Financial and Economic Attitudes Revealed by Search

THE ROLE OF INVESTOR SENTIMENT IN DENMARK

Master's Thesis

11th of May 2022

COPENHAGEN BUSINESS SCHOOL

Applied Economics and Finance

Eskild Lind Benestad (141808)

Fredrik Pettersson (141807)

Supervisor: Oguzhan Cepni

92 pages

206 361 characters

ABSTRACT

Behavioural finance argues that market inefficiency can be explained by irrational behaviour amongst investors. In contrast to the neoclassical theories, which emphasize that markets are indeed efficient, the behavioural models introduce the concept of “noise” in the information available for investors. The noise trader approach suggests that animal spirits amongst investors will be incorporated into asset prices through trading. A phenomenon that has later been named investor sentiment. To measure the impact of investor sentiment in the Danish market, we construct a measure of sentiment following the Financial and Economic Attitudes Revealed by Search (FEARS) methodology, introduced by Da, Engelberg, and Gao (2015). The FEARS methodology applies the online search behaviour for economic words by households as a direct measure of investor sentiment. The application of the Google Trends service allows us to extract historical search volume index data (SVI), and thus observe the household search behaviour in Denmark. Contrary to Da et. al (2015), our FEARS index proxies for investor optimism. The index is dynamic as it consists of the 20 terms exhibiting most positive correlation with stock returns, updated every six months.

Furthermore, we test the relationship between our index and average and cross-sectional stock market returns, governmental and corporate bonds, as well as stock market volatility. Our findings suggest that increased optimism amongst investors relates to an increase in contemporaneous stock market returns, followed by a complete reversal the next day and a subsequent positive spike on lead-day two. Moreover, our research of the sentiment impact on the cross-section of stocks suggests that sentiment has a larger predictive power for value and large stocks than for growth and small stocks. The bond analysis returns ambiguous results. Here, the impact of FEARS on 10-year maturity government bonds and Investment Grade corporate bonds are economically negligible. However, for 5-year maturity government bonds, the FEARS exhibit a delayed positive relationship. Finally, the FEARS is modelled with realized volatility as an ARFIMA model, suggesting a positive relationship after the Financial Crisis of 2007-2008.

Our findings are thus in line with theory of temporary mispricing in stocks caused by investor irrationality. Moreover, the results suggest that household Google search behaviour serves as a good measure of investor sentiment in Denmark. In addition, most of our findings support the results of Da et. al (2015), proving the transferability of their methods. Thus, we can accept our main hypothesis of an apparent relationship between investor sentiment and stock market movements in Denmark.

We would like to thank our academic supervisor, Oguzhan Cepni, for valuable guidance, motivation, and continuous support throughout our research.

TABLE OF CONTENTS

1 INTRODUCTION

Since John Maynard Keynes introduced the notion of animal spirits in 1936 (Keynes, 1936), economists have been determined to explain what the key drivers of stock market movements are. Keynes' theory was built on the assumption that information is dependent on investor attention for it to be processed and incorporated into asset prices through trading. Attention, however, is a limited resource (Kahneman, 1973). Behavioural models of finance suggest that the limited attention of investors, causes prices to deviate from their fundamental value, and thus cause market inefficiency. Financial scholars have later attempted to disprove the theories suggesting that asset prices are driven by irrational investors subject to limited attention.

Rejecting the possibility of inefficient markets, the efficient market hypothesis (EMH) postulates that the value of an asset should reflect all available information regarding its value (Fama, 1970). According to the EMH, a security should be priced equal to its fundamental value, that is, the present value of expected future cash flows based on the currently available information (Marsh and Merton, 1986). Observing financial markets through this lens of economics, noise trading is suggested to be unimportant in the pricing of assets. That is, the EMH suggest that trading by rational arbitrageurs will eventually drive prices to their fundamental value (Friedman, 1953; Fama, 1965a), annulling the impact of irrational investors. The EMH thus excludes behavioural biases, such as loss aversion, pessimism, and herding, which leaves a hole in explaining certain frictions in the market.

The absence of investor attention invites the presence of "noise" in available information. Noise trader theory divides the market participants into two categories: noise traders and information traders (Black, 1986; Shefrin and Statman, 1994). A noise trader is an investor who does not trade based on information, but rather "noise" which is perceived as information (Black, 1986). Thus, supporting the presence of animal spirits in the decisions of investors. It is suggested that investors trade on noise as if it were information, implying that noise will be incorporated into the prices of stocks (Black, 1986).

The presence of animal spirits was later defined as sentiment (De Long, Shleifer, Summers & Waldman, 1990). Neoclassical theories would suggest that the role of investor sentiment is economically negligible, as these theories ignore the existence of biased or emotional investors, arguing that irrational behaviour is offset by rational arbitrageurs. However, behavioural finance theories later introduced the concept of limits to arbitrage. The limits to arbitrage are perhaps the most important channel for this sentiment to impact asset prices (Pontiff, 1996, Shleifer and Vishny, 1997). This theory suggests that arbitrage capital moves slowly to take advantage of the irrational beliefs of sentiment investors. Delong et.al. (1990) portray how these limits to arbitrage along with investor rationality can be used to predict movements in the equity markets, formalizing the role of investor sentiment.

Investor sentiment has later been a large focus area among financial scholars. Introducing one of the most established measurements of investor sentiment in 2007, Baker and Wurgler are still viewed as leading researchers within the field. Baker and Wurgler (2007) modify the EMH through the lens of behavioural finance, by making two assumptions. Assumption one, as described by Delong et.al. (1990), explains how investors are subject to sentiment. That is, a belief about future cash flows and investment risks that are not justified by the facts at hand (Baker and Wurgler 2007). The second assumption, emphasized by Shleifer and Vishny (1997), describes how betting against sentimental investors is both risky and costly. Thus, rational investors are not as aggressive in forcing prices to their fundamental values as the neoclassical model has suggested.

Supporting the behavioural models of measuring investor sentiment, Tetlock (2007) also divides market participants into two categories; noise traders who hold random beliefs about the future and rational arbitrageurs who hold Bayesian beliefs (Tetlock, 2007). Tetlock contributed to the investor sentiment literature mainly through the application of a linguistic sentiment measure. This sentiment proxy was able to measure sentiment directly, rather than the previously established indirect measures. His findings suggested that sentiment can indeed predict that short-horizon returns will be reversed in the long run, opposing the information theory, which predicts that they will persist indefinitely (Shannon, 1948).

It is thus safe to believe that markets are not perfectly efficient, and that noise plays a large part in the decision-making of irrational investors. To catch the essence of the animal spirit, one must thus construct a valid measure of investor sentiment, this, however, is difficult (Baker and Wurlger, 2006). Various measurements of sentiment have been applied in financial analysis since the dawn of the term in 1990.

Inspired by the likes of Tetlock (2007), Da, Engelberg, and Gao (2015) built a sentiment index based on the Google search behaviour of U.S households, namely the Financial and Economic Attitudes Revealed by Search (FEARS) index. Their FEARS index consisted of negatively loaded economic terms, as to proxy for investor pessimism. Da et.al (2015) suggested that their FEARS index was able to predict short-term return reversals, temporary increases in volatility and mutual fund flows of equity funds into bond funds. In addition, their findings suggest that this effect is stronger for stocks that are attractive to retail investors and thus harder to arbitrage, supporting the results of Baker and Wurgler (2007). The methods of Da et al. (2015) have later been adopted by several scholars and tested towards a variety of asset classes and other markets, proving its flexibility (Aroul, Sabherwal and Saydometov 2020; Gao, Ren and Zhang 2020; Ito, Masuda, Naito and Takeda, 2020).

The prediction power of a term-based measure of sentiment is thus viable (Tetlock, 2007; Tetlock, Saar-Tsechansy and Macskassy, 2008; Gao, Da, Engelberg, and Gao, 2015; Aroul, Sabherwal and Saydometov 2020; Gao, Ren and Zhang 2020;). The idea behind such a sentiment proxy is that individual investors will search for negatively loaded words in bad times and positively loaded words in good times. Take for instance the word "recession", when the market outlook is bad, one might observe a general fear among investors as the search occurrence of the term could increase. Thus, behavioural biases, such as pessimism, herding, and loss aversion, can have a significant impact on financial markets during a crisis.

Da et.al. (2015) construct their index based on Google search data extracted from the Google Trends tool, provided by Google Inc. Google is the most popular search engine in the world, providing a platform for information available to all. Google Trends is a service provided by Google for analytical purposes. This service enables the collection of time-series data from different geographical areas and timeframes.

Nasdaq OMX Copenhagen, formerly known as Copenhagen Stock Exchange (CSE) hosts several internationally recognized companies. To proxy for the Danish stock market returns, the FEARS index will be constructed based on the Copenhagen General Index (Copenhagen GI), consisting of all 127 public companies available on Nasdaq OMX Copenhagen. The Copenhagen GI is dominated by a few large companies, mainly within the healthcare sector, followed by the Industrial and Financial sectors.

In an effort to measure the presence of animal spirits in the Danish stock market, we aim to construct a robust measure of investor sentiment. The sentiment index to be constructed will have a basis on Google searches by Danish households, following the approach of Da et al. (2015). As per the authors of this paper's knowledge, there is no paper applying the FEARS methodology to test a possible relationship between investor sentiment and Danish stock returns. Moreover, the flexibility of the FEARS methodology toward other financial measures and asset classes will also be tested. The research to be conducted in this paper thus contributes to an area where prior evidence is scarce, testing the application of noise trader models in Denmark.

1.1 HYPOTHESIS AND MOTIVATION

Our hypothesis is based on the noise trader approach to finance. Primarily we want to test whether current measures of investor sentiment can predict future stock returns. As we are interested in the Danish equity market, we have limited our research to Nasdaq OMX Copenhagen. Additionally, we want to test whether the sentiment index has predictive power in other areas of interest, that is, within the cross-section of stock returns, towards realized volatility and within the bond markets. We thus seek to either accept or rejct the following hypothesis:

Does investor sentiment have predictive power on Danish equity returns?

Along with the additional sub-questions:

1) Does the FEARS relationship vary within the cross-section of returns?
2) Does FEARS have predictive power on Danish Bond Market returns?
3) Is there a significant relationship between FEARS and Realized Volatility?

Research on the relationship between investor sentiment and financial markets provides interesting links between financial theory and real market frictions. That is, it allows us to indirectly observe the impact of noise traders and noise trader risk on asset prices and volatility. Although the previous literature on investor sentiment provides sufficient evidence of apparent sentiment effects, most studies regarding investor sentiment are concentrated in the U.S or other large world markets. Evidence of sentiment effects in other markets does not provide proof of apparent effects in Denmark. That is, like most capital markets, the Nasdaq OMX Copenhagen bears its characteristics, making it unique. In addition, the frequency of Google searches also differs vastly among countries. The application of sentiment-based measures of noise trader impact in Denmark thus yields an interesting study of previously untapped waters.

Moreover, as this paper seeks to replicate the methodology of Da et.al. (2015), further results could also provide evidence of whether this methodology is transferable to the Danish stock market. Proving the FEARS methodology in Denmark could open a whole new area of study for the Scandinavian countries, where there is a general paucity of literature on investor sentiment.

1.2 DELIMITATIONS AND SCOPE OF THESIS

The scope of this paper is to test some, not all, aspects of the sentiment impact on Danish financial markets. As previously mentioned, we hope to contribute to prior literature as we move the methods proven to work in the United States, to Denmark. The potential outcome of this paper is twofold. First, the findings could contribute towards understanding more of the role that google searches have in the pricing of stocks, potentially observing a relationship between sentiment and stock market movements. Secondly, there are also practical implications of a successful stock return prediction. That is, if the FEARS methodology proves the ability to catch the impact of investor sentiment on Nasdaq OMX Copenhagen stock returns, the results of the thesis could be applied to aid in constructing a sentiment-based trading strategy.

An important delimitation of the paper is that the paper does not try to provide a definitive explanation of why there is a relationship between the FEARS index and equity returns. The main purpose of the paper is to test whether there is a significant sentiment effect from Danish retail investors trading towards stock market movements. However, discussions and interpretations of results will be provided, giving the findings theoretical reasoning, and comparing them with prior literature.

1.3 STRUCTURE OF PAPER

This paper is divided into 11 sections, where the first section (1) is the introduction. This is then followed by a section (2) which explains the scientific approach to be used throughout the research. The third section (3) provides some background information on investor sentiment, Copenhagen Stock Exchange and Google. Section four (4) presents the main literature and the fundamental theories revised in advance of any analysis. The data gathering and methodology are in the next section (5) elaborated. Furthermore, section six (6) outlines the empirical strategy to be conducted in the analysis. The seventh (7) will present the results following the analysis prepared in the prior sections. In section eight (8), several robustness tests will be utilized to check the robustness of our chosen methodology. In section nine (9), these results will be discussed and compared to previous research on the topic. Section ten (10) seeks to address some concerns and weaknesses of the research. Finally, in the last section (11), the hypotheses of the paper will be concluded, and the paper finalized.

2 SCIENTIFIC METHOD

The scientific method applied in this paper has a deductive nature (Holm, 2018). All hypotheses have a basis in previous studies. These hypotheses are further tested on the Danish market, for us to either confirm or reject them. As a result, critical rationalism and a falsification method are applied in the testing process (Holm, 2018). To build a foundation for the hypotheses and collection of data, we initially conducted a systematic literature review. The next step saw us gathering available data, to test the deduced hypotheses. This was followed by a thorough analysis which enabled us to either reject or accept the hypotheses.

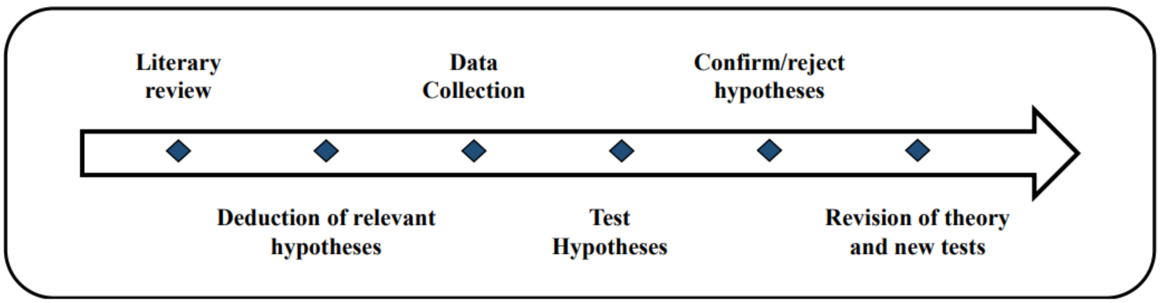

Figure 1: Illustration of the Scientific approach applied in this paper. Source: Own creation based on Bryman and Bell (2011).

Bryman and Bell (2011) define three important criteria in evaluating research: replicability, reliability, and validity. Reliability relates to the accuracy of data collection and analytical methods. All data applied must be adjusted correctly, i.e., there should be no econometric concerns when the data is finally applied in the analysis. As all data applied in this paper is quantitative, it is vulnerable to biases by nature. Having this knowledge, the selection and treatment of data throughout the gathering and analysis have been executed thoroughly, to increase the reliability of our results. Replicability concerns whether it is possible to replicate the findings of the paper. The applied method must be transparent and explained thoroughly for others to be able to test the reliability of the paper. All methods applied in this paper are deduced from well-documented methods of reliable researchers and theory to ensure replicability. Finally, the validity of the paper explains the relationship between the hypotheses and the data. The methodology applied should reflect the scope of the research. External validity refers to whether the results of the paper can be replicated beyond the specific research in focus. As the scope of this paper is to test the replicability of previously applied methods in a new market, we indirectly test the external validity of previous work. Internal validity explains the extent to which there is a credible causal relationship between two variables. This criterium for research is highly in focus throughout the paper and reflected in the proposed hypotheses. Bryman and Bell's three criteria in evaluating research have thus been essential in writing this paper.

3 BACKGROUND

This paper centres around the sentiment effects on the Nasdaq OMX Copenhagen. However, for the further conclusions to be intuitive, we view it as necessary to provide some background information on the practical notion of sentiment, the Danish equity market and Google. Thus, the first section seeks to explain what investor sentiment is, along with its implications[1]. Followed by background information on the market characteristics of Nasdaq OMX Copenhagen, Google Inc., and the Google Trends service.

3.1 SENTIMENT

The term *Investor Sentiment* relates to the general feeling among investors in the market, whether it is optimism (bullish) or pessimism (bearish). An optimistic (pessimistic) view of the real economy and stock markets is the belief that the future economic situation will be favourable (unfavourable) and future stock returns increasing (decreasing).

Investor sentiment may be an important explanatory factor in why we have booming periods in the stock markets. In such periods, one can observe that market participants are willing to take more risk, for instance by leveraging trades more heavily, betting on more risky stocks, or the willingness to pay higher prices for stocks. Moreover, investor sentiment can be observed in daily frequency data, whereby stocks move, without any relevant news emerging. Thus, investor sentiment may be an explanatory factor in retail investor trading activity, a phenomenon denominated as noise-trading, which we will revisit in the literary review.

Furthermore, investor sentiment has several plausible definitions. Baker and Wurgler (2006) propose two definitions in their paper on *Investor Sentiment and the Cross-Section of Stock Returns.* The first definition they provide is that sentiment relates to speculation in the stock markets, whereby increased sentiment may increase the demand for speculative investments (Baker and Wurgler, 2006). Secondly, they argue that sentiment might simply be optimism or pessimism in the stock markets.

Whether a good measurement of sentiment exists should also be questioned. The measurements proposed by prior theoretical and empirical studies are broad. One modern approach to measure sentiment directly is the method later applied in this paper, that is, sentiment as measured by household Google Search behaviour.

[1] This section does not concern the prior literature on sentiment, which will be elaborated in section 3.1

3.2 MARKET

The Nasdaq OMX Copenhagen, formerly known as Copenhagen Stock Exchange ("København Fondsbørs", in Danish), first opened in 1808 as Copenhagen Securities Exchange ("Mæglerforordningen"), as a non-profit organization (Nasdaq OMX Nordic, n.d.a). In 1988, the exchange transitioned to an electronic trading and information system (translated from Sørensen, 2011), and derivates were available for trading, such as options and futures. In 1996, Copenhagen Stock Exchange changed to a joint-stock company, where the ownership was dispersed between the current members, and issuers of stocks and bonds. The Copenhagen Stock Exchange was sold to the Swedish and Finish exchange company OMX AB in 2005 and was then a part of the OMX Nordic Exchange. In 2007, the American giant Nasdaq acquired OMX AB, and the still-standing Nasdaq OMX Group, in which most of the Nordic countries are included, was born (translated from Sørensen, 2011).

The Nasdaq OMX Group provides several indices for Copenhagen. Most famously, the Nasdaq Copenhagen 25 (OMX C25 for short). The OMX C25 index consists of the 25 most traded blue-chip stocks, and hosts several companies of international recognition, for example, Carlsberg, A.P. Møller Mærsk, Novo Nordisk, Ørsted (formerly Dong Energy) and Vestas Wind Systems (Nasdaq OMX Nordic, n.d.b).

Moreover, Nasdaq provides an index consisting of all 127 public companies, named the Copenhagen General Index (Copenhagen GI). This is an uncapped all-share index, whereby the index weights are determined by dividing each index security's market capitalization by the aggregate market capitalization of all Index securities (Nasdaq OMX Nordic, 2022b). The All-Share Index is broken down into industry segments in Figure 2. As can be observed from the figure (2), the healthcare industry is dominant in the Danish stock market. Nasdaq OMX Copenhagen hosts the very large Novo Nordisk, which is one of the main reasons for this industry's market share.

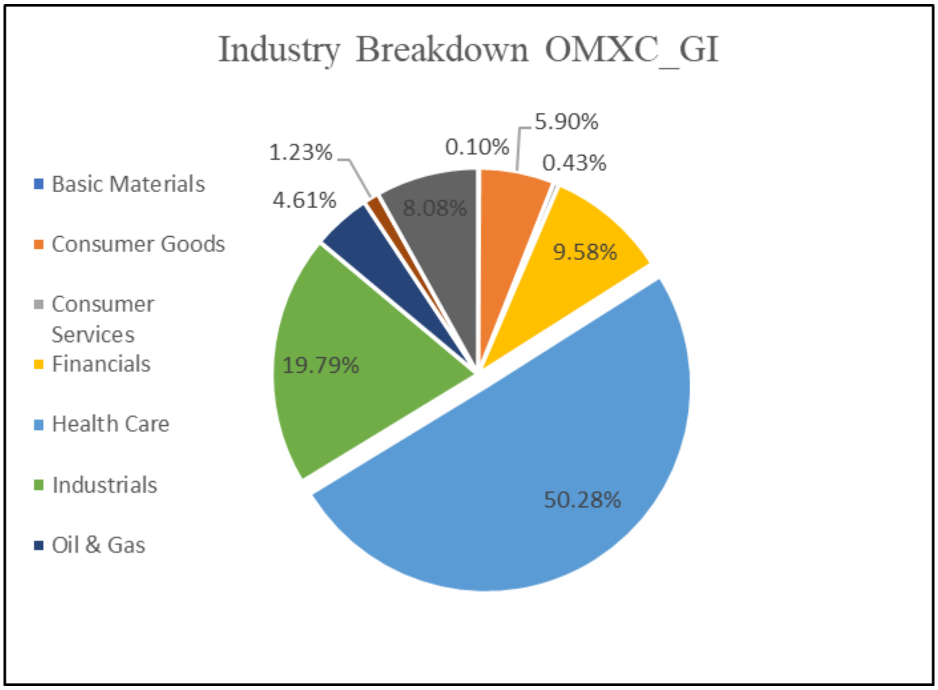

Figure 2: Industry breakdown of the Copenhagen General Index 27/04/2022 (Copenhagen Allshare Index). Own creation, based on Nasdaq Index information. Source: Nasdaq OMX Nordic, 2022a.

3.3 GOOGLE

3.3.1 About Google

Google Inc. was founded by Larry Page and Sergey Brin in 1998, to create a search engine that could manoeuvre the growing data available on the World Wide Web. Specifically, a search engine that used links to determine the importance of individual pages on the World Wide Web (Google, n.d.a).

The company had its Initial public offering on August 19th, 2004, valued at $23 billion. In 2014, the company performed a two-to-one stock split. In 2015, the company undertook a restructuring whereby Google, instead of being the parent company, reorganized under a new parent company Alphabet. (Yahoo Finance, 2020). Today, Alphabet, and its main company Google, is known as an enormous media and technology conglomerate, providing hundreds of different services, from the original Google Search Engine, YouTube, and services for Android phones.

3.3.2 Market Share

Google has been able to whisk away early competitors, most famously Yahoo, and later Bing. The Search Engine has consistently held a huge market share over its contenders. In 2021, the Google Search Engine averaged a 92% market share worldwide (Statcounter GlobalStats, 2022a). Its closest competitor was Microsoft's Bing with a 3.1% market share. The same comparison for Europe reveals a market share averaging 93% for Google, vs. 3% for Bing (Statcounter GlobalStats, 2022b). An even more dominant market share holds for Denmark. In 2009, Google's market share in Denmark was above 97% (Statcounter GlobalStats, 2022c). In 2021, Google still holds a market share of above 95% (Statcounter GlobalStats, 2022d).

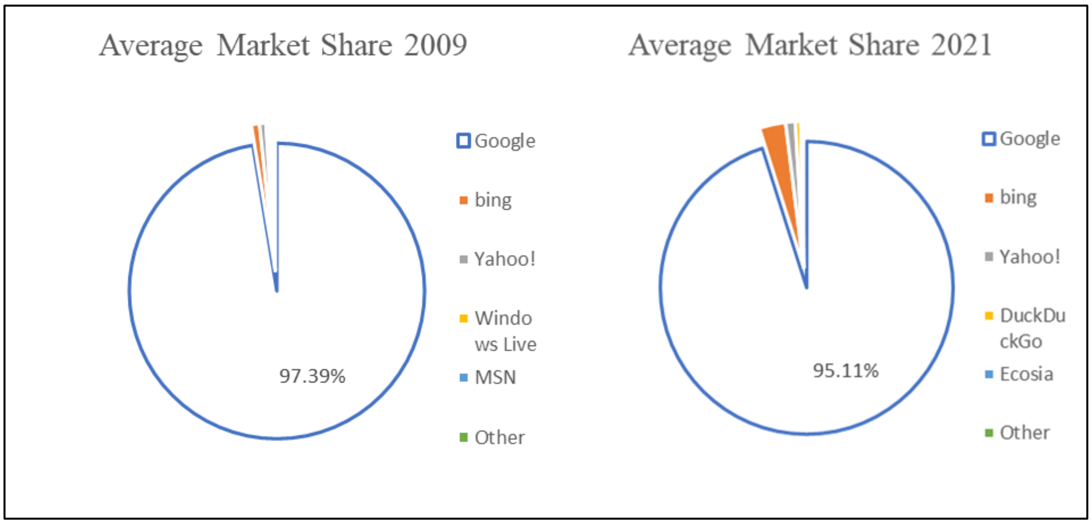

Figure 3: The figure illustrates the average market share of the Google Search Engine in Denmark for 2009 (left-hand side) and 2021 (right-hand side). The figure is an own creation based on the data from Statcounter GlobalStats, (2022c and 2022d)

3.3.3 Google Trends

Google Trends is a service provided by Google, which allows users to observe analytics of relative search term popularity in the Google Search Engine. Specifically, it allows for visualization and analysis of the relative popularity of a google search in the search engine down to city-level geography. The service is anonymized, categorized, and aggregated (Google, n.d.b). Providing two different samples to the public, Google Trends is a frequently applied service in analytics. The first sample consists of real-time data, covering the seven prior days. The second sample provides non-real-time data which goes back to 2004. The non-real-time data, which is used for historical analysis, is normalized, and presented relative to other searches, as opposed to presenting the absolute occurrences of searches for a search term. Specifically, each data point is divided by the total Google searches, within a selected time frame and specific region, thus presenting relative popularity (Google, n.d.b). The resulting numbers are then scaled on a range of 0 to 100 based on a topic's portion to all searches on all topics (Google, n.d.b).

4 REVISION OF LITERATURE AND THEORY

The paradigm of modern finance is the neoclassical discipline. Opposed by the behavioural financial models with inefficient markets, the neoclassical lens of finance emphasizes market efficiency. The noise trader approach to finance was later introduced to explain how prices deviate from their fundamentals, through investor irrationality. This approach gave rise to what is now called investor sentiment.

Haugen (1999) as cited in Ramiah, Xu, and Moosa (2015), describe the evolution of finance by identifying three schools of thought: old finance, modern finance and new finance. In old finance, market participants focused primarily on financial statement analysis, and the nature of financial claims, for example, dividends. In modern finance, participants focus on the very popular asset pricing and valuation models. Due to several strict assumptions associated with these neoclassical models, they have been challenged in later years. The "new" discipline of finance, especially the behavioural finance models, focuses on better understanding the role of market participants under relaxed rationality assumptions. In the following, we introduce the neoclassical financial theories, the behavioural financial theories, and lastly, the investor sentiment literature and its role in the new finance era.

As this paper aims to provide evidence of market anomalies that cause market inefficiency, a thorough revision of literature and theory was conducted before any analysis or data gathering, with the most important presented in the following.

4.1 NEOCLASSICAL THEORY

Statman (1999) as cited in Ramiah et al. (2015) summarizes the main contributions of the neoclassical financial theories, as being "the arbitrage principles of Miller and Modigliani, the portfolio principles of Markowitz, the capital asset pricing model of Sharpe, Lintner, and Black, and the option-pricing theory of Black, Scholes, and Merton. Ramiah (2015) argues that there are four lessons to be learned from the neoclassical finance paradigm. Firstly, the market value of an asset should be determined by its fundamental value, that is, the asset-specific financials and related information about that asset. Secondly, following the efficient market theory as later elaborated upon in section 4.1.2, the market should react quickly to information. Thirdly, prices should follow a random walk process because the arrival of news follows a random walk process. Lastly, no market participants can earn abnormal returns consistent with risk over time. We present two of the most important contributions of neoclassical finance literature in the following.

4.1.1 The Capital Asset Pricing Model (CAPM)

The CAPM is an asset pricing equilibrium model, concerning the expected returns of assets. It was developed by Sharpe, Litner, and Mossin. The CAPM is based on the idea that not all risks should affect asset prices (Perold. A., 2004). Specifically, a risk that can be diversified away in a portfolio is not actually a risk (Perold. 2004). Sharpe (1964) developed the equilibrium model and introduced the capital market line (CML), which further developed to the security market line (SML) and individual asset equilibrium pricing. He assumes that any individual views the outcome of any investment in probabilistic terms and is willing to act on two parameters – expected value and standard deviation. He further assumes that individuals maximize their total utility function:

$$U = f(E_w, \sigma_w) \tag{1}$$

“Where E_w is expected future wealth and σ_w is the predicted standard deviation of possible divergence of actual future wealth from E_w“. (Sharpe, 1964, p. 428). Sharpe now introduces a set of new assumptions. Firstly, individuals prefer a higher wealth to lower wealth; secondly, individuals exhibit risk-aversion. The model postulates that investors will then choose between a set of investment opportunities, the one which maximizes their utility. These investment opportunities involve long or short positions in assets, and lending or borrowing at joint rates for all market participants. Now, one last assumption which allows the equilibrium, is the homogeneity assumption. “All investors are assumed to agree on the prospectus of various investments – the expected values, standard deviation, and correlation coefficients” (Sharpe, 1964, p. 433). The implication of his theorem is an equilibrium capital market line, where all investors maximize their utility, based on choosing an optimal combination of risky assets and a risk-free asset. The slope of the CML is the additional expected return per unit of risk, which we now call the Sharpe ratio.

Mossin (1966) states that this capital market line is no way to measure whether an investor behaves rationally, but rather “it is a way of summarizing the result of rational behaviour, and nothing more” (p. 779). Mossin postulates that without the rationality assumption the whole foundation of the model is destroyed. That is, both the concept of an equilibrium and the capital market line no longer hold (Mossin, 1966). Therefore, it can be argued that the concept of the CAPM and the market line relies heavily on its rationality assumption.

4.1.2 Efficient Market Hypothesis (EMH)

The efficient market hypothesis states that all available information should be reflected in financial asset prices, following the assumption that market participants are rational processors of information (Ramiah et. al. 2015). Fama (1965b) states that "in an efficient market, competition among the many intelligent participants leads to a situation where, at any point in time, actual prices of individual securities already reflect the effect of information" (p. 76), for both historic, and expected future events. The premise is that all available historical information is already traded upon, and any arrival of information will be reflected instantaneously in actual prices. The EMH explains how stock prices should only move in response to new information, through three different types of market efficiency.

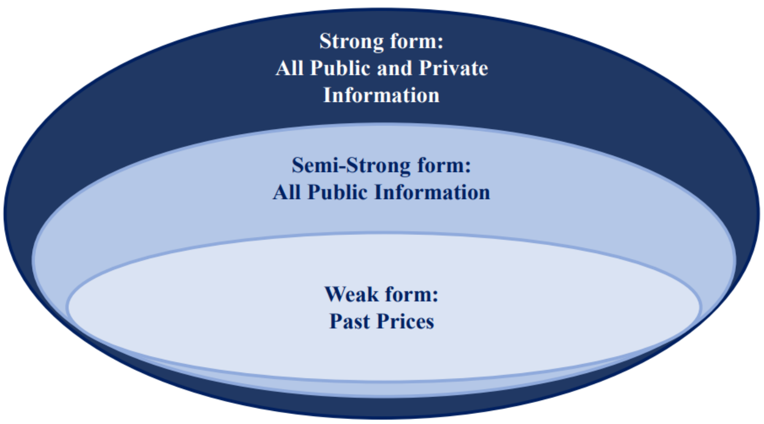

Figure 4: Summary of the three stages of market efficiency. Source: Own creation

The three forms are described as weak, semi-strong, and strong efficiency. In the weak form of market efficiency, the hypothesis assumes that all information about past stock prices is reflected in the stock price of today so that no technical analysis can aid investors in generating abnormal returns. However, under this form of market efficiency, it is possible to generate abnormal returns if you obtain other information than historical prices. Secondly, the semi-strong form assumes that all public information is reflected in the prices instantaneously. That is, public information is described as all information that can affect stock prices, such as macro figures, firm-specific risks, and the financials reported by the firm. The semi-strong form implies that it is not possible to apply either technical or fundamental analysis to outperform the market. Lastly is the strong form of market efficiency. Under this assumption, all information, both public and private, is reflected in the current stock price. The inclusion of private information (insider information) implicates that it is not possible to attain abnormal returns in the stock markets, regardless of your access to information (legal or illegal). Thus, the three stages differ in their assumptions of how much information is reflected in today's stock prices and agree in that information is already traded upon by market participants. The implication for a trader is that one cannot be sure whether the information one is trading upon is already reflected in the stock price.

Moreover, the EMH, like the CAPM, relies heavily on the assumption that a large number of rational market participants can close any arbitrary mispricing gaps instantaneously, due to competition. Stock prices will under this assumption only move with the arrival of news in the market. Because news follows a random pattern, a so-called "random walk", as will stock prices, and thus no pattern should reoccur and be predictable. By extension, no pattern should be able to profit upon.

4.2 BEHAVIOURAL FINANCE

Although economist John Maynard Keynes made the notion of animal spirits known to the world in 1936, financial scholars have since then developed equilibrium models in opposition to this. Keynes argued that the economy is subject to irrationality, as well as rationality. According to Keynes, a large part of economic activity is governed by *animal spirits*, and these animal spirits influence the stock markets (Keynes, 1936). Keynes was thus a forerunner of the discipline that would later become to be known as behavioural finance.

Behavioural finance can be defined as the application of psychology to explain anomalies in the market (Ramiah et al. 2015). Contributions to the behavioural finance literature can be found as early as in 1912, when Selden suggested that stock price movements depended heavily on the mental attitude of market participants (Selden, 1912). The most significant results, however, arose approximately 65 years later when Tversky and Kahneman, developed the heuristics of availability, representativeness, anchoring and framing and most importantly the prospect theory which explains that losses loom larger than gains (Tversky and Kahneman, 1979). The behavioural finance economists are in opposition to the neoclassical finance discipline, suggesting that a model with efficient markets bears little to reality (Shiller, 1981).

"People are rational in neoclassical finance; they are normal in behavioural finance" -Statman (1999)

According to Ramiah et al. (2015) the assumptions of behavioural finance models are similar to those of traditional models, but contain five significant differences: (i) investors are influenced by non-statistical characteristics as well as mean-variance configurations (ii) investors perceive trends where there is no obvious pattern (iii) heterogeneity of investors allow for imperfect information (iv) different investors have different investment opportunities (v) markets are not necessarily in equilibrium, and arbitrage opportunities may be subject to market sentiment.

4.3 NOISE TRADING

Contradictory to neoclassical models of finance, behavioural finance thus allows for market inefficiencies on the basis that investors are subject to common human errors that arise from heuristics and biases. In this view, there are two categories of traders in the market: information traders and noise traders (Black ,1986; Shefrin and Statman, 1994). Defined by Fischer Black in 1986, a noise trader is an investor who does not trade based on information, but rather "noise" which is perceived as information. Black suggested that noise induces irrational investors to make bad trades, as they believe they are trading on information. These trades will then be of benefit to the holders of actual information.

"Noise makes financial markets possible, but also make them imperfect"-Black (1986)

Furthermore, Black (1986) suggests that noise is reflected in the prices of stocks, as well as the true information that stock prices reflect. Thus, as the amount of noise increases, it becomes more profitable for rational traders with information to trade, as the stock prices increase with the noise. However, as Black points out, because noise traders must believe they are trading on information, there will always be an ambiguity of who is trading on information and who is trading on noise.

In later years, noise trading has been identified as a major source of market anomalies, introducing what has come to be known as "noise trader risk" (De Long et al, 1990). Assuming that arbitrageurs are likely risk-averse and have reasonability short horizons, Delong et al. (1990) describe noise trader risk as "the risk that noise traders' beliefs will not revert to their mean for a long time and might in the meantime become even more extreme" (Delong et al., 1990). This risk must be borne by arbitragers exhibiting the abovementioned assumptions, limiting the effectiveness of the arbitrage. Thus, rational investors buying stocks when noise traders depress prices, will drive prices towards fundamentals, but not all the way, as proposed by the Neoclassical models.

4.4 INVESTOR SENTIMENT

The animal spirits introduced by Keynes were later defined as sentiment by De Long et al. (1990). In addition to the introduction of noise trader risk, Delong et.al. (1990) formalized the role of investor sentiment in financial markets. Suggesting that if uninformed noise traders base their trading decisions on their sentiment and risk-averse arbitrageurs encounter limits to arbitrage, changes in sentiment will induce noise trading, excess volatility, and more mispricing.

One year later Lee, Shleifer and Thaler (1991) attempted to capture the behaviour of noise traders by studying an area where many small investors do business, namely the Close-End funds (CEF, hereafter). More specifically, the discount on CEFs served as a proxy for investor sentiment. This proxy was built on economic theory which implies that discounts on various funds move together and that these discounts fluctuate together with prices of securities affected by the same investor sentiment. Their findings suggested that the CEF discounts narrowed when small stocks (often held by individual investors) did well, supporting the assumption that small investors trading on noise alters systematic risk.

Revolutionizing the methods to measure sentiment, Baker and Wurgler (2007) constructed a sentiment index different from previous researchers. Their measure averaged six widely accepted proxies of investor sentiment, namely trading volume, dividend premium, closed-end fund discount, the number and first-day returns of IPOs, and the share of equity in new issues. Their results highly contributed to the sentiment literature, in addition to providing support for the previous findings of Lee et.al (1991). Baker and Wurgler highlight two previously suggested assumptions of sentiment to explain why investor sentiment affects stock prices. The first assumption, originally described by Delong et al. (1990), explains how investors are subject to sentiment. Assumption two, originally emphasized by Shleifer and Vishny (1997), describes how betting against sentimental investors is both risky and costly. Thus, rational investors are not as aggressive in forcing prices to their fundamental values as the neoclassical model has suggested.

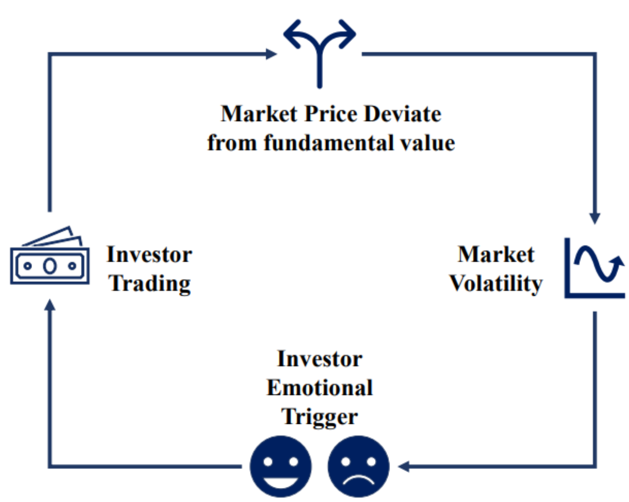

Figure 5: Summary of how investor sentiment causes noise trading to affect stock prices. Source: own creation.

The studies of Baker and Wurgler reveal two key results for investor sentiment. The first suggests that the movement of stock returns can be explained by investor sentiment. When sentiment is low in one period, returns in the subsequent period are high, and vice versa. This finding indicates a relative overpricing in periods with observed high sentiment. Their second finding relates the impact of sentiment to the characteristic of a specific stock, stating that the impact of sentiment is larger on stocks that are characterized by being either small, young, highly volatile, unprofitable, non-dividend paying or distressed. These characterizations of stocks consistently yield higher returns following periods of low sentiment.

4.4.1 SENTIMENT ON THE CROSS-SECTION OF RETURNS

As a direct measurement of investor sentiment is hard to identify (Baker and Wurgler, 2007), various researchers apply some consumer confidence index (CCI) as a proxy for sentiment. These indices can measure sentiment directly, contrary to the indirect measure of sentiment applied by Lee, Shleifer and Thaler (1991).

Fisher and Statman (2000) apply a CCI, measured as the monthly sentiment from surveys on three different groups of market participants: newsletter writers, wall street strategists and individual investors. Their findings suggested a significant correlation between the newsletter writers and the individual investors, but no correlation between the strategist and the two other groups. Additionally, they found that positive sentiment amongst the retail investors had negative predictive power on Standard & Poor 500 in the following month. Moreover, their findings suggest that the correlation with individual investor sentiment changes is higher for large stocks than for small stocks.

Like Fisher and Statman (2000), Brown & Cliff (2005) applied a direct survey measure of investor sentiment. Their findings suggest a weak short-horizon relationship between sentiment and returns, but a strong correlation between sentiment and long-horizon returns. Moreover, they emphasize the concept of institutional sentiment in addition to individual sentiment, where they reject the conventional belief that sentiment should only affect small stocks and is driven primarily by individual investors. When analysing the cross-section of stocks, they find a stronger effect for growth stocks than for value stocks.

However, the findings of Brown and Cliff (2005) have not met a consensus. Kumar and Lee (2006) used the trading record of individual investors to investigate the effect of retail trading on the cross-section of stock returns. They suggested that retail investors tend to overweight value stocks relative to growth stocks in their portfolios. In addition, they report a contemporaneous correlation between the buy-sell imbalance of retail investors and the returns on high book-to-market [2]stocks, however, they find no correlation on low book-to-market stocks, suggesting that the impact of individual investor sentiment is stronger for value than for growth stocks.

Lemmon and Portniaguina (2006) applied a CCI as a proxy for investor optimism aimed to identify the forecasting power of investor sentiment in predicting the cross-section of stock market returns. They find that the sentiment component of confidence has forecasting power over time, on stocks predominantly held by individuals. That is stocks with low institutional ownership. They further provide evidence that sentiment effects are significant for value stocks, and not for growth stocks. This supports the findings of Kumar and Lee (2006), who show that retail investors tend to overweight value stocks relative to growth stocks, and that the trading by such investors correlates with movement in returns for value stocks.

Schmeling (2009) also applies a CCI index, analysing the presence of noise traders and the predictability of CCIs on value, growth, and small and large stocks. This was done for a sample of 18 industrialized countries. His findings concerning the value and growth stocks, were consistent (Kumar and Lee, 2006; Lemmon and Portniaguina, 2006) as he observed a significant effect for both value and growth stocks, but the effect being larger for value stocks. Moreover, he suggested that there is a significant effect of the CCI on the returns of small but not for large stocks, which is consistent with most prior findings (Lee et.al., 1991; Brown and Cliff, 2005; Kumar and Lee, 2006; Lemmon and Portniaguina, 2006; Baker and Wurgler, 2006, 2007). Moreover, Schmeling also investigated why the effect of noise traders is stronger in some countries than others and find that sentiment has a larger effect on returns in countries that are culturally more prone to herd-like investment behaviour, following the hypothesis of Chui, Titman and Wei (2008), and for countries that lack market integrity and efficient regulatory institutions.

4.4.2 TERM-BASED MEASURE OF SENTIMENT

Introducing a new measure of investor sentiment, Tetlock (2007) was the first financial scholar to use news media to predict movements in the stock market. More specifically, he utilized a popular column in the Wall Street Journal to create a proxy for sentiment. Distinguishing between the effect of positive and negative sentiment, he classified all words in the chosen column to either have positive or negative sentiment. The findings of his research indicate that high pessimistic sentiment predicts negative daily returns on the Dow Jones Industrial Average, which is then reversed to its fundamental value. Secondly, he

[2] The book value of a stock divided by the market value. High B/M stocks are often labelled value stocks.

finds that when sentiment is high, this predicts a high trading volume, regardless of whether it is negative or positive.

Adopting the linguistic term-based proxy for investor sentiment applied by Tetlock (2007), Da, Engelberg & Gao (2015) construct an index consisting of the daily occurrence of google searches in the United States. Their index is named the FEARS (Financial and Economic Attitudes Revealed by Search) index. This sentiment index consists of the most negative correlated sentiment-loaded economic terms, with respect to S&P500 returns. Da et al. (2015) apply the Google Trends service to extract data on U.S household search patterns. Their findings suggest that the FEARS predict short-term return reversals, temporary increases in volatility (elaborated on in section 4.4.5) and mutual fund flows of equity funds. In addition, they found that this effect is stronger for stocks that are attractive to retail investors and thus harder to arbitrage, supporting prior results (Lee et.al., 1991; Kumar and Lee, 2006; Baker and Wurgler, 2007).

The methods of Da et al. (2015) have later proved to be transferable to other geographical areas, and applicable to other markets. Aroul, Sabherwal & Saydometov (2020) construct a FEAR-index that captures negative sentiment and analyses its relationship to US housing returns. They find that the housing market has an inverse relationship to changes in the FEARS-index, and that the strength of this relationship is affected by different market characteristics. Additionally, they find that cold markets under financial distress are most affected by sentiment. Gao, Ren & Zhang (2020) advance the methods of Da et al. (2015) by constructing a composite search-based index of investor sentiment for 38 countries. Thus, contrary to Da et al. (2015), they use both positive and negative sentiment in their indices. The scope of their paper is to validate whether the search-based approach of measuring sentiment can be used as a contrarian predictor for country-level market returns. Additionally, they investigate whether sentiment travels across the global markets and measure the return prediction of global sentiment. Their findings suggest that sentiment yields an inverse prediction of country-level market returns. The FEARS methodology has also been applied in predicting exchange rates. Ito et al. (2020) construct two different FEARS indices, one for the USA and one for China, and apply them in their attempt to forecast exchange rates. More specifically, they attempt to forecast the USD/JPY exchange rates including FEARS in different time series and structural models. Their results suggest that the inclusion of the FEARS index improves all of their proposed forecasting models, indirectly proving the FEARS methodology's vast areas of use.

4.4.3 INVESTOR SENTIMENT AND GOVERNMENT BONDS

Research on a relationship between investor sentiment and bond returns has also been a focus area within the sentiment literature, however, the evidence is scarcer than for equities. As the above sections established, methods for examining the sentiment impact on equities are apparent in the cross-section (Lee et.al., 1991; Kumar and Lee, 2006; Baker and Wurgler, 2007) and over time (Lemmon and Portniaguina 2006). The application of sentiment measures is however not limited to that of equity markets. Although bonds are primarily held by institutional investors (Nayak, 2010), there are well-documented information spill overs between the equity and bond markets (Kwan 1996; Downing, Underwood and Xing, 2009; Bethke, Gehde-Trapp, and Kempf, 2017), as well as a high correlation documented between the two (Collin-Dufresne, Goldstein & Martin, 2001). This insinuates that the observed relationship in the equity markets, could be transferable to the bond markets.

Baker and Wurgler (2012) study the relationship between sentiment and the co-movement between government bonds and bond-like stocks. They find that when the investor sentiment index is high and subsequent returns on bond-like stocks are expected to outperform speculative stocks, bond returns are also expected to be positive. Laborda and Olmo (2013) incorporate a sentiment factor constructed from the set of variables introduced by Baker and Wurgler (2007), to proxy for investor sentiment. The scope of their research is to investigate the relationship between market sentiment and the existence of a risk premium in bond markets, i.e., the predictive power of investor sentiment in describing bond risk premia at different maturities. Their findings suggest that market sentiment helps explain systematic deviations in bond prices, that are related to waves of market sentiment. In addition, they find that high investor sentiment favours the excess returns of longer maturity bonds rather than the short-term bonds.

Also having basis in Baker and Wurgler's (2006) composite sentiment index, Fang, Yu and Huang (2018) study the influence of investor sentiment on the time-varying long-term correlation between the U.S stock market and bond markets. Their findings suggest that investor sentiment has a significant positive influence on the long-term correlation between stocks and bonds. Moreover, they show that the shock following a crisis, decreases the average correlation between the two markets, while there is no significant decrease in the sentiment effect.

4.4.4 INVESTOR SENTIMENT AND CORPORATE BONDS

Nayak (2010) explores the impact of sentiment measures on corporate bond yield spreads using a large sample of bond transactions over an 11-year period. He finds evidence that investor sentiment is a significant factor in the determination of corporate bond yield spreads. Moreover, Nayak suggests that when preceding sentiment is high, subsequent spreads are high. In addition, he shows that high-yield bonds demonstrate greater susceptibility to mispricing due to sentiment, thus being more prone than low-yield bonds. Thus, implying that sentiment appears as a prominent factor in bond prices.

Huang, Rossi and Wang (2015) examine whether the effect of sentiment in equity markets spills over to corporate bonds, and to what extent this spill over is related to the high correlation between different assets after the crisis of 2007-2008. They link the equity market sentiment to corporate bond valuations through bond returns, credit spreads and credit spread changes. They find a negative relationship between sentiment and corporate bond spreads, especially after the Financial Crisis. In addition, they find that sentiment helps explain the integration between the equity market and the corporate bond market. Lastly, they find that sentiment is more likely to affect bonds where fundamental risk and liquidity frictions are higher, suggesting that sentiment affects the behaviours of bond investors directly.

4.4.5 INVESTOR SENTIMENT AND VOLATILITY

The establishment of "noise trader risk" in behavioural financial models, also gave rise to research on the relationship between investor sentiment and market volatility. The relationship between noise trading activity and volatility was first conceptualized by Black (1986). Black (1986) argued that the volatility of asset prices would exceed the volatility of asset value (asset fundamental value), in the presence of noise traders.

Brown (1999) tested the implication of De long et al. (1990), that if noise traders affect stock prices, the additional noise created in the market is sentiment, and the risk the noise traders cause is volatility. That is, sentiment should be correlated with volatility. Brown constructed a direct investor sentiment measure, using a survey collected for the American Association of Individual Investors, whereby the participants were asked whether they were bullish, bearish, or neutral in their market expectations for the coming six months. Brown then regressed his investor sentiment index against Closed-End-Funds (CEFs) close-to-close volatility as a proxy for S&P 500 Volatility. His findings reported strong evidence of a relationship between investor sentiment and volatility, suggesting that increased sentiment will lead to an increase in volatility.

Verma and Verma (2007) further contributed to the literature as they decomposed the term investor sentiment, which was previously viewed as fully irrational, into rational and irrational factors. By constructing an investor sentiment measure for both individual and institutional investors, they found that both sentiments are driven by rational and irrational factors. For example, both groups of investor sentiment were suggested to be significantly related to dividend yield, SML and HML factors[3]. However, Verma and Verma (2017) concluded in favour of the expectation of De Long et al. (1990), that for the irrational

[3] From the Fama French five-factor pricing model

sentiment they constructed, investor error (comparable to noise trading) is a significant determinant of stock volatility.

The research by Da et al. (2015) elaborated in section (4.4.2), also contained an extensive study of sentiment on various stock market volatility measures. They tested the FEARS index against daily realized volatility on the SPDR S&P 500 ETF, the VIX, and VIX futures. These volatility measures are controlled for the long-memory component, which often can be found in volatility data. Realized volatility is computed with high-frequency, 15-minute intraday period transactions, following the estimation procedure of Andersen, Bollerslev, Diebold and Ebens (2001). Furthermore, Da et al. (2015) stated that to their knowledge, no prior work had examined the relationship between sentiment measures and market-level volatility at high frequency. Da. et al. (2015) found a contemporaneous relationship, concluding that there is a transitory impact of the FEARS index on stock market volatility, for SPY[4] and aggregate market volatility for the VIX and the VIX futures.

4.5 ONLINE SEARCH BEHAVIOUR

The application of online search data in financial analysis is increasingly popular. Choi and Varian (2012) suggest that the first paper applying web search data in forecasting statistics was by Ettredge, Gerdes, and Karuga (2005), who examined the US unemployment rate. Ettredge et al. (2005) found a positive, significant association between the job-search search term variables, and unemployment data. Soon followed Cooper, Mallon, Leadbetter, Pollack and Peipins (2005) who analyzed cancer-related topics using internet search volume. Cooper et al. (2005) hypothesized that members of the U.S population would use search engines to locate online health information. In their paper, they study the Yahoo! search activity related to the 23 most common types of cancer in the United States. This data was not used to forecast any medical data of incidents but to study the coverage and attention to cancer-related topics. They found that media coverage plays a powerful role in prompting the online searches for cancer information, and that internet search activity can be used as a tool for surveillance of health information-seeking behaviour. In the medical scene, specifically the field of epidemiology, the use of search data as a predictor expanded in the following years. Polgreen, Chen, Pennock, and Nelson (2008) studied the relationship between searches for influenza and actual influenza occurrence, using search data from the Yahoo! search engine. At this point, they state that "an estimated 113 million people in the United States use the internet to find health-related information" (Polgreen et al. 2008, p. 1443).

[4] SPDR S&P 5000 ETF (NYSEARCA:SPY)

By regressing on influenza occurrences, they find that their model predicts an increase in cultures positive for influenza 1-3 weeks in advance of when they occurred. The study implied that search-term surveillance may provide an additional tool for disease surveillance (Polgreen et al. 2008). Choi and Varian (2012) illustrate in their paper how the inclusion of Google Trends variables in simple seasonal AR models outperform models that exclude these variables by 5-20% in forecasting of automobile sales, unemployment claim, travel destination planning and consumer confidence, illustrating a practical use of the information content search term data may contain.

Academics started developing the methodology for the usage of search engine analytics to gain a deeper understanding of stock market dynamics. The earliest was Antweiler and Frank (2004) who studied the claim that Internet stock message boards could move stock markets. Using data from Yahoo! Finance and Raging Bull[5], they constructed bullishness measurements by using linguistics methods. Their data was tested against the 45 companies in the Dow Jones Industrial Composite Index. They found that stock messages have information content useful in predicting market volatility, and that a high volume of messages has a significant, yet small effect on stock market returns.

5 DATA AND METHODOLOGY

To ensure high reliability in our findings, the process of collecting data has been executed thoroughly, through sources found to be valid. The specific data sources applied will in the following section be specified, and any special considerations elaborated. Moreover, all analyses conducted in this paper will have a basis in daily observations to achieve comparable and precise results. If not else stated, all analyses are conducted from 2004, as this is when the earliest observations are available on Google trends data.

Furthermore, most analyses to be conducted in this paper will have basis in the OLS regression, an explanation of the OLS framework is therefore included in Appendix (1) for convenience. Moreover, the following section will see us define and prepare all variables to be applied in the analysis. In addition, the main part of this section will focus on every step of constructing our FEARS index, as we view this as the most important part in the methodology. Nonetheless, to maintain an uncomplicated structure, the construction of the FEARS will follow the preparation of the search term data (SVI[6], hereafter).

[5] RagingBull.com is a website for financial literacy and day trading.
[6] Search Volume Index

Additionally, some restrictions and delimitations must be considered working with our data to ensure the highest possible validity:

- ❖ Google calculates SVI from a random subset of actual historical data. This causes SVIs to be slightly different when downloading terms several times. However, in support of Da et al. (2015), we believe the impact of these minor changes is economically neglectable.
- ❖ As the empirical analysis of this paper is concentrated in Denmark with the country's corresponding financial markets, data can be limited in comparison to the U.S. Although we seek to replicate the methods of Da et al. (2015), some parts of their research are thus perceived to be out of reach, as the limited data would call for extensive research.

5.1 STOCK RETURN DATA

5.1.1 Average returns

The main analysis of this paper is carried out on the aggregate market index for Denmark, namely the OMX Copenhagen General Index ($DKallshare$, hereafter). $Dkallshare$ will thus serve as a proxy for the Nasdaq OMX Copenhagen average return. This is a value-weighted index consisting of all stocks (127, including A and B shares) listed on Nasdaq Copenhagen. Historical data on $DKallshare$ is extracted directly from Nasdaq to ensure high validity (Nasdaq OMX Nordic, n.d.). The analysis towards average returns will be conducted in the sample period spanning from 2004-2022 (01.01.04 - 31.12.21), as Google Trends data is limited to 2004. Figure 6 depict the raw data of $DKallshare$ for the chosen sample period.

Figure 6: The figure presents a graphical output of daily Copenhagen General Index observations over the period 2004-2022. Source: Nasdaq OMX Nordic, n.d..

From the figure 6, we observe a significant drop following the Financial Crisis in 2007-2008. However, the $DKallshare$ data exhibit an overall increasing trend during the sample period. Such data characteristics will be accounted for in section (5.1.3) prior to any analysis.

5.1.2 The Cross-Section of stock returns

The sentiment analysis in the Danish equity markets extends to the cross-section of returns, mainly focusing on value and growth stocks. Price indices for growth and value stocks, individually, will be applied in the further analysis. These indices are provided by Standard and Poor (S&P, hereafter). To achieve greater depth in the later presented results, the analysis will focus on value and growth indices within three different market segments: Large-Mid Cap[7] stocks, Small-Cap stocks and finally the Broad Market stocks. All data on value and growth stocks is extracted from the S&P Global database *Capital IQ*, within the chosen sample period of 2004-2022.

To proxy for value and growth stocks within the broad market segment in Denmark, we apply the S&P Denmark BMI Value Index (*BMIvalue*, hereafter) and S&P Denmark BMI Growth Index (*BMIgrowth*, hereafter), respectively. The S&P BMI[8] indices are subindexes of the S&P Global BMI that includes all companies domiciled in Denmark (S&P Global, n.d.).

Secondly, to proxy for value and growth stocks within the Large and Mid-cap segment, we apply the S&P Denmark LargeMidCap Value Index (*LMvalue,* hereafter) and S&P Denmark LargeMidCap Growth Index (*LMgrowth*, hereafter), respectively. The S&P LM[9] indexes are subindexes of the S&P Denmark BMI with the scope to track the top 85% of market cap stocks in Denmark.

And finally, proxying for stocks within the small-cap segment for Denmark, we apply the S&P Denmark SmallCap Value Index (*SMLvalue,* hereafter) and S&P Denmark SmallCap Growth Index (*SMLgrowth*, hereafter), respectively. The S&P SML[10] indexes are subindexes of the S&P Denmark BMI with the scope to track the bottom 15% of market cap stocks in Denmark.

[7] Short for market capitalization
[8] Short for Broad Market Index
[9] Short for Large and Mid cap
[10] Short for Small-cap

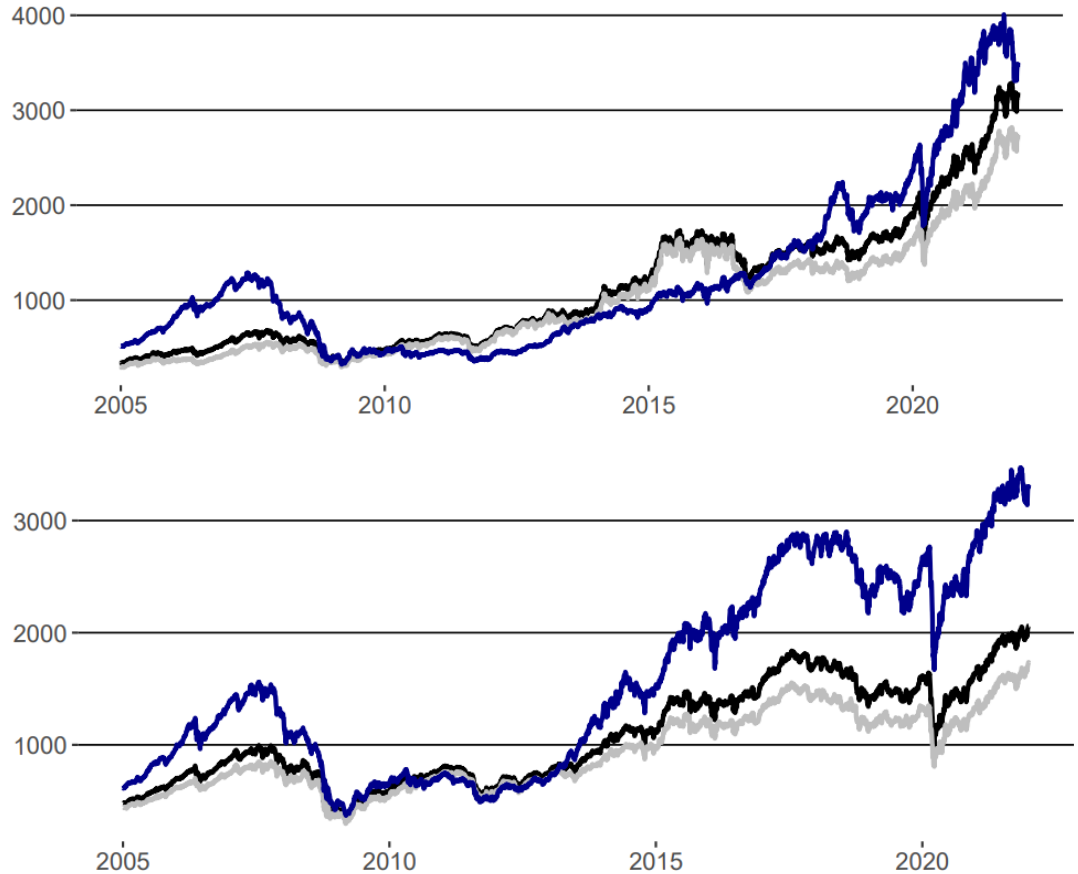

Figure 7: Figure 7 depicts the raw returns of the six indices $BMIvalue$, $BMIgrowth$, $LMvalue$, $LMgrowth$, $SMLvalue$ and $SMLgrowth$ for the sample period from 2004-2022. The blue, grey and black lines represent SML, LM and BMI indices respectively. The top (bottom) panel shows the graphs for the growth (value) indices. Source: S&P Global, n.d.

From figure 7 we observe that the value and growth indices exhibit similar patterns as the $DKallshare$. In addition, the graph suggests that the BMI and LM indices are highly correlated. To provide further insight into whether this is the case, Table 1 provides a correlation matrix for all stock return variables to be analysed in this paper. From table 1, we observe that the $LMvalue$ and $BMIvalue$ are perfectly correlated, which also hold for $LMgrowth$ and $BMIgrowth$. This high correlation might indicate that most stocks in the BMI indices are large and mid-cap stocks.

	Allshare	BMIvalue	BMIgrowth	SMLvalue	SMLgrowth	LMvalue	LMgrowth
Allshare	1	0.92	0.99	0.93	0.95	0.91	0.98
BMIvalue	0.92	1	0.90	0.98	0.82	1.00	0.90
BMIgrowth	0.99	0.90	1	0.90	0.92	0.89	1.00
SMLvalue	0.93	0.98	0.90	1	0.87	0.97	0.88
SMLgrowth	0.95	0.82	0.92	0.87	1	0.79	0.88
LMvalue	0.91	1.00	0.89	0.97	0.79	1	0.90
LMgrowth	0.98	0.90	1.00	0.88	0.88	0.90	1

Table 1: This table shows the correlation matrix for all return variables to be applied in the paper, namely the Allshare, BMIvalue, BMIgrowth, SMLvalue, SMLgrowth, LMvalue and LMgrowth. The correlations are calculated based on historical stock prices for each individual index, spanning from 2004-2022.

Table 2 depicts the summary statistics for all return variables which will be applied in further analysis, prior to any adjustments. The table highlights that $LMvalue$ exhibits the lowest maximal value, and that $SMLgrowth$ has the highest maximal value.

Index	**Mean**	**Std.Dev**	**Min**	**0.25**	**Median**	**0.75**	**Max**
DKallshare	891.4	557.1406	257.3	466.5	630.2	1288.0	2631.9
BMIvalue	1079.6	440.4615	326.2	713.5	937.9	1455.5	2063.3
BMIgrowth	1122.9	692.9381	325.7	554.8	854.7	1582.4	3287.8
LMvalue	928.5	344.6019	297.7	652.9	810.6	1217.8	1736.3
LMgrowth	991.0	594.9823	289.6	485.1	788.7	1395.5	2820.0
SMLvalue	1567.1	860.5191	369.4	710.3	1386.1	2379.1	3472.2
SMLgrowth	1218.0	870.8616	321.9	524.0	938.2	1540.6	4007.8

Table 2: Shows the descriptive statistics for all stock variables to be applied in this paper. The numbers are based on historical stock prices for each individual index, spanning from 2004-2022.

5.1.3 Preparing the stock returns

The previous section highlighted the presence of some concerns in the stock returns variables, which will be accounted for in the following section. As the return variables were heavily correlated, and following similar patterns, all variables will be prepared equally in the following. For simplicity, the following adjustments will be made on a general return variable, $r_{j,t}$.

We start by log transforming the timeseries so that changes in the transformed series represent percentage changes in the original series of $r_{j,t}$ (Stock and Watson, 2014, p. 315). In addition, the observed trend in Figure 7 could also cause potential problems for our further analysis. The most reliable way to handle a trend in the data series is to transform the series. That is, if the series exhibit a trend, the first difference of the series does not (Stock and Watson, 2014, p. 572). To account for the observed trend in the returns data, we initially take the first differences of the log $r_{j,t}$:

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

Figure 8 now depicts data on $\Delta r_{j,t}$, which is the detrended product of $DKallshare_{j,t}$, to serve as an example. From figure 8 we initially observe what might look like yearly seasonality, especially during the Financial Crisis of 2007-2008 and the covid-19 pandemic during in 2020. For $\Delta r_{j,t}$ to be applicable in further analysis, we regress $\Delta r_{j,t}$ on monthly and yearly dummies and extract the residuals removing any seasonal effect.

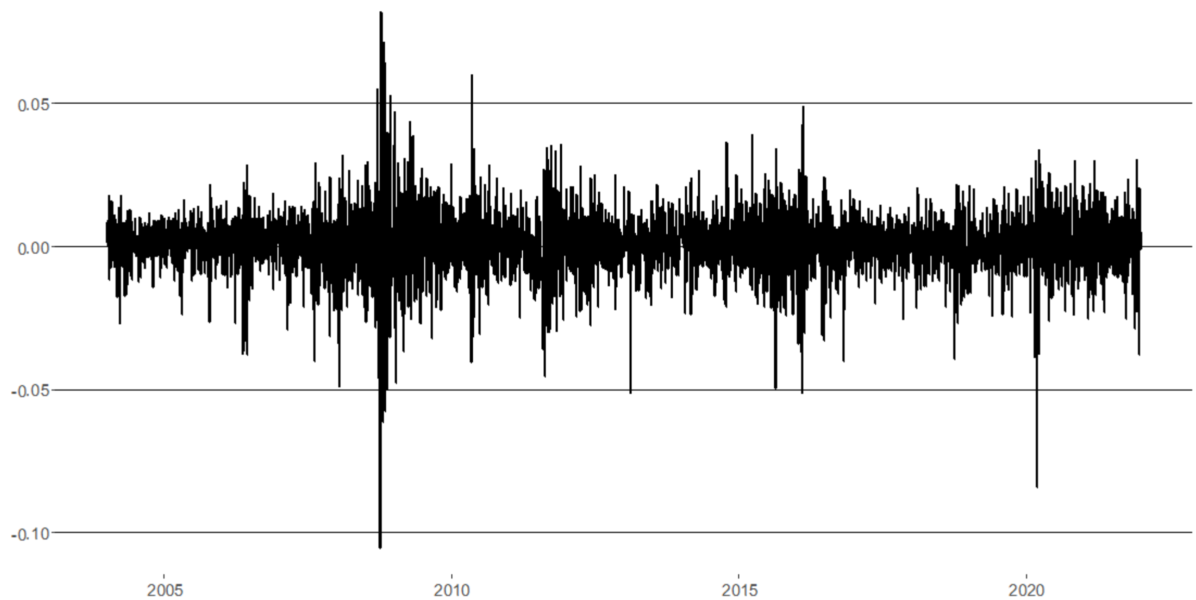

Figure 8: The graph depicts the Logged and first-differentiated of DKallshare for the sample period 2004-2022. This serves as an example for how the adjusted data looks graphically.

Figure 8 also highlight some extreme values, which could decrease the statistical power of our analysis (Stock and Watson, 2014 p. 71). Thus, to remove the apparent outliers, all return data ($\Delta r_{j,t}$) is winsorized at the 5% level (2.5% at each tail). This leaves us with the cleaned return variables depicted in table 3.

Variable
$\Delta DKallshare_{j,t}$
$\Delta BMIvalue_{j,t}$
$\Delta BMIgrowth_{j,t}$
$\Delta LMvalue_{j,t}$
$\Delta LMgrowth_{j,t}$
$\Delta SMLvalue_{j,t}$
$\Delta SMLgrowth_{j,t}$

Table 3: The table depicts the all adjusted variables defined in section 5.1.1 and 5.1.2.

5.2 SVI DATA

The following sections will explain in detail every step of constructing our sentiment index. As our main hypothesis depends on whether the methods of Da et. al (2015) can be applied in Denmark, we aim to follow their methodology. However, there are significant differences between the Danish and US financial markets, which might force us to make some deviations from their methods during the process, in which case these will be stated.

The sentiment index constructed by Da et. al (2015) is called the Financial and Economic Attitudes Revealed by Search (FEARS, hereafter) index. The FEARS name can easily be misinterpreted to proxy for negative sentiment. However, in section 5.3.1, we will explain why we in this paper construct a FEARS index proxying for investor optimism, rather than pessimism. As we are following the methods of Da et al. (2015) in constructing our index, we stick with the name FEARS. This serves as a disclaimer to simplify further interpretation of our results.

5.2.1 Google Trends

The process of collecting historic data from Google Trends is tedious. For data further back than the last 9 months, Google only allows for weekly and monthly frequencies to be downloaded from the website. To obtain daily frequency on the historic data all the way back to 2004, we use what can be called an unofficial API in the R software. A problem of this "unofficial API" is that Google has security measures in place to defer large amounts of data downloading, which might increase the workload of the Google's hardware. Specifically, Google only allows historic data downloads of 7-8 search terms per 24 hours per IP-address. This means that for historic analysis, a lot of time is required to procure the data. However, once a relationship has been established (that is, if our hypothesis is confirmed), any FEARS can be constructed with daily frequencies data downloaded directly from the Google Trends website, which allows for a practical usage of Google Trends data.

With the unofficial API discussed, Google Trends provides us with data on search frequencies for any term, in any timeframe, and within any area of our choice. Since we are interested in the Danish stock market, we have limited the google trends area to Denmark and the Danish language. Figure 9 has the purpose of showing how google trends output data, where we have selected the words "stock" and "gold", respectively "aktie" and "guld" in Danish, as examples. Table 4 shows the summary statistics for the same terms.

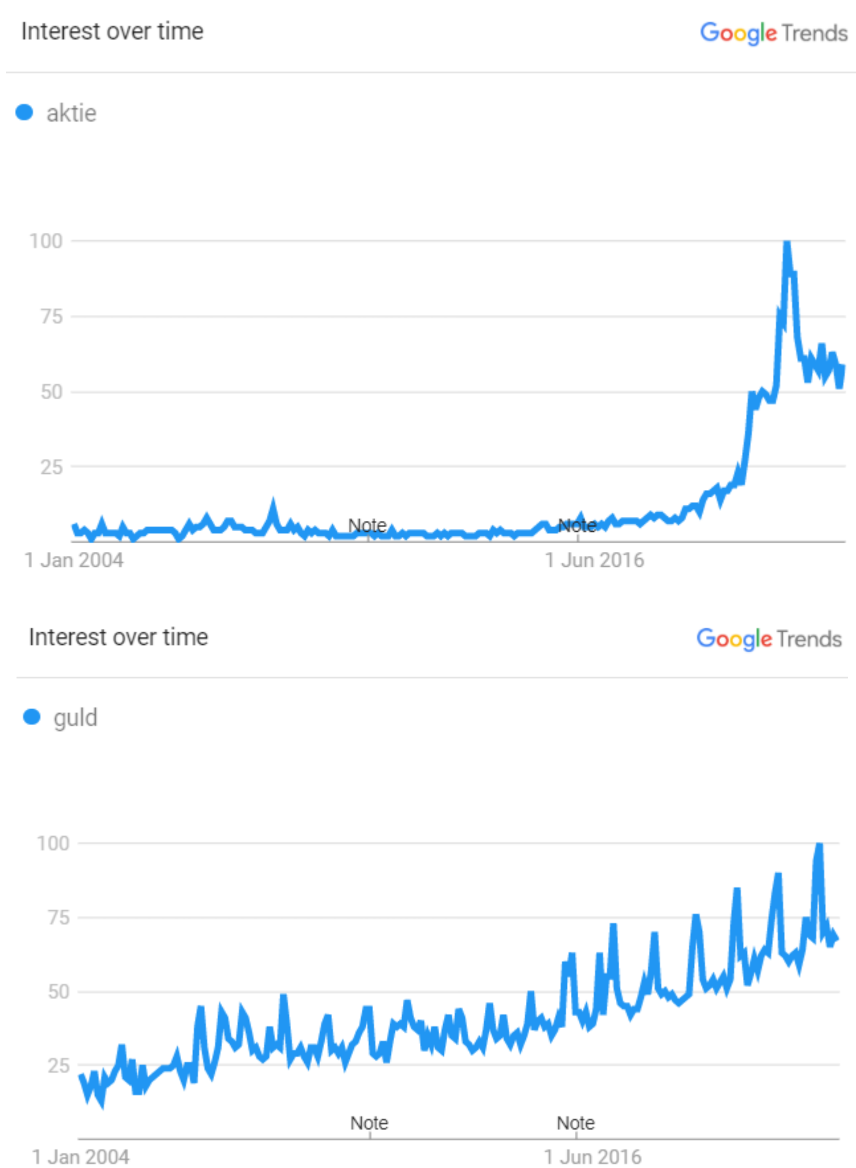

Figure 9: The figures represent a graphical presentation of weekly aggregate search frequency (SVI) from google trends. The top (bottom) panel plots weekly SVI for the term "aktie" ("guld") in Denmark, spanning from 2004 until 2022. The SVI is the weekly search volume scaled by the maximum value for the period. Source: Google Inc, n.d..

As mentioned in section (3.3.3), Google Trends does not provide the absolute number of search occurrences for a given search term. The website returns scaled numbers, that is, the SVI.

"Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means that there was not enough data for this term" (Google, 2022)

The appearing values can thus be interpreted as the total searches of a term relative to the total number of searches on Google over time. That is, a decreasing graph means that a search terms relative popularity is decreasing, but not necessarily the total number of searches for that term.

So, the values depicted in Figure 9 are not the actual search occurrences of "guld" and "aktie", they represent the changes from period to period, with respect to the highest observed search term for the period in a selected region. As we are not interested in the absolute search volume of each term but the relative changes, this way of indexing is beneficial for our research. From Table 4, we observe that the series of "guld" and "aktie" have different mean and median, as well as small differences in the deviations. As the table is based on the daily data, in contrast to the graphs presented in the graph in figure 9, the minimum and maximum values are as expected 0 and 100.

Term	Mean	Std.Dev	Min	0.25	Median	0.75
Aktie	5.977	12.823	0.000	0.720	1.660	3.920
Guld	24.650	14.719	0.000	14.040	23.040	33.540

Table 4: shows the descriptive statistics of daily observations for the two terms "aktie" and "guld" in the period between 2004-2022.

5.2.2 Term Selection

The next step in constructing the sentiment index, is the selection of which terms to be included. To construct a search term list of words, the methods of Da et al. (2015) are replicated.

	Process description	Total terms	Example terms
1	Extract English terms with the "economic" and "positive" or "negative" tags from The Laswell Value & The Harvard IV-4 Dictionaries.	133 terms	*"Depression"* *"Lay"*
2	Translate English terms into Danish and remove duplicate and insufficient data.	83 terms	*"Depression"* *"Lægge"*
3	Remove Danish terms that does not hold for the criteria: "economic" and "positive" or "negative".	55 terms	*"Depression"*
4	Extract the five most related queries to each Danish term.	191 terms	*"Depression"* *"Angst"* *"Great Depression"*
5	Remove related queries that does not hold for the criteria: "economic" and "positive" or "negative".	108 terms	*"Depression"* *"Great Depression"*
6	Remove Danish terms with less than 1000 days of observed values.	86 terms	*"Depression"*

Figure 10: This figure explains every step in the selection process, as well as highlighting the number of terms included in each step. In addition, the two terms "Depression" and "Lay" are included as examples of how terms are processed. The resulting 86 terms are the ones applied in our further analysis.

This approach of obtaining search term data is based on the two dictionaries: The Laswell Value Dictionary & The Harvard IV-4[11]. These dictionaries are utilized as they contain advantageous categories in the sorting of terms. Both dictionaries categorize English terms into a vast set of categories, and as we are interested in financial or economical words, we extract words under the "economic" tag. Our list of *743* terms, include words such as "inflation", "gold", "rich and "unemployed".

Furthermore, we are interested in words that are likely to have some positive or negative sentiment. We thus filter our economic list of terms to the additional tags "positive" or "negative", which yields a primitive list of *133* terms. In the next step, Da et al.(2015) apply google trends to extract the 10 most related queries for each term, however, as we are interested in the Danish search patterns, and thus the Danish language, we need to adjust the approach in adding a translation section.

Achieving a reliable method of translating the economic terms, is important for the terms to be applicable in the index construction. The translation process has thus been thoroughly conducted, applying reliable

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

translation sources, and cross checking of each term. Considering it a valid translation tool, Gao, Ren & Zhang (2020) apply Google Translate to build their FEARS index in different languages. Following their methods, we translate the primitive list (133 terms provided in appendix 2) using Google Translate. When each term is translated, to account for any possible weakness in the Google translation tool, two other[12] translation services are applied to cross check each translation, and thus the accuracy of Google translate.

The primitive list of English words is thus translated into Danish, and now consist of 83 words. The English words "inflation" and "bankruptcy" are for instance translated into "inflation" and "konkurs". This method of constructing a Danish term list cause for some duplicates words. The terms "precious" and "costly" are for instance both translated into the Danish word "kostbar". In addition, some of the translated terms does not have sufficient google trends data in Denmark. The word "capitalize", for example, translated into "kapitalisere" in Danish, does not have enough search occurrences for google trends to provide sufficient data. We thus remove for duplicates and insufficient terms.

The following step consists of removing Danish terms that does not hold for our criteria of being economic, and positive or negative. Removing such terms yields a list of approximately 60 terms. Moreover, we want to gather the "most related queries" [13]of each term from Google Trends. For example, when the word "spare" (*save*) is searched for, most people search for "lån" "spareribs" "fonde" as well. In this step, Da et al. (2015) extract the ten most related queries. As Danish Trend data is limited for several of the terms on the primitive list, only the five most related queries are extracted to ensure that some terms does not achieve vastly heavier weighting, due to the translation process. This extends the list to approximately 200 terms, as shortage of data on some terms makes it impossible to extract as much as five related queries.

When gathering related queries, trend data does not account for our criterions that the words have to be economic and either positive or negative, the list of 200 terms thus consist of many words not relevant for our research. For example, searching for the word "købe" (*buy*) in Danish, results in related searches "købe bil" (*buy a car*), "købe hus" (*buy a house*), "købe aktier" (*buy stocks*) and "købe guld" (*buy gold*). The searches for buying stocks and buying gold are included in our list, while the former searches are excluded.

Finally, to ensure that the terms have sufficient data for further analysis, we set the criterion of every term to have at least 1000 observations (Da et. al (2015). Every term which has less than 1000 observations of daily trend data in the chosen sample period are thus removed. This leaves us with a list of 86 terms, which will make up the foundation for the final FEARS index.

[12] DeepL.com, Ordbogen.com

[13] Google Trends provide a list of the most related queries for all Google searches

5.2.3 Preparing the data

Time series data of the final list of Danish terms [14]is now obtained. The occurrence of each term *j* in a given period *t*, will from now on be denominated as $SVI_{j,t}$, which is short for Search Volume Index. From figure 9 in section (5.2.1), we observed some clear trend in the SVI data. To account for this, we start our cleaning process by log transforming the difference between each observation, obtaining data on the relative change:

$$\Delta SVI_{j,t} = LN\left(SVI_{j,t}\right) - LN(SVI_{j,t-1}) \tag{3}$$

Figure 11 depicts the $\Delta SVI_{j,t}$ for the terms "guld" and "aktie", for two different time periods of 3 months.

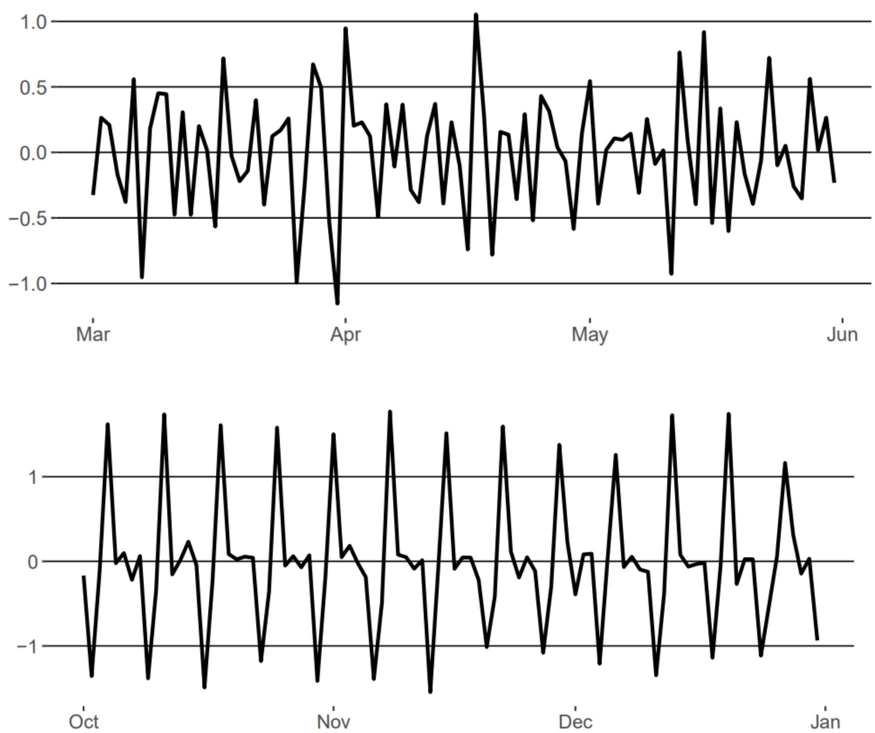

Figure 11: Show the daily log change observations of "guld" in the period (March-June 2010) and daily log change observations of "aktie" in the period (10/2021-12/2021). Source: Google Inc., n.d.

[14] aktie, aktier, arbejdsløshed, arbejdsløshed danmark, arbejdsløshed i danmark, bankruptcy, bonds, commodities, crisis, afgift, afskrivning, afskrivninger, afvikling, aktiekurser, aktiver, arbejdsløs, auktioner, børnepenge, dagpenge, danmark tilskud, egenkapital, eu, fallit, fattig, fattigdom, fattigdom i danmark, fonde, fordel, fordele, forrentning, fradrag, skat, gæld, gevinst, guld, guld pris, guldpriser, indtægt, inflation danmark, investering, købe, købe aktier, konkurser, korruption, koste, koster, likviditet, moms, nasdaq omx, nordnet, økonomisk krise, omx copenhagen, overskud, passiver, penge, penge tilbage i skat, recession, rig, rigdom, skat, skattelettelser, sølv, spar bank, spare, tarif, tilskud, udgift, udgifter, underskud, BNP, boom, charity, donate, donation, equity, fattige, feriepenge, goodwill, taric, velgørenhed, depression, arveafgift, dyrt, billig, lån og spar, værd, lån.

The figure (11) highlights a clear seasonal pattern, especially depicted for "aktie". The SVI change rises in the weekend and after reaching its peak around Monday, it starts a decline. The log changes plotted in figure 11 reflect the 7-day hump pattern following this seasonal effect. Moreover, Figure 11 also highlight the difference in variance of the two ΔSVI ("guld", "aktie"). We know that the SVI are plotted on the same scale, so heteroscedasticity is indeed apparent. Figure 11 also highlight the presence of some extreme values in "guld", which might impact further analysis.

To account for the outliers, seasonality, and the apparent heteroscedasticity, we adjust the data in the following manner. Any concerns outliers might cause in the data, is mitigated by winsorizing each of the series at the 5% level, 2,5% at each tail. The seasonal effects, both monthly and weekly, is removed by regressing each of the series on monthly and weekday dummies, keeping the residuals. Finally, to remove the apparent heteroscedasticity, we follow Baker and Wurgler (2006), standardizing the timeseries by scaling each by the time-series standard deviation. Changes in the occurrence of terms is thus weighted equally when we construct our indices. We now have the adjusted change in daily search volume, denoted as $\Delta ASVI_t,$ for each of our 86 terms.

5.3 CONSTRUCTING THE FEARS

The $\Delta ASVI_t$ is now prepared and ready for application. This led us to the final step of the composition of the sentiment measure, which is determining the search terms to be included in the aggregated index. Following Da et al. (2015), the FEARS will be constructed based on the historical relationship between each $\Delta ASVI_t$ and the $DKallshare$. The purpose of constructing a daily sentiment index based on historical returns, is to test whether investor sentiment has predictive power towards equity returns.

Thus, Following the methodology of Da et al. (2015), we allow the data to identify the search terms that exhibit the strongest historical correlation with $DKallshare$. To catch the impact of $\Delta ASVI_{j,t}$ on $DKallshare_{j,t}$, an expanding rolling regression is applied. The expanding rolling regression is designed as a loop, where every iteration expands the dataset by 120 trading days (6-month window). Furthermore, T-values are then collected to determine which search terms will be used to construct the FEARS for the subsequent 120 trading days. By example, this means that our July 2021 to December 2021 search terms, are selected on the historical contemporaneous relationship from January 2004-June 2021, the January 2021 to June 2021 search terms selected on the historical contemporaneous relationship January 2004- December 2020, and so forth.

Thus, the following OLS regression is run every 120 trading days:

$$\Delta DKallshare_t = \beta_0 + \beta_1 \Delta ASVI_{j,t} + u_t \tag{4}$$

There are primarily three reasons why the expanding rolling regression is advantageous. First, this regression allows the term selection to only have basis in historical data. Running a regression for the historical relationship over the entire sample period would implicate that the search terms selected for a period may have not actually exhibited a relationship with returns prior to its inclusion in the FEARS, but later. It is an important principle to attain, using only historical information, if the index should have any practical application. Secondly, this approach allows for a dynamic search term selection, whereby the search terms included are dynamically selected based on which exhibited the most significant historical relationship with returns. That is, a term can be highly relevant in one period, but ten years later, have no significance. The search term selection is also rebalanced every 120 trading days. Finally, as with Da. et al. (2015), this method is found to increase the statistical power of the regression.

5.3.1 Positive or Negative Sentiment

Economic terms with a negative return correlation, are suggested to have larger impact on stock returns than terms exhibiting a positive relationship (Tetlock, 2007; Da et.al 2015). However, when running regression (9), it becomes clear that terms with both positive and negative relationships impact $DKallshare$ significantly. Contrary to Da et al. (2015) and Tetlock (2007), we find that words with positive relationships provide higher t-values, in absolute terms, than the negative ones.

There are several possible explanations of why our positive words have a stronger relationship with returns than the negative. It makes sense following the findings of Miller (1977), who suggest that positive sentiment plays a more important role than negative sentiment in the capital markets, because pessimistic investors stay on the sidelines because of apparent short-selling constraints. Moreover, Hong & Stein (2007) which emphasize that market prices are driven by optimistic traders, i.e. even if the pessimistic investors become as pessimistic as the optimistic investors are optimistic, prices go up. Supporting this, Yu and Yuan (2011) claim that positive sentiment may be more valuable than negative when emphasizing the impact of noise trading.

Finally, the inclusion of both positive and negative terms would make our results hard to interpret. Thus, we exclude negative terms from our index, building our FEARS index on the $\Delta ASVI$s that are most positively correlated with $DKallshare$.

5.3.2 Defining the FEARS variable

When defining the FEARS variable, the number of terms to be included in the index must be determined. Da et al. (2015) choose a pre-determined amount of 30 words to be included in the index as they consider it to be the minimum number of observations needed to diversify away idiosyncratic noise (Da et.al, 2015). However, Da et al. (2015) construct their sentiment measure for the US, thus facing much higher frequencies in SVI observations, that is, less missing observations or zero-values. Studying the Danish market, and the behaviour of Danish retail investors, a cut-off of 20 terms is thus viewed a better fit.

The inclusion of 20 words instead of 30, potentially allows for more idiosyncratic noise, which in turn will make the index more exposed to irrelevant factors. For example, the search occurrence of the term “rig” might increase if a movie with the name “rig” has been released in the same period. The more terms included in the index at such a point in time would then diversify away the noise from this irrelevant event to a greater extent. As our sample period consist of 18 years of daily observations, this idiosyncrasy will be less relevant over time, and we conclude with choosing a cut-off of 20 search terms for each period. Robustness to alternative cut-off choices will be tested in section 8.

The FEARS index on day t can now be defined as:

$$FEARS_t = \sum_{i=1}^{20} R^i(\Delta ASVI_t) \tag{5}$$

Where $R^i(\Delta ASVI_t)$ is the $\Delta ASVI_t$ for a search term that had a t-statistic rank of *i* in the period January 2004 throughout the following six months, where the ranks range from the least significant (86) to most significant (1) for the term list. We then rank each word from most to least positive. The FEARS in one period, thus represent the average of the 20 most positive words in relation to $DKallsare$.

#	Search Term	T-Statistic
1	nordnet	3.294
2	arbejdsløshed i danmark	2.508
3	afskrivninger	1.932
4	nasdaq omx	1.913
5	aktie	1.796
6	fattigdom	1.693
7	rig	1.671
8	arveafgift	1.588
9	fallit	1.532
10	goodwill	1.494
11	fradrag skat	1.456
12	danmark tilskud	1.333
13	aktier	1.330
14	værd	1.329
15	spare	1.269
16	guldpriser	1.211
17	depression	1.142
18	koster	1.123
19	aktiekurser	1.119
20	korruption	1.111

Table 5: This table reports 20 search terms derived from words of economic sentiment in the Laswell and Harvard dictionaries (section 5.2.2), that exhibit the largest positive correlation with the market for the entire sample period. The terms are ordered from most positive ("nordnet") to least positive ("korruption").

Table 5 depicts the 20 search terms with the largest positive correlation with $DKallshare$ over the entire sample period. Recall that the search terms are updated every six months when constructing the actual sentiment index, i.e., the exact terms shown in table 5 are not relevant for our further results but serve as an example. In addition, figure 12 (*top panel*) depicts the historical performance of the FEARS index. From the figure, we observe what might look like a trend-stationary process in the time-series data, where the FEARS is increasing during the sample period. This might indicate an increasing trend in the use of Google as a search platform in Denmark.

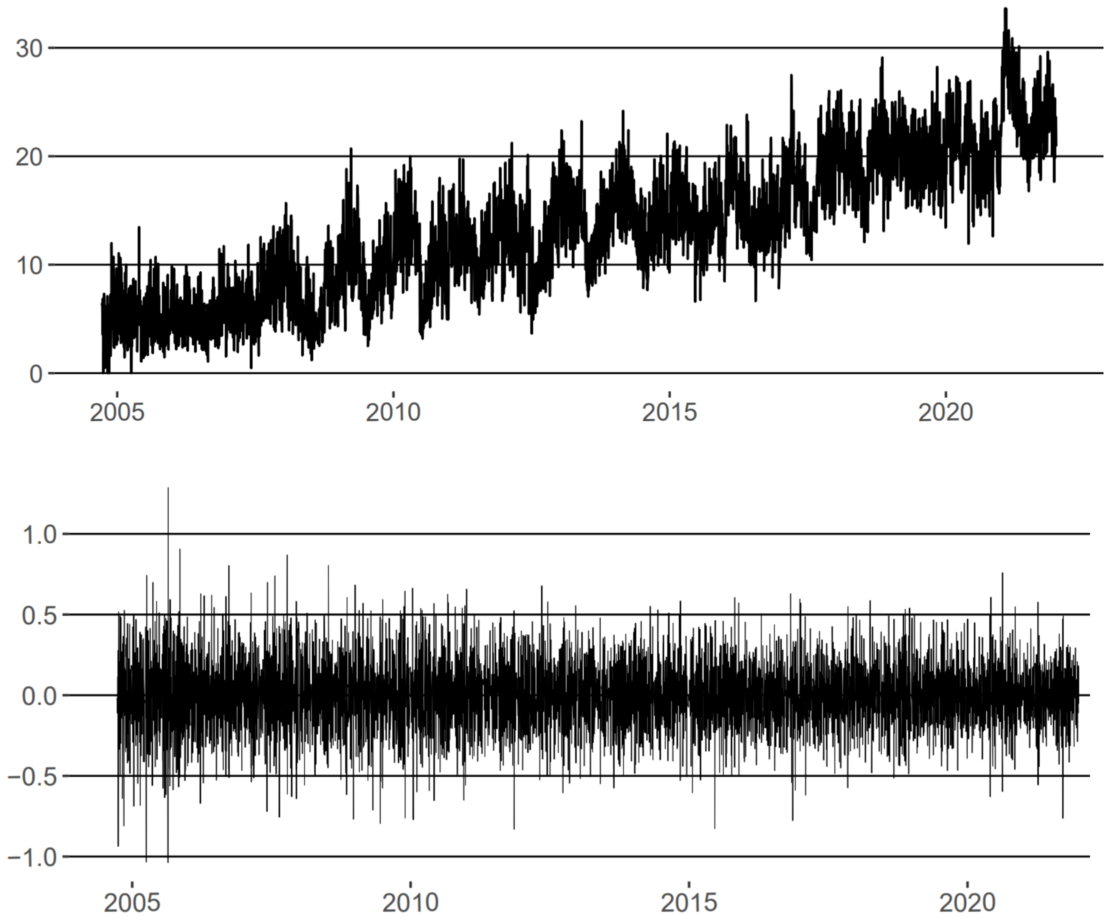

Figure 12: The top panel depicts the unadjusted historical FEARS time series, which is constructed based on SVI raw data. The bottom panel depicts the adjusted historical FEARS time series, which is constructed based on $\Delta ASVI$ data.

Data	Mean	Std.Dev	Min	0.25	Median	0.75	Max
FEARS	0.000	0.234	-1.038	-0.153	-0.003	0.152	1.286

Table 6: The Table above depicts the summary statistics for the cleaned FEARS index. The FEARS has a mean of zero and a standard deviation of 0.234, due to the correlation between the normalized SVI search terms.

5.4 OTHER DATA

The stock return variables as well as the FEARS variable, are now defined and the timeseries data prepared. However, as presented in section (1.1), this paper will also test the application of FEARS on realized volatility and the bond market. Thus, the following sections will prepare the data to be applied for this additional analysis. Moreover, we include control variables to enhance the internal validity of research.

5.4.1 Government bonds

Extending the sentiment analysis, the application of our FEARS will also be tested in the Danish bond market. This analysis will be focused on long term Danish government bonds with maturities 5 and 10 years. Historical data on 10 Year ($10Ybond_{j,t}$) and 5 Year ($5Ybond_{j,t}$) Government bonds are extracted from Investing.com (Investing.com, n.d.) Due to limitations in data, the government bond analysis starts in 2005. Figure 13 shows the raw data on $10Ybond_{j,t}$ and $5Ybond_{j,t}$.

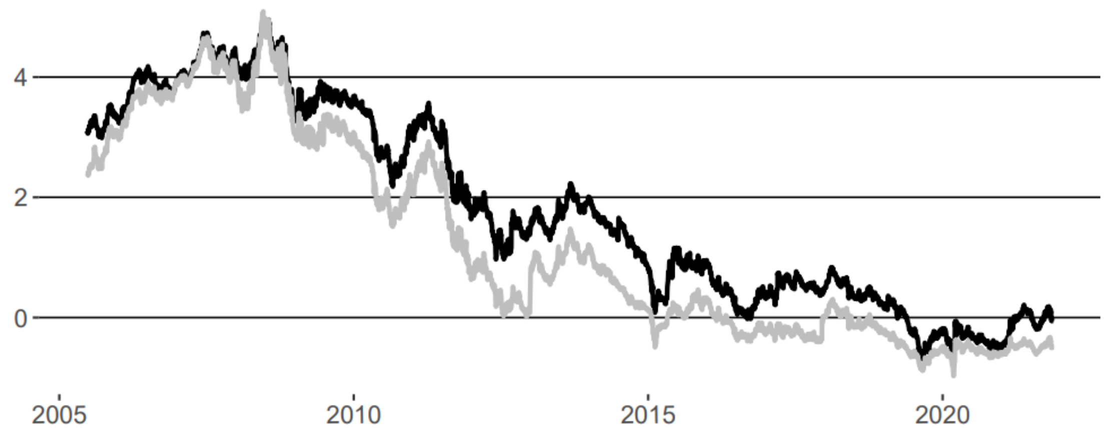

Figure 13: 10Ybond (black) and 5Ybond (grey) returns over the sample period 2005-2022.

Index	Mean	Std.Dev	Min	0.25	Median	0.75	Max
10Ybond	1.8090	1.616441	-0.79900	0.383	1.542	3.4330	4.994
5Ybond	1.2657	1.698317	-0.9730	-0.2105	0.6350	2.9360	5.0910

Table 7: This table shows the summary statistics for the unadjusted variables $5Ybond$ *and* $10Ybond$*, for the sample period 2005-2022.*

Looking at figure 13, we observe declining trends in both $5Ybond$ and $10Ybond$. In section 5.2.2, we observed that the stock variables experienced an increasing trend throughout the sample period. As the stock prices increase, we expect to see a declining trend in bond yields. That is, there is ample evidence that bond prices and stock prices exhibit a high correlation (Collin-Dufresne et.al, 2001; Fang et al, 2018). As the graph above depicts bond yields, which exhibit an inverse relationship with bond price (Bodie, Kane & Marcus, 2014, s. 516), the trend observed follows our expectation. In the same manner, as for the stock return data, this trend will be accounted for by taking the logarithm and first difference of both $10Ybond_{j,t}$ and $5Ybond_{j,t}$:

$$\Delta 10Ybond_t = \ln(10Ybond_t) - \ln(10Ybond_{t-1}) \tag{6}$$

$$\Delta 5Ybond_t = \ln(5Ybond_t) - \ln(5Ybond_{t-1}) \tag{7}$$

Figure 14 depicts the first-differenced log variables spanning from 2005-2022 with the purpose of identifying any concerns in the data.

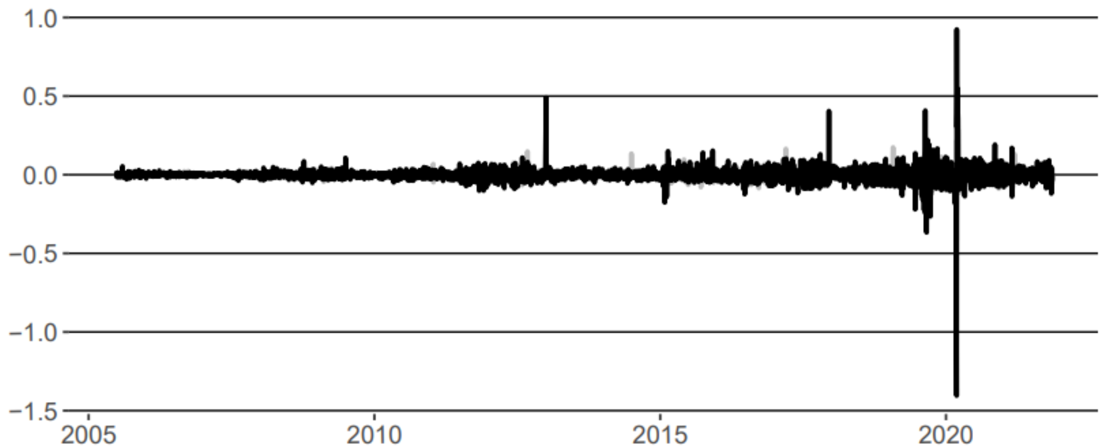

Figure 14: the figure shows data for $\Delta 5Ybond (Grey)$ and $\Delta 10Ybond$ $(Black)$ from 2005-2022.

From figure 14, we observe some yearly seasonality, such as in 2020. In the same year, we also observe a significant negative spike. As monthly seasonality is evident in our stock return data, we expect that the bonds exhibit some monthly seasonality as well. To account for this, we apply year and month dummies. Moreover, the apparent outliers will also be accounted for as we winsorize both the $\Delta 10Ybond_{j,t}$ and $\Delta 5Ybond_{j,t}$ at the 5% level (2.5% at each tail).

5.4.2 Corporate bonds

Although our main analysis for the bond market regards the government bonds in Denmark, this subsection has the scope of adding weight to the analysis of the sentiment impact on bonds. Nayak (2010) suggest that there is more significant predictive power for sentiment regarding speculative-grade bonds (high yield), rather than investment-grade bonds (low yield). However, the analysis towards the corporate bond market will have basis in Danish Investment Grade Bonds (IG bonds), provided by S&P, as the data on high yield bonds are limited for Denmark. The IG Bond index provided by S&P is called the "S&P Denmark Investment Grade Corporate Bond Index" ($IGBond$, hereafter), containing high rated "Investment Grade" bonds. Furthermore, Historical data on the $IGbond$ is extracted from S&P Capital IQ (S&P Global, n.d.)

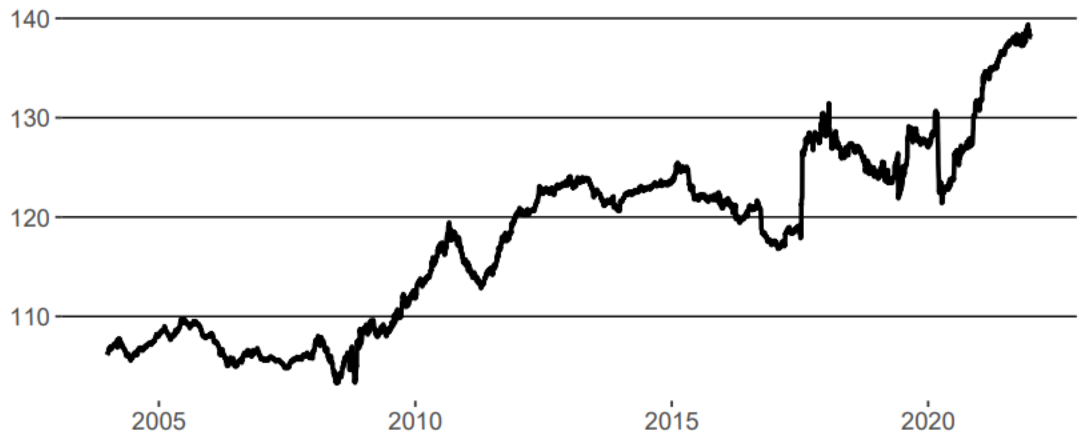

Figure 15: This graph depicts the historical prices of S&P Denmark Investment Grade Corporate Bond Index, spanning from 2004-2022. Source: S&P Global, n.d.

Not surprisingly, as the graph in figure 15 depicts the $IGbond$ price, we observe an increasing trend, opposite to that of the government bond graph (figure 14). Furthermore, the corporate bond data is adjusted in the same manner as the $10Ybond$ and $5Ybond$ variables, as IG Bonds are driven mainly by returns on treasuries (Nayak, 2010). Table 8 depicts the summary statistics for the raw $IGbond$ data.

Index	Mean	Std.Dev	Min	0.25	Median	0.75	Max
IGbond	118.0000	8.951105	103.3000	108.5000	120.3000	123.8000	139.4000

Table 8: This table shows the summary statistics for the S&P Denmark Investment Grade Corporate Bond Index.

Variable
$\Delta 5Ybond_{j,t}$
$\Delta 10Ybond_{j,t}$
$\Delta IGbond_{j,t}$

Table 9: The table highlights the adjusted bond variables. Although the variable are denoted as "$\Delta bond$" after adjustment, they will be denoted as "$bond$" in the discussion section 9.

5.4.3 Volatility

We examine the relationship of investor sentiment and realized volatility as a direct measure of stock market volatility. The realized volatility estimation procedure was developed av Andersen et al. (2001) and Andersen, Bollerslev, Diebold, and Labys (2003). They use continuously recorded transaction prices (at high intraday frequencies), to construct estimates of realized daily volatility return. They state that "volatility estimates so constructed are model-free, and as the sampling frequency of the returns approaches infinity, they are also, in theory, free from measurement error" (Andersen et al. 2001, p. 45).

We obtain our realized volatility data from an online, free database named "Oxford-Man Institute's realized library" (Oxford-Man Institute of Quantitative Finance, n.d.). This data is made available for a project organized by the Oxford-Man Institute of Quantitative Finance at the University of Oxford. The Realized Library data is based on high frequency data, obtained through Thomson Reuters Tick History. The database's makers focus on the methodology of Andersen et al. (2001) as well as Barndorff-Nielsen, Kinnebrock, and Shepard (2008) to create time-series data on Stock market return realized volatility for numerous countries and indices, ranging on different time frequencies and estimator models. From the webpage's description, the econometric method is as follows:

The database contains daily (close to close) financial returns

$$r_1, r_2, \dots r_T \tag{8}$$

and a corresponding sequence of daily realized measures

$$RM_1, RM_2, \dots, RM_T \tag{9}$$

whereby the realized measure of realized variance is

$$RM_t = \sum x_{j,t}^2 \tag{10}$$

Where

$$x_{j,t} = X_{t_{j,t}} - X_{t_{j-1,t}} \tag{11}$$

and $t_{j,t}$ are the times of trades or quotes (or a subset of them) on the t-th day. The realized variance output is annualized, so that we know have a direct measure of stock market volatility equal to the one in Da et al. (2015) paper, that is:

$$RM_t = 250 \sum x_{j,t}^2 \tag{12}$$

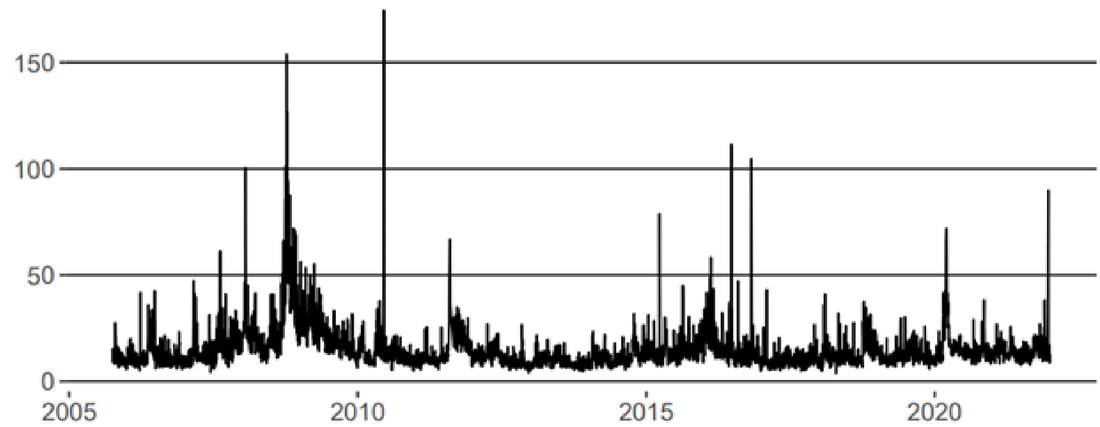

Figure 16: The figure depicts raw daily frequency development of Realized Volatility, measures from 10-minute interval intraday return data on the OMX C20. It is constructed for this thesis, based on data from the Oxford-Man Institute of Finance (Oxford-Man, n.d.).

We obtain the daily frequency time-series datasets from the database: OMX Copenhagen 20 Index Annualized Realized Volatility (10-min). We clean the data as we did with the search term data for the FEARS index, computing the log realized volatility, winsorize at the 5% level (2.5% at each tail) and remove possible seasonal effects by regressing on weekday and monthly dummies. Finally, we are left with a time series of the residuals, which are now winsorized, seasonal-adjusted log time series of realized volatility (Adjusted Realized Volatility).

Data	**Mean**	**Std.Dev**	**Min**	**0.25**	**Median**	**0.75**	**Max**
Adjusted Realized Volatility	0.0000	0.432	-0.80290	-0.3105	-0.05678	0.2462	1.23885

Table 10: Summary statistics for the Adjusted Realized Volatility data.

5.4.4 Control Variables

Control variables can introduce as well as remove biases. In our paper, Control variables are included in our regressions to mitigate upward bias in regression coefficients, because of omitted variable bias. However, in selection of additional regressors we also need to be wary of other biases, that is, the overcontrol bias and endogenous selection bias.

The argument is that the true causal relationship between two variables cannot be identified without including all relevant variables. Failing to include the correct variables will cause an omitted variable bias in the regression model. (Wooldrigde, 2009, as cited in Li, 2021). Coock & Campbell (1979), as cited in Li (2021) stated that in business research, Mill’s three criteria was the philosophy of causal effect that stood out. That is, covariation between independent and dependent variables, the temporal precedence of the independent variable, and the need to use control variables to rule out alternative explanations. Li (2021) argues that while using control variables may at times rule out alternative explanations, it is equally possible that adding control variables introduces overcontrol and endogenous selection biases, creating alternative interpretations rather than ruling them out.

Our aim with control variables is to strengthen the causal arguments of our FEARS index by ruling out alternative explanations. Li (2021) calls this the purification role of control variables. The fear is that by not including relevant control variables, our findings will be subject to the omitted variable bias, and upward bias our regression coefficients. However, when including our control variables, we need to be wary of the overcontrol bias. Overcontrol bias occurs when including an additional variable, which sits on the causal path of other independent variables and the dependent variable. This could be a mediator or an intermediator (Li, 2021). Overcontrol bias might bias our findings up or down. Furthermore, we will need to be wary of introducing the endogenous selection bias to our models. Endogenous selection bias is introduced when conditioning on a collider variable, which may create alternative explanations and spurious causal effects (Li, 2021).

Thus, we will need a strong theoretical fundament for the causality of included control variables, as well as look out for strong collinearity, whereby there is a high correlation between the control variables. As we are trying to determine whether our FEARS index have any statistically significant explanatory power or predicting power of returns and volatility, we should include control variables which represent the alternative explanations for movements in stock market returns and realized volatility movements. Starting our analysis, we propose the following control variables: lagged values of the dependent variable, which in our case will be lagged returns, the Citigroup Economic Surprise Index, the Economic Policy Uncertainty Index, and the implied volatility index VSTOXX. These control variables match closest to the control variables selected by Da. et al. (2015), which increases the similarity between the regressions used for testing, and the methodology.

5.4.4.1 Citigroup Economic Surprise Index

The efficient market thinking suggests that asset prices should instantaneously reflect changes in the underlying fundamentals (Andersen, Bollerslev, Diebold, Vega, 2007). In line with this, Andersen et al. (2007) finds that announcement surprises produce conditional mean jumps. Regardless of the nuances of how macroeconomic news and surprises influence returns, is it safe to say that macroeconomic fundamentals are priced in at some point into the stock market. Fundamentals such as GDP and inflation are important to cash flow forecasting, as well as important indicators of expansion or recession periods. Thus, macroeconomic news influences stock prices. Furthermore, surprises in macroeconomic fundamentals or news should have a correctional effect on the stock market, as the market participants need to alter their view of the economic situation. We argue that in our goal to reduce the omitted variable bias, and thus purify our FEARS variable, a measure of economic surprise is needed.

Our primary source article by Da et al. (2015) includes changes in the Aruoba-Diebold-Scotti (ADS) business conditions index as a control variable when running their FEARS-index on returns and volatility. The ADS business conditions index tracks real business-conditions at high observation frequencies (Aruoba-Diebold-Scotti Business Conditions Index, n.d.). The ADS index is only available for the US, we therefore need to find an alternative, preferably closer to our markets.

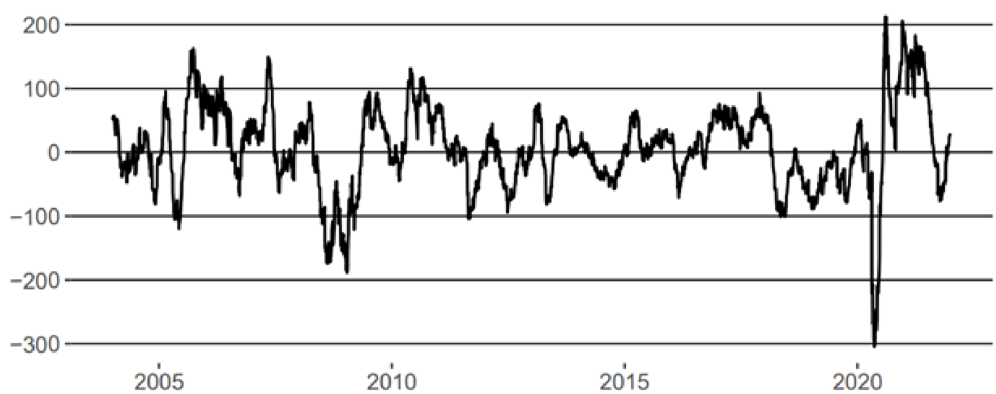

Figure 17: The figure depicts raw daily frequency development of the Citigroup Economic Surprise Index. It is constructed for this thesis, based on Citigroup data (Refinitiv Datastream, n.d.)

The Citigroup Economic Surprise Index (CESI) represents how occurring, real economic data measures relative to the consensus forecast of market economist (Boesler, 2013). The Citigroup Economic Surprise Indices are objective and quantitative measures of economic news. They are defined as "weighted historical standard deviations of data surprises" (Actual releases vs Bloomberg survey median) (Bloomberg, cited in (Boesler, 2013). In short, the Citigroup index displays the difference between forecasted and actual releases of economic news. Citigroup however, do not publish a CESI index for Denmark, so we will make do with the CESI index for the euro zone. The index is obtained from the DataStream, and is recorded in daily frequencies (Refinitiv Datastream, n.d.).

From the raw data, which is illustrated in Figure 17, the data is log-differenced, winsorized at the 5% level (2.5% of both tails) and regressed on weekday and monthly dummies to remove possible seasonality. We are left with a time-series of residuals. Summary statistics on the cleaned data in presented in Table 11.

5.4.4.2 Economic Policy Uncertainty Index

The article from Da et al. (2015) also includes a measure of policy uncertainty as a control variable. They include data from an index called the Economic Policy Uncertainty (EPU) index. The index aims to serve as a proxy for movements in economic policy-related uncertainty. The index is developed for the US-based on newspaper coverage from 10 different large newspapers in the U.S., including the Washington Post, The New York Times and the Wall Street Journal. The index is further based on federal tax code provisions set to expire, drawn on reports by the Congressional Budget Office (CBO), and the disagreement among forecasters drawn from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters (Baker. Bloom. & Davis. n.d.)

The developers of the index write about three strands of literature to which policy uncertainty relates, two of which are relevant in our case. Firstly, uncertainty in policymaking may impact growth and investment. Bernanke (1983) as cited in Baker, Bloom, and Davis (2016), argues that uncertainty in economic policy incentivizes a delay of firm investments and hiring, because investment projects are expensive to undertake, and workers are costly to both hire and fire. Second, as cited in Baker et. al (2016), Friedman (1968) amongst others consider the effects of monetary, fiscal, and regulatory policy uncertainty. Furthermore, Baker et al. 2017 cite Pastor and Veronesi (2012, 2013) who model the theoretical links among fluctuations, policy uncertainty, and stock market volatility. Uncertainty in economic policy will increase the cash flow risk of a firm, reducing the investment appetite, and the valuation.

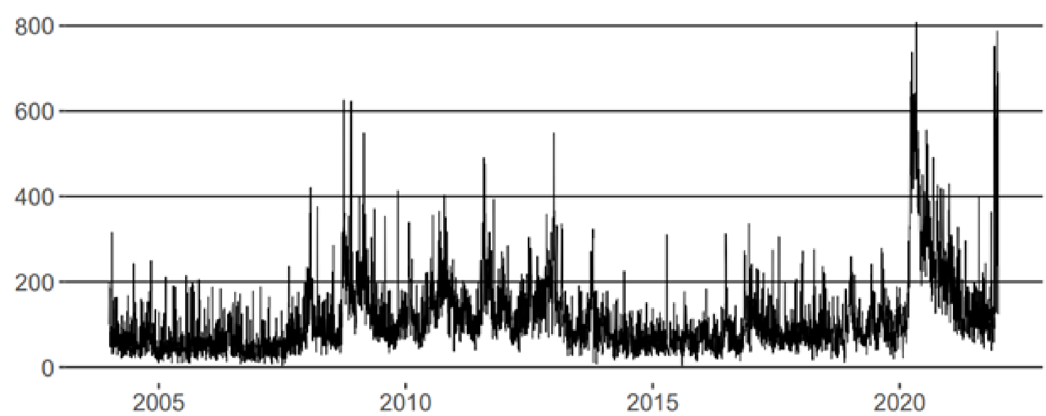

Figure 18: The figure depicts raw daily frequency development of the Economic Policy Uncertainty Index. It is constructed for this thesis, based on data from the St. Loius FED (Refinitiv Datastream, n.d.)

Baker et al. (2017) finds in their paper that elevated policy uncertainty may have harmed macroeconomic performance in the U.S and Europe in recent years. The results also point to sizeable effects of policy uncertainty on the cross-sectional structure of stock price volatilities, investment rates, and employment growth. The index is only reported for the U.S. in daily observations, which is a weakness is using the EPU as a control variable. However, we feel it important to include a measure of policy uncertainty, which may have explanatory power in the Scandinavian models as well, when shocks occur in returns and volatility. From the raw data illustrated in Figure 18, we also winsorize differenced log variables, and regress on weekday and monthly dummies and keep the residuals. The summary statistics are reported in Table 11.

5.4.4.3 VSTOXX

The most famous and traded index on volatility is the VIX index. The VIX index measures constant, 30-day expected volatility on the U.S stock market, derived from real-time, mid-quote prices of S&P 500 Index (SPX) call and put options (Making Sense of the VIX Index: An Indicator of Expected Market Volatility, n.d.). This is also what is used in Da et al. (2015)'s paper. Likewise, we will use the European volatility index Euro Stoxx 50 Volatility (VSTOXX).

The VSTOXX indices are based on EURO STOXX 50 real-time option prices (EURO STOXX 50® Volatility (VSTOXX®), n.d.). The index is calculated by measuring the square root of the implied variance of all options of a given time to expiration. Contrary to the VIX, the VSTOXX is calculated by call and put options on the European stock index *EURO STOXX 50*, which is the index for 50 leading blue-chip stocks in the Euro-zone.

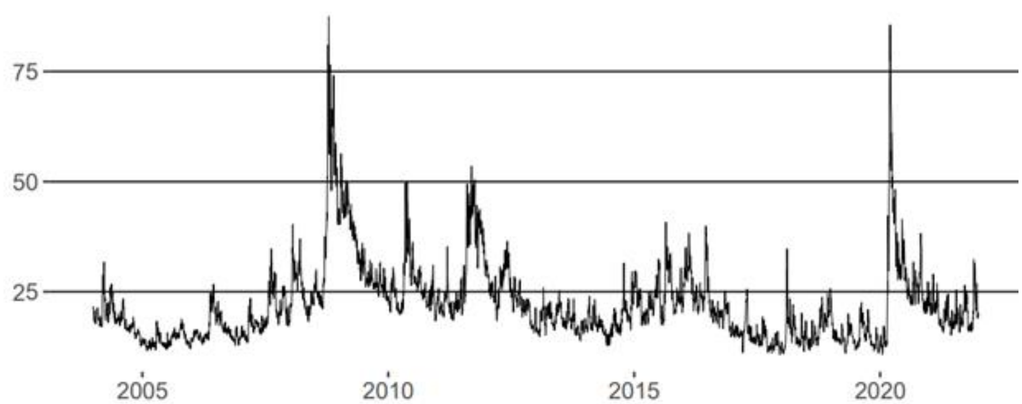

Figure 19: The figure depicts raw daily frequency development of the VSTOXX implied volatility Index calculated from option data on the EURO STOXX 500. It is constructed for this thesis, from Datastream ((Refinitiv Datastream, n.d.)

The VSTOXX, as the comparable alternative to the VIX in the US, is generally viewed as an indicator a measure of uncertainty in the markets and should have a negative relationship with stock market returns, be it contemporaneously or in future stock returns. We clean the data by taking the first difference of the natural logarithm, and regress on weekday and monthly dummies, keeping the residuals.

Tables 11 and 12 report the summary statistics for the control variables defined above. All our control variables are based on daily observations, as all other data.

Data	Mean	Std.Dev	Min	0.25	Median	0.75	Max
CESI-EUR	-0.0001	0.0497	-0.9794	-0.0583	-0.0003	0.0059	1.3800
EPU-US	-0.0569	0.5125	-3.1483	-0.3650	-0.0569	0.2460	3.0325
VSTOXX	0.0002	0.0661	-0.4344	-0.0387	-0.0480	0.0318	0.4857

Table 11: This table depicts the summary statistics of the cleaned control variables, that is CESI-EUR, EPU-US and VSTOXX.

	CESI-EUR	EPU-US	VSTOXX
CESI-EUR	1.0000		
EPU-US	0.0006	1.0000	
VSTOXX	0.0164	-0.0352	1.0000

Table 12: This table depicts a correlation matrix of cleaned control variables, that is CESI-EUR, EPU-US and VSTOXX.

The correlation matrix shows that the cleaned control variables have low correlation coefficients. We find that the Citigroup Economic Surprise Index, and the Economic Policy Uncertainty Index are correlated positively, whereas the Eurozone VSTOXX are correlating negatively with the EPU-US. High correlations between the Control Variables could lead to collinearity issues, but with the results depicted in the correlation matrix, we are confident that no collinearity issues will occur.

6 EMPIRICAL STRATEGY

The empirical strategy seeks to explain how the proposed hypotheses will be tested empirically. More specifically, the following sub-sections will provide evidence from prior research within each area of discussion. This is done to give further insight into why we conduct the tests. Secondly, the methodology to be applied in the testing will be explained in detail, providing more basis for interpreting the results in section 7. For the convenience of the reader, the main hypothesis of the paper is again stated:

Does investor sentiment have predictive power on Danish equity returns?

Moreover, additional tests will be conducted to answer the sub-questions presented in section 1.1, which emphasize whether our FEARS index has predictability towards the cross-section of stocks and the Danish bond market, and finally, a relationship with the realized volatility of stock returns.

The results provided by the regression analyses will be interpreted as basis points increase in the dependent variable, by one standard deviation increase in FEARS. This is made possible by standardizing all search terms included in the aggregate index so that they exhibit a mean of zero and a standard deviation of one. However, our aggregate FEARS index will not have a standard deviation of one due to correlation among the 20 search terms included in the index.

6.1 AVERAGE RETURN

The relationship between investor sentiment and equity returns is well-documented and provides indirect evidence of the market anomalies which lead to inefficiency (Baker and Wurgler 2007). As stated by Baker and Wurgler (2007):

"Now, the question is no longer, as it was a few decades ago, whether investor sentiment affects stock prices, but rather how to measure investor sentiment and quantify its effect".

The originators of the FEARS methodology proved that the sentiment index can predict short-term return reversals in the U.S stock market. That is, they find that the effect of a spike in their negatively loaded FEARS index, results in lower contemporaneous returns, and higher returns (reversal) the following few days (Da et.al, 2015). This observed reversal pattern is consistent with sentiment-induced mispricing of stocks (Delong et.al, 1990). As our constructed sentiment index proxies for investor optimism, these results indicate that we might observe a negative stock return reversal effect, following a positive spike in our fears.

Thus, our main hypothesis will be tested. As explained in section 5.1.1, the $\Delta DKallshare$ will serve as our stock return measure. To measure the relationship between FEARS and the $\Delta DKallshare$, we run the following regression:

$$\Delta DKallshare_{t+k} = \beta_0 + \beta_1 FEARS_t + \textstyle\sum_m^n \gamma_m Control_t^m + u_{t+k} \tag{13}$$

Whereby the $\Delta DKallshare_{t+k}$ denotes the seasonally adjusted log-returns on date i of Denmark's All Share Index on day $t+k$, $FEARS_t$ denotes our constructed FEARS index. Control Variables included are the ones prepared in section 5.4.4, as well as up to four lagged values of returns.

6.2 THE CROSS-SECTION OF RETURNS

There is ample evidence that the effect of investor sentiment varies within the cross-section of stocks. (Lee et.al., 1991; Kumar and Lee, 2006; Brown and Cliff, 2005; Lemmon and Portniaguina 2006; Baker and Wurlger, 2006,2007; Schmeling, 2009). Theory suggests that growth stocks are harder to value and thus harder to arbitrage than historically performing stocks, thus being more prone to shifts in sentiment (Baker and Wurgler, 2006). This is supported by Brown and Cliff (2005) who find a stronger effect of sentiment on growth than for value stocks in the US. Results on whether the effect is larger for value or growth stocks are not uniform. As suggested by Kumar and Lee (2006), retail investors tend to overweight value stocks relative to growth stocks in their portfolios. They report a contemporaneous correlation between the buy-sell imbalance of retail investors and the returns on high book-to-market stocks, but no correlation on low book-to-market stocks, suggesting that the impact of sentiment on individual investors is more significant for value than for growth stocks. Supporting the results of Kumar and Lee, Lemmon and Portniaguina (2006) provide evidence that value stocks alone, tend to respond to changes in sentiment. Baker and Wurgler (2006) however, report that sentiment effects are of similar size for both value and growth stocks, both being significant. Schmeling (2009) reported findings that value stocks are more affected than growth stocks, supporting Baker and Wurgler in finding both significant. Applying different measures of sentiment, the results on whether value or growth stocks are most prone to sentiment differ.

Several tests will be conducted to determine whether any of the prior results of sentiment effects within the cross-section of stocks are transferable to the Danish equity markets. As previously explained, this analysis will be conducted using three different market segments: The broad market, large-mid cap stocks, and small-cap stocks. The variables included in this analysis are defined in section 5.4.1 and 5.4.2, summarized in Table 9. The analysis is split into three sections for a more thorough analysis, taking different stock characteristics into account when discussing the results.

The initial analysis concerns the impact of FEARS on the broad market, that is, the $\Delta BMIvalue$ and $\Delta BMIgrowth$. The following regressions are thus run:

$$\Delta BMIvalue_{t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m^n \gamma_m Control_t^m + u_{t+k} \quad (14)$$

$$\Delta BMIgrowth_{t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m^n \gamma_m Control_t^m + u_{t+k} \quad (15)$$

Where the $\Delta BMIvalue_{t+k}$ denotes the $\Delta BMIvalue$ return on day $t+k$. And the $\Delta BMIgrowth_{t+k}$ denotes the $\Delta BMIgrowth$ returns on day $t+k$. Control variables ($Control_{i,t}^m$) are same as previously explained which include CESI-EUR, EPU-US and four lagged variables of returns.

The next step regards the large and mid-cap stocks, that is, the $LMvalue$ and $LMgrowth$. We thus run the following regressions:

$$\Delta LMvalue_{t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m^n \gamma_m Control_{i,t}^m + u_{t+k} \tag{16}$$

$$\Delta LMgrowth_{t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m^n \gamma_m Control_{i,t}^m + u_{t+k} \tag{17}$$

Where the $\Delta LMvalue_{t+k}$ denotes the $\Delta LMvalue$ return on day $t+k$. And the $\Delta LMgrowth_{t+k}$ denotes the $\Delta LMgrowth$ returns on day $t+k$. Control variables ($Control_{i,t}^m$) are same as previously explained which include CESI-EUR, EPU-US and four lagged variables of returns.

Finally, we test the impact of FEARS on the small-cap stocks of Nasdaq Copenhagen, that is, the $\Delta SMLvalue$ and $\Delta SMLgrowth$. The following regressions are thus run:

$$\Delta SMLvalue_{t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m^n \gamma_m Control_{i,t}^m + u_{t+k} \tag{18}$$

$$\Delta SMLgrowth_{t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m^n \gamma_m Control_{i,t}^m + u_{t+k} \tag{19}$$

Where the $\Delta SMLvalue_{t+k}$ denotes the $\Delta SMLvalue$ return on day $t+k$. And the $\Delta SMLgrowth_{t+k}$ denotes the $\Delta SMLgrowth$ returns on day $t+k$. Control variables ($Control_{i,t}^m$) are same as previously explained which include the CESI-EUR, EPU-US, VSTOXX and four lagged variables of stock returns.

6.3 GRANGER CAUSALITY

We contribute to the investor sentiment literature by performing a Granger Causality test of the FEARS on stock market returns. The aim is to determine whether the FEARS index is useful in forecasting stock market returns, adding weight to the $DKallshare$ analysis explained in section 6.1.

The intuitive explanation of the relationship between the FEARS and stock market returns is built on the expectation that our FEARS index is successful in measuring investor sentiment and proxying for noise trader activity. An alternative explanation is that stock market activity affects the contemporaneous investor sentiment measure, by increasing the popularity of certain search terms which are included in our index. By running a Granger Causality test, the direction of causality may be determined, possibly allowing for a forecasting application of our FEARS index.

A Granger Causality test is statistical hypothesis test, with the aim of determining whether the precedence and information of one time series, is useful in predicting another time series beyond the prediction powers of the dependent variables' own lagged values (Cromwell, Hannan, Labys & Terraza, 1994). If successful, it can be stated that our FEARS "Granger-Causes" stock market return. It should be noted however, that this statement should not be interpreted as "stock market returns are a result of the FEARS index", as Granger Causality does not test for causality in the common term usage, that is, as cause and effect. The Granger Causality Test measures the information and precedence of one time series over another. (Cromwell et al. 1994)

Specifically, Granger Causality reflects the extent to which the lag process in one variable explains the current values of another variable (Cromwell et al. 1994). The Granger Causality test is conducted by estimating an autoregression of the dependent variable ($DKallshare$[15]) with l lags, augmented by including lagged values of our independent variable (the FEARS index) with l number of lags. The following equations are estimated:

$$DKallshare_t = \alpha_1 DKallshare_{t-1} \dots \alpha_t DKallshare_{t-l} + \epsilon_t \quad (R) \tag{20}$$

$$DKallshare_t = \alpha_1 DKallshare_{t-1} \dots \alpha_t DKallshare_{t-l} + \beta_1 FEARS_{t-1} \dots \beta_l FEARS_{t-l} + \epsilon_t \; (UR) \tag{21}$$

The aim of a Granger Causality test is to test whether the lagged values of the FEARS index are helpful in forecasting $DKallshare$ beyond the usefulness of the lagged values of $DKallshare$. The joint hypothesis is therefore:

$$\beta_1 = \beta_2 = \dots = \beta_l = 0 \tag{22}$$

That is, our null hypothesis is that the FEARS does not Granger-Cause Returns. The hypothesis is rejected if the beta-coefficients are statistically significantly different from zero, by measure of the p-value of the F-statistic. If the null hypothesis is rejected, it is concluded that FEARS Granger-Causes Returns.

In order to determine if nature and direction of the relationship between stock market returns and the FEARS index, we repeat the same procedure in the opposite direction, that is, the following equations are estimated.

$$FEARS_t = \alpha_1 FEARS_{t-1} \dots \alpha_l FEARS_{t-l} + \epsilon_t \qquad (R) \tag{23}$$

$$FEARS_t = \alpha_1 FEARS_{t-1} \dots \alpha_l FEARS_{t-l} + \lambda_1 DKallshare_{t-1} \dots \lambda_l DKallshare_{t-l} + \epsilon_t \quad (UR) \tag{24}$$

And the joint hypothesis is:

$$\lambda_1 = \lambda_2 = \dots = \lambda_l = 0 \tag{25}$$

[15] The dependent variable is the $\Delta DKallshare$, shortened to fit the equations.

That is, the null hypothesis is that stock market returns do not Granger-Cause the FEARS index. If we find the "causal" relationship to be bi-directional, meaning that both time series are helpful in forecasting each other, the practical application areas of the FEARS is reduced.

The test statistic is calculated by the following F-statistic formula:

$$F = \frac{(ESS_R - ESS_{UR})/q}{ESS_{UR}/(n-k)} \tag{26}$$

Where ESS_R and ESS_{UR} are the error sum of squares for the Equation 23 and Equation 24, which is the restricted (R) and the unrestricted (UR) equation respectively. q refers to the number of restrictions applied, n is the number of observations and k the number of parameters in the unrestricted model (including the constant) (Cornwell et al. 1994). We accept the alternative hypothesis when the calculated F-statistic is greater than the critical value:

$$F > F_{q,n-k} \tag{27}$$

Where $F_{q,n-k}$ is the critical value, where the statistic is distributed as an F with q degrees of freedom in the numerator and $n - k$ degrees of freedom in the denominator (Cornwell et al. 1994). We run the Granger tests for lags l up to five lags to maintain simplicity in the model, reducing the probability for spurious results.

6.4 BONDS

Although several researchers have transferred the well-documented sentiment effects of equity markets to the bond market, the evidence regarding bonds is scarcer. Stocks and bonds have common risk factors (Fama and French, 1993) and are both joint claims on the assets of a firm (Merton, 1974). However, they also differ in that bonds have guaranteed cashflows, institutional holdings and limited liquidity (Nayak, 2010). Prior literature suggests that one should expect some effect of sentiment on bond markets, following the results on equity markets, as there are well-documented information and thus sentiment spill overs between the equity and bond markets (Kwan 1996; Downing et. al 2009; Bethke et al. 2017), in addition to a high documented correlation between the two markets (Collin-Dufresne et.al, 2001; Fang et al, 2018). Moreover, the bond market is also suggested to co-move with the world markets based on news and investor sentiment (Barberis, Shleifer and Wurgler (2005)). Thus, the following analysis has the scope of testing whether the FEARS index has predictive power in the Danish Government and Corporate bond markets.

6.4.1 Government bonds

As various previous studies focus on the sentiment impact on US bond markets (Laborda and Olmo 2013; Fang et al. 2018), we find it interesting to test whether our measure of sentiment following the methods of Da et.al (2015), has some predictability of the Danish bond market. To test this, we run the following regressions:

$$\Delta 5Ybond_{t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m^n \gamma_m Control_{i,t}^m + u_{t+k} \tag{28}$$

$$\Delta 10Ybond_{t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m^n \gamma_m Control_{i,t}^m + u_{t+k} \tag{29}$$

Where the $\Delta 5Ybond_{t+k}$ denotes the $\Delta 5Ybond$ return on day $t+k$. And the $\Delta 10Ybond_{t+k}$ denotes the $\Delta 10Ybond$ return on day $t+k$. Control variables ($Control_{i,t}^m$) are same as previously explained which include CESI-EUR, EPU-US, VSTOXX and four lagged variables of the dependent variable.

6.4.2 Corporate Bonds

The paucity of research regarding the sentiment effects on corporate bond markets makes the potential findings of this paper harder to interpret. However, some studies suggest that the sentiment effect on equity markets has some spill over effect on corporate bonds. As suggested by Huang et al. (2015) equity market sentiment can affect corporate bond valuations directly through the behaviour of corporate bond investors, or indirectly through the activity of dedicated arbitrageurs who take advantage of the large deviation between equity and credit valuations (Huang et al (2015). The results of Nayak (2010) and (Rossi et al, 2008) show that there is substantial predictive power in the cross-section of corporate bonds, with the impact being strongest for the speculative-grade bonds. These empirical results give us motivation to test whether the impact of investor sentiment also hold for the Danish corporate bonds. The following regression is thus the basis for the following analysis:

$$\Delta IGbond_{t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m^n \gamma_m Control_{i,t}^m + u_{t+k} \tag{30}$$

Where the $\Delta IGbond_{t+k}$ denotes the $\Delta IGbond$ price on day $t+k$. Control variables ($Control_{i,t}^m$) are same as previously explained which include CESI-EUR, EPU-US, VSTOXX and four lagged variables of $\Delta IGbond_t$.

6.5 VOLATILITY

As mentioned in the literary review, Da et. al (2015) tests their FEARS index on three volatility measures in the United States. Realized Volatility of the S&P 500 ETF, the VIX index, which calculates implied volatility from option-data, and lastly, contemporaneous and futures movements of VIX Futures. For the Danish market, we are only able to replicate the study on realized volatility, using available daily frequency realized volatility data, calculated using the methodology of Anderslev. et al. (2001) and Barndorff-Nielsen et al. (2008), as well as Da. et al. (2015), as explained in section 5.4.3

An important feature of the noise trading theory, and implication of the noise trader model of De Long et al. (1990), is the role of investor sentiment on volatility. As presented in the literary review, the implication of uninformed noise trades, deviating the stock market prices from its fundamental values, should be increased volatility as well. Thus, we expect a positive relationship between our FEARS index and volatility. In the following, we explain the empirical strategy in testing this relationship.

We adjust for the longevity-traits of volatility. In time series analysis, "[a] traditional assumption is that the coupling between values at different time instants decreases rapidly as the time difference or distance increases" (Liu, Chen, and Zhang, 2017). However, time series data may exhibit a phenomenon called long memory or long-range persistence. Long-range persistence in times series indicates that the autocorrelation function (ACF) is algebraic in its decay, and thus slower than exponential decay. The implication is that the area under the function curve is infinite (Engle and Patton, 2001). Da et al. (2015) also acknowledges the long persistence properties of volatility time series. This implies that volatility shocks will influence the expectation of volatility for many periods (Engle and Patton, 2001). Engle and Patton (2001) precisely define volatility persistence, where we let expected value of the variance of return k periods in the future as

$$h_{t+k|t} \equiv E_t[(r_{t+k} - m_{t+k})^2] \tag{31}$$

The ARFIMA process is one of the most recognized classes of long-memory models (Liu et. al (2017). The ARFIMA (p, d, q) processes are widely used in modeling long-memory time series (Liu et. al 2017). As in Da et. al (2015), we adjust for the volatility persistence exhibited by modelling the long-range dependence through a fractional integrated autoregressive moving average model, ARFIMA.

$$(1 - L)^d \left(adj.rv_t - \beta_1 FEARS_t - \textstyle\sum_m \beta_m Control_{i,t}^m\right) = (1 - L)\varepsilon^t \tag{32}$$

where the left-hand side denotes the fractional AR (autoregressive) specification of the data, the right-hand side denotes the MA (moving average) specification on the residuals, and the $(1 - L)^d$ being the long-memory fractional process. In the fractional integrated model, the d is modelled between -0.5 and 0.5.

Using a fractional Integration parameter relaxes the assumption in ARMA and ARIMA models that the time series must be fully stationary or integrated. In an ARIMA model with stationarity, i.e., $d = 0$, the model only allows for short-memory or short persistence of exogenous shock effects before the data reverts to the mean. For fully integrated data, i.e., $d = 1$, exogeneous shocks affect the data for the full period, and there is no mean reversion. The fractionally integrated processes are characterized by the long memory, but the effect of an exogeneous shock does not persist infinitely (Box-Steffensmeier, Freeman, Hitt, & Pevehouse, 2014).

We aim to estimate the parameters of p, and q – the autoregressive and moving average specifications, by use of the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The fractional integration parameter is determined by the computer program and selected on log likelihood and AIC selection criteria.

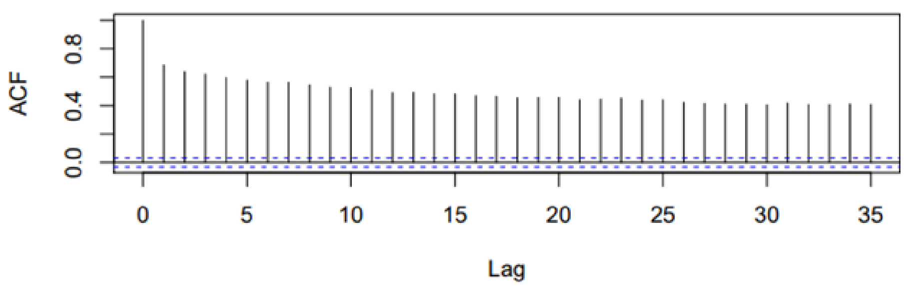

Figure 20: depicts the Autocorrelation Function of Realized Volatility

From the Autocorrelation function (ACF), we find a result that is quite hard to interpret. Firstly, we can interpret the time series as not being fully stationary. This is to be expected, as we see that the ACF of the Adjusted Realized Volatility time series does not exhibit an exponential decay, and thus may exhibit long-memory. This is the reason why we choose to model the time series as an ARFIMA. It does however make the parameter selection harder. The ACF shows similarities with results we would expect from non-stationary time series with trend, or non-stationary random walk time series, except that the spikes drop quite significantly at the first lag.

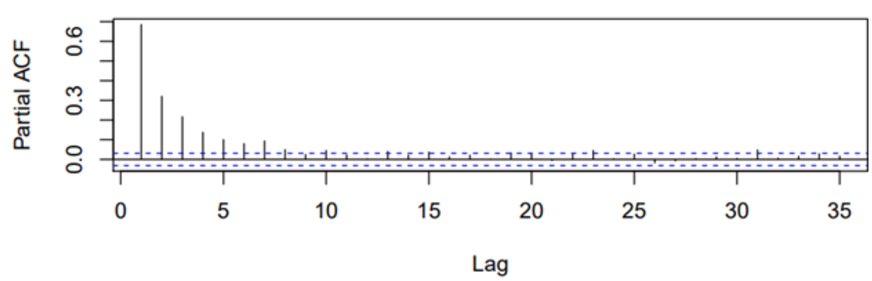

Figure 21: depicts the Partial Autocorrelation Function of Adjusted Realized Volatility

The Partial Autocorrelation Function (PACF) is depicted in Figure 21. This shows that there is a partial autocorrelation in the time series back to 8 lags. In co-interpretation with the ACF, this also looks like a random walk process, or a non-stationary trend process. For a steady interpretation, we allow the software to simulate a good fit for the model.

We start by testing our FEARS index on the seasonally adjusted realized volatility. The program determines the model ARFIMA (2,d,1) to be a good fit, where the fractional integration parameter d is between -0.5 and 0.5. Our FEARS index, the CEI-EUR and EPU-US are included as external regressors. Due to the uncertainty in determining p- and q- parameters, we perform a detailed robustness check where we study the results given different several different parameter selections.

For volatility, we run the test on four different time-intervals: Full period, before the Financial Crisis, during the Financial Crisis, and after the Financial Crisis. The Financial Crisis was a period of great turmoil, which is not representative for the full period sample. One of several implications of the turmoil was the presence of structural breaks in realized volatility. For the S&P 500 realized volatility, Song and Baek (2019) found three structural breaks coinciding with three waves of the Financial Crisis. We aim to test for what periods our investor sentiment index successfully captures volatility development.

6.6 ECONOMETRIC CONCERNS

The time series data might reveal some concerns in our variables defined in section 5, as the data might exhibit undesirable characteristics that could potentially produce unreliable and spurious results, which would make the interpretation and understanding of our future results difficult. Thus, the following section highlight some econometric concerns working with the variables. Although all data included in the construction of each variable is previously adjusted, tests are run to check whether all variables are ready for analysis. The results from the following tests will be reported if results if they yield interesting findings.

For the regression analysis to produce reliable results, the time series data must be stationary. That is, the mean, variance and autocorrelation should be constant over time (Hill, Griffiths and Lim, 2011). Thus, to test whether our variables are stationary, we first run an Augmented Dickey fuller (ADF) test (Stock and Watson, 2014 p. 603). The results produced from the ADF test contain no evidence of apparent unit-root in our variables, suggesting that all variables are stationary, except for realized volatility, which exhibits long memory, and will be fractionally integrated in the tests where it is included.

Furthermore, we test for the homoscedastic assumption in our variables. If the variables do in fact exhibit heteroscedasticity, i.e., the variance not being constant over time, there could be a loss in efficiency in applying the OLS regression, and the biases in estimated standard errors may lead to invalid conclusions (Breusch and Pagan, 1979). To test for apparent heteroscedasticity in our variables, we run the Breusch Pagan (BP) test (Breusch and Pagan, 1979). Although there is little evidence of heteroscedasticity in the results yielded from the BP test, we assume some heteroscedasticity which will be accounted for in the standard errors.

Moreover, we want to identify any possible serial correlation problem in our timeseries variables. That is, the regression errors are autocorrelated (Stock and Watson, 2014 p. 375). To do this, we apply the Durbin Watson (DW) test (Durbin &Watson, 1951). This test allows us to identify any serial correlation problem in our variables. Following the DW test, serial correlation does not seem to be a problem.

Nevertheless, we find it safe to apply HAC[16] standard errors (Stock and Watson, 2014 p. 646) in all OLS regressions. Even if there is no heteroscedastic or serial correlation problems in our variables, this will provide robustness to the results, as the standard errors will become conventional standard errors if there is no heteroscedasticity or serial correlation to be found. [17]

[16] Heteroscedasticity and Autocorrelation-Consistent

[17] Regressions are tested without the HAC standard errors and the results not altered

7 RESULTS

In the following section, the results yielded from our analysis will be presented. This section serves the purpose of providing a clear overview of the findings from our empirical method (section 6), which will in section 9 be discussed. Raw output of each regression can be found in the Appendix[18]. Recall from section 6, that the coefficients for our independent variables will be interpreted as basis points increase in returns by one standard deviation increase in FEARS. Moreover, all tests are conducted using the cleaned variables[19] as explained in section 6. However, in the following section, as well as in the discussion (9), we will address the variables as originally presented[20] in section 5, to simplify.

7.1 AVERAGE RETURNS

Table 13 depicts the results for our average returns regressions (equation 13). The results are reported for tests up to lead day $T + 5$. We find a positive contemporaneous relationship between the FEARS index and the $DKallshare$ returns, significant at the 1% significance level. As our index is a positive sentiment index, we expected a positive relationship before any analysis. The relationship corresponds to a 4.7 basis points increase in the contemporaneous $DKallshare$ for one standard deviation increase of FEARS. Additionally, we can observe that equally significant is the stock market one-day lag returns, and the VSTOXX in explaining contemporaneous returns. The US Economic Policy Uncertainty index is significant at the 10% significant level and has a negative sign. Both VSTOXX and the EPU-US have a negative sign on their coefficient, aligned with what we expected.

From column (2), we can observe a negative coefficient-sign on the FEARS variable, indicating a return reversal. This result is significant at the 5% significance level. The reversal corresponds to a 4.7 basis point decrease in the first lead following one standard deviation increase in the FEARS index. Thus, the contemporaneous upswing is short-lived, as the effect of FEARS is completely reversed within the next day. In addition, following the expectation when regressing on future stock returns, our adjusted R-squared has dropped significantly, from 31.3% to 0.3%, indicating that our regression model struggles in explaining the variance of the lead stock returns.

An interesting finding can be observed in the third (3) column. After the completed reversal in $T + 1$ (2), we observe another spike, equal to the contemporaneous positive effect. The net effect over the

[18] Starting in Appendix 3
[19] For example, “$\Delta DKallshare$" not “$DKallshare$"
[20] For example, “$BMIvalue$" not $\Delta BMIvalue$"

contemporaneous results, and the two first lead days, is an increase of 4.7 basis points, with a one standard deviation increase in the FEARS index.

In Table 13 we also study the effects of sentiment up to the fifth lead day, $DKallshare\, T+5$. However, there is no significance in the coefficients after T+2, and the point estimates are small and economically negligible.

Moreover, two of the control variables included report significant relationships with contemporaneous stock market returns. In Table 13, column (1), we can observe that the 1-day return lag is significant at the 1% level. This implies that past values have explanatory power of present values, which is also why we included the control variable. Additionally, in both columns (1) and (2), we can observe that our control variable VSTOXX is significant at the 1% and 5% levels respectively. This result indicates that the implied volatility calculated from Euro STOXX call and put options not only have a significant contemporaneous relationship with Danish stock market returns, but also short-term future values. In both regressions, we find negative coefficients, in line with expectation. The fact that VSTOXX is significant at the 5% level also indicates the direction of the causality, namely that VSTOXX can influence stock prices in the short-term future.

The results further report significant coefficients on the control variables for lead-day return beyond that of the second day lead. We can observe a positive coefficient on the EPU-US variable in regression (4), either indicating that there is a three-day lag before a movement in the EPU-US affects stock returns, or this is a spurious result that is hard to interpret. This argumentation also holds for the significant VSTOXX coefficients in regressions (5) and (6).

Panel A: FEARS and DKallshare

	DKallshare					
	(1) DKallshare	(2) DKallshare T+1	(3) DKallshare T+2	(4) DKallshare T+3	(5) DKallshare T+4	(6) DKallshare T+5
FEARS	0.0020***	-0.0020**	0.0020**	-0.0005	0.0003	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
DKallshare Lag 1	0.0470***	-0.005	-0.022	-0.012	-0.034	-0.033
	(0.016)	(0.021)	(0.020)	(0.022)	(0.020)	(0.018)
DKallshare Lag 2	-0.05	-0.021	-0.012	-0.031	-0.032	-0.007
	(0.016)	(0.020)	(0.023)	(0.020)	(0.019)	(0.020)
DKallshare Lag 3	-0.011	-0.011	-0.032	-0.031	-0.007	-0.026
	(0.016)	(0.020)	(0.023)	(0.019)	(0.019)	(0.020)
DKallshare Lag 4	-0.008	-0.032	-0.0210*	-0.007	-0.028	0.003
	(0.018)	(0.019)	(0.018)	(0.019)	(0.020)	(0.019)
VSTOXX	-0.0840***	-0.0050**	0.004	0.004	0.0050*	0.0050**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
EPU-US	-0.0004	0.0001	0.0000	0.0010***	-0.0002	-0.0003
	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
CESI-EUR	-0.004	-0.001	0.005	0.003	0.001	0.002
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Observations	4311	4311	4311	4311	4311	4311
Adjusted R-Squared	0.313	0.003	0.003	0.002	0.002	0.002
F-Statistic	247.645**	2.568**	2.982**	2.548**	2.152**	1.922**

*Table 13: The table depicts the relation between Nasdaq OMX Copenhagen (DKallshare) index daily returns to our FEARS index over the sample period 2004-2022. The dependent variables are contemporaneous returns (column (1)), future index daily returns in the five next lead days (2)-(6). The FEARS index is the independent variable. The control variable included are the lagged returns up to four lags, the European volatility index Euro Stoxx 50 Volatility (VSTOXX), changes in a news-based measure of economic policy uncertainty (EPU-U.S), , and changes in the Citigroup Economic Surprise (CESI-EUR) index. The standard errors applied are the robust standard errors constructed in section 6.6. ***,**,* denote significance at 1%, 5% and 10% levels, respectively.*

7.2 THE CROSS-SECTION OF RETURNS

Recall the review in section 4.4.1, where the effect of sentiment on asset prices is shown to have varying predictive power within the cross-section of returns. In this section examine the sentiment effect on the cross-section of stocks, namely on value and growth stocks. In section 5.1.2 we established that this analysis will be divided into three distinct categories of stocks, namely LargeMid-Cap stocks, Small-Cap stocks, and the Broad Market stocks. Within these three categories, indices for both value stocks and growth stocks are utilized, each for every category. As the cross-sections of returns are divided into three, the results will be presented in three different sections. The initial results concern the Broad Market indices, followed by the Large and Mid Cap, and finally the results for the Small-cap indices will be presented.

7.2.1 Broad Market Results

Table 14 below depict the results of the analysis using equation 14 and 15. Panel A and B shows the results for BMIgrowth and BMIvalue, respectively.

From Table 14, Panel A, column (1), we observe a positive contemporaneous relationship between FEARS and BMIgrowth. The coefficient is significant at the 5% level. Specifically, the relationship indicates a 4.7 basis point increase in BMIgrowth for a one standard deviation increase in the FEARS index. Moreover, we can observe an adjusted R^2 of 0.243, indicating that 24.3% of the variance in BMIgrowth can be explained by the FEARS, along with the control variables. Moreover, several of the control variables specified in the first regression (1) also exhibit a significant relationship with the contemporaneous movements of broad market returns. The first lag of BMIgrowth has a positive coefficient of 0.067 significant at the 1% level, which has both a higher coefficient and a higher significance than the FEARS variable. Additionally, the VSTOXX and EPU-US are both significant at the 1% level, with negative coefficients in line with expectations. The findings indicate that the control variables are the primary indicators of stock movements, exhibiting higher explanatory power than FEARS.

The short-term future values of BMIgrowth follow the noise trading patterns observed in section 7.1 of FEARS on Average Returns. In (2) we observe a complete reversal of the initial contemporaneous spike, before a subsequent positive spike in (3). The adjusted R^2 is greatly reduced to a level of 0.002 (2) and 0.003 (4), relating to the difficult nature of forecasting returns. However, the VSTOXX control variable still maintains an explanatory power in (2) at the 10% significance level.

Panel A: FEARS and BMI Growth

	BMI Growth Stocks		
	(1)	(2)	(3)
	BMIgrowth	BMIgrowth T+1	BMIgrowth T + 2
FEARS	0.002**	-0.002**	0.002**
	(0.001)	(0.001)	(0.001)
Lag 1	0.067***	-0.010	-0.028
Lag 2	-0.013	-0.025	0.026
Lag 3	-0.008	0.023	-0.032
Lag 4	0.028	-0.032	-0.049*
VSTOXX	-0.094***	-0.007*	0.005
EPU US	-0.001***	0.0001	0.000
CESI EUR	-0.005	-0.003	0.002
Adjusted R2	0.243	0.002	0.003
F-statistic	170.023***	2.070**	2.747***

Panel B: FEARS and BMI Value

	BMI Value Stocks		
	(1)	(2)	(3)
	BMIvalue	BMIvalue T+1	BMIvalue T + 2
FEARS	0.002***	-0.003***	0.002**
	(0.001)	(0.001)	(0.001)
Lag 1	0.088***	-0.003	-0.017
Lag 2	-0.006	-0.015	-0.013
Lag 3	0.004	-0.013	-0.029
Lag 4	-0.009	-0.028	-0.033
VSTOXX	-0.110***	-0.012***	0.003
EPU US	-0.001*	-0.0002	0.0003
CESI EUR	0.0003	0.007*	0.008
Adjusted R2	0.344	0.00600	0.002
F-statistic	276.838***	4.085***	2.163**

Table 14: The table depicts the relation between S&P Denmark BMI indices daily returns to our FEARS index over the sample period 2004-2022. Panel A relates FEARS to the $BMIgrowth$*, whilst Panel B relates FEARS to* $BMIvalue$*. The dependent variables are contemporaneous returns (column (1)), future index daily returns in the next two days (columns (2) and (3), respectively). The FEARS index is the independent variable. The control variable included are the lagged returns of each index up to four lags, the European volatility index Euro Stoxx 50 Volatility (VSTOXX), changes in a news-based measure of economic policy uncertainty (EPU-U.S), and changes in the Citigroup Economic Surprise (CESI-EUR) index. The standard errors applied are the robust standard errors constructed in section 6.6. ***,**,* denote significance at 1%, 5% and 10% levels, respectively.*

In Table 14, Panel B, a similar relationship with FEARS can be observed, that is, between FEARS and BMIValue. However, comparing the two panels we find that Panel B confirms a higher significance level on the FEARS variable, and a higher adjusted R^2 overall. The contemporaneous relationship between FEARS and BMIvalue (1) is measured as significant at the 1% level. Specifically, we observe a 4.7 basis point increase for one standard deviation increase. A negative relationship with VSTOXX can also be observed, significant at the 1% level, together with a negative coefficient for the EPU-US variable, significant at the 10% level. The adjusted R^2 is at 0.344, revealing that the specified model explains 34.4% of the variance in BMIvalue.

The lead variables of BMIvalue reveal the noise trader pattern observed in section 7.1. However, the first-day lead (2) indicates a reversal stronger than the initial contemporaneous effect, relating to a 7 basis point decrease for one standard deviation increase in the FEARS variable. The VSTOXX variable is still significant at the 1% level, indicating a robust relationship with one-day lead returns. We can also observe that the adjusted R^2 is reduced dramatically, which follows the pattern we have observed in earlier tests. The second-day lead (3) indicates a new upwards spike following the reversal in (2). Specifically, a 4.7 basis point increase following a one standard deviation increase in the day T FEARS.

7.2.2 Large Mid-cap Results

Table 15 below depict the results of regression explained in equations 16 and 17, on both LMgrowth and LMvalue, in panel A and B, respectively.

Similar to the results of Average and Broad Market returns, Table 15, Panel A reveals a positive contemporaneous relationship between the FEARS and LMgrowth (1). The coefficient is significant at the 5% level, corresponding a 4.7 basis point increase in contemporaneous LMgrowth for a one standard deviation increase in FEARS. Also, the first lag of LMgrowth holds a significant coefficient at the 5% level, revealing the relationship between contemporaneous and the last-day return movements. In line with expectation and earlier findings, both VSTOXX and EPU-US holds a negative coefficient significant at the 1% level. In regression (1), FEARS is thus not the most significant or economically important variable. Lastly, we can observe an adjusted R^2 of 0.216, revealing that 21.6% of the variance in LMgrowth can be explained by the FEARS and the control variables.

Regression (2) and (3) reveal the relationship between FEARS on day T with lead day 1 $(T+1)$ and lead day 2 $(T+2)$ results. We can observe the same noise trading pattern, following our expectations from earlier results. Specifically, we see a full reversal in (2), that is, a 4.7 basis point decrease for one standard deviation increase in FEARS, subsequently following a new upwards spike in (3). The VSTOXX and EPU variables exhibit no significant result with short-term future results om LMgrowth. Additionally, the adjusted R^2 falls dramatically after the first regression (1).

In Table 15, Panel B, the first regression (1) reveals a positive contemporaneous relationship between FEARS and LMvalue, corresponding to a 4.7 basis point increase for one standard deviation of FEARS.

This result is significant at the 1% level. Also, here we find that the first lag has a significant relationship and a higher, positive coefficient than that of the FEARS. Additionally, the VSTOXX and EPU-US follow our expectations and return significant negative coefficients. The model corresponds to an adjusted R^2 of 0.318, revealing that 31.8% of the variation in LMvalue can be explained by regression (1).

Panel A: FEARS and LM Growth

	LM Growth Stocks		
	(1)	(2)	(3)
	LMgrowth	LMgrowth T+1	LMgrowth T + 2
FEARS	0.002**	-0.002*	0.002**
	(0.001)	(0.001)	(0.001)
Lag 1	0.055**	-0.018	-0.033
Lag 2	-0.0210	-0.030	0.029
Lag 3	-0.0140	0.026	-0.032
Lag 4	0.0300	-0.032	-0.052*
VSTOXX	-0.093***	-0.006	0.006
EPU US	-0.001***	0.0001	0.000
CESI EUR	-0.0050	-0.003	0.001
Adjusted R2	0.216	0.002	0.003
F-statistic	146.292***	1.874*	2.792***

Panel B: FEARS and LM Value

	LM Value Stocks		
	(1)	(2)	(3)
	LMvalue	LMvalue T+1	LMvalue T + 2
FEARS	0.002***	-0.003***	0.002*
	(0.001)	(0.001)	(0.001)
Lag 1	0.075***	-0.013	-0.026
Lag 2	-0.0160	-0.022	-0.013
Lag 3	-0.0030	-0.014	-0.030
Lag 4	-0.0100	-0.029	-0.034
VSTOXX	-0.111***	-0.011***	0.004
EPU US	-0.001**	-0.0002	0.0003
CESI EUR	0.0030	0.008*	0.009
Adjusted R2	0.318	0.005	0.002
F-statistic	245.840***	3.860***	2.241**

*Table 15: The table depicts the relation between S&P Denmark LargeMidCap indices daily returns to our FEARS index over the sample period 2004-2022. Panel A relates FEARS to the $LMgrowth$, whilst Panel B relates FEARS to $LMValue$. The dependent variables are contemporaneous returns (column (1)), future index daily returns in the next two days (columns (2) and (3), respectively). The FEARS index is the independent variable. The control variable included are the lagged returns of each index up to four lags, the European volatility index Euro Stoxx 50 Volatility (VSTOXX), changes in a news-based measure of economic policy uncertainty (EPU-U.S), and changes in the Citigroup Economic Surprise (CESI-EUR) index. The standard errors applied are the robust standard errors constructed in section 6.6. ***,**,* denote significance at 1%, 5% and 10% levels, respectively.*

The FEARS also exhibits significant coefficients in the lead returns of LMvalue. In the first lead day ($T + 1$) regression (2), we observe a reversal larger than the initial spike in (1). Specifically, a one standard deviation increases in the day T FEARS corresponds to a 7 basis point decrease in the LMvalue T+1 returns. The VSTOXX-variable still has a negative relationship at the 1% significance level. In the second lead day (3), we find a subsequent spike following the reversal in (2), as we expect from our earlier findings. Lastly, we again observe that the adjusted R^2 falls dramatically for the lead value of returns.

7.2.3 Small-cap Results

Table 16 below depict the results from equations 18 and 19, on SMLgrowth and SMLvalue, in panel A and B, respectively.

Contrary to the pattern of our earlier findings, Table 16 Panel A does not report a significant contemporaneous result between the FEARS index and SMLgrowth (1). The day T SMLgrowth is mostly explained by its first lag, and the VSTOXX control variable, both significant at the 1% level. The results of regression (1) depict a positive relationship with the first lag, and a negative relationship with VSTOXX, in line with expectations and earlier results. Furthermore, we can observe an adjusted R^2 of 0.288, revealing that 28.8 % of the variance in SMLgrowth can be explained by our regression.

For the short-term future return values, we can observe the same pattern as earlier. We see a negative coefficient of the FEARS in regression (2) and positive in regression (3). Specifically, a one standard deviation increases in the day T FEARS corresponds to a 4.7 basis point decrease in lead day 1 (2) and a 4.7 basis point subsequent increase in lead day 2 (3). The VSTOXX variable still holds a negative significant coefficient at the 1% level for the lead day 1 variable (2). Finally, we can again observe a sharp decrease in the adjusted R^2 as we progress from contemporaneous values of SMLgrowth to lead variables.

In Table 16 Panel B, the relationship between SMLvalue and FEARS reveals a familiar pattern. The FEARS has a positive coefficient at the 5% level, corresponding to a 2.3 basis point in SMLvalue following a one standard deviation increase in FEARS. Again, we see that both the positive first lag variable coefficient and the negative VSTOXX coefficient are significant at the 1% level. Additionally, the adjusted R^2 is at 0.33, revealing that 33% of the variation in SMLvalue can be explained by regression (1).

Panel A: FEARS and SML Growth

	SML Growth Stocks		
	(1)	(2)	(3)
	SMLgrowth	SMLgrowth T+1	SMLgrowth T + 2
FEARS	0.001	-0.002***	0.002**
	(0.001)	(0.001)	(0.001)
Lag 1	0.144***	0.038	0.003
Lag 2	0.0330	0.004	0.007
Lag 3	0.0220	0.006	-0.035
Lag 4	0.0120	-0.034	-0.025
VSTOXX	-0.097***	-0.013***	0.003
EPU US	-0.0002	-0.0001	0.000
CESI EUR	-0.0070	-0.002	0.004
Adjusted R2	0.288	0.007	0.001
F-statistic	213.308***	4.627***	1.671

Panel B: FEARS and SML Value

	SML Value Stocks		
	(1)	(2)	(3)
	SMLvalue	SMLvalue T+1	SMLvalue T + 2
FEARS	0.001**	-0.002***	0.002***
	(0.001)	(0.001)	(0.001)
Lag 1	0.141***	0.042	0.015
Lag 2	0.0380	0.013	-0.008
Lag 3	0.0310	-0.008	-0.015
Lag 4	-0.0010	-0.013	-0.011
VSTOXX	-0.103***	-0.015***	-0.001
EPU US	-0.0002	-0.0003	0.00004
CESI EUR	-0.0040	0.006*	0.006
Adjusted R2	0.330	0.008	0.001
F-statistic	260.586***	5.506***	1.6600

*Table 16: The table depicts the relation between S&P Denmark Small-Cap indices daily returns to our FEARS index over the sample period 2004-2022. Panel A relates FEARS to the $SMLgrowth$, whilst Panel B relates FEARS to $SMLValue$. The dependent variables are contemporaneous returns (column (1)), future index daily returns in the next two days (columns (2) and (3), respectively). The FEARS index is the independent variable. The control variable included are the lagged returns of each index up to four lags, the European volatility index Euro Stoxx 50 Volatility (VSTOXX), changes in a news-based measure of economic policy uncertainty (EPU-U.S), and changes in the Citigroup Economic Surprise (CESI-EUR) index. The standard errors applied are the robust standard errors constructed in section 6.6. ***,**,* denote significance at 1%, 5% and 10% levels, respectively.*

Moreover, the short-term future values of SMLvalue yield the familiar pattern of the earlier results. In the first lead day SMLvalue T+1 (2), we find a significant reversal at the 1% level, corresponding to a 4.7 basis point decrease for a one standard deviation increase in the day T FEARS. Additionally, the positive second-day lead coefficient in observed in column (3) is significant at the 1% level, which has not been observed before. The result corresponds to a 4.7 basis point increase in the second lead day dependent variable following a one standard deviation increase in the day T FEARS. The VSTOXX variable exhibits a negative relationship at the 1% level in column (2) but loses its significance for the second lead day column (3). Lastly, we observe a notable drop in the adjusted R^2 as expected by earlier results.

7.3 GRANGER CAUSALITY

In Table 17, the results are reported for the Granger Causality test for lags from 1 to 5 lags. For all lags up to the fourth lag, the p-statistic is below 5%, reporting a significant F-statistic at the 5% significance level. For the fifth lag, the test returns a p-value of the F-statistic of 0.061, which is significant at the 10% level. Thus, we can reject the null hypothesis that the FEARS index does not Granger Cause Stock market returns, at the 5% percent level from the first to fourth lag, and at the 10% level for the fifth lag. Also, we are not able to reject the null hypothesis of the reversed test, that is, that Stock Market Returns does not Granger Cause the FEARS index, at any significance level reported.

Lags	**1**		**2**		**3**		**4**		**5**	
Null Hypothesis	F-Statistic	Pr(>F)	F-Statistic	Pr(>F)	F-Statistic	Pr(>F)	F-Statistic	Pr(>F)	F-Statistic	Pr(>F)
FEARS does not Granger Cause Returns	6.485**	0.011**	4.621**	0.010**	3.106**	0.025**	2.397**	0.048**	2.113*	0.061*
Returns does not Granger Cause FEARS	1.0885	0.2969	0.614	0.5412	0.9201	0.4302	0.6097	0.6557	0.0683	0.6475

*Table 17: Results for 1st to 5th lag. Significance is denoted ** relates to significance of the 5% level, and * for the significance at the 10% level. Th null hypothesis is states in the left-most column.*

It is concluded that the FEARS index Granger Causes stock market returns. As discussed in Section 6.3, we conclude that this means the FEARS contains information that is useful in explaining returns. An important additional conclusion is that stock market returns do not Granger Cause the sentiment index. That is, the Granger Causality analysis reports a unidirectional Granger Causality.

7.4 BONDS

In section 4.4, we emphasized the widespan of sentiment measures previously applied towards the equity market. Although the evidence of the impact of sentiment is highly concentrated towards stocks, the application of these methods are flexible and not at all limited to stock returns. Following the empirical strategy explained in section 6.4.1, this section will yield results on our analysis towards the bond market, and more specifically, the previously defined variables: 5Ybond, 10Ybond and IGbond. Moreover, we want to emphasize that the data applied for these variables is return data for the two former variables, and price data for the latter. This is stated to make interpretation of the following results more intuitive, as bond returns and prices exhibit an inverse relationship (Bodie et al., 2014, s. 516). Thus, when interpreting the results, one should interpret the signs for the government bond variables opposite of the corporate bond variable. Table 18 depict the results for the bond market analysis. Panel A reports results on the prediction of FEARS on the 5Ybond and 10Ybond, while Panel B focuses on impact of the FEARS on IGbond.

7.4.1 Government Bonds

The results are shown in Table 18, below. From Panel A columns (1) we observe a positive coefficient of FEARS on contemporaneous 10Ybond returns. However, the coefficient is not significant and economically negligible. We can also observe that the lag values of 10Ybond have no significance in explaining current values. The adjusted R^2 for the regression is 0.063, which in turn means that the 6.3% of the variance 10Ybond can be explained by the model. Thus, the model has significantly less explanatory power than the equity returns regressions of section 7.1 and 7.2, indicating that the specification, that is, the FEARS and the control variables are less fit to model 10Ybond.

The same results hold true for the lead variables of 10Ybond, that is, regression (2) and (3). We observe a negative coefficient in one day lead returns (2), and a positive coefficient in second-day lead returns (3), but both coefficients are non-significant and economically negligible. In regression (3) we find a significant coefficient in the $T-4$ lag variable, indicating a negative relationship between returns and a distant lag value. We interpret this as a spurious relationship, which needs no further elaboration. Finally, we observe very low adjusted R^2 values for the lead regression, indication a poor fit for the model.

Regarding the 5Ybond, however, we see that the coefficients are positive for the both the contemporaneous and 1 and 2-day lead variables (columns 4, 5 and 6, Panel A, Table 18). Although the two former variables reveal no statistical significance, in column (6), we observe a significant positive relationship between FEARS and the second-day lead, that is, a one standard deviation increase in the FEARS corresponds to 4.7 basis points increase in 5Ybond T+2. For both regression specification (2) and (3) we find an adjusted R^2 of 0.1%, indicating a low fit of the model.

7.4.2 Corporate bonds

In Table 18, Panel B, we can observe a negative coefficient on the FEARS variable, indicating a negative relationship with FEARS and contemporaneous IGbond. The coefficient is, however, economically negligible, and not significant. The only significant variable is the three-day lag value of IGbond, which indicates a positive relationship between IGbond and its own past value. The adjusted R^2 is only 0.017, revealing that 1.7% of the variance in the dependent variable can be explained by the model (1).

In regression (2) we find a positive coefficient, whereas in regression (3) we find a negative coefficient of the FEARS variable. The FEARS is non-significant and economically negligible for both regressions. The IGbond T+2 (3) results have a significant coefficient in the 2-day lag variable, similar to the result of regression (1). The future short-term regressions also suffer from low model fit, indicated by an adjusted R^2 of 0.3% (2) and 0.4% (3). The low fit indicates that the model is not sufficiently specified to capture the variation in the dependent variable.

Panel A: FEARS and Government Bonds

	10-year Government Bond			5-year Government Bond		
	(1)	(2)	(3)	(4)	(5)	(6)
	10Y	10Y T + 1	10Y T + 2	5Y	5Y T+1	5Y T+2
FEARS	0.001	-0.0001	0.0003	0.001	0.00002	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Lag 1	0.024	0.016	-0.013	-0.032	0.004	-0.005
Lag 2	0.010	-0.013	0.031	-0.002	-0.005	0.016
Lag 3	-0.013	0.029	0.006	-0.006	0.015	0.017
Lag 4	0.027	0.006	-0.035*	0.015	0.018	-0.031
Other Controls	YES	YES	YES	YES	YES	YES
Adjusted R2	0.063	-0.00001	0.002	0.059	0.001	0.001

Panel B: FEARS and Corporate Bonds

	IG Bond		
	(1)	(2)	(3)
	IGbond	IGbond T+1	IGbond T+2
FEARS	-0.0001	0.00003	-0.00003
	(0.001)	(0.001)	(0.001)
Lag 1	0.007	0.025	0.043*
Lag 2	0.024	0.041	0.021
Lag 3	0.039*	0.023	-0.016
Lag 4	0.021	-0.017	0.008
Other Controls	YES	YES	YES
Adjusted R2	0.017	0.003	0.004

*Table 18: The table depicts the relation between the bond indices daily returns to our FEARS index over the sample period 2004-2022. Panel A relates FEARS to the 10 and 5-Year government bonds (column (1), (2) and (3), and (4), (5) and (6), respectively), whilst Panel B relates FEARS to IG Bonds (column (1), (2) and (3)). The dependent variables are contemporaneous returns (column (1) and (4)), future index daily returns in the next two days (columns (2), (3), (5) and (6), respectively). The FEARS index is the independent variable. The control variable included are the lagged returns of each bond index up to four lags, the European volatility index Euro Stoxx 50 Volatility (VSTOXX), changes in a news-based measure of economic policy uncertainty (EPU-U.S), , and changes in the Citigroup Economic Surprise (CESI-EUR) index. The standard errors applied are the robust standard errors constructed in section 6.6. ***,**,* denote significance at 1%, 5% and 10% levels, respectively.*

7.5 VOLATILITY

The following section highlights the results from our ARFIMA model constructed in section 6.5. The findings are reported in Table 19 below. In regression (2) we find a statistical relationship between FEARS and the contemporaneous Adjusted Realized Volatility movements. The result holds a p-value of 0.047 and is significant at the 5% confidence level. This regression relates to the period after the ending of the Financial Crisis of 2008, that is, starting at July 2009 and to 31.12.2021. The coefficient is positive and relates to a 90-basis points increase in Adjusted Realized volatility, for a one standard deviation increase in FEARS.

Panel A: ARFIMA (2,d,1) on seasonal-adjusted log realized volatility

	Jan 2005 - Dec 2021	July 2009 - Dec 2021	Jan 2005 - Dec 2007	Dec 2007 - June 2009
	(1)	(2)	(3)	(4)
p 1	0.8994***	0.8448***	-0.07784	-0.4663775
	(0.03792)	(0.03165)	(0.38328)	(0.614653)
p 2	0.058915 .	0.1426***	-0.0658	-0.10009
	(0.035185)	(0.02058)	(0.061685)	(0.106911)
q	0.9186***	0.9677***	0.037597	-0.2887554
	(0.0360316)	(0.04688)	(0.39749)	(0.6185467)
d	0.3437***	0.4992***	0.43608***	0.4823***
	(0.053507)	(0.0000)	(0.05427)	(0.023093)
FEARS	0.0203197	0.03875*	-0.02181	0.0064005
	(0.016837)	(0.01926)	(0.04591)	(0.0461628)
CESI-EUR	-0.1518*	-0.01620 '	-0.0417*	0.037015
	(0.0077)	(0.0091)	(0.0181)	(0.0257)
EPU-US	-0.02782	-0.3868***	0.670819	-0.49976 '
	(0.0087)	(0.0900)	(0.5447)	(0.2788)
Observations	3946	3035	535	376
AIC	-9939	-7718	-1183	-991
Log Likelihood	4978.6	3868.4	600.5	504.9

*Table 19: depicts the results of testing the FEARS and Realized volatility relationship, modelled as a ARFIMA (2,d,1) model. The for tests (1)-(4) refers to different intervals in time, surrounding the breakpoint of the Financial Crisis. The dependent variable in all tests is the log-seasonally adjusted Realized Volatility measure. The independent variables are denoted as follows: AR = p, whereas the number, e.g., p 2 corresponds to the th-lag. MA = q, with the number corresponding to the th-lag. The fractional integration parameter is denoted d. Control variables are included in all tests and refers to the CESI-EUR and EPU-US variable. AIC results refers to the Akaike selection parameter, and Log Likelihood refers to the Maximum Log Likelihood selection parameter. Significance is denoted "***" for p-value < 0.01, "**" for significance at the 1% level, "*" for the 5% level, and " ' "for the 10% significance level.*

Focusing on the results displayed in test (2), we find that our chosen AR and MA-parameters p and q are statistically significant at the 1% level, so that our parameters, determined by the software, seem to be a good fit. Furthermore, we find that the fractional integration parameter is chosen by the model to be 0.499, and statistically significant at the 1% level. Allowing for fractional integration of the d parameter, we allow for a more relaxed view on the assumption that data is either stationary (i.e., $d = 0$) or integrated ($d = 1$). We interpret the significant d value of 0.49 as an indication that the realized volatility time series data is covariance stationary but exhibits long memory. This means that external shocks have a long-term impact before the data returns to its mean value, slower than exponentially, as it would for a fully stationary time-series whereby the d parameter holds the value of zero. In other words, distant observations are correlated. The reasoning for this relationship is quite intuitive. As explained by the established relationship between FEARS and Average returns in section 7.1, FEARS induces noise traders to push stock prices upwards and away from their fundamental values. This action increases the movements in stocks, and on the index level, the stock return volatility. As this result follows the expectation of noise trader theory, it implies that the FEARS is good at proxying for noise trader behaviour.

For the full period results (1), we observe the expected sign of the coefficient, but no significance. The results yield a p-value of 0.16. Panel B presents a further breakdown of the period, (3) being the period before the Financial Crisis, and (4) during the Financial Crisis. Neither test (3) nor (4) returns significant results. In addition, test (3) returns a negative coefficient, indicating a negative relationship between our positive sentiment index and stock market volatility, contrary to our theoretical expectations. Because of a high p-value of 0.63 (reported in Appendix 11.3) we bear little weight to this result. However, it helps us to understand why we are not able to find a significant relationship for the full period test (1).

8 ROBUSTNESS CHECKS

The results presented above are products of decisions made throughout the research, from constructing the FEARS index, to testing its application. Although the methodology of Da et.al (2015) serve as a guideline in our research, some deviations from their methods are made where found necessary. These deviations and other decisions could have significant alterations on our results. Thus, providing robustness to our model, alternative methods will be applied in the following section. We will check whether these alternative decisions will alter the above discussed results, identifying any possible improvement in our methodology and disprove any indication of data mining.

One of the most important decisions in the construction of our FEARS index is the number of search terms we include in our aggregate index each period. Da et. al (2015) uses a 30-search term average for their FEARS index, stating that this number is sufficient to alleviate idiosyncratic noise. However, as explained in section 5.3.2, we have chosen a cutoff of 20 terms in our FEARS. Moreover, to check whether the amount of search terms included would have altered results, we construct two new FEARS indices, aggregating 10 search term and 30 search term averages. These are then regressed on DKallshare. Table 20, Panel A present the results of these tests.

Comparing the test results in Table 20, Panel A, and the results in Table 13, Panel A, we find that the 20-search term and 30-search term FEARS indices yield very similar results, whereas the 10-search term FEARS does not return a significant relationship with the 1-day lead returns (2). Arguing against Da et al. (2015), we suggest that the inclusion of an optimal number of search terms (30) in the aggregate FEARS index is not that important. Moreover, the results indicate that the inclusion of additional search terms yield better results than allowing for more idiosyncratic noise in the model.

Furthermore, we want to check whether the exclusion of our control variables would alter results. Additional tests are thus run, where we exclude the individual control variables, one by one. The results are reported in Table 20, Panel B. The results suggest that the exclusion of the variables one by one, or all together does not affect the coefficient of FEARS on DKallshare. However, when excluding the VSTOXX variable, the significance of this contemporaneous relationship increases from the 5% to the 1% level. In addition, the Adjusted R^2 drops dramatically. These observations imply that VSTOXX is an important variable in explaining the variance of stock market returns, and should be included, as it is. This is also illustrated by its continuous significant relationship with DKallshare at the 1% level. For the other control variables, however, we find that inclusion or exclusion yields little effect on either significance of the relationship, nor the changes in the adjusted R^2. We do, however, find some significance at the 10% level, and choose to include them in our tests.

Considering the results yielded from the robustness checks on search term and control variable inclusion, the model is proven robust as it is. Although some deviations are made from the methodology proposed by Da et al. (2015), our choices now have more grounding. That is, deviating from the choices we have made on the number of search terms, and the inclusion of our control variables, would in large parts not alter our results or conclusions.

Robustness Check

Panel A: FEARS on Returns for Different quantity of Google Search Terms on returns

	Top 10			Top 30		
	(1)	(2)	(3)	(3)	(4)	(5)
Variables	Ret (t)	Ret (T+1)	Ret (T+2)	Ret (t)	Ret (T+1)	Ret (T+2)
FEARS 10	0.001**	-0.001	0.001*			
	(0.0004)	(0.0005)	(0.0004)			
FEARS 30				0.0020***	-0.0020**	0.0020**
				(0.001)	(0.001)	(0.001)
Return Lag 1	0.0470***	-0.004	-0.022	0.0480***	-0.05	-0.022
	(0.016)	(0.021)	(0.020)	(0.016)	(0.021)	(0.020)
Return Lag 2	-0.005	-0.021	-0.012	-0.006	-0.021	-0.012
	(0.016)	(0.020)	(0.023)	(0.016	(0.020)	(0.023)
Return Lag 3	-0.011	-0.11	-0.032	-0.01	-0.011	-0.032
	(0.016)	(0.023)	(0.020)	(0.016)	(0.023)	(0.020)
Return Lag 4	-0.009	-0.031	-0.031*	-0.008	-0.031*	-0.031
	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)	(0.018)
VSTOXX	-0.0840***	-0.006**	0.004	-0.0840***	-0.0060**	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
EPU-US	-0.001*	-0.0001	-0.0001	-0.0004	-0.0001	-0.0001
	(0.0002)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)
CESI-EUR	-0.00001	0.00001	0.000	-0.00001	0.00001	0.0000
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(-0.0002)
Observations	4311	4311	4311	4311	4311	4311
Adjusted R-Squared	0.313	0.002	0.003	0.314	0.002	0.003

Panel B: FEARS on Returns when Omitting Control Variables

	Omitting VSTOXX		Ommitting EPU-US		Omitting CESI-EUR		Omitting All Controls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	Ret (t)	Ret (T+1)	Ret (t)	Ret (T+1)	Ret (t)	Ret (T+1)	Ret (t)	Ret (T+1)
FEARS 20	0.002**	-0.002**	0.0020***	-0.0020**	0.002***	-0.002**	0.002**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Return Lag 1	0.03	-0.006	0.005***	-0.005	0.047***	-0.005	0.03	-0.006
	(0.019)	(0.021)	(0.016)	(0.021)	(0.016)	(0.021)	(0.019)	(0.021)
Return Lag 2	-0.005	-0.022	-0.004	-0.022	-0.005	-0.021	-0.005	-0.022
	(0.021)	(0.020)	(0.016)	(0.020)	(0.016)	(0.020)	(0.021)	(0.020)
Return Lag 3	-0.021	-0.011	-0.011	-0.011	-0.011	-0.011	-0.0021	-0.011
	(0.020)	(0.023	(0.016)	(0.023)	(0.016)	(0.023)	(0.020)	(0.023)
Return Lag 4	-0.012	-0.032	-0.008	-0.0032	-0.008	-0.032	-0.012	-0.032
	(0.023)	(0.019)	(0.018)	(0.019)	(0.018)	(0.019)	(0.023)	(0.019)
VSTOXX			-0.083***	-0.006**	-0.084***	0.006**		
			(0.003)	(0.003)	(0.003)	(0.003)		
EPU-US	0.000*	0.0001			-0.0004*	0.0001		
	(0.0003)	(0.0003)			(0.0002)	(0.0003)		
CESI-EUR	-0.0060*	-0.00001	-0.004	-0.001				
Observations	4311	4311	4311	4311	4311	4311	4311	4311
Adjusted R-Squared	0.002	0.002	0.314	0.003	0.314	0.003	0.002	0.002

Table 20: This table report results from robustness checks relation to the construction of the index, that is, number of search terms included, and Control variables included. The dependent variables are contemporaneous and lead/future Denmark All Share daily returns. In Panel A, we test the two alternative construction quantities, and each test include control variable, that is, lagged return up to lag 4, the European implied volatility index VSTOXX, the Economic Policy Uncertainty Index EPU-US, and the Citigroup Economic Surprise Index CESI-EUR. Panel B test the composition of the regression by excluding control variables. In Panel B, only the original FEARS 20 is tested and used as the independent variable.

Moreover, we perform Robustness Checks for the parameter selection concerning the ARFIMA model constructed in section 6.5. The purpose of the robustness tests is to determine whether alteration of the AR and MA parameters will have a significant impact on the results reported in section 7.5, Table 17. In the reported $ARFIMA(2, d, 1)$ where $d \, \epsilon(-0.5{,}0.5)$, we found significant parameters and a significant relationship between realized volatility and the FEARS.

In Table 21, Panel A, we report eight test results on different sets of AR- and MA-parameters. The software itself determines a good fit for the fractional integration parameter d, which is set between $d \, \epsilon(-0.5{,}0.5)$. We can observe constant positive coefficients for FEARS in the eight tests depicted horizontally in Table 21, where the most positive coefficients have values of 0.0330 – 0.03371. The exception is test (5), modelled as a $ARFIMA\ (3, d, 1)$. Here, no significance can be reported in the AR- and MA- parameters, but a stronger significance in the FEARS coefficient, which is also considerably larger than for the other results.

Moreover, we observe that the fractional integration parameter d, is stable above 0.45-0.49 throughout the various tests, except for test (6), which is modelled $ARFIMA\ (3, d, 3)$. Similar to test (5), which also returned spurious results, test (6) also reports AR- and MA-parameters without significance. We find that most other results return similar results to our chosen $ARFIMA(2, d, 1)$, in both the FEARS coefficient, significant AR- and MA-parameters, in addition to a similar coefficient on the fractional integration parameter. Moreover, one can observe a stable Log-Likelihood through the tests, and a similar AIC.

In summarization, for many different sets of AR- and MA-parameter specifications, the conclusion, and interpretation would be the same, for both the relationship of FEARS on realized volatility and the long-memory properties that the time series exhibits.

Panel A: Robustness Check on parameter-selection for the post-financial crisis, seasonality-adjusted log realized volatility

	(0,d,1)	(0,d,2)	(1,d,0)	(2,d,0)	(3,d,1)	(3,d,3)	(4,d,2)	(4,d,3)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p 1			-0.1744***	-0.1529***	0.251	-0.1109	-0.4049***	0.6862***
			(0.07905)	(0.02731)	(0.9233)	(0.32093)	(0.02114)	(0.0678)
p 2				-0.0454*	-0.04	0.8674***	0.6529***	0.7908***
				(0.02188)	(0.1144)	(0.0519)	(0.02255)	(0.0698)
p 3					-0.006088	0.10252	0.1261***	-0.7654***
					(0.07858)	(0.29835)	(0.01958)	(0.0555)
p 4							0.062437	-0.1142***
							(0.01819)	(0.0266)
q 1	0.147***	0.1561***			0.37143	-0.08289	-0.2353***	0.8192***
	(0.0304)	(0.02824)			(0.8892)	(0.38691)	(0.01683)	(0.05775)
q 2		0.02123				0.8440***	-0.7641***	0.7150***
		(0.02025)				(0.07415)	(0.01683)	(0.06843)
q 3						0.0247349		-0.8997***
						(0.32634)		(0.0476)
d	0.468***	0.4796***	0.4996***	0.4759***	0.4997***	0.3528***	0.4949***	0.4562***
	(0.02219)	(0.02129)	(0.06333)	(0.02049)	(0.0006)	(0.09277)	(0.00697)	(0.0233)
FEARS	0.03330 '	0.0324 '	0.0331 '	0.03242 '	0.0500**	0.0349 '	0.0371 '	0.0344 '
	(0.01964)	(0.01963)	(0.01959)	(0.01963)	(0.01909)	(0.01966)	(0.01960)	(0.01965)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	3035	3035	3035	3035	3035	3035	3035	3035
AIC	-7782	-7781	-7769	-7782	-7738	-7782	-7774	-7786
Log Likelihood	3899	3898	3899	3898	3879	3902	3899	3906

*Table 21: reports results from robustness checks on the selection of the Autoregressive (AR) and Moving Average (MA) parameters. The Table hosts results of 8 separate combinations of AR and MA- parameters. The dependent variable in all tests is the log-seasonally adjusted Realized Volatility measure. The independent variables are denoted as follows: AR = p, whereas the number, e.g., p 2 corresponds to the th-lag. MA = q, with the number corresponding to the th-lag. The fractional integration parameter is denoted d. Control variables are included in all tests and refers to the CESI-EUR and EPU-US variable. AIC results refers to the Aikake selection parameter, and Log Likelihood refers to the Maximum Log Likelihood selection parameter. Significance is denoted "***" for p-value < 0.01, "**" for significance at the 1% level, "*" for the 5% level, and " ' " for the 10% significance level.*

9 DISCUSSIONS

In this section, the findings presented in section 7 will be discussed, following the empirical analysis explained in section 5. This part of the paper will provide further insight into whether we should accept or reject the main hypothesis, along with the additional sub-questions. A discussion of our results in comparison with prior literature will serve the purpose of highlighting the consensus of our findings. In addition, interpretations in line with theory will be drawn to give the results some intuitive reasoning.

9.1 AVERAGE RETURN

As forementioned in the literary review, the sentiment measures constructed by Baker and Wurgler (2007) stands as a pillar in the investor sentiment literature. Their findings indicate an explanatory power of sentiment on stock prices, supporting the notion that noise trading will impact asset prices (Black 1986; Delong et al., 1990).

As the FEARS and $DKallshare$ exhibit a significant relationship, both in the contemporaneous and in the lead variables, the findings are in line with the expectation. That is, our findings suggest that sentiment is an indicator of contemporaneous and future stock returns. Specifically, when investor optimism increases, noise traders will buy the stock on what is perceived as information. However, this information contains noise not reflected in the fundamental value of the stock. Thus, by trading on this noise, the price of the $DKallshare$ deviates from its fundamental value. This gap is then corrected by rational arbitrageurs the following day, as illustrated by the reversion in column (2) Table 13. From a theoretical perspective, this reversal makes intuitive sense as a positive relationship in the one-day lead variable would imply that noise trading permanently shifts prices away from fundamentals, which oppose the theorem of temporary mispricing.

As this paper replicates the methodology of Da et.al. (2015), it is natural to compare our results with their findings in the U.S stock market. Recall that our FEARS index was constructed based on search terms that had a positive correlation with stock market returns, in opposition to Da et. al (2015), who used the historic negative t-statistics to construct the FEARS. Their findings suggested that an increase in FEARS (investor pessimism increases) is followed by a contemporaneous decrease in stock returns. As an increase in our FEARS index (investor optimism increases) is followed by a contemporaneous increase in stock returns, our results are similar and the conclusion the same. They too find a temporary contemporaneous relationship in line with noise trader theory, intuitively explained as, negative sentiment should inflict noise traders to sell (or short) stocks, whereas in our paper, with opposite signs, we interpret it inversely, but the same way: positive sentiment should inflict noise traders to buy stocks.

A difference can be found in the duration of the contemporaneous effect. Da et. al (2015) finds that the initial contemporaneous effect is almost completely reversed within the two first days. In our results for the Danish stock market, however, we can observe a complete reversal already the next day, that is $Returns\ (T+1)$. Furthermore, we find that the reversal then allows for a new spike on the second lead day. This is in contradiction to their results. While we expect that the second spike in lead day 2 will be offset by a second reversal at $Returns\ T+3$, our investor sentiment index is not able to confirm this pattern. Intuitively, one could speculate that the second day (T+1) reversal ignites noise trading the subsequent day, by the belief that the correction was too immediate and harsh. That is, the expectation might be that stock prices will “bounce” following a potential overcorrection.

Although the signs in the lead variables differ, the significant relationship between FEARS and returns lasts equally long. That is, the impact on stock returns following a spike in FEARS diminishes after two days. In economic terms, this diminishing significance of FEARS on $DKallshare,$ might suggest that the noise trading effects die out two days following a spike in FEARS. However, this could also be a characteristic of the constructed model, as the relationship observed by Da et.al (2015) lasts equally long. The FEARS model may not be able to observe a relationship with a higher duration than two days. Such a limit in the model could have significance for our results, as it is unknown if the positive spike in $DKallshare\ T+2$ will be reversed on the following day (T+3).

Furthermore, the presence of sentiment effects on the days following a change in FEARS provide stronger arguments in favour of our sentiment measure, than the contemporaneous returns. That is, the stock return data ends at 5 pm local time, when trading close. Google, however, will continue to record search data until the very next day, which yields an additional seven hours of data collection. This implies that stock market information arriving after the trading day ends, could cause shifts in sentiment. This yields uncertainty in the causal relationship between contemporaneous returns and FEARS, as the stock returns could cause movements in sentiment. For the lead variables however, this concern is not relevant, as all sentiment effects prior will be reflected in any movements.

9.2 THE CROSS-SECTION OF STOCK RETURNS

A different picture naturally emerges when analysing the sentiment effects on the cross-section of stocks, as previous literature does not provide consistent results for this topic. Recall that the analysis toward the cross-section was divided into three segments, the discussion will thus focus on each of the three separately and then summarize the overall results. Interestingly, the results from our analysis also introduce some discussions regarding the cross-section of returns. That is, the results also enable us to compare the sentiment relationship between small and large stocks, using the LM indices as Large Cap stock proxies.

Concerning the broad market analysis, our findings on contemporaneous returns are consistent with Baker and Wurgler (2006) who find an equally significant effect on both value and growth stocks. However, as the reversal in T+1 is more significant for BMIvalue than BMIgrowth, the findings differ from theirs, suggesting that the broad market value stocks are more prone to sentiment. These results, however, are in line with the findings of Schmeling (2009) who found a significant sentiment effect for both value and growth stocks, and stronger for value stocks. In addition, the findings coincide somewhat with the findings of Lemmon and Portniaguina (2006), who suggest that the effect of sentiment is stronger for value stocks, however, they find no significance for growth stocks.

Furthermore, the LM stock analysis provides similar results. For the contemporaneous and one-day lead variables, the impact of a spike in FEARS is more significant for value stocks. However, from the two-day lead variables, (Table 15, Panel A (3) and Panel B (3)) the positive spikes which occurs two days following an increase in FEARS is more significant for LM growth stocks, than LM value stocks. This can be interpreted as a more delayed effect of sentiment for the LM growth than the LM value stocks. Overall, the large-cap results also provide support for the findings of Schmeling (2009).

The situation changes in the SML analysis. Here, an increase in FEARS have no significant impact on the contemporaneous growth returns. The findings on SMLvalue, however, are similar to those presented for the BMI and LM indices, where we find significance in both the contemporaneous returns and the lead returns. Thus, summing the findings for the small-cap stocks, the results suggest that also here, value stocks are more prone to investor sentiment.

Behavioural theories suggesting that the noise caused by individual investors is reflected in the prices of stocks might help explain why value stocks are more prone to sentiment changes than growth stocks. Kumar and Lee (2006) suggest that value stocks exhibit higher retail investor concentration, implying that the noise caused by individual investors will have more spill overs to value stocks rather than growth stocks. Our findings are thus in line with noise trader theory, which implies that individual investors are the ones driving stock prices, conditional on the transferability of Kumar and Lee's (2006) findings to the average Danish retail investor.

Another interesting line to draw from the cross-section results is the comparison between the SML and the LM indices. Most prior literature provides evidence in support of Baker and Wurgler who suggest that small stocks should be more prone to sentiment than large stocks (Baker Wurgler 2006). Moving the findings of Kumar and Lee (2006) toward the small and large stock analysis, they suggest that the stocks most prone to sentiment are those with the highest retail concentration, in this case small stocks. Although the consensus seems to be that small stocks are more vulnerable to sentiment changes, Fisher and Statman (2000) claim that the correlation between changes in individual investor sentiment is higher concerning large stocks. So there is some inconsistence in results here as well.

However, as most literature are in support of Baker and Wurgler (2006), the expectation should be that the SML indices are more prone to sentiment than the LM indices. However, our results suggest that there is little difference in the relationship between small and large stocks in Denmark. From Tables 15 and 16, one can observe that the LM stocks exhibit higher coefficients than the SML stocks, although the lead variables for the SML exhibit more significance. These results are somewhat ambiguous but suggest that the contemporaneous relationship between FEARS and large stocks is statistically more significant than for the small stocks. Observing the lead variables, however, the large stocks exhibit higher coefficients (in absolute terms), but lower significance.

Moreover, the significant relationship between FEARS and LM stocks can have some cause in the characteristics of Nasdaq OMX Copenhagen. Recall that the Danish stock market is dominated by a few large companies. From the correlation matrix in Table 1, we observed that the LM and BMI indices are almost perfectly correlated, supporting the assumption that most of the BMI indices contain LM stocks, rather than SML stocks. Amateur retail investors in the United States are subject to a diverse stock market, thus being able to construct diverse small-cap-weighted portfolios. In Demark, however, the number of stocks available on the market is limited in comparison, suggesting that the average Danish investor might weigh large stocks heavier in their portfolios. Following the notion that market prices are affected by retail investor trading, large stocks should thus be more vulnerable to sentiment changes in Denmark than in the USA, further understating the difficulties in comparing results across markets.

Summing up, we thoroughly believe that our results provide good insight into what stock characteristics makes them more prone to investor sentiment. However, it would be premature to conclude that the abovementioned interpretations of our results hold for certainty, as there is some ambiguity in the SML and LM results. Our interpretations simply serve as intuitive explanations of why our FEARS index might impact value and large stocks to a greater extent than growth and small stocks, in Denmark.

9.3 GRANGER CAUSALITY

Providing even stronger evidence of sentiment effects on average market returns, our results from section 7.3 determine that FEARS Granger Causes the $DKallshare$ returns. To the authors of this papers knowledge, no one has previously reported any Granger Causality results for this relationship in Denmark. Our findings reveal that when 1-4 lags are applied, the FEARS Granger Causes average returns at the 5% significance level, and for 5 lags at the 10% level. These results are unidirectional, meaning that in the reverse Granger Test, average returns do not Granger Cause the FEARS.

The implication of this result is an important one. Along with the relationship established between the FEARS and short-term future returns, the Granger Test results also strengthen the direction of the casual relationship. We determine that the relationship is unidirectional, which alleviates some of the concerns that the FEARS may be a result of stock market movements, rather than the stock market movement being a result of FEARS.

9.4 BONDS

The bond market is proven to co-move with the world markets based on news and investor sentiment (Barberis, Shleifer and Wurgler, 2005). In addition, there are also evidence of sentiment spill overs from equity markets to bond markets (Kwan 1996; Downing et. al 2009; Bethke et al. 2017). As section 7.1 proved that there is a significant relationship with between FEARS and the DKallshare, we would expect to find a relationship between FEARS and the Danish bonds. However, as many financial scholars have pointed out, the sentiment effect is expectedly lower on bonds than stocks (Baker and Wurgler, 2012). In this section, the investigation following our regression models will thus be discussed in comparison to previous literature, and the findings interpreted.

9.4.1 Government Bonds

Starting with the government bond regressions from Table 18, Panel A, the regressions on 10Y government bonds yielded results in line with expectation. That is, the signs resemble the relationship between FEARS and $DKallshare$. An increase in FEARS could thus be interpreted as a contemporaneous increase in the bond yield, which was reversed the next day. However, this observed relationship is not statistically significant, suggesting that the predictive power of FEARS is not as strong for 10Y bond returns, as for the stocks.

Interestingly, in the bond regression for the 5Y government bonds, the coefficient for FEARS is positive both on contemporaneous and the one- and two-day-lead returns, while the only significance observed is found on $'5Y\ T+2'$ (Table 18, Panel A, column 6). This delayed impact of FEARS might indicate that investors leave their bond positions one day following a significant withdrawal from stocks[21]. Moreover, in contrast to what is expected following the stock market results, there is no significant reversal observed for the 5Ybond. However, as discussed in section 9.1, it is possible that the FEARS methodology is only capable of capturing the relationship two days following a spike in sentiment, suggesting that one might have had observed a reversal in T+3.

9.4.2 Corporate bonds

The proven relationship between investor sentiment and stock market movements, can also be interpreted as a reflection of irrational beliefs about the outlook of a company. As both stocks and corporate bonds are claims on firm assets (Merton, 1974), one should expect that the relationship observed in the stock market also hold for the corporate bond market.

[21] Refers to the reversal observed in DKallshare T+1

However, the regression analysis on IG Bonds yields the same relationship as for the 10Y Bonds, depicted with inverse signs[22]. As the IGbond variable is constructed based on high rated (low yield) corporate bonds, this is not a surprise finding. That is, returns on investment grade bonds are driven mainly by returns on treasuries (Nayak, 2010). These results are somewhat in support of prior research, which suggest that low yield corporate bonds are less prone to sentiment in comparison to high yield bonds (Nayak, 2010; Huang et al. 2015). Thus, the Danish IG Bonds may not be as prone to sentiment as Danish high yield bonds could potentially be. This suggest that our findings are insufficient in concluding a significant relationship between FEARS and the Danish corporate bond market.

Summing up the findings for the corporate and government bond regressions, the results are somewhat in support of Baker and Wurgler (2012) in that the sentiment effect is lower for bonds than stocks in general. Although the signs observed in 10Ybond and IGbond are in line with expectation, there was no observed statistical significance. However, the two-day relationship between FEARS and the 5Y bonds highlight what can be interpreted through the theory of "flight-to-quality" (Beber, A., Brandt, M., & Kavajecz, K., 2009). This theory explains that investors will escape to "safe havens", that is low risk assets such as bonds, in bad times. As our index proxies for optimism, this must be viewed inversely, as a "flight-from-quality". That is, when there is increasing optimism amongst investors, they enter the stock market, and with some delayed effect, leave their positions in the bond market, as suggested by the $5Ybond$ results.

9.5 VOLATILITY

Regarding the analysis on FEARS and volatility explained in section 6.5, the discussion will in large compare our results in section 7.5 to prior research, to shed light on the consensus of our findings.

Most comparable is the study of Da. et al (2015), from which the methodology we use stem from. Da et al. (2015) find a highly significant relationship between their FEARS index, and their volatility measures, that is, realized volatility, the VIX and VIX futures. Due to limitation in Denmark data, we are only able to replicate the study on realized volatility. There are three differences between our volatility analysis and theirs. Firstly, their FEARS captures pessimism, our captures optimism. Secondly, they construct the realized volatility measure using 15-minute intraday intervals on their data, whereas we use already-calculated data, based on 10-minute intraday intervals. The effect of this on the results is not likely to be important. Thirdly, Da et al. (2015) construct their index based on a 30-word average, whereas we use 20 words.

[22] Recall that the IG Bond is a price index, while the 10Ybond is constructed from yield data.

Our test (2) results are in line with both the results of Da et. al (2015), and the noise trader theory. The intuition is that for an increase in positive sentiment, noise traders will trade bullishly (optimistically), and push stock market prices. This trading behaviour increases the deviation of a stock price to its fundamental value and increases the overall price volatility in the market. Stock return volatility increases with the volatility of our sentiment measure, following a simple intuition. The more sentiment contributes to stock prices deviating from fundamental values, the more excessive price volatility will be.

However, the results of Da et al. (2015) are perceived as more robust. Specifically, our findings highlight a significant relationship in the period after the Financial Crisis (July 2009) to the end of our sample period (December 2021). Da et al. however, report a highly significant relationship for their full sample period, which spans from January 2004-December 2011. Nevertheless, as our results shows that investor sentiment has a significant relationship with the realized volatility post Financial Crisis, we conclude that our FEARS is able to measure volatility in Denmark.

Moreover, Verma & Verma (2007) test the effects of noise trading on conditional volatility, building on the theoretical approach of De Long et al. (1990). Our empirical strategy contains some distinct differences from the methods of Verma & Verma. Specifically, they model indices for both rational and irrational investors, as well as focusing on both bearish and bullish signals. One implication of the google trends-based FEARS index is that it relates to retail investors who trade without news relating to fundamentals, which can be referred to as irrational investors. Thus, there is an expiation that rational or institutional investors find their information elsewhere. Additionally, we only test for bullish sentiment and signals.

Despite these differences, our two papers are similar in the interpretation of results. They too find that investor error (noise trading) driving stock prices away from fundamental values is a significant determinant of stock market volatility. Thus, both ours and their research find evidence that the relationship between noise traders and volatility as proposed by De Long et al. (1990) holds in our respective markets.

10 CONCERNS AND WEAKNESSES

So far, we have established the predictability of investor sentiment on short-term returns and contemporaneous stock market volatility. In the following we highlight alternative explanations for our results, using earlier literature and intuitive reflections to mitigate concerns, as well as shed light on weaknesses in our methods.

10.1 ENDOGENOUS SEARCH

One alternative explanation of our results, as mentioned by Da. et al (2015), is that spikes in our FEARS index are endogenous to macroeconomic events. That is, spikes or movement in the FEARS index (Figure 12) itself may correlate or alternatively result from macroeconomic events. Da. et al (2015) disregard this concern, as a correlation between the FEARS and the arrival of news and macro events should be expected. That is, investor sentiment is not immune to macro events.

As mentioned in section 4.1.2, the arrival of news follows a random walk, as does large macroeconomic events. However, it is unlikely that the investor sentiment index should be explained by such news-pattern to a large extent. Additionally, following the methodology of Da. et al (2015), our control variables should be able to capture the impact macroeconomic events. Although our control variables are not explicitly concerning Denmark, but rather Europe and the U.S., this might be an advantage. That is, our control variables should capture both American and European distress stemming from large macroeconomic events, which might have a small lag before hitting the Danish market.

10.2 REVERSE CAUSALITY

A second explanation might be that the Google search behaviour is not motivated by sentiment, but rather expectations of an event. For expected news events, expected movements in the stock market, or expectations in general, the google engine might record data on searches motivated by the expectation of such events. One very convincing argument provided by Da. et al (2015), is that the observed reversal patterns, makes reversed causality unlikely. From a behavioural point of view, it is unlikely that market participants who expect a downturn in stock returns, as we find in lead one-day returns (Returns T+1 in Table 13), should search for optimistic terms today. However, in the construction of the FEARS, we objectively study the historic relationship between the FEARS and stock returns. All search terms are thus not necessarily optimistic in nature throughout the testing period. Therefore, we are not able to disregard this alternative explanation completely, through the arguments of Da et al. (2015).

10.3 SEARCH ENGINE BEHAVIOUR

Whether the search patterns captured by FEARS reflect optimism or pessimism, or if it just captures random activity, is a concern in the methodology. That is, households might not search for economic search terms because they are especially optimistic or worried, but rather for information gathering and curiosity. The literature on search term analytics disregards such concern. Search engine analytics has been illustrated over time as a predictor in several fields of study. As mentioned in Section 4.5, Polgreen et al. (2008) used search engine analytics to predict an increase in influenza cases. Additionally, our optimistic words exhibit a positive relationship with contemporaneous stock returns, which is a relationship that is hard to interpret as coincidental.

However, we want to check this claim of random search activity, to ensure that our positive words index can capture optimism. Thus, we construct a new FEARS Index, whereby the search terms that we include are selected at random. That is, instead of choosing search terms based on historical positive relationship with the market, we use a random number generator to determine the 20 words to include. All other parts of the construction phase are identical. For example, the index is still dynamic, and changes terms every 120 trading days. The results are reported in appendix 13, and reveals no relationship between the random FEARS and stock market returns.

10.4 FEARS AND AGGREGATE VOLATILITY

One final clarification point concerns the causal relationship between the investor sentiment index and market volatility. Regarding the realized volatility, which is calculated on intraday data for the OMX C20 index, we interpret the FEARS as simply having a causal relationship through increased or decreased noise trading activity. As mentioned in section 7.5, increased stock trading pushes stock prices upwards away from fundamental values. The results further reveal a reversal in the one-day lag, and a subsequent spike in day $T+2$. An increased frequency of up and down movement naturally implies more volatility in stock market returns.

However, other interpretations of the causal relationship exist. Da et. al (2015) states that it is possible that the FEARS index works as a proxy for time-varying risk aversion, whereby negative sentiment (as they measure) increases risk aversion. Implying that in our case, positive sentiment should reduce it. One weakness in our analysis, is that we do not have a measure of aggregate uncertainty in the Danish market. We are limited to the European VSTOXX, which relates to implied volatility on the EURO STOXX 50 index. Thus, in our volatility section, we are not able to test the FEARS on aggregate volatility for any potential relationship. By extension, we cannot determine a causal relationship between our measure of Danish investor sentiment, and aggregate volatility. Our test refers to stock return volatility, in which we find that the observed significant relationship is in line with noise trader theory.'

11 CONCLUSIONS

In this paper, we use household google search behaviour to construct a daily sentiment index for Denmark. Following the approach presented by Da, Engelberg and Gao (2015), we aggregate search terms like “rig”, “aktie” and “investering”, (“rich”, “stock” and “investment”, respectively) to construct a Financial and Economic Attitudes Revealed by Search (FEARS) index. Our FEARS index is constructed based on the economic terms exhibiting the largest historic positive correlation with the Nasdaq OMX Copenhagen average returns within the period 2004-2021, thus proxying for investor optimism.

Our findings prove that the FEARS index is a predictor of returns in the Danish stock market. Our results are in support of Da et al. (2015), where we find a contemporaneous relationship between sentiment and stock returns. Intuitively, this indicates that positive sentiment induces noise traders to buy stocks, which is in line with theory. Moreover, this positive effect is then completely reversed the next day, consistent with the theories of temporary mispricing in stocks (Delong et.al. 1990) and previous findings (Tetlock 2007, Da et al. 2015). However, contrary to the findings of Da et.al. (2015), we find a positive relationship with returns two days following an increase in FEARS. Our FEARS index is not able to determine whether this second-day spike will again be reversed the following day. Intuitively, one could interpret the second-day effect as a reaction to the reversal the previous day.

The observed reversal pattern contributes to the robustness of the methodology, as it aids in determining the causal direction of the relationship between FEARS and stock market returns. Although it slightly deviates from our expectations, the reversal is nevertheless consistent with sentiment and noise trader models. In support of this, the Granger Test determines a unidirectional relationship, implying that FEARS contains information that helps explain future values of stock returns, but not the other way around. That is, FEARS Granger Causes returns.

With the scope of adding to existing research of the sentiment effect on the cross-section of stocks and gaining further insight into what characteristics of a stock makes it more vulnerable to sentiment changes, we conduct analysis within the cross-section of stocks. Our results indicate that the sentiment effect is strongest amongst stocks under the label “value”, rather than “growth”. As there is little consensus to be found in prior literature regarding this relationship, the findings are in support of some, but not all previous studies. However, following noise trader theory, we perceive the findings to make sense as value stocks are suggested to exhibit higher retail investor concentration, than growth stocks (Kumar and Lee, 2006). In addition, analysing the cross-section enabled us to examine whether FEARS has more predictive power towards small or large stocks. Here our findings suggest that large stocks are arguably more prone to sentiment than small stocks, contrary to most previous findings (Kumar and Lee, 2006; Baker and Wurgler 2006, 2007; Schmelling 2009) but in support of some (Fisher and

Statman, 2000). However, as most of the prior research is conducted in the U.S, where the stock market characteristics differ significantly, the results are not directly comparable.

Moreover, a different picture arises within the relationship between FEARS and the bond market. Regarding the government bonds, there is no significant impact of FEARS on those with a 10-year maturity. However, for 5-year maturity bonds, the FEARS have some delayed prediction. This finding indicates that investors flee from their bond positions, two days following an increase in FEARS, consistent with an inverse flight-to-quality (Beber et al. 2009; Nayak, 2010). Moreover, extending the analysis toward the corporate bond market, the IG Bond analysis yield the same results as for the 10-year bonds, which makes intuitive sense since IG bonds are driven mainly by returns on treasuries (Nayak, 2010).

Furthermore, we constructed an ARFIMA model to test whether sentiment has any significant relationship with realized market volatility. Our findings are in support of Da. et al. (2015), suggesting that there is a significant relationship between FEARS and the realized volatility, also in Denmark. However, the robustness of our results is not as strong as theirs, as we only observe a significant relationship after the end of the Financial Crisis. This finding implies, as illustrated in Figure 5, that noise traders are present in the Danish market, and that the relationship between noise traders and volatility, as developed by De Long et al. (1990), holds according to expectation and theory.

Thus, our main hypothesis can be accepted, as there is an apparent relationship between investor sentiment and stock market movements. As for our sub-questions, we prove that the effect of sentiment differs within the cross-section of stocks, being strongest for value and large stocks. Moreover, our evidence towards the bond markets is not sufficient to accept the hypothesis. Finally, our results concerning volatility are in line with prior research and expectation, where we find a significant prediction of FEARS on volatility, and thus accept the hypothesis.

11.1 IMPLICATIONS

This paper is one of the first to document the relevance of investor sentiment impact on the Danish stock market, as per the authors of this paper's knowledge. From a theoretical perspective, this paper provides empirical evidence of apparent sentiment effects in Denmark. Moreover, our research builds on the behavioural finance models and the notion of "noise trading". Our findings can be interpreted as indirect evidence on the impact of investor irrationality in financial markets through noise trading, as proposed by Black (1986) and De Long et al. (1990).

Additionally, we provide evidence of short-term market inefficiency. This allows for further understanding of what impacts stock prices in the short run. Since similar mispricing and subsequent reversal patterns are observed for the cross-section of different stocks, we also provide salient evidence on the joint dynamics of stocks that exhibit different characteristics. Moreover, some indication of a reverse flight-to-quality can be observed from our findings on the bond markets, although they are ambiguous. As seen from the 5-year treasury bonds, investors tend to liquidate their bond positions when optimism increases, adding an interesting perspective on the proposed safe havens (Beber et al. 2009).

Furthermore, this paper proves the transferability of the FEARS methodology to previously untapped waters. That is, by successfully replicating the methodology of Da et al. (2015) for research in Denmark, we provide further robustness to their model. In addition, as our FEARS index is constructed based on positive search terms, we are the first ones, as per the authors knowledge, to apply the FEARS index as a proxy for investor optimism.

Just as important, the methods applied in this paper follow a new strand of sentiment literature, underscoring the usefulness of SVI data in financial analysis. SVI data has the potential to reveal the underlying beliefs of household populations objectively and directly to empirics (Da et al, 2015). Da et al. (2015) suggest that financial models link beliefs to equilibrium outcomes such as volume or returns, thus, search behaviour has the potential to provide more accurate tests of economic models. As the research conducted has a basis in the Google search behaviour of Danish individuals, our findings give further grounding on the role of Google searches in Denmark.

As mentioned in section 1.2, the value of this paper is twofold. The findings also contribute in some practical perspective. That is, our findings can help illustrate how one could effectively construct a sentiment-based investment strategy for the Nasdaq OMX Copenhagen. In section 5.3, we highlighted that SVIs, and Danish equity returns exhibit a historical correlation. As the FEARS methodology only relies on historical data in the construction, there are few limits to the extent one can apply the approach in designing a profitable trading strategy. That is, if observing a significant increase in our FEARS index (investor optimism increases), one could go short stocks, on the following day, taking advantage of the return reversal highlighted in section 7.1.

Finally, our results from section 7.3 suggest that investor sentiment Granger Causes stock market movements for up to five lags. A practical implication of our results is thus that sentiment can be incorporated as a factor in forecasting models for returns, having proven a significant relationship.

11.2 FURTHER RESEARCH

There are few limits for future research within the sentiment literature. Having established a proven transferability of the FEARS methodology to Denmark, a natural starting point for further research would be to apply the methods of Da et al. (2015) to identify a relationship between investor sentiment and other Scandinavian equity markets. As the methodology is proven to work in predicting returns on the Nasdaq OMX Copenhagen, research on Norway, Sweden and Finland would allow for some comparison between the four.

Concerning the specific discussions of this paper, a continuation of the research could provide further analysis of the sentiment effects within the Danish market. A natural extension could be that of testing whether the predictive power of our constructed FEARS holds for other asset classes, not considered in this paper. As the analysis on corporate bonds faced limitations in this paper, an interesting extension would be to apply the FEARS methodology towards more speculative-grade bonds, constructing indices containing high and low yield Danish corporate bonds, as speculative-grade bonds are suggested to have a stronger relationship with sentiment (Nayak, 2010).

Likewise, the successful replication of the FEARS on Danish equity returns could induce research using FEARS for other applications. For instance, following Aroul et al. (2020), further research could test FEARS predictive power on Danish housing returns, extending the robustness of the sentiment index in Denmark. Moreover, our analysis of the cross-section of returns introduces the notion of FEARS predictive power across different stocks. As the relationship with FEARS varies within the cross-section of returns, an interesting extension of the paper could see the application of FEARS in a thorough sector analysis in Denmark, observing how the sentiment impact different sectors, providing further insight into what stock characteristics correlate with sentiment.

Furthermore, one of the key predictions of the noise trader literature of De Long et al. (1990), and Black (1996), is that investor sentiment should increase the number of trades, i.e., should increase the liquidity in the market. An interesting question for further research is if the FEARS methodology can establish a significant relationship with a measure of stock market liquidity, on Nasdaq OMX Copenhagen. The expectation, for a positive sentiment index, would be that the increased positivism should increase the number of traders or the number of trades each noise trader engages in. Moreover, such a study would need to find a good measure of market liquidity for a broad selection of stocks or an index. The construction of a broad liquidity measure could for instance be the bid-ask spread, calculated at the stock level, and aggregated for an entire stock market index. This requires a lot of computational work, as there are a lot of stocks in a broad index. More advanced liquidity measures can thus be considered. From a theoretical perspective, we would expect a bid-ask spread to narrow during periods of increased optimism.

As mentioned in section 10, this thesis is not able to, nor does it aim to establish a causal relationship between household search engine behaviour, and the aggregate volatility levels within the market. Our realized volatility measure has a causal unidirectional relationship with returns, whereas the implied volatility measures, such as the VSTOXX, exhibit predictability towards stock market returns. For future research, one might study the causal relationship between-, and the predictive powers of investor sentiment on aggregate volatility. This research question is limited to markets whereby a VIX or VSTOXX index exists, or by the computational skills required to produce a convincing measure of volatility.

Moreover, this paper proves the presence of a positive relationship between investor optimism and contemporaneous stock returns, however, it disregards pessimism in the market. Constructing a FEARS index in Denmark based on search terms that exhibit a negative correlation with historical stock returns would provide further insight into the theories of asymmetric sentiment effects (Brown and Cliff, 2005). Brown and Cliff suggest (2005) suggest that limits to arbitrage, and more specifically, short-selling constraints, can make it difficult for rational investors to prevent market prices from exceeding their fundamental value in optimistic periods. Even the risk from short-selling alone may prevent investors of selling short even if they would be able to do so. No similar friction, however, holds in pessimistic periods, where arbitrageurs are free to take the necessary long positions. Thus, a pessimistic sentiment index would provide interesting insight into whether this asymmetry has significance.

Finally, it would be interesting to test the practical implication of the FEARS index predictive power. Further research could test the predictability of the FEARS on different theoretical portfolios, such as high-beta, low-beta, industry-specific, momentum-portfolios, etc. It would be interesting to study if the FEARS could contribute to building a trading strategy that generated alpha, that is, abnormal returns. However, we suspect that the FEARS methodology would quickly be implemented by rational arbitrageurs, who, by increasing the arbitrary capital, would fill the information gap, and strangle the alpha-generating possibilities.

12 BIBLIOGRAPHY

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001, July). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*(1), 43-76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003, March). Modelling and Forecasting Realized Volatility. *Econometrica: Journal of the econometric society*, *71*(2), 579-625.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Vega, C. (2007, November). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics*, *73*(2), 251-277.

Antweiler, W., & Frank, M. Z. (2004, June). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Jounral of Finance*, *59*(3), 1259-1294.

Aroul, R. R., Sabherwal, S., & Saydometov, S. (2020). FEAR Index, city characteristics, and housing returns. *Real Estate Econ.*, *50*, 173-205.

Aruoba-Diebold-Scotti Business Conditions Index. (n.d.). Federal Reserve Bank of Philadelphia. Retrieved 2022, from https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/ads

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*, 1645-1680.

Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, *21*(2), 129-152. 10.1257/jep.21.2.129

Baker, S. R., Bloom, N., & Davis, S. J. (2016, November). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, *131*(4).

Baker, S. R., Bloom, N., & Davis, S. J. (n.d.). *Methodology*. Economic Policy Uncertainty. https://www.policyuncertainty.com/methodology.html

Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, *75*(2). https://doi.org/10.1016/j.jfineco.2004.04.003

Barndorff-Nielsen, O. E., Kinnebrock, S., & Shepard, N. (2008). Measuring Downside Risk - Realised Semivariance. *CREATES Research paper, 42*.

Beber, A., Brandt, M., & Kavajecz, K. (2009). Flight-to-quality or flight-to-liquidity? Evidence from the euro-area bond market. *Review of Financial Studies*, *22*, 925-957. 10.1093/rfs/hhm088

Bethke, S., Gehde-Trapp, M., & Kempf, A. (2017). Investor sentiment, flight-to-quality, and corporate bond comovement. *Journal of Banking and Finance*, *82*, 112-132. http://dx.doi.org/10.1016/j.jbankfin.2017.02.007

Black, F. (1986). Noise. *Journal of Finance*, *41*(3), 529-543.

Bodie, Z., Kane, A., & Marcus, A. L. (2014). *Investments* (Tenth Edition. ed.). New York: McGraw-Hill Education.

Boesler, M. (2013, December). The State of the World as seen through Citi's Economic Surpise Indices. *Business Insider*. https://www.businessinsider.com/citi-economic-surprise-index-2013-12?r=US&IR=T

Box-Steffensmeier, J., Freeman, J. R., Hitt, M. P., & Pevehouse, J. C.W. (2014). Selections on Time Series Analysis. In *Time Series Analysis for the Social Sciences* (pp. 175-178). Cambridge University Press.

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 1287-1294.

Brown, G. W. (1999, January). Volatility, Sentiment, and Noise traders. *Financial Analysts Journal*, *55*(2), 82-90.

Brown, G. W., & Cliff, M. T. (2005). Investor sentiment and asset valuation. *Journal of Business*, *78*, 405-440.

Bryman, A., & Bell, E. (2011). *Business Research Methods* (Vol. 3). Oxford: Oxford University Press.

Choi, H., & Varian, H. (2012, June). Predicting the Present with Google Trends. *The Economic Record*, *99*, 2-9.

Chui, A. C.W., Titman, S., & Wei, J. (2008). Individualism and momentum around the world. *AFA 2006 Boston Meetings Paper*.

Collin-Dufresne, P., Goldstein, R. S., & Martin, J. S. (2001). The determinants of credit spread changes. *Journal of Finance*, *56*, 2177-2207. 10.1111/0022-1082. 00402

Cooper, C. P., Mallon, K. P., Leadbetter, S., Pollack, L. A., & Peipins, L. A. (2005). Cancer Internet Search Activity on a Major Search Engine, United States 2001-2003. *Journal of Medical Research*.

Cromwell, J. B., Hannan, M. J., Labys, W. C., & Terazza, M. (1994). *Multivariate Tests for Time Series Models*. Sage Publications.

Da, Z., Engelberg, J., & Gao, P. (2015). The Sum of All Fears: Investor Sentiment and Asset Prices. *The Review of Financial Studies*, *28*(1), 1-32. 10.1093/rfs/hhu072

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*, 703-738.

Downing, C., Underwood, S., & Xing, Y. (2009). Relative informational efficiency of stocks and bonds: An intraday analysis. *Journal of Financial and Quantitative Analysis*, *44*, 1081-1102. 10.1017/S0022109009990305

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression II. *Biometrika*, *38*, 159-178.

Engle, R. F., & Patton, A. J. (2007). 2 - What good is a volatility model? In *Forecasting volatility* (pp. 47-63). Quantitative Finance.

Ettredge, M., Gerdes, J., & Karuga, G. (2005, November). Using Web-based Search Data to Predict Macroeconomic Statistics. *Communications of the ACM*, *48*.

EURO STOXX 50® Volatility (VSTOXX®). (n.d.). Qontigo. Retrieved 2022, from https://www.stoxx.com/index-details?symbol=V2TX

Fama, E. (1965a). The behaviour of stock market prices. *Journal of Business*, *38*, 34-105.

Fama, E. F. (1965b, September-October). Random Walks in Stock Market Prices. *Financial Analysts Journal*, 55-59.

Fama, E., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *33*, 3-56. 10.1016/0304-405X(93)90023-5

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Works. *The Journal of Finance*, *25*(2), 383-417. 10.2307/2325486

Fang, L., YU, H., & Huang, Y. (2018). The role of investor sentiment in the long-term correlation between U.S. stock and bond markets. *Int. Rev. Econ. Finance*, *58*, 127-139. http://dx.doi.org/10.1016/j.iref.2018.03.005

Fisher, K. L., & Statman, M. (2000). Investor Sentiment and Stock Return. *Financial Analysts Journal*, *56*(2), 16-23. 10.2469/faj.v56.n2.2340

Friedman, M. (1953). The case for flexible exchange rates. *Essays in Positive Economics*. University of Chicago Press, Chicago.

Gao, Z., Ren, H., & Zhang, B. (2020). Googling investor sentiment around the world. *Journal of Financial and Quantitative Analysis*, *55*, 549-580.

Google. (n.d.). *Our story*. About Google. https://about.google/our-story/

Google. (n.d.). *Trend Answers*. Google Support. https://gs.statcounter.com/search-engine-market-share/all/europe/2021

Google Inc. (n.d.). *Google Trends*. Google Trends. https://trends.google.com/trends/?geo=DK

Hill, R. C., Griffiths, W. E., & Lim, G. C. (2011). Principles of econometrics. *Hoboken, N.J: Wiley*, *4*.

Holm, A. B. (2018). *Videnskab i virkeligheden* (Vol. 2). København: Samfundslitteratur.

Hong, H., & Stein, J. C. (2007). Disagreement and the Stock Market. *Journal of Economic Perspectives*, *21*, 109-128.

Huang, J. Z., Rossi, M., & Wang, Y. (2015). Sentiment and corporate bond valuations before and after the onset of the credit crisis. *Journal of Fixed Income*, *25*, 34-57.

Indexes - Nasdaq. (2022). Nasdaq Nordic. Retrieved May 3, 2022, from http://www.nasdaqomxnordic.com/indexes

Ito, T., Masuda, M., Naito, A., & Takeda, F. (2020). Application of Google Trends-based sentiment index in exchange rate prediction. *Journal of Forecasting.*, *40*, 1154-1178. 10.1002/for.2762

Investing.com. (n.d.). *Denmark 10 Years Government Benchmark* [TRXVDKGOV10D]. Investing.com. https://www.investing.com/indices/tr-dk-10-y-government

Investing.com. (n.d.). *Denmark 5 Years Government Benchmark* [TRXCDKGOV5D]. Investing.com. https://www.investing.com/indices/tr-dk-5-y-government

Kahneman, D. (1973). *Attention and effort*. PRENTICE-HALL, INC., Englewood Cliffs, New Jersey.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, *47*(2), 313-327.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. London, UK: Macmillan.

Kumar, A., & Lee, C. M. (2006). Retail Investor Sentiment and Return Comovements. *Journal of Finance, 61*(5), 2451-2486. 10.1111/j.1540-6261.2006.01063.x

Kwan, S. H. (1996). "Firm-Specific Information and the Correlation Between Individual Stocks and Bonds." *Journal of Financial Economics*, *40*(1), 63-80.

Laborda, R., & Olmo, J. (2014). Investor sentiment and bond risk premia. *Journal of Financial Markets*, *18*, 206-233. http://dx.doi.org/10.1016/j.finmar.2013.05.008

Lee, C. M., Shleifer, A., & Thaler, R. H. (1991). Investor Sentiment and the Closed-End Fund Puzzle. *The Journal of Finance, 46*(1), 75-109. 10.2307/2328690

Lemmon, M., & Portniaguina, E. (2006). Consumer confidence and asset prices: some empirical evidence. *Review of Financial Studies*, *19*, 1499–1529.

Li, M. (2021). Uses and abuses of statistical control variables: Ruling out or creating alternative explanations? *Journal of Business Research*, *126*, 472-488.

Liu, K., Chen, Y., & Zhang, X. (2017, June). An Evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average) Programs. *axions*.

Making Sense of the VIX Index: An Indicator of Expected Market Volatility. (n.d.). Cboe Global Markets. Retrieved 2022, from https://www.cboe.com/tradable_products/vix/

Marsh, T., & Merton, R. C. (1986). Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices. *The American Economic Review*, *76*(3), 483-498.

Merton, R. K. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance, 29*, 449-470. DOI: 10.2307/2978814

Miller, E. M. (1977). Risk, Uncertainty, and Divergence of Opinion. *Journal of Finance*, *32*, 1151-1168.

Mossin, J. (1966, October). Equilibrium in a Capital Asset Market. *Econometrica*, *34*(4), 768-783.

Nasdag OMX Nordic. (n.d.). *Index*. Index Breakdown. https://indexes.nasdaqomx.com/Index/Breakdown/OMXCGI

Nasdag OMX Nordic. (n.d.). *Index Methodology*. Methodology Nordic Allshare. https://indexes.nasdaqomx.com/docs/Methodology_Nordic_AllShare.pdf

Nasdaq OMX Nordic. (n.d.). *About us*. About us. http://www.nasdaqomxnordic.com/about_us

Nasdaq OMX Nordic. (n.d.). *Index Activity*. Indexes. http://www.nasdaqomxnordic.com/indexes/index_activity?Instrument=DK0060775369

Nayak, S. (201. Investor sentiment and corporate bond yield spreads. *Rev. Behav. Finance*, *2*, 59-80. https://doi.org/10.1002/rbf.9.

Oxford-Man Institute of Quantitative Finance. (n.d.). *OMX Copenhagen 20 Index Annualized Realized Volatility (10-min)*. https://realized.oxford-man.ox.ac.uk/data/visualization

Pontiff, J. (1996). Costly arbitrage: Evidence from closed-end funds. *Quarterly Journal of Economics*, *111*(4).

Perold, A. F. (2004). The Capital Asset Pricing Model. *Journal of Economic Perspectives*, *18*(3), 3-24.

Polgreen, P. M., Chen, Y., Pennock, D. M., Nelson, F. D., & Weinstein, R. A. (2008, December). Using Internet Searches for Influenza Surveillance. *Clinical Infectious Diseases*, *47*(11), 1443-1448.

Ramiah, V., Xu, X., & Moosa, I. A. (2015, October). Neoclassical Finance, behavioral finance and noise traders: A review and assessment of the litterature. *International Review of Financial Analysis*, *41*, 89-100.

Refinitiv Datastream. (n.d.). *EKCEIR* [CESI - EUROZONE - ECONOMIC SERIES]. Citigroup. https://www.refinitiv.com/en/products/datastream-macroeconomic-analysis?elqTrackId=80457d96caa94f51befc004f92d0ca04&elqaid=8607&elqat=2

Refinitiv Datastream. (n.d.). *USEPUINDXD* [Economic Policy Uncertainty Policy Index for the United States, Daily , Not Seasonally Adjusted]. Federal Reserve Bank of St. Lois. https://www.refinitiv.com/en/products/datastream-macroeconomic-analysis?elqTrackId=80457d96caa94f51befc004f92d0ca04&elqaid=8607&elqat=2

Refinitiv Datastream. (n.d.). *VSTOXXI* [VSTOXX Volatility Index]. https://www.refinitiv.com/en/products/datastream-macroeconomic-analysis?elqTrackId=80457d96caa94f51befc004f92d0ca04&elqaid=8607&elqat=2

S&P Global. (n.d.). *S&P Denmark BMI Growth Index*. Capital IQ. https://www.spglobal.com/marketintelligence/en/campaigns/sp-capital-iq-pro?utm_source=google&utm_medium=cpc&utm_campaign=CIQPro_Search_Google&utm_term=capital%20iq&utm_content=542966230023&_bt=542966230023&_bk=capital%20iq&_bm=e&_bn=g&_bg=128384020444&gcli

S&P Global. (n.d.). *S&P Denmark BMI Value Index*. Capital IQ. https://www.spglobal.com/marketintelligence/en/campaigns/sp-capital-iq-pro?utm_source=google&utm_medium=cpc&utm_campaign=CIQPro_Search_Google&utm_term=capital%20iq&utm_content=542966230023&_bt=542966230023&_bk=capital%20iq&_bm=e&_bn=g&_bg=128384020444&gcli

S&P Global. (n.d.). *S&P Denmark Investment Grade Corporate Bond Index*. Capital IQ. https://www.spglobal.com/marketintelligence/en/campaigns/sp-capital-iq-pro?utm_source=google&utm_medium=cpc&utm_campaign=CIQPro_Search_Google&utm_term=capital%20iq&utm_content=542966230023&_bt=542966230023&_bk=capital%20iq&_bm=e&_bn=g&_bg=128384020444&gcli

S&P Global. (n.d.). *S&P Denmark LargeMidCap Growth Index*. Capital IQ. https://www.spglobal.com/marketintelligence/en/campaigns/sp-capital-iq-pro?utm_source=google&utm_medium=cpc&utm_campaign=CIQPro_Search_Google&utm_term=capital%20iq&utm_content=542966230023&_bt=542966230023&_bk=capital%20iq&_bm=e&_bn=g&_bg=128384020444&gcli

S&P Global. (n.d.). *S&P Denmark LargeMidCap Value Index*. Capital IQ. https://www.spglobal.com/marketintelligence/en/campaigns/sp-capital-iq-pro?utm_source=google&utm_medium=cpc&utm_campaign=CIQPro_Search_Google&utm_term=capital%20iq&utm_content=542966230023&_bt=542966230023&_bk=capital%20iq&_bm=e&_bn=g&_bg=128384020444&gcli

S&P Global. (n.d.). *S&P Denmark SmallCap Growth Index*. Capital IQ. https://www.spglobal.com/marketintelligence/en/campaigns/sp-capital-iq-pro?utm_source=google&utm_medium=cpc&utm_campaign=CIQPro_Search_Google&utm_term=capital%20iq&utm_content=542966230023&_bt=542966230023&_bk=capital%20iq&_bm=e&_bn=g&_bg=128384020444&gcli

S&P Global. (n.d.). *S&P Denmark SmallCap Value Index*. Capital IQ. https://www.spglobal.com/marketintelligence/en/campaigns/sp-capital-iq-pro?utm_source=google&utm_medium=cpc&utm_campaign=CIQPro_Search_Google&utm_term=capital%20iq&utm_content=542966230023&_bt=542966230023&_bk=capital%20iq&_bm=e&_bn=g&_bg=128384020444&gcli

Schmeling, M. (2009). Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*, *16*, 394-408. 10.1016/j.jempfin.2009.01.002

Selden, G. C. (1912). *Psychology of the stock market: Human impulses lead to speculative disasters*. New York: Ticker Publishing.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*.

Sharpe, W. F. (1964, September). Capital Asset Prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19*(3), 421-442.

Shefrin, H., & Statman, M. (1994). . Behavioral capital asset pricing theory. *Journal of Financial and Quantitative Analysis*, *29*, 323-349.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, *71*, 421-436.

Shleifer, A., & Vishny, R. (1997). The limits of arbitrage. *Journal of Finance, 52,* 35-55.

Song, J., & Baek, C. (2019, June). Detecting structural breaks in realized volatility. *Computational Statistics & Data Analysis*, *132,* 58-75.

StatCounter GlobalStats. (2022). *Seach Engine Market Share Europe 2021*. Search Engine Market Share. https://gs.statcounter.com/search-engine-market-share/all/europe/2021

Statcounter GlobalStats. (2022). *Search Engine Market Share Denmark*. Search Engine Market Share. https://gs.statcounter.com/search-engine-market-share/all/denmark/2009

Statcounter GlobalStats. (2022). *Search Engine Market Share Denmark 2021*. Search Engine Market Share. https://gs.statcounter.com/search-engine-market-share/all/denmark/2021

Statcounter GlobalStats. (2022). *Search Engine Market Share Wordwide 2021*. Search Engine Market Share. https://gs.statcounter.com/search-engine-market-share/all/worldwide/2021

Stock, J. H., & Watson, M. W. (2014). *Introduction to Econometrics*. Pearson.

Sørensen, B. G. (2011). *Københavns Fondsbørs*. Den Store Danske. https://denstoredanske.lex.dk/K%C3%B8benhavns_Fondsb%C3%B8rs

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, *62*(3), 1139-1168. 10.2139/ssrn.685145

Tetlock, P. C., Saar-Tsechansy, M., & Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance*, *63*(3), 1437-1467.

Verma, R., & Verma, P. (2007, July). Noise trading and stock market volatility. *Journal of Multinational Financial Management*, *17*(3), 231-243.

Yahoo Finance. (2020). *This Day in Market History: The Google IPO*. Yahoo Finance. https://finance.yahoo.com/news/day-market-history-google-ipo-105500069.html

Yu, J., & Yuan, Y. (2011). Investor Sentiment and the Mean-Variance Relation. *Journal of Financial Economics*, *100*, 367–381.

13 APPENDIX

13.1 The OLS Estimator and the Adjusted R^2

The OLS Estimator

Most of the tests run in this paper utilize a multivariable OLS regression model, also called the Multiple Regression Model. The Multiple Regression model determined the values of the coefficients using the OLS (Ordinary Least Squares) Estimator. For a multiple regression $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$, the coefficients are calculated in the following steps (Stock and Watson, 2014, p. 239). First, we let $b_0, b_1 \ldots b_k$ be estimates of $\beta_0, \beta_1 \ldots \beta_k$. The predicted value of Y_i is calculated using $b_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$. The mistake in the prediction of Y_i is then

$$Y_i - (b_0 + b_1 X_1 \ldots \beta_k X_k) = (Y_i - b_0 - b_1 X_1 - \cdots - b_k X_k) \tag{X.1}$$

From this, we can calculate the squared prediction mistakes, given by:

$$\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_1 - \cdots - b_k X_k)^2 \tag{X.2}$$

The estimators of $\beta_0, \beta_1, \ldots, \beta_k$, is the coefficients that minimizes the sum of squares mistakes in equation (X.2) (Stock and Watson, 2014, p.239). The predicted value of Y_i given the variables X_1 to X_k is then denoted

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k + \hat{u}_t \tag{X.3}$$

Where $\hat{u}_t$ is OLS residuals, given by the difference actual Y_i values, and the OLS predicted $\hat{Y}_i$ values, that is:

$$\hat{u}_i = Y_i - \hat{Y}_i \tag{X.4}$$

The Adjusted R-Squared ($Adjusted\ R^2$)

The R^2 is the fraction of the sample variance of Y_i explained by (or predicted by) the regressors. The R^2 is given by

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} \tag{X.5}$$

where the ESS (explained sum of squares) = $\Sigma_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$ (The $\bar{Y}$ is the average of Y), the TSS (Total sum of squares) = $\Sigma_{i=1}^{n} (Y_i - \bar{Y})^2$, and SSR (sum of squared residuals) = $\Sigma_{i=1}^{n} \hat{u}_i^2$ (Stock and Watson, 243) . The problem is that for all added regressors (except regressors where the estimates coefficient is equal to zero), the R^2 increases, which makes it less than ideal for determining if including additional regressors increases the fit of the model. The reason is that if the OLS estimator determined the coefficient to any value different from zero, it must mean that the OLS estimators find that the error term is less than in the regression without the additional regressor, decreasing the SSR (Stock and Watson, 2014, p. 243). From equation (X.5), we see that this actually increases the R^2. Therefore, when using a multiple regression model, we need to study the Adjusted R^2, given by:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} \tag{X.6}$$

Whereby, the $\frac{n-1}{n-k-1}$ is always greater than 1, and increases for an additional regressor. Thus, whether the adjusted R^2 increases or not, is determined on which of the two effects, that is, the SSR decrease effect, or the $\frac{n-1}{n-k-1}$ effect is stronger (Stock and Watson, 2014, p. 243)

13.2 The Primitive Search Term list

BANKRUPT, BANKRUPTCY, BEGGAR, BLACKMAIL, BRIBE, BROKE, BUM, CHEAP, COMMONER, CORRUPT, COST, COSTLINESS, COSTLY, DEBTOR, DEFAULT, DEFICIT, DEPRECIATION, DEPRESSION, DESTITUTE, EXPENSE, EXPENSIVE, EXTRAVAGANT, FINE, FIRE, GAMBLE, HOLE, HUSTLE, HUSTLER, INFLATION, JOBLESS, LAID, LAY, LIQUIDATE, MISER, OWE, POOR, POVERTY, RECESSION, SQUANDER, TARIFF, UNDERWORLD, UNECONOMICAL, UNEMPLOYED, UNPROFITABLE, VAGANBOND, VAGRANT, WASTE, ACCRUE, AFFLUENCE, AFFLUENT, AFLOAT, ALLOWANCE, ARISTOCRACY, ARISTOCRAT, ARISTOCRATIC, ASSOCIATE, BACKER, BARGAIN, BENEFACTOR, BENEFICIARY, BENEFIT, BENEVOLENCE, BENEVOLENT, BEQUETH, BETROTH, BETROTHAL, BINUS, BOOM, BREADWINNER, BUY, CAPITALIZE, CHARITABLE, CHARITY, COMPENSATE, CONTRIBUTE, COOPERATIVE, DONATE, DONATION, ECONOMIZE, ENDOW, ENTREPRENUERIAL, EQUITY, FRUGAL, GAIN, GENEROSITY, GOLD, GUIDE, INEXPENSIVE, INHERIT, INVALUABLE, LUCRATIVE, LUXURY, MERITORIOUS, NOBILITY, NOBLEMAN, PARNTER, PARNTERSHIP, PATRON, PATRONAGE, PRECIOUS, PRICELESS, PRIVILEGED, PRODUCTIVE, PRODUCTIVITY, PROFIT, PROFITABLE, PROSPER, PROSPEROUS, RECOMPENSE, REWARD, RICH, RICHES, RICHNESS, SAVINGS, SECURITY, SUBSIDIZE, SUBSIDY, THRIFT, THRIFTY, TREASURE, VALUABLE, WORTH, INTERVENTION, RUIN, POLLUTION, SUCCESS, GIFT, PARTNERSHIP, GHETTO, SHORTAGE, FELLOWSHIP, COLONY, LIQUIDIATION

13.3 Output FEARS and Average Returns

FEARS and Control Variables regressed on $\Delta DKallshare$

	Dependent variable:					
	Returns (1)	`Return T+1` (2)	`Return T+2` (3)	`Return T+3` (4)	`Return T+4` (5)	`Return T+5` (6)
`FEARS 20`	0.002***	-0.002**	0.002**	-0.0005	0.0003	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
`Return Lag 1`	0.047***	-0.005	-0.022	-0.012	-0.034*	-0.033*
	(0.016)	(0.021)	(0.020)	(0.022)	(0.020)	(0.018)
`Return Lag 2`	-0.005	-0.021	-0.012	-0.031	-0.032*	-0.007
	(0.016)	(0.020)	(0.023)	(0.020)	(0.019)	(0.019)
`Return Lag 3`	-0.011	-0.011	-0.032	-0.031*	-0.007	-0.026
	(0.016)	(0.023)	(0.020)	(0.019)	(0.019)	(0.020)
`Return Lag 4`	-0.008	-0.032	-0.031*	-0.007	-0.028	0.003
	(0.018)	(0.019)	(0.018)	(0.019)	(0.020)	(0.019)
VSTOXX	-0.084***	-0.006**	0.004	0.004	0.005*	0.005**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
`EPU-US`	-0.0004*	0.0001	0.0001	0.001***	-0.0002	-0.0003
	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
`CESI-EUR`	-0.004	-0.001	0.005	0.003	0.001	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Constant	-0.00001	0.00001	0.00000	0.00005	-0.00003	-0.00003
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Observations	4,311	4,311	4,311	4,311	4,311	4,311
R2	0.315	0.005	0.006	0.005	0.004	0.004
Adjusted R2	0.314	0.003	0.004	0.003	0.002	0.002
Residual Std. Error (df = 4302)	0.008	0.010	0.010	0.010	0.010	0.010
F Statistic (df = 8; 4302)	247.645***	2.568***	2.982***	2.584***	2.152**	1.992**

13.4 FEARS and the cross section of returns – Broad Market

FEARS and Control Variables regressed on $\Delta BMIgrowth$ and $\Delta BMIvalue$

	Dependent variable:					
	`BMI Growth` (1)	`BMI Growth T+1` (2)	`BMI Growth T+2` (3)	`BMI Value` (4)	`BMI Value T+1` (5)	`BMI Value T+2` (6)
`FEARS 20`	0.002**	-0.002**	0.002**	0.002***	-0.003***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
`Return Lag 1`	0.067***	-0.010	-0.028	0.088***	-0.003	-0.017
	(0.022)	(0.030)	(0.026)	(0.020)	(0.033)	(0.027)
`Return Lag 2`	-0.013	-0.025	0.026	-0.006	-0.015	-0.013
	(0.023)	(0.025)	(0.029)	(0.024)	(0.026)	(0.031)
`Return Lag 3`	-0.008	0.023	-0.032	0.004	-0.013	-0.029
	(0.020)	(0.029)	(0.027)	(0.020)	(0.032)	(0.028)
`Return Lag 4`	0.028	-0.032	-0.049*	-0.009	-0.028	-0.033
	(0.023)	(0.027)	(0.027)	(0.024)	(0.027)	(0.029)
VSTOXX	-0.094***	-0.007*	0.005	-0.110***	-0.012***	0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
`EPU-US`	-0.001***	0.0001	0.0003	-0.001*	-0.0002	0.0003
	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
`CESI-EUR`	-0.005	-0.003	0.002	0.0003	0.007*	0.008
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.007)
Constant	-0.00005	0.00000	0.00002	-0.00002	-0.00001	0.00000
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Observations	4,209	4,209	4,209	4,209	4,209	4,209
R2	0.245	0.004	0.005	0.345	0.008	0.004
Adjusted R2	0.243	0.002	0.003	0.344	0.006	0.002
Residual Std. Error (df = 4200)	0.011	0.013	0.013	0.010	0.012	0.012
F Statistic (df = 8; 4200)	170.023***	2.070**	2.747***	276.838***	4.085***	2.163**

13.5 FEARS and the cross section of returns – Small Cap Market

FEARS and Control Variables regressed on $\Delta SMLgrowth$ and $\Delta SMLvalue$

	Dependent variable:					
	`SML Growth` (1)	`SML Growth T+1` (2)	`SML Growth T+2` (3)	`SML Value` (4)	`SML Value T+1` (5)	`SML Value T+2` (6)
`FEARS 20`	0.001	−0.002***	0.002**	0.001**	−0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
`Return Lag 1`	0.144***	0.038	0.003	0.141***	0.042	0.015
	(0.023)	(0.031)	(0.027)	(0.021)	(0.027)	(0.024)
`Return Lag 2`	0.033	0.004	0.007	0.038*	0.013	−0.008
	(0.024)	(0.026)	(0.029)	(0.020)	(0.024)	(0.027)
`Return Lag 3`	0.022	0.006	−0.035	0.031	−0.008	−0.015
	(0.021)	(0.029)	(0.025)	(0.019)	(0.027)	(0.024)
`Return Lag 4`	0.012	−0.034	−0.025	−0.001	−0.013	−0.011
	(0.023)	(0.025)	(0.028)	(0.021)	(0.024)	(0.025)
VSTOXX	−0.097***	−0.013***	0.003	−0.103***	−0.015***	−0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
`EPU-US`	−0.0002	−0.0001	−0.0002	−0.0002	−0.0003	0.00004
	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0004)
`CESI-EUR`	−0.007	−0.002	0.004	−0.004	0.006*	0.006
	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.005)
Constant	0.00001	0.00001	−0.00001	0.00000	−0.00000	0.00000
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Observations	4,209	4,209	4,209	4,209	4,209	4,209
R2	0.289	0.009	0.003	0.332	0.010	0.003
Adjusted R2	0.288	0.007	0.001	0.330	0.008	0.001
Residual Std. Error (df = 4200)	0.010	0.012	0.012	0.010	0.012	0.012
F Statistic (df = 8; 4200)	213.308***	4.627***	1.671	260.586***	5.506***	1.660

13.6 FEARS and the cross section of returns – Large Cap Market

FEARS and Control Variables regressed on $\Delta LMgrowth$ and $\Delta LMvalue$

	Dependent variable:					
	`LM Growth` (1)	`LM Growth T+1` (2)	`LM Growth T+2` (3)	`LM Value` (4)	`LM Value T+1` (5)	`LM Value T+2` (6)
`FEARS 20`	0.002**	−0.002*	0.002**	0.002***	−0.003***	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
`Return Lag 1`	0.055**	−0.018	−0.033	0.075***	−0.013	−0.026
	(0.023)	(0.030)	(0.027)	(0.021)	(0.034)	(0.028)
`Return Lag 2`	−0.021	−0.030	0.029	−0.016	−0.022	−0.013
	(0.024)	(0.026)	(0.029)	(0.025)	(0.028)	(0.033)
`Return Lag 3`	−0.014	0.026	−0.032	−0.003	−0.014	−0.030
	(0.021)	(0.030)	(0.028)	(0.021)	(0.033)	(0.029)
`Return Lag 4`	0.030	−0.032	−0.052*	−0.010	−0.029	−0.034
	(0.024)	(0.028)	(0.027)	(0.026)	(0.029)	(0.031)
VSTOXX	−0.093***	−0.006	0.006	−0.111***	−0.011***	0.004
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
`EPU-US`	−0.001***	0.0001	0.0004	−0.001**	−0.0002	0.0003
	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0004)
`CESI-EUR`	−0.005	−0.003	0.001	0.003	0.008*	0.009
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.009)
Constant	−0.0001	0.00000	0.00003	−0.00003	−0.00002	0.00000
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Observations	4,209	4,209	4,209	4,209	4,209	4,209
R2	0.218	0.004	0.005	0.319	0.007	0.004
Adjusted R2	0.216	0.002	0.003	0.318	0.005	0.002
Residual Std. Error (df = 4200)	0.012	0.013	0.013	0.011	0.013	0.013
F Statistic (df = 8; 4200)	146.292***	1.874*	2.792***	245.840***	3.860***	2.241**

13.7 Granger Causality Test output – FEARS on Average Returns

Granger Causality Test R-Output for the null hypothesis that FEARS does not Granger Cause $\Delta DKallshare$ for up to five lags.

```
> print(test1)
Granger causality test

Model 1: data$Returns ~ Lags(data$Returns, 1:1) + Lags(data$`FEARS 20`, 1:1)
Model 2: data$Returns ~ Lags(data$Returns, 1:1)
  Res.Df Df      F  Pr(>F)
1   4317
2   4318 -1 6.4854 0.01091 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> print(test2)
Granger causality test

Model 1: data$Returns ~ Lags(data$Returns, 1:2) + Lags(data$`FEARS 20`, 1:2)
Model 2: data$Returns ~ Lags(data$Returns, 1:2)
  Res.Df Df      F   Pr(>F)
1   4314
2   4316 -2 4.6213 0.009889 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> print(test3)
Granger causality test

Model 1: data$Returns ~ Lags(data$Returns, 1:3) + Lags(data$`FEARS 20`, 1:3)
Model 2: data$Returns ~ Lags(data$Returns, 1:3)
  Res.Df Df      F  Pr(>F)
1   4311
2   4314 -3 3.1055 0.02547 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> print(test4)
Granger causality test

Model 1: data$Returns ~ Lags(data$Returns, 1:4) + Lags(data$`FEARS 20`, 1:4)
Model 2: data$Returns ~ Lags(data$Returns, 1:4)
  Res.Df Df      F  Pr(>F)
1   4308
2   4312 -4 2.3966 0.04817 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> print(test5)
Granger causality test

Model 1: data$Returns ~ Lags(data$Returns, 1:5) + Lags(data$`FEARS 20`, 1:5)
Model 2: data$Returns ~ Lags(data$Returns, 1:5)
  Res.Df Df      F   Pr(>F)
1   4305
2   4310 -5 2.1128 0.06096 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

13.8 Granger Causality Test output –Average Returns on FEARS

Reversed Granger Causality Test R-output for the null hypothesis that $\Delta DKallshare$ does not Granger Cause FEARS.

```
> print(revtest1)
Granger causality test

Model 1: data$`FEARS 20` ~ Lags(data$`FEARS 20`, 1:1) + Lags(data$Returns, 1:1)
Model 2: data$`FEARS 20` ~ Lags(data$`FEARS 20`, 1:1)
  Res.Df Df      F Pr(>F)
1   4317
2   4318 -1 1.0885 0.2969
```

```
> print(revtest2)
Granger causality test

Model 1: data$`FEARS 20` ~ Lags(data$`FEARS 20`, 1:2) + Lags(data$Returns, 1:2)
Model 2: data$`FEARS 20` ~ Lags(data$`FEARS 20`, 1:2)
  Res.Df Df     F Pr(>F)
1   4314
2   4316 -2 0.614 0.5412
```

```
> print(revtest3)
Granger causality test

Model 1: data$`FEARS 20` ~ Lags(data$`FEARS 20`, 1:3) + Lags(data$Returns, 1:3)
Model 2: data$`FEARS 20` ~ Lags(data$`FEARS 20`, 1:3)
  Res.Df Df      F Pr(>F)
1   4311
2   4314 -3 0.9201 0.4302
```

```
> print(revtest4)
Granger causality test

Model 1: data$`FEARS 20` ~ Lags(data$`FEARS 20`, 1:4) + Lags(data$Returns, 1:4)
Model 2: data$`FEARS 20` ~ Lags(data$`FEARS 20`, 1:4)
  Res.Df Df      F Pr(>F)
1   4308
2   4312 -4 0.6097 0.6557
```

```
> print(revtest5)
Granger causality test

Model 1: data$`FEARS 20` ~ Lags(data$`FEARS 20`, 1:5) + Lags(data$Returns, 1:5)
Model 2: data$`FEARS 20` ~ Lags(data$`FEARS 20`, 1:5)
  Res.Df Df      F Pr(>F)
1   4305
2   4310 -5 0.6683 0.6475
```

13.9 FEARS and Government bonds– 5 Year Bond results

FEARS and Controls regressed on $\Delta 5Ybond$

	Dependent variable:		
	`5Y` (1)	`5Y T+1` (2)	`5Y T+2` (3)
`FEARS 20`	0.001	0.00002	0.002*
	(0.001)	(0.001)	(0.001)
`5Y lag 1`	-0.032	0.004	-0.005
	(0.026)	(0.022)	(0.020)
`5Y lag 2`	-0.002	-0.005	0.016
	(0.021)	(0.023)	(0.019)
`5Y lag3`	-0.006	0.015	0.017
	(0.023)	(0.020)	(0.018)
`5Y lag4`	0.015	0.018	-0.031
	(0.020)	(0.019)	(0.022)
VSTOXX	-0.055***	0.008*	0.005
	(0.005)	(0.005)	(0.004)
`EPU-US`	-0.0005	-0.001	0.0005
	(0.0004)	(0.0004)	(0.0004)
`CESI-EUR`	0.019***	0.005	-0.002
	(0.005)	(0.005)	(0.004)
Constant	-0.00003	-0.00003	0.00003
	(0.0002)	(0.0002)	(0.0002)
Observations	4,085	4,085	4,085
R2	0.061	0.002	0.003
Adjusted R2	0.059	0.001	0.001
Residual Std. Error (df = 4076)	0.015	0.015	0.015
F Statistic (df = 8; 4076)	33.062***	1.264	1.516

13.10 FEARS and Government bonds– 10 Year Bond results

FEARS and Controls regressed on $\Delta 10Ybond$

	Dependent variable:		
	`10Y` (1)	`10Y T+1` (2)	`10Y T+2` (3)
`FEARS 20`	0.001	-0.0001	0.0003
	(0.001)	(0.001)	(0.001)
`10Y lag 1`	0.024	0.016	-0.013
	(0.026)	(0.029)	(0.027)
`10Y lag 2`	0.010	-0.013	0.031
	(0.027)	(0.030)	(0.021)
`10Y lag3`	-0.013	0.029	0.006
	(0.027)	(0.021)	(0.022)
`10Y lag4`	0.027	0.006	-0.035*
	(0.021)	(0.021)	(0.021)
VSTOXX	-0.049***	0.002	0.008**
	(0.004)	(0.004)	(0.004)
`EPU-US`	-0.0002	-0.001*	0.0004
	(0.0004)	(0.0004)	(0.0004)
`CESI-EUR`	0.019**	0.0002	-0.002
	(0.007)	(0.004)	(0.005)
Constant	0.00000	-0.00003	0.00003
	(0.0002)	(0.0002)	(0.0002)
Observations	4,089	4,089	4,089
R2	0.065	0.002	0.004
Adjusted R2	0.063	-0.00001	0.002
Residual Std. Error (df = 4080)	0.013	0.014	0.014
F Statistic (df = 8; 4080)	35.577***	0.995	2.100**

13.11 FEARS and Corporate Bond

FEARS and Controls regressed on $\Delta IGbond$

	`IG Resid` (1)	`IG Resid T+1` (2)	`IG RESID T+2` (3)
`FEARS 20`	-0.0001	0.00003	-0.00003
	(0.0001)	(0.0001)	(0.0001)
`IG lag 1`	0.007	0.025	0.043*
	(0.025)	(0.023)	(0.024)
`IG lag 2`	0.024	0.041	0.021
	(0.023)	(0.025)	(0.024)
`IG lag 3`	0.039*	0.023	-0.016
	(0.023)	(0.026)	(0.019)
`IG lag 4`	0.021	-0.017	0.008
	(0.023)	(0.019)	(0.020)
VSTOXX	0.003***	-0.001	-0.001***
	(0.0005)	(0.001)	(0.0005)
`EPU-US`	0.00001	0.0001	-0.00002
	(0.00004)	(0.00005)	(0.00004)
`CESI-EUR`	-0.001*	-0.0004	-0.0001
	(0.001)	(0.0004)	(0.0003)
Constant	0.00001	0.00001	0.00000
	(0.00002)	(0.00003)	(0.00002)
Observations	4,312	4,312	4,312
R2	0.018	0.005	0.006
Adjusted R2	0.017	0.003	0.004
Residual Std. Error (df = 4303)	0.002	0.002	0.002
F Statistic (df = 8; 4303)	10.131***	2.643***	3.178***

13.12 Volatility testing output

Full Volatility testing R-output

(1) Full period

```
Mode 1 Coefficients:
             Estimate   Std. Error  z-value   Pr(>|z|)
phi(1)     0.89940955  0.03792097 23.71800 < 2.22e-16 ***
phi(2)     0.05895146  0.03518522  1.67546   0.093844 .
theta(1)   0.91868462  0.03604164 25.48953 < 2.22e-16 ***
d.f        0.34370503  0.05350701  6.42355 1.3313e-10 ***
X1         0.02319701  0.01683746  1.37770   0.168295
X2        -0.01518101  0.00773886 -1.96166   0.049802 *
X3        -0.02782041  0.08692526 -0.32005   0.748931
Intercept -0.00126994  0.17663522 -0.00719   0.994264
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
sigma^2 estimated as 0.080228; Log-likelihood = 4978.63; AIC = -9939.27; BIC = -9882.74

Numerical Correlations of Coefficients:
          phi(1) phi(2) theta(1) d.f   X1    X2    X3    Intercept
phi(1)     1.00  -0.89  -0.11    -0.67 -0.01  0.00  0.01  0.00
phi(2)    -0.89   1.00   0.53     0.87  0.01  0.00 -0.01  0.00
theta(1)  -0.11   0.53   1.00     0.76  0.00  0.00 -0.01  0.00
d.f       -0.67   0.87   0.76     1.00  0.01 -0.01 -0.02  0.00
X1        -0.01   0.01   0.00     0.01  1.00  0.04 -0.01  0.00
X2         0.00   0.00   0.00    -0.01  0.04  1.00 -0.01  0.00
X3         0.01  -0.01  -0.01    -0.02 -0.01 -0.01  1.00  0.00
Intercept  0.00   0.00   0.00     0.00  0.00  0.00  0.00  1.00
```

(2) After the Financial Crisis

```
arfima::arfima(z = data1, order = c(2, 0, 1), numeach = c(1, 1), xreg = xreg.m)


Mode 1 Coefficients:
             Estimate   Std. Error      z-value   Pr(>|z|)
phi(1)     8.44771e-01  3.16544e-02     26.68731 < 2.22e-16 ***
phi(2)     1.42622e-01  2.05815e-02      6.92962 4.2198e-12 ***
theta(1)   9.67573e-01  4.68828e-02     20.63810 < 2.22e-16 ***
d.f        4.99156e-01  6.32376e-07 789333.73256 < 2.22e-16 ***
X1         3.87521e-02  1.92666e-02      2.01136   0.044287 *
X2        -1.61999e-02  9.05759e-03     -1.78854   0.073689 .
X3        -3.83681e-01  9.00944e-02     -4.25865 2.0566e-05 ***
Intercept -2.86002e-02  9.80459e+00     -0.00292   0.997673
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
sigma^2 estimated as 0.0780914; Log-likelihood = 3868.4; AIC = -7718.8; BIC = -7664.63

Numerical Correlations of Coefficients:
          phi(1) phi(2) theta(1) d.f   X1    X2    X3    Intercept
phi(1)     1.00  -0.88   0.82    -0.01 -0.01 -0.01 -0.04  0.00
phi(2)    -0.88   1.00  -0.49     0.00  0.02  0.01  0.01  0.00
theta(1)   0.82  -0.49   1.00    -0.01 -0.01  0.00 -0.06  0.01
d.f       -0.01   0.00  -0.01     1.00  0.00  0.00  0.00  0.00
X1        -0.01   0.02  -0.01     0.00  1.00  0.05 -0.02  0.00
X2        -0.01   0.01   0.00     0.00  0.05  1.00 -0.01  0.00
X3        -0.04   0.01  -0.06     0.00 -0.02 -0.01  1.00  0.00
Intercept  0.00   0.00   0.01     0.00  0.00  0.00  0.00  1.00
```

(3) Before the Financial Crisis

```
Mode 1 Coefficients:
           Estimate Std. Error  z-value   Pr(>|z|)
phi(1)    -0.0778420  0.3832754 -0.20310   0.839059
phi(2)    -0.0658068  0.0616847 -1.06683   0.286050
theta(1)   0.0375978  0.3974921  0.09459   0.924642
d.f        0.4376075  0.0542771  8.06247 7.4765e-16 ***
X1        -0.0218104  0.0459130 -0.47504   0.634759
X2        -0.0416338  0.0180942 -2.30095   0.021395 *
X3         0.6708193  0.5447018  1.23154   0.218123
Intercept -0.0470413  0.3113679 -0.15108   0.879913
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
sigma^2 estimated as 0.106288; Log-likelihood = 600.506; AIC = -1183.01; BIC = -1144.47

Numerical Correlations of Coefficients:
          phi(1) phi(2) theta(1) d.f   X1    X2    X3    Intercept
phi(1)     1.00   0.60   0.98     0.21 -0.03  0.03 -0.04  0.01
phi(2)     0.60   1.00   0.54    -0.26 -0.03  0.04 -0.01 -0.01
theta(1)   0.98   0.54   1.00     0.34 -0.02  0.03 -0.04  0.01
d.f        0.21  -0.26   0.34     1.00  0.01 -0.01  0.01  0.03
X1        -0.03  -0.03  -0.02     0.01  1.00  0.03 -0.08  0.00
X2         0.03   0.04   0.03    -0.01  0.03  1.00 -0.05  0.00
X3        -0.04  -0.01  -0.04     0.01 -0.08 -0.05  1.00  0.00
Intercept  0.01  -0.01   0.01     0.03  0.00  0.00  0.00  1.00
```

(4) During the Financial Crisis

```
Mode 1 Coefficients:
            Estimate  Std. Error  z-value Pr(>|z|)
phi(1)    -0.46637757  0.61465437 -0.75876 0.447994
phi(2)    -0.10009419  0.10691118 -0.93624 0.349151
theta(1)  -0.28875540  0.61854668 -0.46683 0.640622
d.f        0.48233194  0.02309329 20.88624  < 2e-16 ***
X1         0.00640055  0.04616276  0.13865 0.889725
X2         0.03701530  0.02568852  1.44093 0.149605
X3        -0.49975613  0.27883687 -1.79229 0.073087 .
Intercept  0.57529625  0.58847260  0.97761 0.328268
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
sigma^2 estimated as 0.0687474; Log-likelihood = 504.89; AIC = -991.779; BIC = -956.389

Numerical Correlations of Coefficients:
          phi(1) phi(2) theta(1) d.f   X1    X2    X3    Intercept
phi(1)     1.00   0.88   1.00     0.12  0.13 -0.07 -0.01  0.00
phi(2)     0.88   1.00   0.86    -0.04  0.11 -0.08  0.01  0.00
theta(1)   1.00   0.86   1.00     0.16  0.12 -0.06 -0.01  0.00
d.f        0.12  -0.04   0.16     1.00 -0.01  0.01 -0.01 -0.01
X1         0.13   0.11   0.12    -0.01  1.00 -0.06  0.06  0.00
X2        -0.07  -0.08  -0.06     0.01 -0.06  1.00  0.04  0.00
X3        -0.01   0.01  -0.01    -0.01  0.06  0.04  1.00  0.00
Intercept  0.00   0.00   0.00    -0.01  0.00  0.00  0.00  1.00
```

13.13 Randomly Generates FEARS on Average Returns

Table for the random-search term FEARS and Controls regressed on $DKallshare$

Panel A: Random Search Terms FEARS and Returns			
	Returns		
	(1)	(2)	(3)
	Returns	Returns (T+1)	Returns (T+2)
FEARS	-0.0010	0.0005	0.0000
	(0.0005)	(0.001)	(0.001)
Lag 1	0.047***	-0.040	-0.022
Lag 2	-0.0500	-0.021	-0.012
Lag 3	-0.0110	-0.011	-0.032
Lag 4	-0.0080	-0.031	-0.031*
VSTOXX	-0.083***	-0.006	0.004
EPU US	-0.001***	0.0001	0.000
CESI EUR	-0.0001	0.000	0.005*
Adjusted R2	0.312	0.002	0.002

The table depicts a test run on a "search terms randomly generated" FEARS index, where the 20 search terms are randomly selected out of the 87 possible search terms. The index is dynamic, as is the original FEARS index, and the search terms included, changes every 120 trading days. As can be seen, the results return no significant coefficients on the FEARS, and coefficients are economically negligible. The results aim to alleviate the concern that our FEARS index represents neutral or random activity.