when measured on a monthly basis. The results are the same for average returns, and small and large stocks analysed separately. We do not find any evidence that any other variables can predict the sentiment of newspapers in a subsequent period neither.

The lack of any convincing results can be due to different reasons. Either our selected method is not the best to capture monthly newspaper sentiment. Alternatively, the newspaper sentiment measured on a monthly basis is simply too stable and unrelated to stock returns.

4.3 Monthly Google Search Index

We now proceed to the results of our monthly Google search index. Recall that we have a sample period limited to 127 monthly observations. The sample period ranges from May 2008 until November 2019.

4.3.1 Average return

Table 20 shows results for models with monthly return as dependent variables. We see that FEARS coincides with returns the current month, and that it isolated predicts positive returns in the first subsequent month. This implies that whenever the searches for negative terms increase, it predicts increased stock prices the next month.

The numbers suggest that a one standard deviation increase in FEARS is associated with an approximately 1.2% higher return the following month. The result is significant at the 0.1% level and is robust to changes in the inclusion of control variables. We find similar results when using *OBX*, *EW* and *VW* in the construction of FEARS and as dependent variables.

This does not accord with the idea that a negative sentiment predict declining returns. It does however accord with findings in Fisher and Statman (2000), who find that the monthly sentiment of individual investors serves as a contrarian indicator. This will be discussed more under subchapter 4.5.2.

	(1) Allshare Return(t)	(2) Allshare Return(t+1)	(3) Allshare Return(t+1)	(4) Allshare Return(t+1)	(5) Allshare Return(t+2)	(6) Alshare Return(t+3)	(7) Allshare Return(t+3)	(8) Allshare Return(t+4)
FEARS_Allshare	0.0297**	0.0371***	0.0452***	0.0406**	0.0188	-0.0275*	-0.00668	-0.0226
	(3.33)	(3.54)	(4.62)	(3.09)	(1.39)	(-2.29)	(-0.52)	(-1.74)
VIX	-0.0606***	-0.149***	-0.158***		-0.0454	0.0614**	0.0416	0.0304
	(-3.88)	(-8.10)	(-8.77)		(-1.93)	(2.95)	(1.79)	(1.31)
Spread	-0.00421	0.0455			0.115*	0.104*		0.0164
	(-0.12)	(1.13)			(2.24)	(2.27)		(0.33)
Brent	0.146***	0.0557			-0.0625	-0.124*		-0.0497
	(3.68)	(1.19)			(-1.01)	(-2.26)		(-0.83)
Consumer Confidence	0.00557	0.00473	0.00698	0.0122	0.00923	0.0111	0.0220	0.0107
	(1.03)	(0.76)	(1.26)	(1.10)	(0.93)	(1.36)	(1.90)	(1.09)
Corporate Confidence	-0.0307	-0.0192	0.0553***	0.00748	-0.00269	-0.0306	0.107***	-0.0710
	(-1.40)	(-0.76)	(3.40)	(0.28)	(-0.07)	(-0.95)	(3.52)	(-1.88)
Fund Flow	0.000797	-0.000700			-0.000260	-0.000552		-0.00108
	(0.99)	(-0.76)			(-0.19)	(-0.48)		(-0.82)
Volatility	-0.634***	-0.268	-0.313***	-0.270*	0.400*	0.343	-0.0868	0.233
	(-5.71)	(-1.87)	(-3.38)	(-2.20)	(2.01)	(1.98)	(-0.50)	(1.20)
Yearly Dummy Variables	YES	YES	NO	NO	YES	YES	NO	YES
Return(t-3)	0.0815	-0.00576	-0.0869		-0.0637	-0.201*	-0.283**	-0.0773
	(1.15)	(-0.07)	(-1.29)		(-0.67)	(-2.37)	(-3.29)	(-0.84)
Return(t-2)	-0.466***	-0.0278	0.00101		-0.0207	-0.117	-0.245*	-0.254**
	(-6.43)	(-0.31)	(0.01)		(-0.21)	(-1.33)	(-2.49)	(-2.74)
Return(t-1)	0.228**	-0.278**	-0.302***		0.0523	-0.00786	-0.118	-0.0653
	(2.63)	(-2.95)	(-3.39)		(0.47)	(-0.08)	(-1.11)	(-0.60)
Return(t)		0.0985	0.250**		-0.0897	-0.00577	0.00748	0.0538
		(0.95)	(2.73)		(-0.73)	(-0.05)	(0.07)	(0.45)
_cons	0.0224	-0.0414*	0.0226***	0.0196*	-0.158***	-0.192***	0.0136	-0.127***
	(1.30)	(-2.09)	(3.86)	(2.46)	(-5.16)	(-7.53)	(1.20)	(-4.15)
N	126	126	126	126	125	124	124	123
R^2	0.822	0.773	0.737	0.134	0.485	0.634	0.215	0.402
adj. R^2	0.786	0.725	0.717	0.105	0.374	0.555	0.153	0.270

*Table 20: The table relates Allshare monthly returns to FEARS Allshare. The dependent variables are contemporaneous returns (model 1) and future Allshare monthly return in the next 4 months (model 2-8). The independent variable is FEARS Allshare. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2008. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

We see that the R-squared of the models are high. There are more than one reason for this, and some of these implies that we should interpret the value with scepticism. Monthly VIX contributes significantly towards predicting monthly returns. More concerning is the fact that we have a limited sample period of only 126 monthly observations ranging from 2008 until 2018. In the years following the financial crisis of 2007/ 2008, OSE experienced first a continuous fall, followed by a period of continuous strong returns. This is visible in figure 1 under chapter 3.1. This is manifested in significant yearly dummy variables in our models. In model 3, where the yearly dummy variables are omitted, it is manifested in significant results for the lagged monthly return variables. The effect of having a limited sample during this specific period thus contributes towards high values for R-squared, when we include yearly dummies and lagged returns as control variables. This is visible for instance in model 4, where both the yearly dummy variables and lagged returns are omitted. The R-squared of this model is significantly lower.

Our concern is to get unbiased and valid estimates for FEARS. The results for model 2, 3 and 4 show that both the coefficients and the significance of FEARS are largely unaffected by the effects that inflate R-squared. This does however highlight the fact that our results are based on a limited sample.

Table 21 shows results for the monthly FEARS and weekly returns. Week 1 refers to the first week following and so on. Model 5 uses aggregated returns for the four subsequent weeks as dependent variable. This corresponds to the first subsequent month, as used as dependent variable in model 2 in table 20. The monthly data uses calendar months, and the weekly data uses calendar weeks. The two ways of calculating monthly return is thus not completely equal, yet the results are almost identical.

We see that the second and fourth week are significant in the direction we expect from the monthly finding. We found that an increase in FEARS is associated with positive returns in the first following month. The weekly analysis indicates that it is primarily the second and fourth week that accounts for the effect.

	(1) Allshare Return(t_{week}+1)	(2) Allshare Return(t_{week}+2)	(3) Allshare Return(t_{week}+3)	(4) Allshare Return(t_{week}+4)	(5) Cumulative Return Week 1-4	(6) Allshare Return(t_{week}+5)	(7) Allshare Return(t_{week}+6)	(8) Allshare Return(t_{week}+7)	(9) Allshare Return(t_{week}+8)
FEARS Allshare	-0.00687	0.0170*	0.00513	0.0194**	0.0376***	0.0171	0.0116	-0.00902	-0.00902
	(-0.97)	(2.18)	(0.60)	(2.64)	(3.66)	(1.78)	(1.47)	(-1.22)	(-1.22)
VIX	-0.103***	-0.0319*	-0.0292	0.0141	-0.146***	-0.00503	-0.0241	-0.0176	-0.0176
	(-8.38)	(-2.35)	(-1.98)	(1.11)	(-8.06)	(-0.30)	(-1.76)	(-1.36)	(-1.36)
Spread	0.160***	-0.150***	-0.0869**	0.116***	0.0447	-0.0309	0.0629*	0.109***	0.109***
	(5.96)	(-5.10)	(-2.68)	(4.15)	(1.13)	(-0.85)	(2.10)	(3.82)	(3.82)
Brent	-0.0289	0.0962**	0.0232	-0.0181	0.0382	-0.0103	0.0290	-0.0686*	-0.0686*
	(-0.89)	(2.68)	(0.59)	(-0.54)	(0.82)	(-0.24)	(0.80)	(-2.04)	(-2.04)
Consumer Confidence	0.00976	0.00935	-0.00542	-0.00895	0.00456	0.0155*	0.00573	-0.00436	-0.00436
	(1.92)	(1.57)	(-0.90)	(-1.65)	(0.75)	(2.29)	(1.00)	(-0.97)	(-0.97)
Fund Flow	-0.000648	-0.000174	-0.00105	0.00117	-0.000524	0.000783	-0.00118	0.000264	0.000264
	(-0.93)	(-0.22)	(-1.26)	(1.60)	(-0.58)	(0.84)	(-1.52)	(0.40)	(0.40)
Corporate Confidence	-0.00490	-0.0454*	-0.0103	-0.00306	-0.0175	-0.0827**	0.0352	-0.0101	-0.0101
	(-0.25)	(-2.01)	(-0.44)	(-0.15)	(-0.71)	(-3.15)	(1.60)	(-0.56)	(-0.56)
Volatility	-0.110	-0.131	-0.339**	0.150	-0.283*	0.165	-0.203	0.243*	0.243*
	(-1.06)	(-1.12)	(-2.73)	(1.39)	(-2.01)	(1.18)	(-1.76)	(2.36)	(2.36)
Yearly Dummy Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES
Return(t-3)	0.0910	-0.147**	0.0195	0.0515	-0.0258	-0.0921	0.0175	-0.0393	-0.0393
	(1.82)	(-2.66)	(0.33)	(0.99)	(-0.34)	(-1.36)	(0.32)	(-0.73)	(-0.73)
Return(t-2)	-0.0196	-0.102	0.00797	-0.0149	0.00367	-0.00862	-0.110	-0.0108	-0.0108
	(-0.38)	(-1.84)	(0.13)	(-0.28)	(0.04)	(-0.12)	(-1.94)	(-0.18)	(-0.18)
Return(t-1)	-0.0148	-0.157*	0.00528	-0.0338	-0.304**	0.0562	-0.103	0.181**	0.181**
	(-0.25)	(-2.43)	(0.07)	(-0.55)	(-3.24)	(0.71)	(-1.57)	(2.75)	(2.75)
Return(t)	0.0670	-0.129	-0.166*	0.0109	0.118	-0.0893	-0.0994	-0.0817	-0.0817
	(1.04)	(-1.83)	(-2.14)	(0.16)	(1.14)	(-1.02)	(-1.39)	(-1.13)	(-1.13)
_cons	-0.0423**	-0.0306	0.0323	-0.0199	-0.0400*	-0.0982***	0.00922	-0.0626***	-0.0626***
	(-2.69)	(-1.67)	(1.73)	(-1.18)	(-2.07)	(-4.69)	(0.52)	(-4.41)	(-4.41)
N	125	125	125	125	125	125	125	125	125
R^2	0.672	0.485	0.333	0.413	0.783	0.405	0.314	0.443	0.443
adj. R^2	0.602	0.374	0.190	0.286	0.736	0.277	0.166	0.323	0.323

*Table 21: The table relates Allshare weekly returns to monthly FEAR Allshare. The dependent variables are future Allshare weekly return in the next 8 weeks. In model 5, the dependent variable is Allshare return over the first four weeks. The independent variable is the monthly FEAR Allshare sentiment index. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2008. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

Table 22 shows results for additional analysis of FEARS. Both model 1 and 2 predict returns in the first subsequent month. They are identical, except that FEARS is omitted in the second model. If the control variables are not altered between the two models, it supports the idea that FEARS is independent of the other variables. We see that some of the control variables are altered to some extent. This indicates that FEARS is somewhat related to *Spread, Fund Flow, Corporate Confidence* and *Volatility*.

Model 3 and 4 use FEARS as dependent variables. We see that FEARS in the current month corresponds to the current month's volatility. This indicates that increased volatility occur together with an increase in searches of negative economic terms, which is not very surprising. We also see that increased FEARS coincides with positive monthly returns.

We also see that current changes in yield spread predict next month's FEARS. As discussed under subchapter 3.3.5, an increase in *Spread* can be due to either increased yield on 10-year bonds or decreased yields on 3-month bills. When 3-month yields decrease, it is often interpreted as a negative shift in the short-term economic outlook (Bodie, Kane and Marcus, 2018, pp. 480). It is not surprising if this is followed by an increase in Google searches for terms like r*ecession*, *bankruptcy* and *default*.

The comparison of model 1 and 2 also show how much R-squared is altered by including FEARS. We see that FEARS increases R-squared by 0.059[19]. We find approximately the same value when excluding yearly dummy variables and lagged returns in the models. A direct interpretation of this value is that FEARS explain as much as 5.9% of next months stock return.

[19] The corresponding value when comparing the adjusted R-squared is 0.068. The values depend on the specification of the model and we do not interpret this as exact values, rather as approximations.

	(1) Allshare Return(t+1)	(2) Allshare Return(t+1)	(3) FEARS Allshare(t)	(4) FEARS Allshare(t+1)
FEARS Allshare	0.0371***			
	(3.54)			
VIX	-0.149***	-0.158***	-0.128	-0.0549
	(-8.10)	(-8.02)	(-0.76)	(-0.33)
Spread	0.0455	0.0669	0.594	1.171**
	(1.13)	(1.56)	(1.61)	(3.20)
Brent	0.0557	0.0512	-0.364	-0.398
	(1.19)	(1.01)	(-0.83)	(-0.90)
Consumer Confidence	0.00473	0.00436	-0.0171	-0.0121
	(0.76)	(0.62)	(-0.25)	(-0.19)
Fund Flow	-0.000700	-0.00130	-0.0158	-0.00550
	(-0.76)	(-1.30)	(-1.71)	(-0.61)
Corporate Confidence	-0.0192	-0.0295	-0.0802	0.233
	(-0.76)	(-1.04)	(-0.30)	(0.91)
Volatility	-0.268	-0.176	3.679**	0.691
	(-1.87)	(-1.16)	(2.70)	(0.51)
Yearly Dummy Variables	YES	YES	YES	YES
Return(t-3)	-0.00576	0.0417	0.701	0.736
	(-0.07)	(0.52)	(1.02)	(1.07)
Return(t-2)	-0.0278	-0.0330	1.312	-0.972
	(-0.31)	(-0.38)	(1.87)	(-1.37)
Return(t-1)	-0.278**	-0.213*	0.414	0.647
	(-2.95)	(-2.18)	(0.51)	(0.80)
Return(t)	0.0985	0.107	2.558**	-0.716
	(0.95)	(1.04)	(2.97)	(-0.83)
_cons	-0.0414*	-0.0650**	-0.622**	-0.581**
	(-2.09)	(-3.07)	(-3.10)	(-3.00)
N	126	126	126	125
R^2	0.773	0.714	0.264	0.289
adj. R^2	0.725	0.657	0.116	0.144

*Table 22: The first two models relate Allshare monthly returns to FEARS Allshare. The dependent variables are return one month into the future (model 1 and 2). In the first model, the independent variable is FEARS Allshare. In the second model, we check whether the exclusion of FEARS alters the coefficients of the control variables. The third and fourth models relate FEARS Allshare to other control variables. The dependent variables are FEARS Allshare in the current (model 3) and FEARS Allshare in the following month (model 4). The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2008. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

4.3.2 Cross-sectional analysis

We now proceed to the results of our additional analysis on small and large stocks. We limit the analysis to the return of the first subsequent month, as this is the period that FEARS has shown to predict for average returns.

Table 23 reports our models for crosschecking the returns of *OBX* and *Allshare*. We use *NEFA OBX* as independent variable, and the return of *Allshare* as dependent variable, and opposite. We see that the significance is slightly higher when using Allshare in the construction of FEARS. However, we observe no sign that FEARS predicts the return of either *OBX* or *Allshare* better than the other does. The same is the case when crosschecking *EW* and *VW*.

	(1) OBX Return(t+1)	(2) OBX Return(t+1)	(3) Allshare Return(t+1)	(4) Allshare Return(t+1)
FEAR_OBX	0.0332**		0.0335**	
	(2.93)		(3.15)	
FEAR Allshare		0.0373**		0.0371***
		(3.34)		(3.54)
VIX	-0.161***	-0.160***	-0.150***	-0.149***
	(-8.05)	(-8.14)	(-8.00)	(-8.10)
Spread	0.0395	0.0329	0.0504	0.0455
	(0.90)	(0.76)	(1.23)	(1.13)
Brent	0.0419	0.0539	0.0464	0.0557
	(0.85)	(1.11)	(0.97)	(1.19)
Consumer Confidence	0.00533	0.00586	0.00427	0.00473
	(0.80)	(0.90)	(0.66)	(0.76)
Fund Flow	-0.000917	-0.000762	-0.000857	-0.000700
	(-0.94)	(-0.79)	(-0.91)	(-0.76)
Corporate Confidence	-0.0181	-0.00994	-0.0287	-0.0192
	(-0.67)	(-0.38)	(-1.10)	(-0.76)
Volatility	-0.196	-0.206	-0.264	-0.268
	(-1.36)	(-1.46)	(-1.80)	(-1.87)
Yearly Dummy Variables	YES	YES	YES	YES
Return(t-3)	-0.0224	-0.0392	0.0138	-0.00576
	(-0.29)	(-0.52)	(0.18)	(-0.07)
Return(t-2)	0.0219	0.0149	-0.0295	-0.0278
	(0.25)	(0.17)	(-0.34)	(-0.31)
Return(t-1)	-0.309**	-0.342***	-0.237*	-0.278**
	(-3.26)	(-3.63)	(-2.50)	(-2.95)
Return(t)	0.131	0.145	0.0690	0.0985
	(1.27)	(1.43)	(0.65)	(0.95)
_cons	-0.0506*	-0.0437*	-0.0483*	-0.0414*
	(-2.34)	(-2.02)	(-2.41)	(-2.09)
N	126	126	126	126
R^2	0.775	0.784	0.760	0.773
adj. R^2	0.727	0.738	0.709	0.725

*Table 23: The table relates Allshare monthly returns and OBX monthly return to monthly FEARS Allshare and monthly FEARS OBX. The dependent variables are Allshare return one month into the future (model 1 and 2) and OBX return one months into the future (model 3 and 4). In model 1 and 3 the independent variable is FEARS OBX. In model 2 and 4 the independent variable FEARS Allshare. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2008. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

We now proceed to the analysis of the return of the ten size portfolios. The results are reported in table 24. We see that both the coefficient is larger and the significance is higher for the five portfolios with the largest stocks on OSE. Portfolio 1, 4 and 5 are however also significance at the 5% level. The results are similar when using *OBX* in the construction of FEARS. Recall that both *OBX* and *Allshare* are value weighted and might be biased towards predicting the return of large stocks over small stocks. We therefore also construct an alternative *FEARS EW*, which should not have any bias towards predicting the return of large stocks. The results are however even more convincing. We thus conclude that the results are stronger for large stocks than small stocks. A table showing these results are found in appendix 5.

This suggest that the explanatory power of FEARS on next month's return is primarily present for large stocks. We do however find some evidence that some of the effect is also present for small stocks. This finding is consistent with findings in Fisher and Statman (2000). This is discussed under subchapter 4.5.2.

	(1) Portfolio 1 Return(t+1)	(2) Portfolio 2 Return(t+1)	(3) Portfolio 3 Return(t+1)	(4) Portfolio 4 Return(t+1)	(5) Portfolio 5 Return(t+1)	(6) Portfolio 6 Return(t+1)	(7) Portfolio 7 Return(t+1)	(8) Portfolio 8 Return(t+1)	(9) Portfolio 9 Return(t+1)	(10) Portfolio 10 Return(t+1)
FEARS Allshare	0.0220*	0.0199	0.0209	0.0212*	0.0228*	0.0530***	0.0392***	0.0314*	0.0312**	0.0491***
	(2.55)	(1.56)	(1.94)	(2.15)	(2.18)	(4.69)	(3.94)	(2.53)	(2.86)	(4.14)
VIX	-0.0154	-0.0550*	-0.0689*	-0.0810***	-0.119***	-0.0962***	-0.138***	-0.101***	-0.167***	-0.174***
	(-0.81)	(-2.51)	(-2.19)	(-3.66)	(-4.68)	(-3.88)	(-4.44)	(-3.90)	(-4.36)	(-6.48)
Spread	0.0273	0.0907**	0.0654*	0.0472	0.0757	0.0395	0.109*	0.0467	0.0742	-0.0177
	(0.80)	(2.66)	(2.02)	(0.95)	(1.92)	(0.79)	(2.34)	(0.69)	(1.48)	(-0.38)
Brent	-0.0815	0.0683	0.127	0.138	-0.00117	-0.00922	-0.108	0.0522	0.0836	0.246*
	(-0.78)	(0.58)	(1.48)	(0.93)	(-0.01)	(-0.08)	(-0.84)	(0.44)	(0.63)	(2.33)
Consumer Confidence	-0.00170	0.0342*	0.0162*	0.00759	0.00550	0.00546	-0.000259	0.00700	0.00272	0.00559
	(-0.27)	(2.29)	(2.50)	(1.00)	(0.65)	(1.01)	(-0.05)	(1.07)	(0.57)	(0.72)
Fund Flow	0.000574	0.00108	-0.00188	-0.00119	0.00183	0.000656	0.000308	-0.000325	-0.00143	-0.00186
	(0.60)	(0.97)	(-1.54)	(-1.35)	(1.64)	(0.73)	(0.33)	(-0.35)	(-1.23)	(-1.83)
Corporate Confidence	-0.0285	0.00336	0.0215	-0.0413	-0.0855*	-0.0133	0.0227	0.0446	0.00301	-0.0235
	(-0.97)	(0.09)	(0.79)	(-1.27)	(-2.62)	(-0.37)	(0.82)	(1.40)	(0.11)	(-0.65)
Volatility	-0.455**	-0.418*	-0.583***	-0.390	-0.442	-0.363	-0.398	-0.110	-0.318	-0.467
	(-2.76)	(-1.99)	(-3.40)	(-1.58)	(-1.98)	(-1.44)	(-1.82)	(-0.43)	(-1.37)	(-1.97)
Yearly Dummy Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Return(t-3)	-0.235**	-0.103	0.0191	-0.0261	0.0492	-0.235**	-0.0767	-0.00575	-0.0956	0.00994
	(-2.95)	(-1.06)	(0.20)	(-0.25)	(0.57)	(-2.71)	(-1.08)	(-0.06)	(-1.33)	(0.13)
Return(t-2)	0.152	-0.0478	-0.0938	0.0215	-0.0909	0.136	0.132	0.0918	0.0423	0.0630
	(1.35)	(-0.34)	(-1.39)	(0.26)	(-1.18)	(1.42)	(1.44)	(1.02)	(0.52)	(0.60)
Return(t-1)	-0.424***	-0.345*	0.0352	-0.00337	-0.0637	-0.295**	-0.245*	-0.353***	-0.193*	-0.436***
	(-3.92)	(-2.47)	(0.39)	(-0.03)	(-0.61)	(-2.92)	(-2.40)	(-3.68)	(-2.19)	(-4.67)
Return(t)	0.493***	0.0432	-0.167	-0.0427	0.0699	0.409***	0.257*	0.477***	0.135	0.107
	(5.21)	(0.25)	(-1.71)	(-0.43)	(0.86)	(4.35)	(2.32)	(4.40)	(1.29)	(0.91)
_cons	0.00364	0.00400	0.0147	-0.0317	-0.0331	-0.0165	0.00232	-0.0115	-0.0200	-0.0418
	(0.27)	(0.22)	(0.79)	(-1.40)	(-1.32)	(-0.81)	(0.10)	(-0.44)	(-0.75)	(-1.68)
N	127	127	127	127	127	127	127	127	127	127
adj. R^2	0.487	0.272	0.357	0.433	0.507	0.662	0.619	0.625	0.647	0.738

*Table 24: The table relates monthly returns of 10 size portfolios on OSE to monthly FEARS Allshare. The dependent variables are return one month into the future for each size portfolio, respectively. The independent variable is the monthly FEARS Allshare sentiment index. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models. as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2008. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

4.3.3 Conclusions

We find evidence that monthly occurrence of Google searches isolated predicts the return of large stocks on OSE the following month. The numbers suggest that a one standard deviation increase in FEARS is associated with an approximately 1.2% higher return the following month, all else equal. The effect is primarily present for large stocks, but we also find some evidence that it is present for small stocks.

4.4 Robustness checks

The models we estimate depend on several choices we have made. We have reported some alternative models already. In the following we explain additional tests we have performed to see if our results are sensitive to alternative choices.

An obvious concern is that the results depend on the choice of including 10 terms each period. We can alternatively either choose a different amount of predetermined terms or select a cut-off based on a p-value. In the latter case, the number of terms included will vary, but the exclusion of insignificant terms will ensure that only terms below the predetermined p-value are included. To control if our results are sensitive to this selection, we perform alternative constructions for weekly NEFA and monthly FEARS.

We first test alternative constructions of weekly NEFA. By increasing the amount of terms to 15 we get a slightly lower significance; and by increasing it to 20, the results get insignificant. Alternatively, we can select a p-value as the cut-off criterion and let the amount of terms included vary. When we apply this criterion with a p-value of 0.2 and 0.3, we get very similar results as when predetermining to use 10 terms.

We then proceed to alternative constructions for the monthly FEARS. By increasing the amount of predetermined terms to 15, we obtain similar results, but with slightly lower significance. If we increase the number of terms to 20, we see that the results are insignificant. When testing the alternative selection criteria of a p-value cut-off, we do however obtain strong results. Using a p-value of 0.25, we obtain results that are almost identical to the ones presented.

We thus conclude that it matters how we construct our sentiment indexes, but our findings are not dependent on the specific choice we have made. By testing alternative constructions, we obtain similar results. However, when we start to include insignificant terms, the significance of the indexes disappear, which is not surprising.

As discussed under chapter 3.6, one concern is that the selected methodology can yield results even for random unrelated time series. To test this potential claim we construct an alternative weekly newspaper index using random terms. To ensure this we randomly draw a number between 1 and 397, which corresponds to the number of pages in the dictionary Bokmålsordlista (Wangensteen, 1992). We then select the first term listed on the page and download a time series for the term from the newspaper corpus. Terms that are uncommon in the language found in newspapers are filtered out due to the lack of data. We repeat this random draw until we have time series for 40 random terms. The construction of the index is as explained under subchapter 3.2. The results of this random index is absent. This supports the notion that the methodology only works for time series that have a real historic relation to returns; and that our methodology does not involve any data mining.

4.5 Discussions and comparisons

It is obvious that neither newspapers nor Google searches directly move stock prices - only transactions do. In the previous, we have limited our investigation to the relation between stock prices and the sentiment indexes without too much emphasis on what causes this relationship. For prices to move, the aggregated supply and demand for stocks has to shift. Our two sentiment indexes thereby have some relation to this shift in relative demand.

This relationship can possibly be that what the sentiment index measures causes the shift. This can hypothetically be the case for the newspaper index, where the increasingly negative sentiment in newspapers can cause some investors to decrease their demand for risky stocks.

Alternatively, they do not directly cause a shift in stock prices, yet still have a systematic relation. The sentiment indexes can serve as proxies for beliefs about the future, and there exist some systematic relationship between these beliefs and future stock prices. Another factor can also cause both a shift in the respective sentiment index and a shift in stock prices.

It is also possible that some combination occurs. Some third factor can for instance increase both the occurrence of negative terms in newspapers and move stock prices, and the newspapers might contribute by amplifying the move in stock prices.

Our empirical investigation does not aim at answering these questions, but we are able to relate our findings to existing empirics and theories. In the following, we relate both our findings to the most comparable existing studies, before we discuss how the results should be interpreted.

4.5.1 Comparison to Larsen and Thorsrud (2017)

Our findings on the weekly newspaper index are directly comparable to findings in Larsen and Thorsrud (2017) (from now on LT). LT analyses the relation between newspaper sentiment and OSE, but apply a different method. They limit the basis to the Norwegian newspaper DN (which is also part of our corpus). Instead of measuring the occurrence of terms, they apply a machine learning algorithm to categorize complete texts. They are thus aiming at analysing the text much more as humans perceive it. Our method of counting the occurrence of simple terms has the advantage that it is simple and intuitive. It is thus interesting to see if the results of our alternative method correspond to the findings of LT.

We find that almost only negative terms matter, and simplify our index by excluding positive terms. This differs from LT. who relates also positive themes or topics to returns. A “positive” shift in our index is associated with a decline in the occurrence of negative terms and not an increase in positive terms.

LT analyses daily changes and finds that news topics predict next day return. They further investigate if this effect continues or reverse over time, and find a continuation that peaks after 14 days (Larsen and Thorsrud, 2017, pp. 13). We find that a weekly newspaper index predicts returns in the second week, in the same direction as LT finds. The two significantly different methods thus yield very compatible results.

4.5.2 Comparison to Fisher and Statman (2000)

Several previously discussed papers create sentiment indexes based on Google searches. However, none of these measure changes in a long-term perspective as our monthly index does. DEG does for instance calculate their original FEARS from daily changes in the occurrence of terms. Although the method is the same, the results are not immediately comparable. The most relevant comparison is thus with monthly sentiment indexes created with different methods.

Fisher and Statman (2000) (from now on FS) measures the monthly sentiment of individual investors. This is done with a survey from the American Association of Individual Investors, where they let survey answers from the last week of the month represent the monthly sentiment. They find that a bullish sentiment significantly predicts negative return the next month. This is the same direction as our finding on the monthly Google search sentiment. We find that an increasingly negative sentiment (corresponding to bearish sentiment) predicts positive returns in the next month.

Not only is the direction of our finding compliant with FS. We find that the effect is almost only present for large stocks. This is similar to what FS finds. FS reports the R-squared for the sentiment index prediction on next month return for large stocks to be 0.05. We find that the R-squared of our model on average return next month increases with 0.059 by including FEARS. Our finding on monthly Google search is thus very similar as the findings in FS.

Measuring the sentiment of Google searches is not the same as the direct measurement of individual investors' sentiment that FS applies. However, the sentiment of Google searches can be closely related to the sentiment of individual investors. Several papers argue that internet searches serves as a proxy for the sentiment of unprofessional individual retail investors. Among these are DEG, Joseph, Wintoki and Zhang (2011) and Herve, Zouaoui and Belvaux (2019). The intuition is that only this group of investors expose their biases or sentiment through internet searches. If this assumption is true, then there is a close connection between our finding and that of FS.

The data sample of FS extends from 1987 until 1998. We have a similar length on our sample period, only 20 years later, from 2008 until 2018. The fact that the results correspond so well might indicate that our measure is a good proxy for the sentiment of individual investors, and that the pattern observed in the US is observed 20 years later in Norway.

4.5.3 Discussion of findings on Newspaper Index

Our finding is that the sentiment of newspapers predict negative returns two weeks later.

A substantial tradition focuses on the behaviour of individual retail investors and argues that they move prices due to irrational beliefs and influence from noise (Schleifer and Summers, 1990). This is along the lines that DEG explains their findings.

Our finding on the weekly newspaper index can be interpreted in this tradition. The spirit of newspapers affects retail investors; whenever newspapers use negative economic terms relatively more, the attitude of retail investors is affected in a bearish way, such that they reduce their relative demand for stocks. If this is true, then there exists a causal relation between the occurrence of terms in newspapers and a shift in stock prices.

Some research that hold such interpretations substantiate their claims with findings that the effects are stronger for small stocks that are relatively more held by amateur retail investors (Kumar and Lee, 2006; Barber, Odean and Zhu, 2009). Our findings are however opposite, as the effect of weekly newspaper sentiment are only apparent for large stocks. Our findings do thus not give obvious support to this part of a behavioural interpretation.

This does however not imply that individual retail investors do not play a role for our finding. Recall that a few stocks dominate OSE. This phenomenon might lead to a different role for individual retail investors than what is observed in other countries. We thus cannot exclude the possibility that our finding is due to the behaviour of individual retail investors.

Interpretations in favour of market efficiency and rationality often analyse price deviations not as inefficiencies, but as risk premiums (Bodie, Kane and Marcus, 2018, pp. 356). Whenever some measure of risk increase, investors demand a higher return to hold a risky asset. This can result in falling prices immediately after the risk measure increase, and higher return later on. We observe increasing prices in the same week as the occurrence of negative terms increase, followed by decreasing prices two weeks later. The pattern we observe is thus not obviously compliant with a theory of risk premium. We can however not exclude the possibility that we observe parts of a complex mechanism. Whenever uncertainty gradually increase, this can potentially increase the occurrence of negative terms in newspapers, before it later increases a risk premium.

It would be premature to conclude unambiguously in favour of any of these interpretations. We do however believe that the sentiment of newspapers affect some investors, and that this can lead to a subsequent shift in relative demand for stocks. The perspective of Grossman and Stiglitz (1980), who shows that there are limitations to arbitrage, motivates this view. If a negative newspaper sentiment leads to some less-informed investors selling stocks, then it is quite possible that this is not fully counteracted by other investors.

4.5.4 Discussion of findings on the Google Search Index

As briefly discussed under subchapter 4.5.2, existing literature on internet search sentiment relates it to individual retail investors (Da, Engelberg and Gao, 2015; Joseph, Wintoki and Zhang, 2011; Herve, Zouaoui and Belvaux, 2019). The idea is that internet searches serves as a proxy for the sentiment of less-informed individuals.

This hypothesis fits well with the direction of our findings. We find that whenever the negative sentiment increase, it is associated with positive returns both the current and the next month. As Fisher and Statman puts it, the sentiment of individual investors thus serve as a contrarian indicator for stock returns (Fisher and Statman, 2000). Individual investors are bullish when they should be bearish, and opposite. This supports the idea that individual investors are less informed and underperforms compared to the market, as discussed in chapter 1.

This hypothesis does not imply a causal relationship between Google searching and stock market transactions, only that internet searches serve as a proxy for beliefs or concerns about the future, and that these beliefs are systematically wrong.

We believe our finding fits less with a rational interpretation of market efficiency. We observe that increased searching for negative terms corresponds with increased returns both in the current and subsequent month. We thus do not observe signs of a rightful fear of declining stock prices.

Due to the nature of our empirical investigation, we do not claim to provide the full story on the relation between internet searches and stock returns. Yet the comparison to existing theories suggest that we should follow this simple interpretation. The representative internet user systematically manifest a bearish attitude when they should in fact have a bullish attitude, and we observe this in our data.

5. Conclusions

Our hypothesis was that it is possible to predict future stock returns by measuring current sentiment. Secondarily we studied if there are different implications for small and large stocks. Our two findings can be summarized as follows.

We find evidence that a weekly increase in the occurrence of negative economic terms in newspapers isolated predicts negative returns two weeks later. This effect is only present for large stocks.

We also find evidence that a monthly increase in Google searches of negative economic terms isolated predicts positive returns the following month. This effect is primarily present for large stocks.

Both our findings support the hypothesis that some measure of current sentiment can predict returns on OSE. This is also supported by the fact that both our findings support existing studies that have applied different methods.

Yet the question remains if our two findings are compliant. They stem from the same method of counting the occurrence of negative terms, but use vastly different sources. An increase in the newspaper index predict lower returns after two weeks, whereas an increase in the Google search index predict higher returns the next month. If these were proxies for the same sentiment, they would not be compliant. It is however little reason to expect that newspapers and Google search follow a similar pattern. Under subchapter 3.2.7 we saw that the correlation between monthly FEARS and monthly NEFA is only 0.04. Our findings further support the idea that newspaper and Google searches express unrelated sentiments.

The value of our findings are twofold. It contributes to understand the role of newspapers and Google searches in the financial markets. In addition, it can potentially contribute towards better investment decisions. It is sobering to recall that the R-squared of the sentiment indexes are small. In other words, the indexes explain very little of future returns. Investing based on such an indicator thus provides little reward to risk. However, the success of (short-term) investing depends largely on having marginally better information than others have, and our indexes can potentially be a source to such marginal information.

6. Appendixes

6.1 Appendix 1: Alternative table, weekly NEFA, average return

	(1) OBX Return(t)	(2) OBX Return(t+1)	(3) OBX Return(t+2)	(4) OBX Return(t+2)	(5) OBX Return(t+2)	(6) OBX Return(t+3)	(7) OBX Return(t+4)
NEFA_OBX	0.0109**	0.00213	-0.00897*	-0.00810*	-0.00800*	0.00283	-0.00102
	(2.84)	(0.62)	(-2.30)	(-2.42)	(-2.41)	(0.73)	(-0.26)
VIX	-0.0966***	-0.173***	-0.0359**	-0.0447***	-0.0425***	-0.0134	-0.0141
	(-8.72)	(-14.99)	(-2.65)	(-3.79)	(-3.62)	(-0.97)	(-1.01)
Spread	-0.00220	-0.0362	-0.0774***			0.0425	0.00859
	(-0.12)	(-1.85)	(-3.31)			(1.74)	(0.35)
Brent	0.868	0.800	0.204			0.735	1.152
	(1.18)	(0.93)	(0.19)			(0.64)	(1.01)
Consumer Confidence	0.0138*	0.0174*	0.0284**	0.0173*	0.0171***	0.0303**	0.0244*
	(2.19)	(2.35)	(3.13)	(2.46)	(3.63)	(3.07)	(2.47)
Fund Flow	0.000936**	0.00118***	0.00124**	0.00141***	0.00126**	0.000963*	0.000401
	(3.12)	(3.39)	(2.91)	(3.46)	(3.30)	(2.09)	(0.87)
Corporate Confidence	-0.879	-0.815	-0.224			-0.746	-1.168
	(-1.19)	(-0.95)	(-0.21)			(-0.65)	(-1.02)
Volatility	-0.0151	-0.0207	0.0309	0.0461	-0.00754	0.0705	0.0612
	(-0.48)	(-0.57)	(0.69)	(1.23)	(-0.26)	(1.46)	(1.27)
Yearly Dummy Variable	YES	YES	YES	YES	NO	YES	YES
Return(t-3)	0.0794*	-0.0641	0.0198	0.0175	0.0179	-0.0337	-0.0196
	(2.07)	(-1.84)	(0.49)	(0.50)	(0.52)	(-0.82)	(-0.48)
Return(t-2)	-0.233***	0.0248	-0.103**	-0.0703*	-0.0670	0.00280	-0.0328
	(-5.84)	(0.72)	(-2.59)	(-2.03)	(-1.95)	(0.07)	(-0.80)
Return(t-1)	0.453***	-0.0369	-0.0158	-0.0116	-0.00608	-0.121**	0.000703
	(12.06)	(-1.08)	(-0.40)	(-0.34)	(-0.18)	(-3.00)	(0.02)
Return(t)		0.00380	-0.0440	-0.0344	-0.0182	-0.0340	-0.120**
		(0.11)	(-1.11)	(-1.00)	(-0.53)	(-0.85)	(-2.95)
_cons	-0.0222	-0.0183	-0.00388	0.000345	0.00223	-0.0212	-0.0295
	(-1.06)	(-0.75)	(-0.13)	(0.07)	(1.12)	(-0.65)	(-0.91)
N	662	662	662	872	872	661	660
R^2	0.377	0.357	0.116	0.091	0.071	0.071	0.058
adj. R^2	0.354	0.333	0.083	0.064	0.062	0.036	0.022

*Table 25: The table relates OBX weekly returns to weekly NEFA OBX. The dependent variables are contemporaneous returns (model 1) and future Allshare weekly return in the next 4 weeks (model 2-7). The independent variable is weekly NEFA OBX. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX,. changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility. and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year = 2003 for model 1, 2, 3, 6, 7. Base year =1999 for model 4. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

6.2 Appendix 2: Alternative table, weekly NEFA, cross checks

	(1) Value Weighted Return(t+2)	(2) Value Weighted Return(t+2)	(3) Equally Weighted Return(t+2)	(4) Equally Weighted Return(t+2)
NEFA_VW	-0.00903*		-0.00488*	
	(-2.49)		(-2.16)	
NEFA EW		-0.00909**		-0.00510*
		(-2.62)		(-2.37)
VIX	-0.0293*	-0.0305*	-0.0255**	-0.0261***
	(-2.41)	(-2.52)	(-3.29)	(-3.37)
Spread	-0.0629**	-0.0636**	-0.0470***	-0.0475***
	(-2.99)	(-3.03)	(-3.46)	(-3.49)
Brent	0.319	0.269	0.616	0.589
	(0.34)	(0.28)	(0.96)	(0.92)
Consumer Confidence	0.0227**	0.0231**	0.0195***	0.0197***
	(2.79)	(2.85)	(3.57)	(3.62)
Fund Flow	0.00114**	0.00114**	0.000982***	0.000982***
	(2.96)	(2.96)	(3.81)	(3.80)
Corporate Confidence	-0.337	-0.286	-0.628	-0.601
	(-0.35)	(-0.30)	(-0.98)	(-0.93)
Volatility	0.0484	0.0478	0.0906*	0.0907*
	(1.10)	(1.09)	(2.08)	(2.09)
Return(t-3)	0.0146	0.0125	0.0450	0.0460
	(0.36)	(0.31)	(1.13)	(1.15)
Return(t-2)	-0.106**	-0.109**	-0.0483	-0.0513
	(-2.68)	(-2.75)	(-1.21)	(-1.29)
Return(t-1)	-0.0251	-0.0273	-0.0536	-0.0549
	(-0.64)	(-0.69)	(-1.35)	(-1.39)
Return(t)	-0.0399	-0.0389	0.00398	0.00512
	(-1.00)	(-0.98)	(0.10)	(0.13)
Yearly Dummy Variable	YES	YES	YES	YES
_cons	-0.00268	-0.00118	-0.0112	-0.0104
	(-0.10)	(-0.04)	(-0.62)	(-0.58)
N	662	662	662	662
R^2	0.111	0.112	0.177	0.178
adj. R^2	0.077	0.078	0.146	0.147

*Table 26: The table relates VW weekly returns and EW weekly return to weekly NEFA EW and weekly NEFA VW. The dependent variables are VW return two weeks into the future (model 1 and 2) and EW return two weeks into the future (model 3 and 4). In model 1 and 3 the independent variable is weekly NEFA VW. In model 2 and 4 the independent variable is weekly NEFA EW. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2003. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

6.3 Appendix 3: Alternative table, weekly NEFA, size portfolios

	(1) Portfolio 1 Return(t+2)	(2) Portfolio 2 Return(t+2)	(3) Portfolio 3 Return(t+2)	(4) Portfolio 4 Return(t+2)	(5) Portfolio 5 Return(t+2)	(6) Portfolio 6 Return(t+2)	(7) Portfolio 7 Return(t+2)	(8) Portfolio 8 Return(t+2)	(9) Portfolio 9 Return(t+2)	(10) Portfolio 10 Return(t+2)
NEFA Allshare	0.00210	-0.000826	-0.00206	-0.000472	-0.00984***	-0.00333	-0.00383	-0.00622*	-0.00721	-0.0123**
	(1.03)	(-0.32)	(-0.83)	(-0.17)	(-3.70)	(-1.15)	(-1.40)	(-2.04)	(-1.92)	(-3.25)
VIX	-0.0194**	-0.0339***	-0.0306***	-0.0377***	-0.0314***	-0.0346***	-0.0188	-0.0251*	-0.0361**	-0.0517***
	(-2.87)	(-4.00)	(-3.64)	(-4.20)	(-3.34)	(-3.53)	(-1.95)	(-2.39)	(-2.82)	(-4.04)
Consumer Confidence	0.00281	0.0175***	0.00473	0.0220***	0.0161**	0.0178**	0.0170**	0.0243***	0.0178*	0.0186*
	(0.71)	(3.50)	(0.96)	(4.27)	(2.67)	(2.99)	(2.77)	(3.63)	(2.30)	(2.47)
Fund Flow	0.000782***	0.000819**	0.00104***	0.00134***	0.00106**	0.000675	0.00122***	0.00124**	0.00144**	0.00148***
	(3.31)	(2.80)	(3.52)	(4.37)	(2.97)	(1.94)	(3.38)	(3.21)	(3.12)	(3.32)
Volatility	0.0639	0.0985*	0.138***	0.179***	0.189***	0.120*	0.121*	0.154**	0.226***	0.134*
	(1.96)	(2.44)	(3.36)	(4.25)	(3.80)	(2.48)	(2.44)	(2.91)	(3.54)	(2.13)
Yearly Dummy Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Return(t-3)	0.0699*	0.00559	0.00356	-0.00259	0.0446	-0.0153	-0.0295	-0.0148	0.0179	0.0392
	(2.08)	(0.17)	(0.10)	(-0.08)	(1.34)	(-0.45)	(-0.88)	(-0.44)	(0.52)	(1.13)
Return(t-2)	-0.0193	-0.0679*	0.0255	0.00120	-0.0243	-0.0393	-0.0911**	-0.0728*	-0.0154	-0.0704*
	(-0.58)	(-2.03)	(0.75)	(0.04)	(-0.73)	(-1.15)	(-2.72)	(-2.16)	(-0.45)	(-2.04)
Return(t-1)	0.00421	0.0319	0.0137	0.0300	-0.0145	-0.0440	0.0239	0.0361	-0.0625	0.00687
	(0.12)	(0.95)	(0.41)	(0.90)	(-0.44)	(-1.30)	(0.71)	(1.07)	(-1.83)	(0.20)
Return(t)	-0.0623	-0.0431	-0.0361	-0.0507	0.0461	-0.0245	-0.0148	0.0318	-0.00391	-0.0266
	(-1.84)	(-1.28)	(-1.06)	(-1.51)	(1.39)	(-0.72)	(-0.44)	(0.94)	(-0.11)	(-0.78)
_cons	0.00616*	0.00512	0.00735*	0.0116**	0.0000350	-0.000305	-0.00223	-0.000941	-0.00559	-0.00152
	(2.09)	(1.41)	(1.98)	(2.94)	(0.01)	(-0.07)	(-0.50)	(-0.20)	(-0.99)	(-0.28)
N	872	872	872	872	872	872	872	872	872	872
R^2	0.079	0.111	0.098	0.168	0.108	0.098	0.091	0.101	0.085	0.093
adj. R^2	0.052	0.085	0.072	0.143	0.082	0.071	0.065	0.075	0.058	0.066

*Table 27:The table relates weekly returns of 10 size portfolios on OSE to NEFA Allshare. The dependent variables are return two weeks into the future for each size portfolio, respectively. The independent variable is the weekly NEFA Allshare. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread, changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=1999. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

6.4 Appendix 4: Alternative table, weekly NEFA, size portfolios

	(1) Portfolio 1 Return(t+2)	(2) Portfolio 2 Return(t+2)	(3) Portfolio 3 Return(t+2)	(4) Portfolio 4 Return(t+2)	(5) Portfolio 5 Return(t+2)	(6) Portfolio 6 Return(t+2)	(7) Portfolio 7 Return(t+2)	(8) Portfolio 8 Return(t+2)	(9) Portfolio 9 Return(t+2)	(10) Portfolio 10 Return(t+2)
NEFA EW	0.00297	-0.000446	-0.00305	-0.00440	-0.00881***	-0.00220	-0.00572*	-0.00699*	-0.0110**	-0.0116**
	(1.34)	(-0.18)	(-1.21)	(-1.68)	(-3.31)	(-0.71)	(-1.98)	(-2.25)	(-2.82)	(-2.84)
VIX	-0.0194*	-0.0317***	-0.0294***	-0.0385***	-0.0286**	-0.0302**	-0.00982	-0.0212	-0.0406**	-0.0375**
	(-2.53)	(-3.71)	(-3.40)	(-4.54)	(-2.97)	(-2.74)	(-0.76)	(-1.91)	(-2.95)	(-2.59)
Spread	-0.0148	-0.0285	-0.0240	-0.0181	-0.0453**	-0.0613**	-0.0388	-0.0493*	-0.0899***	-0.101***
	(-1.09)	(-1.91)	(-1.61)	(-1.27)	(-2.66)	(-3.15)	(-1.10)	(-2.50)	(-3.76)	(-4.03)
Brent	1.688***	0.673	1.274*	0.884	0.270	0.950	0.765	-0.386	0.129	-0.402
	(3.43)	(1.28)	(2.34)	(1.83)	(0.42)	(1.31)	(1.10)	(-0.53)	(0.15)	(-0.44)
Consumer Confidence	0.00313	0.0216***	0.00650	0.0299***	0.0158*	0.0214**	0.0241*	0.0309***	0.0240**	0.0330***
	(0.63)	(3.83)	(1.17)	(5.55)	(2.36)	(2.81)	(2.43)	(3.84)	(2.65)	(3.37)
Fund Flow	0.000770**	0.000711**	0.000926***	0.00117***	0.00112***	0.000612	0.00117*	0.00114**	0.00118**	0.00121**
	(3.16)	(2.64)	(3.42)	(4.64)	(3.41)	(1.69)	(2.56)	(3.08)	(2.66)	(2.60)
Corporate Confidence	-1.696***	-0.678	-1.283*	-0.894	-0.274	-0.960	-0.779	0.371	-0.150	0.371
	(-3.44)	(-1.28)	(-2.35)	(-1.84)	(-0.43)	(-1.32)	(-1.12)	(0.51)	(-0.17)	(0.41)
Volatility	0.0460	0.0965*	0.0663	0.0891*	0.124*	0.0749	0.0996	0.0970	0.125	0.103
	(1.15)	(2.20)	(1.47)	(2.16)	(2.25)	(1.24)	(1.00)	(1.58)	(1.70)	(1.30)
Yearly Dummy Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
	(3.43)	(1.40)	(2.06)	(2.03)	(0.33)	(1.31)	(1.23)	(-0.24)	(0.19)	(-0.23)
Return(t-3)	0.0336	-0.00307	0.0202	0.00305	0.0740	-0.0590	-0.0125	-0.00222	0.0550	0.0517
	(0.87)	(-0.08)	(0.51)	(0.08)	(1.88)	(-1.47)	(-0.24)	(-0.06)	(1.38)	(1.29)
Return(t-2)	-0.0281	-0.0527	-0.0114	-0.0275	-0.0775*	-0.0338	-0.0901	-0.103**	-0.0345	-0.0964*
	(-0.72)	(-1.41)	(-0.29)	(-0.71)	(-1.98)	(-0.84)	(-1.84)	(-2.60)	(-0.87)	(-2.42)
Return(t-1)	-0.0200	0.0296	-0.0153	-0.0166	-0.0106	-0.111**	-0.0372	-0.0291	-0.0866*	0.00390
	(-0.51)	(0.78)	(-0.39)	(-0.42)	(-0.27)	(-2.80)	(-0.67)	(-0.72)	(-2.20)	(0.10)
Return(t)	-0.0660	-0.0758	-0.0798*	-0.0933*	0.0320	-0.0473	-0.0207	0.0450	-0.0384	-0.0234
	(-1.67)	(-1.93)	(-2.02)	(-2.40)	(0.82)	(-1.19)	(-0.40)	(1.13)	(-0.97)	(-0.59)
_cons	-0.0536**	-0.0190	-0.0314	-0.0223	-0.00284	-0.0277	-0.0238	0.0161	-0.00106	0.0140
	(-2.99)	(-0.98)	(-1.59)	(-1.26)	(-0.12)	(-1.04)	(-0.95)	(0.60)	(-0.03)	(0.42)
N	662	662	662	662	662	662	663	662	662	662
R^2	0.105	0.143	0.134	0.227	0.140	0.126	0.101	0.131	0.128	0.127
adj. R^2	0.071	0.111	0.101	0.198	0.108	0.093	0.067	0.099	0.095	0.094

*Table 28: The table relates weekly future returns of 10 size portfolios on OSE to weekly NEFA EW. The dependent variables are return two weeks into the future for each size portfolio, respectively. The independent variable is weekly NEFA EW. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2003. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

6.5 Appendix 5: Alternative table, monthly FEARS, size portfolios

	(1) Portfolio 1 Return(t+1)	(2) Portfolio 2 Return(t+1)	(3) Portfolio 3 Return(t+1)	(4) Portfolio 4 Return(t+1)	(5) Portfolio 5 Return(t+1)	(6) Portfolio 6 Return(t+1)	(7) Portfolio 7 Return(t+1)	(8) Portfolio 8 Return(t+1)	(9) Portfolio 9 Return(t+1)	(10) Portfolio 10 Return(t+1)
FEARS EW	0.0195	0.0101	0.0247*	0.0151	0.0232*	0.0445***	0.0420***	0.0256*	0.0347**	0.0345**
	(1.94)	(0.91)	(2.45)	(1.82)	(2.26)	(3.62)	(3.79)	(1.99)	(2.82)	(2.64)
VIX	-0.0198	-0.0591*	-0.0740*	-0.0852***	-0.122***	-0.105***	-0.145***	-0.107***	-0.172***	-0.186***
	(-1.06)	(-2.61)	(-2.40)	(-3.85)	(-4.80)	(-4.24)	(-4.69)	(-4.16)	(-4.50)	(-6.66)
Spread	0.0286	0.0929**	0.0674*	0.0529	0.0808	0.0418	0.109*	0.0488	0.0703	-0.0164
	(0.84)	(2.67)	(2.20)	(1.15)	(1.96)	(0.78)	(2.32)	(0.72)	(1.38)	(-0.33)
Brent	-0.0847	0.0682	0.116	0.167	-0.0000609	-0.00693	-0.104	0.0544	0.0842	0.272*
	(-0.81)	(0.57)	(1.39)	(1.16)	(-0.00)	(-0.06)	(-0.78)	(0.45)	(0.63)	(2.31)
Consumer Confidence	-0.00303	0.0340*	0.0142*	0.00791	0.00316	0.00199	-0.00383	0.00531	-0.000290	0.00246
	(-0.47)	(2.33)	(2.23)	(0.90)	(0.39)	(0.30)	(-0.62)	(0.79)	(-0.06)	(0.26)
Fund Flow	0.000588	0.00103	-0.00176	-0.00107	0.00191	0.000640	0.000449	-0.000307	-0.00132	-0.00222*
	(0.60)	(0.93)	(-1.51)	(-1.20)	(1.76)	(0.69)	(0.48)	(-0.32)	(-1.14)	(-2.01)
Corporate Confidence	-0.0293	0.00346	0.0204	-0.0490	-0.0878**	-0.0170	0.0193	0.0429	-0.00185	-0.0428
	(-0.97)	(0.09)	(0.74)	(-1.41)	(-2.70)	(-0.46)	(0.70)	(1.26)	(-0.06)	(-0.99)
Volatility	-0.491**	-0.428	-0.634***	-0.456	-0.493*	-0.451	-0.520*	-0.163	-0.415	-0.590*
	(-2.91)	(-1.98)	(-3.58)	(-1.88)	(-2.13)	(-1.73)	(-2.35)	(-0.61)	(-1.73)	(-2.03)
Yearly Dummy Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Return(t-3)	-0.238**	-0.112	0.0184	-0.0196	0.0640	-0.247*	-0.0748	-0.00424	-0.110	0.0594
	(-2.95)	(-1.09)	(0.19)	(-0.18)	(0.74)	(-2.55)	(-1.07)	(-0.04)	(-1.56)	(0.73)
Return(t-2)	0.165	-0.0506	-0.0877	0.0204	-0.0748	0.175	0.132	0.0890	0.0672	0.0117
	(1.47)	(-0.36)	(-1.34)	(0.24)	(-0.97)	(1.60)	(1.56)	(1.01)	(0.79)	(0.12)
Return(t-1)	-0.422***	-0.352*	0.0374	-0.00719	-0.0795	-0.326**	-0.251*	-0.354***	-0.203*	-0.372***
	(-3.88)	(-2.45)	(0.41)	(-0.06)	(-0.77)	(-2.92)	(-2.52)	(-3.39)	(-2.25)	(-3.93)
Return(t)	0.497***	0.0448	-0.170	-0.143	0.0732	0.413***	0.229*	0.470***	0.116	0.00691
	(5.09)	(0.25)	(-1.78)	(-1.45)	(0.87)	(4.12)	(2.08)	(4.13)	(1.12)	(0.06)
_cons	0.00646	0.00289	0.0191	-0.0332	-0.0297	-0.0119	0.0108	-0.00896	-0.0132	-0.0463
	(0.42)	(0.15)	(0.99)	(-1.34)	(-1.11)	(-0.49)	(0.43)	(-0.31)	(-0.46)	(-1.42)
N	127	127	127	127	127	127	127	127	127	127
R^2	0.574	0.387	0.476	0.492	0.603	0.696	0.674	0.676	0.706	0.723
adj. R^2	0.483	0.258	0.365	0.384	0.519	0.631	0.605	0.607	0.643	0.664

*Table 29: The table relates monthly future returns of 10 size portfolios on OSE to monthly FEARS EW. The dependent variables are return one month into the future for each size portfolio, respectively. The independent variable is monthly FEARS EW. The set of control variables include lagged returns up to three lags, changes in the CBOE volatility index (VIX), changes in yield spread (Spread), changes in Brent Spot oil price (Brent), changes in the consumer confidence index, changes in the corporate confidence index, net flows into stocks (Fund Flow), historical volatility, and yearly dummy variables. We report heteroscedastic robust standard errors in all models, as well as having adjusted for autocorrelation by using a Cochrane-Orchutt procedure. Base year=2008. *, **, and *** denote significance at the 5%, 1% and 0.1% levels, respectively.*

References

Baker, Malcolm & Wurgler, Jeffrey (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, 2007 (Vol. 21, No. 2), pp. 129-152. DOI: 10.1257/jep.21.2.129

Barber, Brad & Odean, Terrance (2008). All that Glitters: The Effect of Attention on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies,* 2008 (Vol. 21, No. 2), pp. 785-818. DOI: 10.1093/rfs/hhm079

Barber, Brad; Odean, Terrance & Zhu, Ning (2009). Do Retail Trades Move Markets? *Review of Financial Studies, 2009* (Vol. 22, Issue 1), pp. 151-186. DOI: 10.1093/rfs/hhn035

Barber, Brad M. & Odean, Terrance (2013). The Behavior of Individual Investors. *Handbook of the Economics of Finance, 2013* (Vol. 2, Part B), pp. 1533-1570. DOI: 10.1016/B978-0-44-459406-8.00022-6

Ben-Rephael, Azi; Kandel, Schmuel & Wohl, Avi (2012). Measuring Investor Sentiment with Mutual Fund Flows. *Journal of Financial Economics, 2012.* (Vol. 104, Issue 2), pp. 363-382- DOI: 10.1016/j.jfineco.2010.08.018

Black, Fisher (1986). Noise. *The Journal of Finance, 1986* (Vol. 41, No. 3), pp. 528-543. DOI: 10.1111/j.1540-6261.1986.tb04513.x

Bodie, Zvi; Kane, Alex & Marcus, Alan J. (2018). Investments. 11th Edition. New York: McGraw-Hill Education

Da, Zhi; Engelberg, Joseph & Gao, Pengjie (2015). The Sum of All Fears: Investor Sentiment and Asset Prices. *The Review of Financial Studies, 2015* (Vol. 28, Issue 1), pp. 1-32. DOI: 10.1093/rfs/hhu072

Fama, Eugene F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Works. *The Journal of Finance, 1970* (Vol. 25, No. 2), pp. 383-417. DOI: 10.2307/2325486

Fang, Lily & Peress, Joel (2009). Media Coverage and the Cross-section of Stock Returns. *The Journal of Finance, 2009* (Vol. 64, No. 5), pp. 2023-2052. DOI: 10.1111/j.1540-6261.2009.01493.x

Fisher, Kenneth L. & Statman, Meir (2000). Investor Sentiment and Stock Return. *Financial Analysts Journal, 2000* (Vol. 56, Issue 2), pp. 16-23. DOI: 10.2469/faj.v56.n2.2340

French, Kenneth R. & Roll, Richard (1986). Stock Return Variances: The Arrival of Information and the Reaction of Traders. *Journal of Financial Economics, 1986* (Vol. 17, Issue 1), pp. 5-26. DOI: 10.1016/0304-405X(86)90004-8

French, Kenneth R; Schwert, G. William & Stambaugh, Robert F (1987) Expected Stock Returns and Volatility. *Journal of Financial Economics, 1987*. (Vol. 19, Issue 1), pp. 3-29. DOI: 10.1016/0304-405X(87)90026-2

Grossman, Sanford J. & Stiglitz, Joseph E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review, 1980* (Vol. 70, No. 3), pp. 393-408. Retrieved from http://www.jstor.org/stable/1805228

Heuer, Justus; Merkle, Christoph & Weber, Martin (2017). Fooled by Randomness: Investor Perception of Fund Manager Skill. *Review of Finance, 2017* (Vol. 21, No. 2), pp. 605-635. DOI: 10.1093/rof/rfw011

Herve, Fabrice; Zouaoui, Mohamed; Belvaux, Bertrand (2019). Noise traders and smart money: Evidence from online searches. *Economic Modelling, 2019* (Fourthcoming). DOI:10.1016/j.econmod.2019.02.005.

Joseph, Kissan; Wintoki, M. Babajide & Zhang, Zelin (2011). Forecasting Abnormal Stock Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search. *International Journal of Forecasting, 2011* (Vol. 27, Issue 4), pp. 1116-1127. DOI: 10.1016/j.ijforecast.2010.11.001

Jung, Jeeman & Schiller, Robert J. (2002). One Simple Test of Samuelsson’s Dictum for the Stock Market. *NBER Working Paper, 2002* (No. 9348). DOI: 10.3386/w9348

Keynes, John Maynard (1936). The general theory of employment, interest and money. London :Macmillan.

Kumar, Alok & Lee, Charles M.C. (2006). Retail Investor Sentiment and Return Comovements. *The Journal of Finance, 2006* (Vol. LXI, No. 5), pp. 2451-2486. DOI: 10.1111/j.1540-6261.2006.01063.x

Kyle, Albert S. (1985). Continuous Auctions and Insider Trading. *Econometrica 1985* (Vol. 53, No. 6), pp. 1315-1335. DOI:10.2307/1913210

Larsen, Vegard & Thorsrud, Leif Anders (2017). Asset returns, news topics, and media effects. *Working Paper 2017.* (No. 17). Norges Bank Research. Retrieved from https://static.norges-bank.no/contentassets/b19465af1c824d478493f22c4623f731/working_paper_17_17.pdf?v=09/20/2017120952&ft=.pdf

Lee, Charles M. C; Shleifer, Andrei. & Thaler, Richard H. (1991). Investor Sentiment and the Closed-End Fund Puzzle. *The Journal of Finance, 1991* (Vol. 46, No. 1), pp. 75-109. DOI:10.2307/2328690

Lillebø, Arild (2005). Økonomileksikon. 6000 oppslagsord om økonomi og arbeidsliv. Oslo: Damm

Norges Bank (2018). *Det Norske Finansielle Systemet. En Oversikt.* (Rapport). Oslo: Norges Bank.

NOU 2018: 5. (2018). *Kapital i omstillingens tid*. Oslo: Nærings- og fiskeridepartementet.

Næs, Randi; Skjeltorp, Johannes A. & Ødegaard, Bernt Arne (2008). Hvilke faktorer driver kursutviklingen på Oslo Børs? *Norsk Økonomisk Tidsskrift, 2008* (No. 123), pp. 36-81. Downloaded from: http://www1.uis.no/ansatt/odegaard/publications/2008_not_faktorer_oslo_bors/faktorer_oslo_bors_des_2008_fra_journal.pdf

Odean, Terrance (1998). Are Investors Reluctant to Realize Their Losses? *The Journal of Finance, 1998.* (Vol. 53, Issue 5), pp. 1775–1798. DOI: 10.1111/0022-1082.00072

Schleifer, Andrei & Summers, Lawrence H. (1990). The Noise Trader Approach to Finance. *The Journal of Economic Perspectives, 1990* (Vol. 4. No. 2), pp. 19-33. DOI: 10.1257/jep.4.2.19

Tetlock, Paul C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance, 2007* (Vol. 62, No. 3), pp. 1139-1168. DOI: 10.2139/ssrn.685145

Tetlock, Paul C.; Saar-Tsechansy, Mantay &Macskassy, Sofus (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance, 2008* (Vol. 63. No. 3), pp. 1437-1467. DOI: 10.1111/j.1540-6261.2008.01362.x

Wangensteen, Boye (ed.) (1992). Bokmålsordlista. Rettskrivingsordliste for skole, hjem og kontor. Oslo, Norway: Universitetsforlaget

Warther, Vincent A (1995). Aggregate Mutual Fund Flows and Security Returns. *Journal of Financial Economics* (Vol. 39, Issue 2-3), pp. 209-235. DOI: 10.1016/0304-405X(95)00827-2

Ødegaard, Bernt Arne (2019). Empirics of the Oslo Stock Exchange.
Basic, descriptive, results 1980-2018. Downloaded from:
http://finance.bi.no/~bernt/wps/empirics_ose_basics/empirics_ose_basics_2019_02.pdf