Lappeenranta-Lahti University of Technology
School of Business and Management
Strategic Finance and Analytics

Predicting OMX Helsinki stock prices using social media sentiment of Finnish retail investors

Author: Jani Karttunen
1st supervisor: Azzurra Morreale
2nd supervisor: Jan Stoklasa

ABSTRACT

Author:	Jani Karttunen
Title:	Predicting OMX Helsinki stock prices using social media sentiment of Finnish retail investors
Faculty:	LUT, School of Business and Management
Master's program:	Strategic Finance and Analytics
Year:	2021
Master's thesis:	Lappeenranta-Lahti University of Technology 56 pages, 14 tables, 6 figures, and 5 appendices
Examiners:	Azzurra Morreale, Jan Stoklasa
Keywords:	sentiment analysis, behavioral finance, machine learning, classifier algorithms, Naïve bayes, VAR models, Granger causality

Sentiment analysis uses machine learning to interpret moods from text. There have been studies to see whether sentiment analysis can be used to measure investor sentiment and forecast asset prices, but their results have been conflicting. This thesis aims to further study the relationship between investor sentiment and stock prices by studying the effect of Finnish investor sentiment towards OMX Helsinki stock prices. The study consists of classifying the sentiment social media posts about individual stocks using Naïve Bayes classifier to create stock-specific investor sentiment time series, which were then used as regressors in VAR models aiming to predict the future stock prices. The results of this study reveal that there is no predictive power in the Finnish investor sentiment as the prediction errors and price direction forecast accuracies do not improve with the inclusion of sentiment in the models. This conclusion was further confirmed with Granger causality analysis, which could not find any predictive power in sentiment towards stock prices.

TIIVISTELMÄ

Tekijä:	Jani Karttunen
Otsikko:	OMX Helsinki osakehintojen ennustaminen piensijoittajien sosiaalisen median sentimentin avulla
Tiedekunta:	LUT, School of Business and Management
Maisteriohjelma:	Strategic Finance and Analytics
Vuosi:	2021
Pro gradu -tutkielma:	Lappeenrannan-Lahden teknillinen yliopisto LUT 56 sivua, 14 taulukkoa, 6 kuviota, ja 5 liitettä
Tarkastajat:	Azzurra Morreale, Jan Stoklasa
Hakusanat:	sentimenttianalyysi, behavioraalinen rahoitus, koneoppiminen, klassifiointi algoritmit, naiivi Bayes, VAR-mallit, Granger-kausaalisuus

Sentimenttianalyysissä koneoppimista käytetään tunnistamaan mielialoja tekstistä. Useat tutkimukset ovat tutkineet voisiko sentimenttianalyysiä käyttää sijoittajien sentimentin mittaamiseen ja investointikohteiden hintojen ennustamiseen, mutta tulokset ovat olleet ristiriitaisia. Tämä tutkielma pyrkii tutkimaan sijoittajasentimentin ja osakehintojen välistä yhteyttä lisää tarkastelemalla suomalaisen sijoittajasentimentin vaikutusta OMX Helsingin osakehintoihin. Tutkimuksessa sosiaalisen median viestien sentimentti tunnistettiin naiivilla Bayesin luokittimella ja koottiin osakekohtaisiksi sentimenttiaikasarjoiksi, joita käytettiin selittävinä muuttujina osakehintoja ennustavissa VAR malleissa. Tutkimuksen tulokset eivät löytäneet mitään ennustavaa voimaa sijoittajien sentimentistä, sillä ennustusvirheet tai hintojen suunnan ennustamisen tarkkuus eivät parantuneet sentimenttimuuttujan mukaanotosta huolimatta. Tämä johtopäätös varmistettiin Granger-kausaalisuusanalyysillä, mikä ei myöskään löytänyt mitään ennustavaa voimaa sentimentistä osakehintoja kohtaan.

TABLE OF CONTENTS

LIST OF APPENDICIES

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Social media has become increasingly prevalent in the daily lives of people and allows them to freely share their opinions and thoughts on wide variety of topics. These social media posts are usually publicly accessible by anyone and incorporate enormous amounts of information about what people are discussing and thinking in almost real-time. However, the number of posts is so vast, and constantly increasing, that manually analyzing even a fraction of them would be impossible for a human. To solve this problem, many automated systems and algorithms have been developed to capture the information content of these posts automatically. One of the most common of these methods is sentiment analysis, which uses machine learning to detect moods, attitudes, emotions, or opinions from text. In the simplest analyses the text is classified as ‘positive’, ‘negative’ or ‘neutral’, but there is no real limit on what emotions could be detected. (Si, Mukherjee, Liu, Li, Li, Deng 2013; Priyani, Madhavi & Singh 2017)

The most popular targets for sentiment analysis have been consumer reviews, news articles, and social media posts. The potential applications for the technology are wide, and can range from business to healthcare, and thus the academic interest towards sentiment analysis has risen exponentially during the past couple of years. The progress is not only caused by the beforementioned increased availability of opinionated data but also by the progress made in artificial intelligence and machine learning, which has created the tools required for automated analysis. This all is made possible by reduced cost of computational power and data storage. (Priyani, Madhavi & Singh 2017)

According to Mittal & Goel (2012) there have been numerous attempts to apply sentiment analysis in stock price prediction even though the generally accepted efficient market hypothesis rejects the possibility of such task being possible. The attempters believe in an alternative school of thought, behavioral finance, which assesses that sentiment of individual investors has an effect on their decision making and thus on the asset prices, at least in the short term. Therefore, sentiment analysis could be used to capture and measure investor sentiment, which in turn could be used to create a predictive model for asset prices. At first, investors sentiment was attempted to be captured and measured by conducing sentiment analysis on online news articles but recently academics have shifted their focus onto more direct source, social media,

which reflects the thoughts of its users, including investors, in almost real time (Yu, Duan & Cao 2013; Nguyen, Shiari, Velcin 2015). The rationale for using social media sentiment in predictions is simple; If the decision making of individuals is affected by their current sentiment, then the sentiment could be used to predict their behavior and thus the asset prices, at least in the short term (Si et al. 2013).

1.1 Motivation

Although academic interest towards the topic has risen in the recent years there still is no definitive answer on whether social media sentiment can be used to predict asset prices. Many studies have found statistically significant relationships between the two but almost as many studies have been unable replicate those results. Results of past research are further elaborated later in the literature review. If the asset prices could be predicted with social media sentiment, this would allow institutional players to use their resources to predict any trends, booms, busts, or irrationalities, and plan their actions accordingly. It would also open opportunities to manipulate asset prices by interfering in the discussion. Therefore, there still is clear need for more studies on the subject. This research aims to bring more evidence for or against the predictive relationship between social media sentiment and asset prices.

Majority of past research has been conducted using English-speaking social media and markets. There are only limited number of studies on sentiment's effect on stock prices in more minor languages and markets like the Finnish language and market, for example. For example, Grigaliūnienė & Cibulskienė (2010) studied the said relationship in all Nordic markets but focused on larger country-wide portfolios instead of individual assets. More importantly, the study relied on using consumer confidence indices as proxies for sentiment, which limited the frequency of data to monthly. Ali, Ahmed & Östermark (2020) studied investor sentiment's effect on Finnish stock market but also had chosen to use volatility index as a proxy as opposed to attempting to use social media sentiment. This gives additional motivation for the study, as social media sentiment has not been used to measure investor sentiment in the Finnish markets. Other benefit social media derived investor sentiment has over the commonly used proxies is that social media posts become available in real-time. This gives more time

to abuse the sentiment information before it is reflected in the asset prices and thus could make social media posts better suited for asset price forecasting.

Finnish stock market itself is also an interesting case for the study because of some of its special characteristics. According to Jakobson & Korkeamäki (2014), Finnish population generally is averse towards investing their money into stock market and as such the number of retail investors is rather low. The market also suffers from low liquidity and high volatility. These characteristics, especially during downturns, have deterred international investors from seeing Finnish stock market as particularly desirable investment target. Thus, the market is rather isolated, and it can be assumed that the market prices are not influenced much by foreigners, allowing the research to focus on only Finnish language social media. Based on the research by Lindén, Jauhianen & Hardwick (2020) the current sentiment analysis methods are able to only reach around 60% accuracy for Finnish language text. However, there is only small number of instances where positive is text was mistaken for negative and vice versa. They also note that even different humans did not agree with the sentiment all the time. It is also interesting to see whether the results from the Finnish market are similar to those gotten from studying the English market.

1.2 Literature review

Antweiler & Frank (2004) were among the first to use sentiment analysis to measure investor sentiment. They collected 1.5 million messages from Yahoo Finance message boards and classified them into ‘buy’, ‘sell’, and ‘hold’ groups using a Naïve Bays classifier they had trained with a small manually labeled sample. The usage of sentiment analysis allowed them to study the intraday effect as previously used proxy indices were usually only available as monthly aggregates. Their research found that net-positiveness could be used to predict volatility and asset prices. The relationship towards asset prices was statistically significant but the economic size was small, leading to believe that at least major excess returns are not achievable with sentiment analysis. Regardless of its results, the study’s methodology has since been used in later studies, which have either simply reused it or at least used it as a basis for more advanced methods.

Rao & Srivastava (2012) also used a Naïve Bayes classifier to capture daily investor sentiment from Twitter posts made between June 2010 and June 2011. Their methodology included Granger causality analysis, which showed that the returns forecast was improved with the inclusion of sentiment data. Ho, Damien, Bu & Konana (2017) and Piñeiro-Chousa, López-Cabarcos, Pérez-Pico & Ribeiro-Navarrete (2018) all independently found evidence for predictive power in sentiment toward asset prices. They all employed similar methodology of using a basic machine learning classifier, or alternatively a dataset with pre-labeled sentiment, and then attempted to create predictive models with the sentiment and prices as variables. Research by Checkley, Higón, Añón & Alles (2017) also confirmed that there is a clear link from investor sentiment to returns, volatility, and trading volume each. They, however, note that the link is stronger with the latter two, with price direction being more difficult to accurately predict using investor sentiment. Audrino, Sigrist & Ballinari (2020) focus their research on sentiment's effect on volatility. They used non-linear classification and HAR models and found out that the sentiment does help predictions in short term. There however was a difference based on the size and type of company; Smaller firms or ones whose stock is held mainly by institutions could not be predicted as well.

Ranco, Aleksovski, Caldarelli, Grčar & Mozetič (2015) measured investor sentiment from Twitter and used a Support Vector Machine classifier to label the data. Using Granger causality analysis, they did not find any significant predictive power in the measured sentiment data and, in addition, the Pearson correlation between the datasets was low. This is contrary to other studies presented but shows how there still is not definite proof whether reliably predicting asset prices with, or without, sentiment is possible.

Baker & Wurgler (2007) state that measurement of investor sentiment correctly is very difficult, and thus the used sentiment set has major impact on the results. Ho et al. (2017) also found evidence that the effect of sentiment depends on the timeframe being studied. In their study the coefficients of sentiment data were more significant during time periods of economic stability. Deng, Huang, Sinha & Zhao (2018) also state that what dataset is chosen has enormous impact on the results of sentiment studies.

Nguyen et al. (2015) compared multiple different classifiers and used them to predict whether prices would go up or down. The classifier-based predictive models were also

compared against one with manually labeled sentiment and a model using past prices as the sole predictor. The results showed that all sentiment-including models outperformed the price only model. However, even the best models did not get average correct sign percentages higher than 54%, which they argue is a number that could also be acquired from a model randomly guessing the direction. There was some variability between stocks and for some individual stocks the model was able to reach correct sign percentage of 70%, which can be considered significantly accurate. Derakhshan & Beigy (2019) employed similar methodology but included a Turkish sentiment dataset in addition to an English one. Their results on the English dataset were similar to those by Nguyen et al. (2015) but when using the Turkish dataset, the average correct sign percentages were found to be slightly lower than those of the English one. Therefore, the accuracy of models in different languages should be expected to be different. This difference could be attributed to different availability and quality of natural language processing tools for different languages or fundamental differences between individual cultures or markets.

1.3 Research question

This thesis studies the effect of Finnish retail investor sentiment towards the future prices of stocks traded on OMX Helsinki. The retail investor sentiment is assumed to be captured by the semantic orientation of social media posts made about these stocks. The study is conducted by building predictive models with past sentiment and past prices predicting the current price and analyzing their predictive power. The research question for the study is:

> “Can social media sentiment be used to predict OMX Helsinki stock prices?”

The retail investor sentiment is going to be captured by analyzing Finnish language posts made on the local financial newspaper Kauppalehti’s discussion forums and aggregating the sentiment into a time series representing the net-positiveness towards a stock. This is in line with the past research by Antweiler & Frank (2004) and others. The classification of posts into positive, neutral, and negative classes is done using a

Naïve Bayes classifier because it has been commonly used in past research and is known to achieve accuracies that can rival those of more advanced and complex classifiers while also being easy to understand and computationally cheap to run. The predictive relationship itself is studied by building models with and without the sentiment regressor and comparing accuracies of their predictions. If the sentiment can be used to predict stock prices the model with sentiment should perform better. This can be further confirmed by using a Granger causality analysis, which tests whether the effect of sentiment on price can be considered statistically significant. Methodology similar to this has been previously used in research by Rao & Srivastava (2012), Ranco et al. (2015) and Checkley et al. (2017) who all studied English-speaking sentiment and markets.

1.4 Limitations

Measuring investor sentiment is challenging and how it is done has great effect on the results of these types of studies (Baker & Wurgler 2007). In addition, majority of natural language processing methods have been developed with English language in mind. This leaves the possibility that when applied on other languages, they are either less accurate in best-case scenario, or completely unusable in the worst-case scenario. Especially since Finnish and English belong to completely different language groups. As no alternatives have yet been developed the possibility for lower classification accuracy has to be accepted but one should bear this in mind when interpreting the results.

The research is also limited to only include social media discussion about stocks traded on OMX Helsinki stock exchange. This filtering is easy to implement as Kauppalehti's forums have divided discussion for each stock into their own threads. This causes any effects sentiment might have towards stocks traded elsewhere or other asset types altogether is not considered for this study. In addition, only the most popular stocks were considered to limit the data needed to be collected. This also removes stocks without any posts where sentiment would be a constant zero, or where it would be affected too much by one or two individuals. In addition, as the social media posts are all collected from a single source, a financial message board, the dataset is more focused on retail investors and does not include much of unrelated noise. However, an

assumption must be made that the opinions of this message board's users reflect the opinions of investors in general. If the assumption is false, the results cannot be generalized. This limitation is necessary to limit the amount of data needed to be collected and preprocessed. Past research has also relied on single sources for their sentiment data. For example, Antweiler & Frank (2004) and Nguyen et al. (2015) both collected data only from Yahoo Finance message boards.

The study also is limited to only use social media posts made in 2019, from the first trading day in January until the last trading day in December. This limitation is also made to limit the amount of data needed to be collected. However, a timeframe of one year or less should be enough that the conclusions of the study can be generalized (Nguyen et al. 2015).

1.5 Structure of the thesis

This thesis is divided into six chapters. The first chapter was the introduction chapter, which introduced the topic, reviewed the past research, and presented the motivation for the thesis, including the research question and limitations. The second chapter presents the theoretical background of this thesis. It includes the economic theories relevant for asset price prediction, efficient market hypothesis and behavioral finance theory, and the introduces other relevant topics such as machine learning and sentiment analysis. The third chapter introduces the datasets that are used in conducting the study. These are the message board dataset from which sentiment is measured, the social media corpus used in classifier training and the stock prices which are attempted to be predicted. The fourth chapter introduces the methodology. This includes the methods used to extract sentiment information from the text, such as vectorization and Naïve Bays classifiers, and the models and tests with which the relationship is evaluated and predictions are made, such as vector autoregressive models and Granger causality analysis. The fifth chapter presents the results of the study starting from presenting the classifier and the sentiment dataset. This is followed by assessment of the predictive models, their predictions and finally the Granger causality analysis. The sixth and final chapter gives a quicky summary of the thesis, conclusions, and suggestion for further research.

2 THEORETICAL BACKGROUND

The relevant theories and academic background for this thesis are presented in this chapter. The first two subchapters present the traditional efficient market hypothesis and the newer behavioral finance schools of thought, that have opposing opinions on feasibility of asset price forecasting. This is followed by a subchapter presenting artificial intelligence and machine learning, which give required background for natural language processing and sentiment analysis that are introduced in the final subchapter. The final subchapter also includes a short history of sentiment analysis' usage in forecasting of financial data.

2.1 Efficient market hypothesis

A long standing believe among economists has been that the markets are efficient. This is the basic idea of efficient market hypothesis, which says that any information about assets is efficiently absorbed by the market participants and thus reflected in the asset prices. This creates a situation where neither technical analysis, which means using the past prices to predict future prices; nor fundamental analysis, which means using the fundamentals, such as earnings, to predict future prices can allow an investor to beat the market and obtain higher return for the chosen level of risk. From a market participant's point of view future prices of assets are completely random. As the price always is a prefect representation of value and reflects all information, it can only change when new information becomes available and as new information cannot be predicted, the new price cannot be predicted either. (Fama 1970; Malkiel 2003)

The efficient markets hypothesis can be divided into three different forms. These are weak form, semi-strong form, and strong from market efficiencies based on what information at least cannot be used to predict future asset prices. Stronger forms of market efficiency always require the weaker forms to hold too. A weak form market efficiency only requires that the future price cannot be predicted using past prices. This form says that the price follows a random walk but allows the price to be predicted with other available information. A semi-strong form assumes that no publicly available information is able to predict future asset price as the information gets absorbed into the asset prices almost instantly and no investor is able to benefit from using the

information. Strong form expands the semi-strong form to also include private information. Thus, not even insider knowledge could lead to abnormal returns. (Fama 1970)

The efficient market hypothesis makes attempting to predict asset prices pointless. In semi-strong and strong form any information publicly available would not allow an investor to create an accurate predictive model. In the best case, weak form market efficiency would make only technical analysis unproductive and would allow other information to be able to be used in prediction. The main argument for existence of efficient markets is that if predicting future prices was possible, why do investment funds that are managed by professionals consistently lose to passive index funds. And even if there was a way to beat the market, it would be abused until the information would eventually get absorbed into the prices, leading to market efficiency in the long-term. The empirical evidence seems to support this notion, as many strategies abusing market anomalies seem to have lost their strength after they have become widespread. (Schwert 2003; Timmermann & Granger 2004; Malkiel 2005)

2.2 Behavioral finance

The efficient market hypothesis has not stopped academics and investors from attempting to create models that could better predict asset prices (Timmermann & Granger 2004). According to Simon (1995) the rational investor, “homo economicus”, assumed by classic financial literature is extremely unrealistic and should be replaced with a more flawed and human investor borrowed from psychology. Investors cannot collect and interpret all available information perfectly and then calculate their own maximum benefit based on their preferences. In fact, investors might not even be completely sure about their own preferences. They tend to make simplifications and seek outcomes that are only “good enough” instead of the absolutely optimal ones. Daniel & Titman (1999) state that many behavioral biases, such as overconfidence, conservatism, and herding are constantly affecting decision-making of investors, which leads to pricing anomalies that can persist for long periods of time. Their research found out that using momentum strategy on U.S. growth stocks allowed abnormal returns. This effect did not dissipate even after momentum strategy became widespread, which made them reject the efficient market hypothesis.

The bounded rationality of an investor as described by behavioral finance leads to them using several heuristics in decision-making, which leads to them making systematic and predictable errors (Tversky & Kahneman 1974). According to Barberis, Schleifer & Vishny (1998) the investor sentiment, the collective expectations investors have for the market and individual assets, is itself unpredictable. They also argue that real investors do not actually believe in the random walk theory, but instead assume that the prices are either mean-reverting or trending. Investors irrationally make assumptions on the future performance of an asset based on one of these modes. Arbitrage might not be able to fix the mispricing as the sentiment can affect prices for longer than it is financially sustainable for arbitrager to hold their position. Therefore, it should be possible to gain abnormal returns by analyzing the market as the prices do not always reflect the correct value of an asset.

2.3 Artificial intelligence and Machine learning

In simplest terms, an artificial intelligence (AI) is machine that is able to alter its behavior in order to achieve a goal it has been given. An AI is able to perceive its environment, sometimes literally with usage of sensory, or more commonly by simply learning from its past experiences. AI applications should be used in tasks where computers can easily outperform humans, these include computational routine tasks. For example, arithmetic calculations or sorting data takes a long time when done by a human, but a computer can do it very quickly. (Poole, Mackworth, Goebel 1998)

Machine learning is a branch of AI, which focuses on machines that learn from their experience as opposed to rules preset by its programmers. According to Jordan & Mitchell (2015) the emergence of machine learning is made possible by the increasing availability of data online, decreased costs of both computing power and data storage, and the development of new state-of-the-art algorithms. The appeal of machine learning is apparent; Instead of giving programmers the near impossible task of creating a ruleset that accounts for every possible situation and completely captures the underlying patterns, the machine is instead given a set of examples which it can use to learn similarly to how humans can learn from their past experiences (Shalev-Shwarz & Ben-David 2014). Machine learning can be divided into three main

paradigms based on how the machine is trained: supervised learning, unsupervised learning, and reinforcement learning.

2.3.1 Supervised learning

Supervised learning is the most popular machine learning paradigm. It is characterized by its training data, which includes both inputs and outputs. Thus, the purpose of supervised machine learning is to create a model that can most accurately reach the output using the inputs it has been given. Supervised learning is used to solve classification problems. For example, supervised machine learning can be used to identify incoming emails as either spam or not spam, by giving it a dataset that includes both spam and non-spam emails. The algorithm would then attempt to find some patterns which it can use to correctly identify spam emails. Common supervised machine learning models include logistic regression, decisions trees and forests, Bayesian classifiers, support vector machines, and neural networks, among others. (Jordan & Mitchell 2015)

2.3.2 Unsupervised learning

Unsupervised machine learning does not have any initial outputs in its training data, which is its main difference from the other machine learning paradigm. Unsupervised learning attempts to find commonalities hidden in the data, which it uses to divide the training data into groups. Unsupervised learning is commonly used in clustering and anomaly detection. An example of the former could be finding distinct target groups among consumers by giving the model, such as k-means, the consumer data and letting the algorithm divide them based on their characteristics. An example of the latter could be having machine learning learn the normal values of a system, and inform the users when it finds abnormalities, which would let problems be fixed long before a human would notice them. For example, finding fraudulent activity from a bank transaction dataset. (Jain, Murty & Flynn 1999; Hodge & Austin 2004)

2.3.3 Reinforcement learning

Reinforcement learning is a compromise between supervised and unsupervised learning. Similar to supervised learning, there exists a correct output that should be reached but the training data can only give hints on what that correct output would be. Therefore, the model is allowed to make sub-optimal decisions and learn by trial-and-error. An example of reinforcement learning is a chess engine, where the input is the current position of the board, and the output is the most optimal move. In addition, every move the engine makes affects the following input it receives, which is the new position reached after the opposing player has made their move. The model explores possible moves and if it manages to win the game, it is rewarded. Eventually, the model learns to choose the move that is most likely to lead to it winning the game. (Shalev-Shwartz, Ben-David 2014; Jordan & Mitchell 2015)

2.4 Natural language processing

Natural language processing (NLP) has been one of the primary parts of artificial intelligence research from the beginning. Its purpose is to create machine that is able to understand languages by extracting the full meaning from text or speech. Initially NLP models were built by studying languages and building rulesets that would help the machine to understand language. However, these models could only work in extremely restricted environments and tasks, which led to NLP research shifting to using machine learning to train models that would learn the rules themselves from corpora they were given. (Brill & Mooney 1997)

Applications of NLP include tasks such as machine translation, summarizing text, retrieving information using queries, answering questions, and speech recognition. Understanding language is difficult for a machine, as there is a lot of ambiguity in the meaning, which must be interpreted from context, tone, or word choices. One example is sarcasm, which machines struggle to recognize. Thus, NLP models still cannot capture all information in text, and some meaning is lost during the process. (Wiriyathammabhum, Summers-Stay, Fermüller & Aloimonos 2016)

To use supervised machine learning based NLP with text, it must be transformed into a form that can be better understood by a machine. Usually this is done by transforming

the text into a vector consisting of tokens. Tokens are usually words, numbers, and special characters that are separated by whitespace. This can create large vectors, which is why the information should be reduced into a bag of words representation, where same words are listed along their counts instead of them repeating. As an alternative for counts, the words can be weighted with methods, which change the words importance based on the number of its occurrences in the specific text. (Loughran & McDonald 2016)

2.4.1 Sentiment analysis

One application of NLP is sentiment analysis, which attempts to extract the semantic orientation – mood, emotion, sentiment, opinion, etc. – from text. Sentiment analysis can be used to acquire range of moods or emotions from text, but usually it is limited to recognizing whether a text is positive or negative. It has usually been applied by marketing experts to study the sentiment towards certain products. Thus, sentiment analysis methods have mainly been developed using product reviews. An advantage of using reviews is that usually they include a rating that indicates the reviewer's semantic orientation, leading to availability of prelabeled training data for the supervised machine learning algorithms. (Nasukawa & Yi 2003)

Sentiment analysis is usually very simple NLP task. At most it might include part of speech tagging but is usually limited to calculation of word counts. The text is transformed into a bag-of-words vector, where tokens are potentially weighted in some manner. The purpose of weighing is to reduce the impact of common tokens, such as stop words like "and", "is", and "the", and on the other hand, reduce the effect of extremely rare tokens. This process assumes that sentiment can be derived from the words themselves, as the information in the structure of the text is lost. (Nasukawa & Yi 2003; Loughran & McDonald 2016; McGurk, Nowak & Hall 2020)

Sentiment analysis can be divided into two different methods: lexicon-based on machine learning based. The lexicon-based method requires at least one pre-determined lexicon of words. These lexicons are then used to acquire a semantic orientation for a word, and then the whole text's semantic orientation is calculated, at its simplest, as a sum of its words' semantic orientations. The lexicon-based models

are simple, but they require for the researcher to create or use a pre-determined lexicon, which might not include all tokens that affect the sentiment. In addition, lexicons tend to not be generalizable, and a new lexicon is required when the context of text is changed as words can have different semantic orientations in different contexts, for example between industries. (Loughran & McDonald 2016; McGurk et al. 2020)

According to McGurk et al. (2020) lexicon-based models are especially bad with social media, which includes short texts where opinions are expressed uniquely by each person. Social media also includes more informal speech and has ever-changing slang, which can make lexicons outdated quickly. These issues are attempted to be fixed with machine learning based sentiment analysis. Instead of using pre-determined lexicons, a supervised machine learning lexicon is instead taught using pre-labeled data, called a corpus. The model interprets the positive and negative words by itself, which makes it more objective and complete compared to researcher manually choosing them.

2.4.2 Sentiment analysis in finance

Baker & Wugler (2007) state that it is indisputable that investor sentiment affects the market. However, they state that correctly measuring the investor sentiment is especially challenging and has a major effect on the results of studies. They themselves created a proxy by combining metrics they thought captured the current sentiment. Nguyen et al. (2015) and McGurk et al. (2020) argue that textual analysis of news or social media would give more accurate and more frequent data with about investors sentiment. The most commonly used types of sentiment analyses in finance-based applications are based on training models with sets of pre-labeled words or sentences. These sets can be acquired by manually labeling part of the collected data similarly to Anteiler & Frank (2004) or by using a corpus consisting of either pre-labeled words or sentences like Checkley et al. (2017) or even by using pre-trained models like Rao & Srivastava (2012).

According to Evans, Owda, Crockett & Vilas (2019) investors use social media and news when deciding what investments to make. The media they frequently use, such

as news articles, finance-focused message boards, and traditional social media like Twitter, can be analyzed to measure investor sentiment. Even if those sources do not actually work as a direct proxy for investor sentiment, the people following them are bound to have their opinions and behavior affected by reading them (Nasukawa & Yi 2003).

The most common classification models tend to perform well in sentiment classification too. For example, in finance-focused research Naïve Bayesian models have been used by Antweiler & Frank (2004) and Rao & Srivastava (2012) among others, while Nguyen et al. (2016), Yang et al. (2016) and Derakhshan & Beigy (2019) had used Support Vector Machines. In addition, according to Nguyen et al. (2016) and Derakhshan & Beigy (2019) financial researchers have also attempted to use more complex models such as convoluted neural networks. However, the complex models do not have significant improvements in accuracy over the traditional ones.

3 DATA

This chapter introduces the datasets used in the study. The first subchapter introduces the message board data, its features, and how it is preprocessed for TF-IDF vectorization and classification, both of which will be presented in detail in chapter 4. The second subchapter introduces the sentiment corpus, which will be used to train the classifier. The trained model will analyze the sentiment of each message, which will be aggregated into daily net positive sentiment. This sentiment time series will be used in a model aiming to forecast the stock prices. The historical stock prices are presented in the final subchapter of this chapter.

3.1 Message board data

The message board dataset is collected from Finnish financial newspaper Kauppalehti's internet message board. Kauppalehti was chosen as it is the most visited financial news website and the 6th most visited news website in Finland by weekly visitors (Karppinen, Nieminen, Markkanen 2011). Thus, it can be assumed to be a local equivalent to Yahoo Finance, which has been used in past studies such as those by Antweiler & Frank (2004), Ho et al. (2017) and Nguyen et al. (2015). The discussion on the website is mainly divided into user made threads for individual stocks, which allows easier identification of what stock each message is about.

The message board data consists of messages made during 2019, more specifically from 2019-01-01 to 2019-12-31. The data was collected by acquiring all threads, which discussed an individual stock traded in OMX Helsinki stock exchange and had their last post made during or after 2019. All messages from these threads where then downloaded and the result was a dataset with 62,676 observations, each consisting of a timestamp, the text content of the post, and the name of the stock the post is about.

The interest of investors seems to be mainly focused only on a couple of companies, with the vast majority receiving very little attention (Appendix 1). The least popular stock had received only 5 posts regarding it in total while the most popular had received 16,871 posts in total. Inclusion of the least talked stocks would lead to creation of time series with a vast majority of datapoints having value of 0 or be biased by opinions of one or two individual users. By focusing on the most popular stocks, this study hopes

to better capture the overall collective investor sentiment and to limit the number of individual time series that have to be forecasted. Thus, the dataset was further filtered to only include the ten most discussed companies, which led to datasets size being reduced to 44,365 posts. It should be noted that the number of companies was chosen completely arbitrarily, and dismissal of less popular stocks has the downside that the results cannot be fully generalized. As can be seen from Figure 1, the number of posts still decreases exponentially from the most popular to the least popular stock.

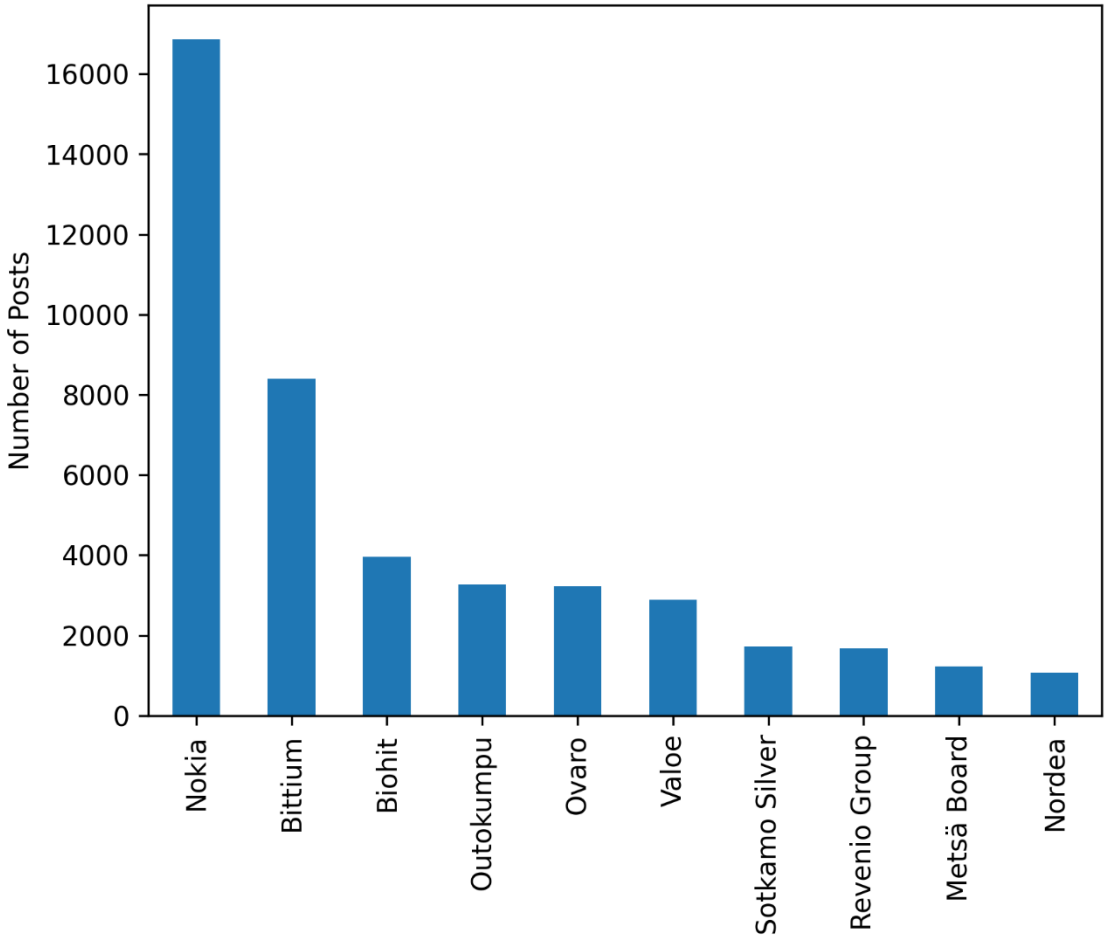


Figure 1. Total posts in 2019

In line with research by Nguyen et al. (2014) and McGurk et al. (2020) daily frequency is chosen for the sentiment analysis. This frequency should be short enough to capture some of the shorter-term effects while also not requiring as much and as frequent data

as intra-day frequency would. Descriptive statistics for the companies regarding daily messages are presented in Table 1.

Table 1. Descriptive statistics for daily posts

Company	Mean	Std	Min	Max
Nokia	46.22192	81.7491	0	698
Bittium	23.013	20.0829	0	219
Biohit	10.835	12.7688	0	143
Outokumpu	9.947	24.16052	0	260
Ovaro Kiinteistösijoitus	8.824	10.296	0	47
Valoe	7.95	1.601	0	78
Sotkamo Silver	4.7916	6.847	0	45
Revenio Group	4.599	8.514436	0	56
Metsä Board	3.4166	7.918368	0	92
Nordea	2.915	8.6278	0	90

The most popular stock, Nokia, has average 46 posts about it each day, which is twice as much as the second popular stock, Bittium. The least popular stock, Nordea, only has three daily posts on average. However, as can be seen from the standard deviations it seems that the most popular stocks also have more volatile number of daily posts. This can be further confirmed by looking at the Figure 2 which has plotted the daily posts for all ten stocks. Some stocks, such as Nokia and Outokumpu especially, have majority of posts about them posted during certain spikes. The mentioned stocks have seen little discussion during the late spring and summer while being relatively discussed in the winter seasons. There seems to be one large spike during the early year, and one slightly smaller one after summer. Though they are also volatile, stocks like Ovaro, Valoe and Sotkamo Silver have their posts more evenly distributed throughout the year.

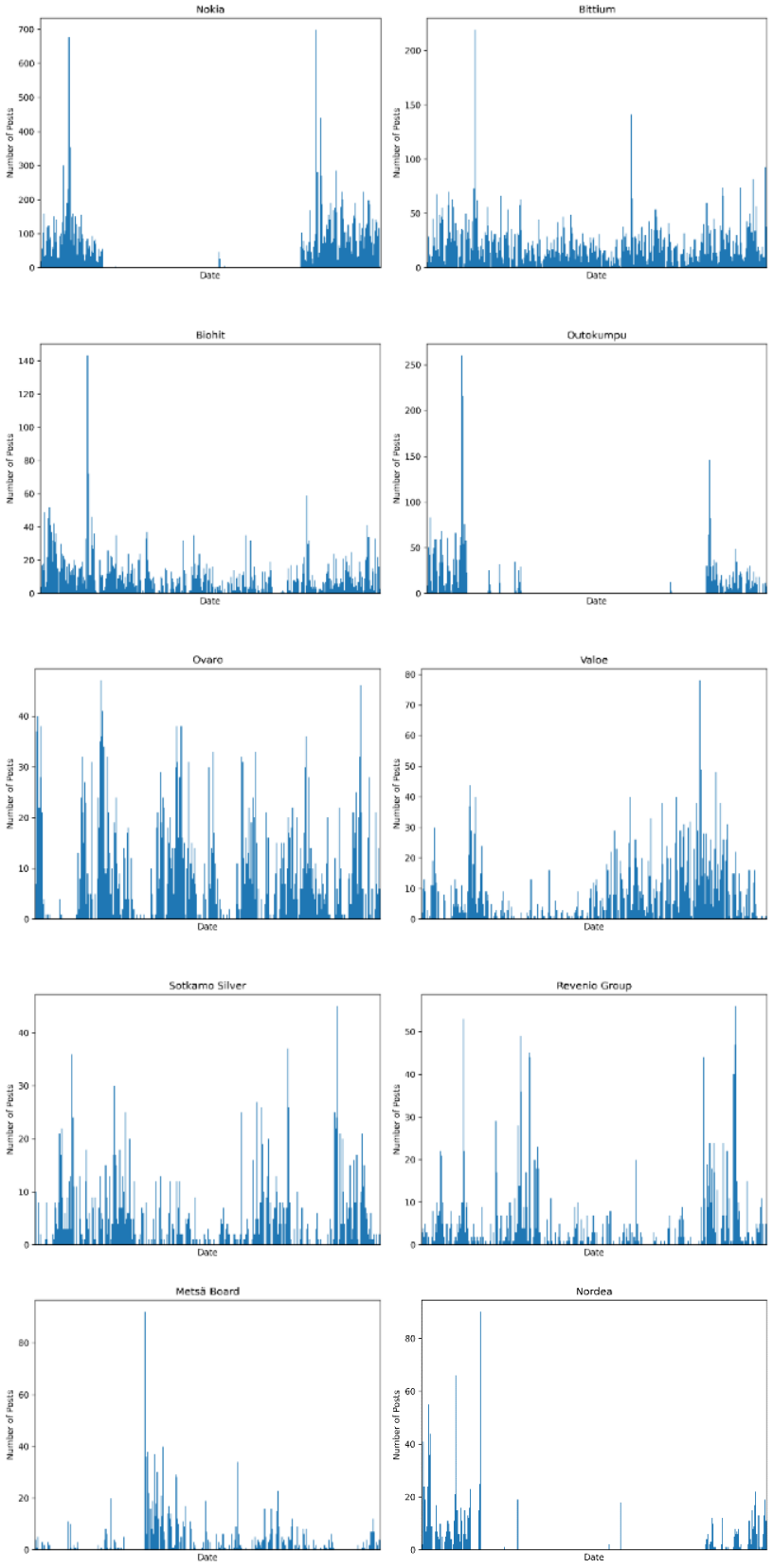


Figure 2. Daily posts

The messages need to be cleaned for them to be usable in sentiment analysis. First, any direct quotations to past posts are removed to decrease redundancy in the messages by not allowing sentiment of one post be taken into consideration more than once. All non-relevant metadata information that is embedded in the posts themselves is removed. This includes timestamps for any edits and notes that the message has been removed by either the user or site moderation. Hyperlinks to websites are removed from the posts as they would only confuse the sentiment analysis algorithm. Finally, all non-alphabetic characters, like numbers and special characters, were removed to limit the tokens used in sentiment analysis to only include words. Posts that became empty during this cleaning process are dropped and the size of the dataset is reduced down to 44,092 total posts.

3.2 Sentiment corpus

Machine learning based sentiment analysis is dependent on existence of training and testing datasets that have prelabeled text with correct semantic orientations. Manually labeling a sufficiently large dataset is time-consuming and even with help of experts correctly labeling data is almost impossible. Thus, the majority of research analyzing sentiment is done using text that has users self-label their sentiment, for example product reviews. However, prelabeled datasets can lose some of their ability to train accurate classifiers when used with text from different context. For example, classifiers trained with formal text like official documents might not work with casual text like social media posts. This further increases the difficulty of finding a training dataset as the context should be as closely related as possible in order for the trained classifier to be accurate. (Liu 2010)

Lindén, Jauhiainen, & Hardwick (2020) have created a corpus of 27,000 sentences that have been prelabeled as either positive, neutral, or negative. The data is collected from a Finnish social media site and has been manually annotated by majority vote of multiple Finnish-speaking people. The distribution of prelabeled sentiment is visible at Table 2. The majority of the sentences in the corpus are neutral, while 15% are negative and 11% are positive.

Table 2. Distribution of prelabeled sentiment

Positive	3066	11 %
Neutral	19825	73 %
Negative	4109	15 %

As the corpus consists of social media posts it should be contextually very similar to the message dataset consisting of social media posts from a financial forum. There might be some issues as the corpus does not account for all possible financial or site-specific slang, which will have some effect on the accuracy and the reliability of the classification. Past research such as studies by Rao & Srivastava (2012) and Li, Shang, & Wang (2019) have made use of generic sentiment corpuses to assess sentiment of their data.

3.3 Price history data

The historical stock prices for each of the ten companies was acquired from Yahoo Finance. The frequency for the historical prices was chosen to be daily to be in line with the collected message board data and past studies. Thus, the dataset consists of 250 adjusted close prices for each of the ten companies included. Adjusted close prices, which remove any effect of dividends, stock splits et cetera, were chosen to make the prices more comparable between the companies and better represent their value for the shareholders.

The prices were further transformed into returns by taking their logarithmic differences, which are plotted in Appendix 2. The plots indicate that the returns seem to have a constant mean, but their variance changed occasionally mainly due to spikes in either direction. Summary statistics for the returns are shown in Table 3, which support the conclusions drawn visually from the plots as the returns have close to zero means and medians.

Table 3. Summary statistics for daily returns

	Median	Mean	Std
Nokia	0.033 %	-0.070 %	0.011
Bittium	-0.071 %	-0.024 %	0.007
Biohit	0.000 %	0.039 %	0.019
Outokumpu	-0.087 %	-0.013 %	0.016
Ovaro Kiinteistösijoitus	0.000 %	-0.028 %	0.006
Valoe	-0.430 %	-0.028 %	0.079
Sotkamo Silver	0.000 %	-0.005 %	0.015
Revenio Group	0.042 %	0.099 %	0.007
Metsä Board	0.076 %	0.042 %	0.011
Nordea	0.000 %	0.014 %	0.007

The standard deviation of daily returns ranges from the 0.007 of Nordea to the 0.079 of Valoe, the latter of which is significantly more volatile than any other stock. Valoe also clearly has the lowest median returns while its mean is more in line with others. Therefore, the volatility seems to correlate more with the median, which is not affected as much by the outlier values.

4. METHODOLOGY

This chapter introduces the methodology used in the study. First subchapters introduce the methodology used in the sentiment analysis, starting from the vectorization and weighting of the text, and ending in the introduction of Naïve Bayes classifiers used in the study. The following subchapter introduces vector autoregressive model, which is going to be used to examine the relationship between daily sentiment and stock returns, and its basic assumptions. The final subchapter introduces the metrics used to evaluate the accuracy of the classifier and the forecasts made by the VAR models.

The study is conducted using Python scripting language and relevant libraries. For example, Scikit-learn is used for classifiers and data preprocessing, such as vectorization and TF-IDF weighting, and statsmodels library is used to build the autoregressive models, which are used to determine the predictive power of the sentiment. All visualizations have been built using either matplotlib library or Microsoft Office software.

The methodology of the study is presented in a system diagram seen in Figure 3, which visualizes the whole process from data collection to the evaluation of results. In the first part of the process all the datasets were collected and cleaned in a way described in the previous chapter. The datasets including textual data, message board dataset and sentiment corpus, are transformed into format that is better understood by machine learning algorithms using TF-IDF vectorizer, which transforms the text into vectors and weights each word with TF-IDF weighting. The corpus dataset is then used to train and evaluate different Naïve Bayes classifier models, out of which the most accurate one is used to classify posts of the message board dataset. The decision to use Naïve Bayes classifiers is in line with previous research by Antweiler & Frank (2004) and Rao & Srivastava (2012), who came to conclusion that Naïve Bayes classifiers can be expected to classify text's sentiment with reasonable accuracy for the purpose of investor sentiment measurement. TF-IDF weighting is used alongside vectorization because it is known to significantly improve classification accuracy (Loughran & McDonald 2016). The sentiment time series is created by classifying each post as either positive (1), neutral (0), or negative (-1), which are then aggerated into daily net-positiveness for each company using the equation by Antweiler & Frank (2004):

$$Day's\ sentiment\ = ln\left(\frac{M_t^{Pos}}{M_t^{Neg}}\right) \quad (1)$$

where M_t^{Pos} is number of positive posts that day,

M_t^{Neg} is number of negative posts that day.

When the number of posts for a day is zero, the sentiment value is also set as zero. Similarly, when the equation 1 is undefined because either the number of positive or negative posts is zero, the sentiment is also set as zero. Therefore, only days with at least one negative and one positive post are considered.

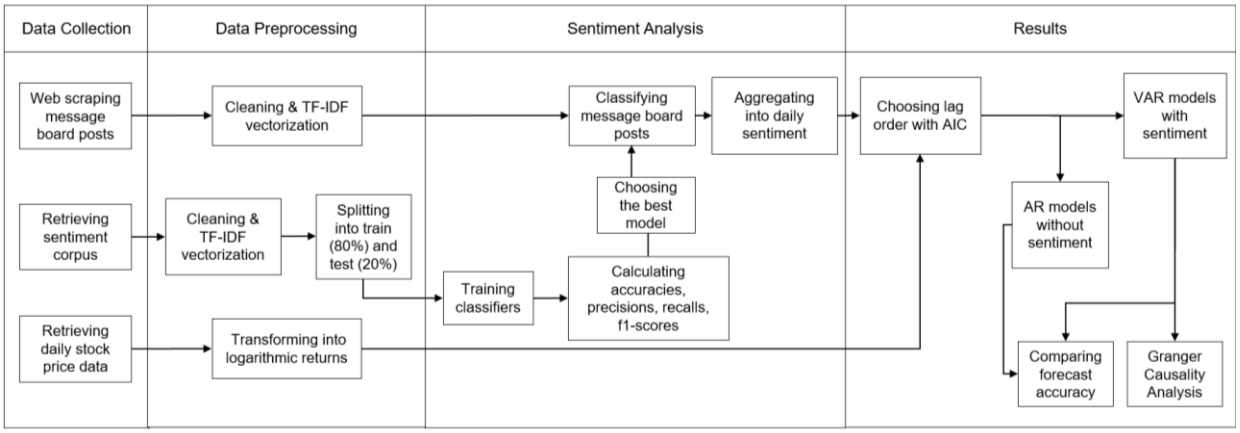


Figure 3. System diagram of the study

The predictive power of sentiment is evaluated by building VAR models where present returns are predicted using past sentiment and past returns. The predictive power of the models is evaluated using metrics introduced later in this chapter, and a Granger causality analysis is conducted to check for existence of any relationship between sentiment and returns. This type of methodology has been employed in studies by Rao & Srivastava (2012), Ranco et al. (2015), Checkley et al. (2017), and Piñeiro-Chousa et al. (2018).

4.1 TF-IDF vectorization

Vectorization is a process of transforming text information into numerical information that can be understood by machine learning algorithms. The most basic form of vectorization is done by taking all words, or terms, in the data and transforming them into variables that can either have a value of true or false depending on whether word is present in the text. However, this has some disadvantages. For example, as the number of variables can easily become large as the size of the dataset increases and every word is assumed to be as important in determining the correct class. (Salton, Fox & Wu 1983)

To fix the problems of the binary frequency vectorization, TF-IDF, term frequency inverted document frequency, vectorization was developed. Instead of giving terms Boolean values, they are instead given weights. This allows more accurate classification as the more important terms are given more weight while the effect of less important one can be taken out. In addition, the complexity of the model can also be brought down. (Salton & Buckley 1988)

TF-IDF weighting is composed of two parts: the term frequency and inverse document frequency. The rationale behind term frequency is that more commonly mentioned terms seem to be more useful in classifying documents and should thus be given more weight (Salton & Buckley 1988). Term frequency is calculated by taking the number of the term in given text and dividing it by the number of terms in that document, which can be expressed as (Salton & Buckley 1988):

$$TF(t,d) = f_{t,d} \Big/ \sum_{t\prime \in d} f_{t\prime,d} \qquad (2)$$

where $f_{t,d}$ is count of term t in text d,

$\sum_{t\prime \in d} f_{t\prime,d}$ is count of all terms in text d.

However, term frequency itself is not sufficient as many of the most common words, like “and”, “is”, and “the” are not important for the classification (Salton, Fox & Wu 1983). Inverse document frequency can be used to reduce the weight of these words (Salton & Buckley 1988). The IDF factor can be calculated by dividing the total number

of texts in the dataset with the number of texts that include the term whose weight needs to be determined (Salton, Fox & Wu 1983; Salton & Buckley 1988):

$$IDF(t) = ln\frac{N}{n_t} \quad (3)$$

where N is number of texts in dataset,

n_t is number of texts with term t.

The term frequency inverted document frequency can then be calculated as the product of term frequency and inverted document frequency weights (Salton, Fox & Wu 1983; Salton & Buckley 1988):

$$TF\ IDF(t,d) = TF(t,d) \cdot IDF(t) \quad (4)$$

where $TF(t,d)$ is TF weight,

$IDF(t)$ is IDF weight.

The TF-IDF is based on term discrimination, which considers that the most important terms for classification are those that are frequent in the text to be classified by have low frequency in the full dataset. This has little theoretical justification, which has led to criticism towards TF-IDF. However, in practice the weighting has performed better than the more complex alternatives. (Salton & Buckley 1988)

4.2 Naïve Bayes classifier

Naïve Bayes classifier is a machine learning model for classification based on Bayes theorem (Domingos & Pazzani 1997). In its simplest form the theorem states that the probability of an event A happening when another event B is known to have occurred can be estimated from the priori probabilities of both events and the conditional probability of B happening in a case where A has already occurred. This can be written as (Maron 1961):

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}, \qquad P(B) \neq 0 \tag{5}$$

where $P(A|B)$ is probability of A when B is true,

$P(B|A)$ is probability of B when A is true,

$P(A)$ is probability of A,

$P(B)$ is probability of B.

According to Maron (1961), in the context of sentiment analysis, the equation 5 represents a simplified case where the text contains only one word which affects its semantic orientation. The theorem estimates the probability of a text belonging in certain class based on the knowledge that a word with semantic orientation is present in it using the priori probability of a text being in said class, priori probability of text having the word with semantic orientation, and the conditional probability of text belonging to this class having said word. As the probability of the word being in said text can be confirmed to be true, the priori probability can be assumed to be 100%. Thus, the equation 5 can be reduced to:

$$P(A|B) = k \cdot P(B|A) \cdot P(A) \tag{6}$$

where $P(A|B)$ is probability of A when B is true,

$P(B|A)$ is probability of B when A is true,

$P(A)$ is probability of A,

k is scaling factor.

In reality the number of words with semantic orientation that affect the classification of the text is larger one. Thus, the equation 6 needs to be expanded to include any n number of words (Maron 1961):

$$P(A|B_1, B_2, \ldots, B_n) = k \cdot P(A) \cdot \prod_{i=1}^{n} P(B_i|A) \tag{7}$$

where $P(A|B_1, B_2, \ldots, B_n)$ is probability of A when all B are true,

$P(B_i|A)$ is probability of B_i when A is true,

$P(A)$ is probability of A,

k is scaling factor,

n is number of words included.

The classifier has an assumption that inclusion of certain words does not have any effect on probabilities of other words appearing. This assumption of independence is the reason why the classifier is Naïve as the assumption would not seem to hold in reality. However, in practice the classifier is observed to tolerate even clear violations of this assumption and being able to outperform more complex classification models. In addition, the Naïve Bayes classifier does not require much processing power and is relatively simple to understand, which has increased its popularity in solving of classification problems. (Domingos & Pazzani 1997)

4.2.1 Multinomial Naïve Bayes

Multinomial Naïve Bayes is an extension of Naïve Bayes classifier, which is specifically used in text classification. It assumes that the probability of event A happening when B has occurred follows multinomial distribution instead of a Gaussian one (Manning, Raghavan & Schuetze 2008). The priori probability of text belonging in certain class is defined as (Manning et al. 2008):

$$P(A) = \frac{N_A}{N} \tag{8}$$

where N_A is number of texts in class A,

N is total number of texts.

The conditional probability of event B happening when event A has occurred is defined as (Manning et al. 2008):

$$P(B|A) = T_{AB} \Big/ \sum_{B' \in V} T_{AB'} \quad (9)$$

where T_{AB} is count of word B in class A,

$\sum_{B' \in V} T_{AB'}$ is count of words in class A.

In text classification context, the priori probabilities are simply how many occurrences of a certain word there are in certain class, while the conditional probabilities are estimated with count of the word in certain class divided by the count of all words. The multinomial model works best with integer word counts but can somewhat reliably be used with fractions like TF-IDF weighted counts. (Manning et al. 2008)

4.2.2 Complement Naïve Bayes

Despite of multinomial Naïve Bayes being commonly used for text classification, its assumption of text following a multinomial distribution is unrealistic. Complement Naïve Bayes model includes some corrections to the multinomial model that improve its accuracy to that of state-of-the-art algorithms. (Rennie, Shih, Teevan, Karger 2003)

According to Rennie et al. (2003) the complement Naïve Bays model calculates the conditional probability of event B happening when event A has occurred differently from multinomial model. Instead of counting the occurrences of word in a class and dividing it with the count of all the words in the class, the count of the word in all other classes is divided by the count of all the words in all other classes. This can be written as:

$$P(B|\tilde{A}) = T_{\tilde{A}B} \Big/ \sum_{B' \in V} T_{\tilde{A}B'} \quad (10)$$

where $T_{\tilde{A}B}$ is count of B in classes other than A,

$\sum_{B' \in V} T_{\tilde{A}B'}$ is count of words in other classes.

The complement model wants to minimize the probability of word appearing in other classes rather than maximize the probability of it appearing in a class. Thus, the conditional probability of event B happening when A has occurred in equation 7 is replaced with the inverse of the probability from equation 10. (Rennie et al. 2003)

The advantage of complement Naïve Bayes is that it works better with skewed training data that is common in text classification and that it is able to better utilize weighted word counts such as TF-IDF. These features allow it to give significantly better performance in text analysis than other models, while keeping the model easy to understand and implement, and its low computing power requirements. (Rennie et al. 2003)

4.3 Vector autoregressive models

Autoregressive (AR) models are commonly used in economic applications to forecast future values of time series such as unemployment, asset prices, or exchange rates for example. The models expect that at least the near future values of these time series can be forecasted using some combination of their present and past values. (Hamilton 1994)

However, according to Lütkepohl (2005) many real-world time series are not only affected by themselves but also by other time series. For example, the above-mentioned exchange rates are affected by interest rates. Vector autoregressive (VAR) models expand the AR models to also use values of other time series. The equation for VAR model with k time series and lag order of p, k-variate VAR(p) model is:

$$Y_t = C + A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + U_t \tag{11}$$

$$Y_{t-p} = \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{k,t-p} \end{bmatrix}, C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}, A_p = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & & \vdots \\ \vdots & & \ddots & \\ a_{k,1} & \cdots & & a_{k,k} \end{bmatrix}, U_t = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

where Y_{t-p} are the p:th lags of k time series,

C are intercept terms,

A_p are estimated coefficients,

U_t	are white noise processes,
$a_{i,j}$	is coefficient in y_i for lag of y_j.

The coefficients in VAR models are estimated using least squares estimation, where the errors of estimated Y_t to real Y_t is minimized. These errors are the white noise processes in U_t of equation 11, and are assumed to be normally distributed, have a zero mean, and variance of one. (Hyndman & Athanasopoulos 2018)

4.3.1 Stationarity assumption

VAR model requires all time series it forecasts to either be stationary or cointegrated. A time series is stationary if its properties are the same no matter at what time point it is observed. In other words, the time series should not have any predictable patterns like a trend or seasonality. This can be confirmed if the time series has a constant mean, variance, and autocorrelative structure. (MacKinnon 1994; Hyndman & Athanasopopulos 2018)

Non-stationary time series are said to have a unit root and as such, removing any unit roots from the data transforms it into a stationary time series. The variance of time series can be stabilized by transforming it, for example into logarithms, and differencing can reduce the effect of time: trend and seasonality. Asset price time series are made stationary by transforming them into returns. (Hyndman & Athanasopopulos 2018)

The stationarity of a time series can be tested using unit root test. The most popular ones are Kwiatkowski, Phillips, Schmidt & Shin's (1992) KPSS test, and Dickey & Fuller's (1979) and Said & Dickey's (1984) augmented Dickey-Fuller test. The KPSS test has a null hypothesis for stationarity, while the augmented Dickey-Fuller test has a null hypothesis for existence of a unit root. Therefore, using both tests can help to confirm stationarity of the data.

KPSS test models the time series as (Kwiatkowski et al. 1992):

$$y_t = r_t + \beta t + \varepsilon_t \tag{12}$$

$$r_t = r_{t-1} + u_{t,}, \quad u_t \; iid \; N(0, \sigma_u^2)$$

where y_t is time series,

r_t is random walk,

βt is deterministic trend,

ε_t is stationary error.

The null hypothesis of KPSS test is that the variance of u_t from equation 12 is zero, which leads to random walk r_t being constant zero. When this is true, the time series is trend-stationary. The test can easily be modified to test for level stationarity by removing the trend component from the equation. (Kwiatkowski et al. 1992)

The augmented Dickey-Fuller test, with constant and trend, can be written as (Said & Dickey 1984):

$$y_t = \rho y_{t-1} + u_t \tag{13}$$

which can be rewritten as:

$$\Delta y_t = \alpha + \theta y_{t-1} + \lambda_t + \sum_{s=1}^{m} a_s \Delta y_{t-s} + v_t \tag{14}$$

where Δy_t is first differences of time series,

α is constant,

λ_t is trend,

θy_{t-1} is the effect of previous value,

$\sum_{s=1}^{m} a_s \Delta y_{t-s}$ is the effect of lagged differences.

Values of stationary time series return to their mean, which means that the effect of previous value should disappear from the model as time passes. Thus, the coefficient rho in the equation 13 should be smaller than one for stationarity, or exactly one for non-stationarity. If rho is larger than one, the effect of pervious values would indefinitely increase as time passes, which is unrealistic and can assumed to never happen. In addition, coefficient rho cannot be larger than one as the effect of past values is

assumed to be unable to increase in real world datasets. Thus, the coefficient theta in equation 14 is rho minus one, and it must be negative when the series is stationary. The null hypothesis is that theta is zero and the series is non-stationary. The augmented Dickey-Fuller test includes the effect of lagged differences in the test equation, which allows the tested time series to be more complex by removing autocorrelation. Like in KPSS test, the constant and term components can be removed if the tested time series does not have them. (Dickey & Fuller 1979; Said & Dickey 1984)

4.3.2 Information criteria

The number of coefficients that need to be estimated in VAR models follows the equation:

$$k + p \cdot k^2 \tag{15}$$

where k is number of time series,

p is lag order.

Thus, the addition of more lags or new time series easily leads to models becoming very complex with large number of coefficients that need to be estimated. OLS estimation also tends to always prefer more complex models as any new predictor variables cannot increase the sum of squared residuals. Therefore, there is a need to find a model that gets the most information out of the data with the least complexity possible. (Hyndman & Athanasopopulos 2018)

Akaike's (1974) information criterion (AIC) is the most commonly used tool in selecting the number of lags for autoregressive models. The AIC can be calculated for all models in consideration, and it helps to choose among them. The AIC is calculated as:

$$AIC = 2 \cdot k + 2 \cdot ln(MSE) \tag{16}$$

where k is number of estimated coefficients,

MSE is mean squared error.

Akaike information criterion gets higher values when the number of estimated coefficients increases and when the mean squared error increases. Thus, the most optimal model can be found by choosing the one with lowest AIC value as that model has the lowest mean squared error while also only including the variables which have significant effect on the predicted values. This is achieved by only including variables which decrease mean squared error more than they increase the penalty function. (Wagenmakers, Farrel 2004)

4.3.3 Granger causality analysis

Granger's (1969) causality analysis is used to determine whether a time series affects values of other time series. It should be noted that Granger causality does not actually mean causal relationship but merely that the predictions of a time series can be improved using other time series. When there is Granger causality changes in one time series are followed by changes in other time series, but it does not mean that there is necessarily a causal relationship.

According to Lütkepohl (2005) Granger causality can be easily understood and implemented using a VAR model. For example, a bivariate VAR(p) model:

$$y_{1,t} = \sum_{s=1}^{p} a_{11,s} y_{1,t-s} + \sum_{s=1}^{p} a_{12,s} y_{2,t-s} + \varepsilon_t \quad (17)$$

$$y_{2,t} = \sum_{s=1}^{p} a_{22,s} y_{2,t-s} + \sum_{s=1}^{p} a_{21,s} y_{1,t-s} + \nu_t \quad (18)$$

where

$y_{1,t}$, $y_{2,t}$	are values of time series at time t,
ε, ν	is a white noise process,
$a_{ij,s}$	is coefficient in y_i for lag s of y_j,
p	is the total number of lags.

The equations 17 and 19 show a situation where Y_1 and Y_2 would Granger cause each other if the inclusion of the other's lags increases the accuracy of the prediction. In other words, if coefficients $a_{i,j}$ are statistically nonzero. (Granger 1969; Freeman 1983)

Granger causality can be tested by comparing a restricted model and unrestricted model where the latter includes the new variable to-be-tested and former does not. This can be done with simple F-test with null hypothesis that inclusion of new variable does not significantly reduce the residual sum of squares. (Freeman 1983; Lütkepohl 2005)

4.4 Model evaluation

The performance of both the chosen classifier and the VAR model must be assessed. The classifier's ability to correctly categorize text into positive and negative sentiment is evaluated using the common supervised machine learning evaluation metrics. The VAR models are evaluated by how accurately they are able to predict returns. This is done by comparing their predictive power against that of AR models without sentiment for their respective stocks.

4.4.1 Classifier evaluation

One of the most popular ways to assess classifier performance is to build a confusion matrix. It compares the classes predicted by the classifier to the actual classes of the data. An example with simple binary confusion matrix is shown in Figure 4. The cases belonging to class A that are correctly classified into class A are true positives (TP) and the cases belonging to class B that are correctly classified into B are true negatives (TN). The binary classifier can also make two types of mistakes, false positives (FP), where cases belonging to class B are classified into class A; and false negatives (FN), where cases belonging to class A are classified into class B. A good classifier maximizes the number of true predictions while minimizing the number of false predictions. (Fawcett 2006)

		Predicted	
		Class A	Class B
Actual	Class A	TP	FN
	Class B	FP	TN

Figure 4. Confusion matrix for binary classification

However, the Naïve Bayes classifier used in this study uses three different classes. An example of this tertiary case is shown in Figure 5. Similar to binary case, the values on the diagonal cells (AA, BB & CC) represent the correctly classified cases while the other values are classified incorrectly in some way. For example, cell AB has the cases belonging to class A, which classified into class B; and cell AC has the cases belonging to class A, which classified into class C. (Beleites, Salzer & Sergo 2013)

		Predicted		
		Class A	Class B	Class C
Actual	Class A	AA	AB	AC
	Class B	BA	BB	BC
	Class C	CA	CB	CC

Figure 5. Confusion matrix for tertiary classification

The values from confusion matrices can be used to calculate various metrics, which can be used to easily assess the performance of the classifier. The most common is accuracy, which is calculated by dividing the number of correctly classified cases, true cases, with the number of total cases. Accuracy alone is not enough to assess the performance, as especially with very imbalanced data it is possible to get high accuracies by simply classifying all cases into the majority class. If the purpose of the model is to identify the minority cases, the beforementioned classifier would have very poor performance in its intended task. (Fawcet 2006; Powers 2020)

According to Beleites et al. (2013) and Powers (2020), other common performance metrics for classifiers are precision, recall, and f_1-score, which combines the former two. Precision tells how many of the cases classified into a class actually belong into it. In a three-class case it is calculated by dividing the number of true cases by all cases

classified into the class. Recall tells how well the classifier recognizes cases belonging into class and is calculated by dividing the number of true cases with number of all cases belonging to the class. Figure 6 showcases visually how these fractions are calculated for class A. F_1-score is simply the harmonic mean of the two, and can be calculated as (Powers 2020):

$$F_1 Score = \frac{2}{(1/PRECISION + 1/RECALL)} \tag{19}$$

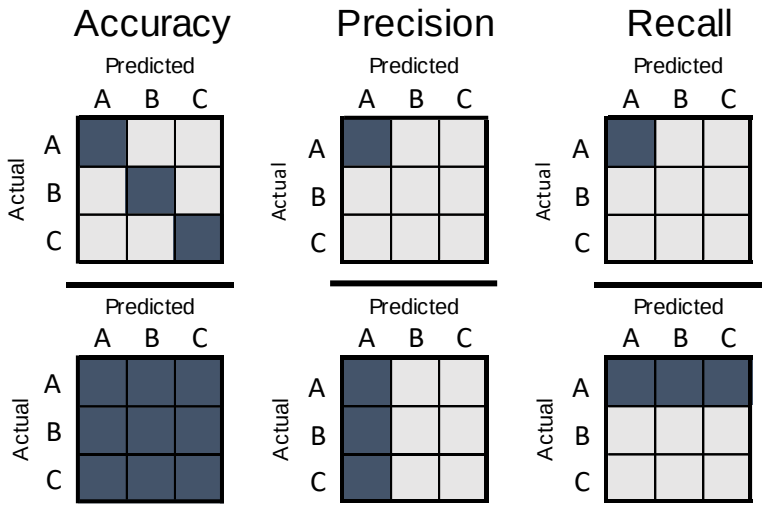


Figure 6. Confusion matrix fractions (class A)

In machine learning, a dataset is usually divided into training and testing sets. The training set is the dataset that is used to train the classifier and usually consists of 80% of the dataset. The 20% not used in the training is a testing set, which can be used to assess how well the model can perform on new data which it has not seen before. This also reduces the change of overfitting the model to any specific training data. All the above-mentioned metrics are usually calculated on the testing data. (Allen 1974; Stone 1974)

In addition, especially when the data is skewed, it is beneficial to make sure that training and testing datasets have equal distributions of all classes. This can be achieved with stratification, which is commonly used whether an unbalanced dataset needs to be divided for machine learning. (Allen 1974; Stone 1974)

4.4.2 Forecast evaluation

Majority of metrics used to evaluate accuracy of forecasting model are based around the error, which is the difference between the predicted value and the real value. The error represents the part of the real value that cannot be predicted using the model. Therefore, minimizing the error logically leads to more accurate predictions as it makes the predicted values closer to the real values. Minimization of the error also leads to minimization of other accuracy measurements. (Lütkepohl 2005; Hyndman & Athanasopopulos 2018)

According to Hyndman & Athanasopopulos (2018) one of the most commonly used error measurement is the root mean squared error (RMSE). It is calculated by taking the square root of mean squared error, which itself is a sum of all errors squared. It is calculated by taking the square root of mean squared error, which itself is calculated by taking the mean of all errors, which are squared so errors with different sign do not cancel each other out. RMSE can be calculated with:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{20}$$

where y is real value,

$\hat{y}$ is predicted value,

N is number of observations.

According to Brooks (2014) an investor might not be especially interested in the model's capacity to correctly estimate the exact future price but rather whether the price is going up or down. Therefore, an error measurement metric to examine how well the model can predict whether the returns are negative or positive is also beneficial. This can be measured with percentage of correct sign predictions, which shows how large part of predictions had the correct sign. This can be calculated with:

$$Correct\ \text{Sign}\ \% = \frac{1}{N}\sum_{i=1}^{N} f(y_i \cdot \hat{y}_i), \qquad f(y_i \cdot \hat{y}_i) = \begin{cases} 1, & (y_i \cdot \hat{y}_i) > 0 \\ 0, & (y_i \cdot \hat{y}_i) \leq 0 \end{cases} \tag{21}$$

where y is real value,

$\hat{y}$ is predicted value,

N is number of observations.

5. RESULTS

This chapter presents the results of the study. The first subchapter presents the accuracy metrics for the classifiers and the justification for choosing one of them for the study. The following subchapter presents the results of using said classifier on the message board data and introduces the created sentiment dataset. In the next subchapter the datasets, both sentiment and returns, are tested for stationarity in order to check whether VAR models can be built for them. The following subchapter introduce the built VAR models, their predictions which are evaluated using previously presented metrics. Finally, the results of Granger causality tests are presented.

5.1 Classifier selection

The prelabeled social media corpus was prepared for training of the model by applying the same preprocessing steps to it that were applied to the original collected dataset as described in the data chapter. These steps included removal of non-alphabetic characters, hyperlinks, and transformation of all characters into lowercase. The sentences from the corpus were also vectorized using TF-IDF vectorization. In order to assess the performance improvement by TF-IDF vectorization a count vectorized alternative was also created. The corpus was split into a train and a test set with stratified split applied. The sizes of the datasets were 80% and 20% for training and testing, respectively.

The training dataset was used to train all types of Naïve Bayes classifiers introduced in the previous chapter, the Gaussian, multinomial, and complement models. Two sets of each model type were trained, one with simple count vectorization and one with TF-IDF weighted vectorization. This led to six models from which to choose the most optimal from.

As can be seen from the confusion matrices shown in Table 4, the Gaussian models seem to be the worst performing and the type of vectorization does not seem to have any major effect on the performance, though TF-IDF weighting seems to reduce the number of correctly classified negative and positives posts. The multinomial models seem to both perform little better, though they run into a problem of classifying most posts as neutral. Here, the using TF-IDF weighting has removed any posts being

falsely classified as positive or negative. However, the total number of posts classified as either positive or negative is lowest of all trained models. The complement models do not seem to perform better than the multinomial ones as the number of falsely classified positive and negative posts is higher. There is, however, some improvement when comparing to Gaussian models. Similar to multinomial models, TF-IDF weighting has increased the number of posts classified as neutral, but this time still predicts some posts incorrectly as either positive or negative when they are not.

Table 4. Confusion matrices for all classifiers

Count vectorization

Gaussian

Actual \ Predicted	Negative	Neutral	Positive
Negative	266	407	149
Neutral	665	2421	879
Positive	56	239	318

Multinomial

Actual \ Predicted	Negative	Neutral	Positive
Negative	52	768	2
Neutral	26	3922	17
Positive	7	475	131

Complement

Actual \ Predicted	Negative	Neutral	Positive
Negative	241	556	25
Neutral	392	3399	174
Positive	49	285	279

TF-IDF vectorization

Gaussian

Actual \ Predicted	Negative	Neutral	Positive
Negative	258	417	147
Neutral	675	2436	854
Positive	56	248	309

Multinomial

Actual \ Predicted	Negative	Neutral	Positive
Negative	8	814	0
Neutral	0	3965	0
Positive	0	542	71

Complement

Actual \ Predicted	Negative	Neutral	Positive
Negative	142	667	13
Neutral	260	3603	102
Positive	31	356	226

To better understand the differences between the models, the values from the confusion matrices are used to calculate accuracies, precisions, recalls, and f_1-scores for each model. All of these are shown in Table 5. Based on accuracies, the usage of TF-IDF weighting does not seem to matter much on Finnish language text. The accuracies of the models themselves range from poor to mediocre, with both Gaussian models reaching 56%, multinomial models reaching 75% and 76%, and complement

models reaching 73% and 74% for count vectorized and TF-IDF vectorized, respectively.

The Gaussian models also have rather poor precisions, recalls and f_1-scores. Whether TF-IDF weighting was used or not did not have large impact on any of the metrics. The complement models got similar values as the Gaussian models apart from having higher precisions, especially for positive class, and better recall for neutral class. The multinomial model differs noticeably from the others by having higher precisions at the expense of having the worst recall values. The TF-IDF model also has more noticeable effect here compared to other model types as using it gives the model precision of 1 for both the positive and the negative class. However, as the recalls are still very poor, the model can hardly be seen as optimal.

Table 5. Evaluation metrics for classifiers

Count vectorization

Gaussian — 56% accuracy

	Precision	Recall	F1-score
Negative	0.27	0.32	0.29
Neutral	0.79	0.61	0.69
Positive	0.24	0.52	0.32

Multinomial — 76% accuracy

	Precision	Recall	F1-score
Negative	0.61	0.06	0.11
Neutral	0.76	0.99	0.86
Positive	0.87	0.21	0.34

Complement — 73% accuracy

	Precision	Recall	F1-score
Negative	0.35	0.29	0.32
Neutral	0.8	0.86	0.83
Positive	0.58	0.46	0.51

TF-IDF vectorization

Gaussian — 56% accuracy

	Precision	Recall	F1-score
Negative	0.26	0.31	0.28
Neutral	0.79	0.61	0.69
Positive	0.24	0.50	0.32

Multinomial — 75% accuracy

	Precision	Recall	F1-score
Negative	1.00	0.01	0.02
Neutral	0.75	1.00	0.85
Positive	1.00	0.12	0.21

Complement — 74% accuracy

	Precision	Recall	F1-score
Negative	0.33	0.17	0.23
Neutral	0.78	0.91	0.84
Positive	0.66	0.37	0.47

Assessment of all classifiers has shown that the normally used sentiment analysis methods perform significantly worse on Finnish language text than on the English one they were designed for. TF-IDF weighting also does not seem to increase accuracy as much as in English text, and in most cases has no effect at all on the performance of the model.

When the message board dataset was classified using the multinomial classifier with TF-IDF weighting, it resulted in 43964 posts being classified as neutral, 51 as positive, and 4 as negative. This number of positive and negative posts is clearly not enough to create sentiment time series for 10 different stocks and would very likely lead to sentiment being zero most of the time. The non-weighted version of the model is not any better, with only 93 positive and 53 negative values, with the rest being neutral. This leaves Gaussian and complement models as the only practical choices, out of which Gaussian should clearly be rejected as its accuracy is significantly worse. The complement model classifies 2647 negatives and 1872 positives without TF-IDF weighting and 767 negatives and 737 positives with it.

The complement model without TF-IDF weighting was chosen in the end because it had higher f_1-scores for both the positive and the negative class. The semantic orientation of collected message board posts can be seen in Table 6. The majority of posts for all stocks are clearly negative. Interestingly, there are more negative posts than positive ones for almost all stocks, which is the opposite of the situation with the test data. This could be caused by users of the message board using words the classifier associates with negative posts more commonly than in the corpus. It could also be caused by poor performance of the model, which is supported by the fact that when using multinomial classifier, the results mirrored the test set more.

Table 6. Semantic orientation of posts

	Positive	Neutral	Negative
Nokia	753	14861	1122
Bittium	451	7376	511
Biohit	154	3590	192
Outokumpu	110	2915	214
Ovaro Kiinteistösijoitus	67	3005	147
Valoe	64	2693	126
Sotkamo Silver	65	1554	101
Revenio Group	77	1515	65
Metsä Board	63	1093	71
Nordea	68	898	98
Combined	**1872**	**39500**	**2647**
	4 %	90 %	6 %

A time series that reflects the daily sentiment for each stock was created in the manner explained in the methodology chapter. Plots of daily sentiment for each stock can be seen from Appendix 3. It should be noted that as it is very likely that the classifier did not capture the investor sentiment properly, and that thus the created sentiment time series does not either. Therefore, it should be expected that there is no or only minimal predictive power in the sentiment variable. However, this might not mean that investor sentiment does not affect stock returns but rather that the classifier failed to capture the investor sentiment.

5.2 Testing for unit roots

Before VAR models are built the time series, both the sentiment and daily returns, are checked for stationarity using unit root testing. The stationarity is determined for all variables using both the Dickey-Fuller test and the KPSS test. The augmented Dickey-Fuller test results for all variables are shown in Table 7. The null hypothesis of existence of unit root is rejected for the returns of all stocks on 1% significance level, which supports the assumption that they are stationary. As for the daily sentiment variable, the null hypothesis can be rejected for all stocks expect for Nokia and Outokumpu. The test statistic for Nokia is extremely close to the critical value and unit root can at least be rejected on a 5% significance level.

Table 7. Augmented Dickey-Fuller test results

	Returns		Sentiment	
Company	Test statistic	p-value	Test statistic	p-value
Nokia	-14.300	1.24E-26	-3.428	0.010
Bittium	-14.995	1.11E-27	-14.682	3.14E-27
Biohit	-19.628	0.00E+00	-9.342	8.75E-16
Outokumpu	-15.221	5.47E-28	-3.093	0.027
Ovaro Kiinteistösijoitus	-6.062	1.21E-07	-18.510	2.11E-30
Valoe	-16.645	1.62E-29	-18.526	2.11E-30
Sotkamo Silver	-6.909	1.23E-09	-12.302	7.42E-23
Revenio Group	-15.793	1.10E-28	-19.848	0.00E+00
Metsä Board	-14.633	3.73E-27	-19.979	0.00E+00
Nordea	-13.647	1.61E-25	-5.608	1.22E-06

The results of KPSS test, shown in Table 8, are used to confirm the findings of augmented Dickey-Fuller test. The null hypothesis of KPSS test is stationarity and therefore the results seem to support the augmented Dickey-Fuller test results. For returns, all the test statistics are high enough to escape the lookup table's range and cannot be rejected at even 10% significance level. The results are not as clear for daily sentiments, with Nokia and Nordea having theirs rejected on 10% significance level, and Outokumpu and Metsä Board even at 5% significance level. However, none are rejected at 1% significance level.

Table 8. KPSS test results

	Returns		Sentiment	
Company	Test statistic	p-value	Test statistic	p-value
Nokia	0.129	0.1*	0.368	0.09
Bittium	0.173	0.1*	0.181	0.1*
Biohit	0.184	0.1*	0.180	0.1*
Outokumpu	0.056	0.1*	0.589	0.03
Ovaro Kiinteistösijoitus	0.109	0.1*	0.078	0.1*
Valoe	0.136	0.1*	0.190	0.1*
Sotkamo Silver	0.086	0.1*	0.233	0.1*
Revenio Group	0.105	0.1*	0.226	0.1*
Metsä Board	0.122	0.1*	0.543	0.03
Nordea	0.140	0.1*	0.378	0.09

* test statistic is out of lookup table's range, actual p-value is larger

Based on the results of both tests, the returns can clearly be considered stationary. The daily sentiment variables for every stock outside of Nokia, Outokumpu, Metsä Board and Nordea can clearly be considered stationary. Even out of these exceptions, only Outokumpu got evidence for non-stationarity from both tests while the rest only got weak or conflicting evidence from them. The cause of this could be the infrequency of daily sentiment as especially these stocks have periods where they are talked about a lot and periods where they are not talked about at all. Overall, the variables seem stationary enough for the building of VAR models to commence.

5.3 VAR models

The lag orders for VAR models were chosen using AIC, which was calculated for each stock from lag length of 1 to 40. It should be noted that the minimum number of lags tested was 1 to ensure that Granger causality analysis could later be done. In addition, all previous lags below the maximum lag length were always included to reduce the number iterations needed. The lag orders are shown in Table 9 and, apart from Outokumpu and Nordea, they all seem to be rather low, and many include only one past price and daily sentiment datapoint. This would seem to imply that the effect of sentiment quickly disappears from the system. Alternatively, the non-considered constant-only models could be the most optimal, which could mean that neither the past prices nor daily sentiments are able to increase the predictive power enough to offset the penalty function for their inclusion.

Table 9. Lag orders

Company	Lag order
Nokia	6
Bittium	1
Biohit	5
Outokumpu	27
Ovaro Kiinteistösijoitu	4
Valoe	1
Sotkamo Silver	1
Revenio Group	1
Metsä Board	3
Nordea	35

VAR model assumes that its residuals, created by a white noise process, have a zero mean, include no autocorrelation, come from a normal distribution, and have no heteroscedasticity. Thus, the residuals are tested for autocorrelation with Ljung-Box (Ljung & Box 1978) test, for normality with Shapiro-Wilk test (Shapiro & Wilk 1965), and for heteroscedasticity with Engle's (1986) ARCH test. The null hypotheses for these tests are "sample has no autocorrelation", "sample comes from normal distribution", and "sample has no ARCH effects", respectively. The p-values for all these tests are collected in the Table 10. The null hypothesis for Ljung-Box cannot be rejected for any model, which supports the chosen lag orders as there is no autocorrelation left in the residuals. Based on Engle's ARCH test results, majority of the models have homoscedastic residuals except for Ovaro and Valoe, which reject the null hypothesis on 1% significance level. The normality is clearly rejected for almost all models apart from Outokumpu. However, based on the Q-Q plots seen in Appendix 4, this could simply be caused by there being more extreme values that have led to longer tails than in normal distribution.

Table 10. P-values for residual tests

Company	Ljung-Box	Shapiro-Wilk	Engle's ARCH test
Nokia	0.998	0.000	0.999
Bittium	0.212	0.000	0.529
Biohit	0.976	0.000	0.999
Outokumpu	0.999	0.167	0.987
Ovaro Kiinteistösijoitus	0.095	0.000	0.005
Valoe	0.957	0.000	0.007
Sotkamo Silver	0.421	0.000	0.207
Revenio Group	0.487	0.000	0.612
Metsä Board	0.864	0.000	1.000
Nordea	1.000	0.000	0.894

Based on these results, the VAR models can be built with these lag orders. The regression results for all the models are shown in Appendix 5. At a glance the models do not seem to be particularly good as majority of their coefficients are statistically insignificant. This includes coefficients for both the daily returns and the created daily sentiment variable. As the past prices seem to be bad predictors this would imply that at least weak-form market efficiency is true. In the case of daily sentiment, the reason can also be the poor performing classifier.

5.3.1 Accuracy of predictions

The predictive power of both the past returns and past daily sentiment for each stock is determined by evaluating the accuracy of their predictions. The predictions are created for the length of the year 2019 starting as soon as possible with the first trading day being the lag furthest in past. The accuracy of the predictions is evaluated by comparing them to those made by models without sentiment variable, which happen to be AR models of same lag length. It should be noted that predicting the same time period as the model was trained on is not optimal to determine the predictive power in a real-world scenario. However, it does give an idea whether the sentiment has had an effect on the returns.

The root mean squared errors for all the models are shown in Table 11. The VAR models of each stock have slightly lower RMSEs compared to AR models. This is expected as accuracy of OLS can only go up or stay the same with introduction of new variables. However, as the effect of adding sentiment seems to range from non-existent to very small, it does not seem to hold major predictive power in it. The largest differences are found in models for Outokumpu, Valoe and Nordea while the smallest differences, which are not even visible at shown precision, are in models for Ovaro and Sotkamo.

Table 11. RMSE for predictions

Company	AR	VAR
Nokia	0.0234	0.0230
Bittium	0.0151	0.0149
Biohit	0.0303	0.0302
Outokumpu	0.0252	0.0236
Ovaro Kiinteistösijoitus	0.0110	0.0110
Valoe	0.0742	0.0738
Sotkamo Silver	0.0171	0.0171
Revenio Group	0.0206	0.0205
Metsä Board	0.0230	0.0229
Nordea	0.0140	0.0121

Being able to predict the exact future price accurately is not always the most desirable feature of predictive model but instead the interest is more in the direction of the price or in other words, whether the return is going to be positive or negative. To assess this, correct sign percentages were calculated. They can be seen in Table 12. The improvement in correct sing percentage is not as uniform as the changes of RSME were. Some models, such as Outokumpu, Metsä Board, and Nordea significantly improved their correct sign percentage after the addition of sentiment, with Outokumpu having the highest improvement from 58% to 68% correctly predicted signs. Two models, Ovaro and Sotkamo, had their correct sign percentage slightly decrease for which the most likely explanation is that the predicted returns were very close to zero. The remaining stocks only achieved minor improvement with the inclusion of sentiment variable.

Table 12. Correct sign % for predictions

Company	AR	VAR
Nokia	0.49	0.53
Bittium	0.54	0.54
Biohit	0.46	0.47
Outokumpu	0.58	0.68
Ovaro Kiinteistösijoitus	0.47	0.46
Valoe	0.52	0.52
Sotkamo Silver	0.52	0.50
Revenio Group	0.51	0.52
Metsä Board	0.50	0.55
Nordea	0.57	0.64

Overall, results from both metrics combined show that the inclusion of sentiment variable seems to somewhat increase the predictive power of the models. However, the improvements in RMSE are so minor that they could be simply caused by the nature of OLS estimation. The correct sign percentage seems to support the idea that sentiment variable improves predictive power, but this is dependent on the stock and the model and cannot alone be used to draw any conclusions. The absolute values of correct sign percentages are also very close to 50%, apart from Outokumpu and Nordea, and thus do not seem to be that useful in practice. These metrics were also calculated on the same data that the model was fitted on, and thus the performance can be a lot worse on new data.

5.3.2 Granger causality results

Granger causality test is used to determine whether there is statistically significant predictive power in the daily sentiment variable. The Granger causality test is set up by having the previously created VAR models which included sentiment as unrestricted models and having models without the sentiment as restricted models. F-test framework is used to determine whether the sentiment regressors have significant effect on the returns.

The results of Granger causality test for whether daily sentiment variable Granger causes returns variable are shown in Table 14. The results are very clear and in almost

all cases the null hypothesis of no Granger causality cannot be rejected. Exception to this was Revenio Group's model, which rejects the null hypothesis on 1% significance level. However, based on the metrics presented in previous subchapter, the model does still not seem to be very good at predicting price or price direction. Overall, the results seem to show that there is no predictive power in sentiment variable towards returns. However, this is expected as many of models' coefficients themselves did not appear statistically significant.

Table 13. F-test for "sentiment Granger causes returns."

Sentiment causes returns	
Company	p-value
Nokia	0.295
Bittium	0.943
Biohit	0.346
Outokumpu	0.832
Ovaro Kiinteistösijoitus	0.111
Valoe	0.177
Sotkamo Silver	0.016
Revenio Group	0.002
Metsä Board	0.049
Nordea	0.682

However, it could be possible that there exists Granger causal relationship in the opposite direction with past returns affecting the sentiment variable. This Granger causality test was conducted in a similar manner as the previous one, but with the variables having swapped places. The results for this test are shown in Table 14. Once again, the null hypothesis cannot be rejected for any stock on 1% significance level.

Table 14. F-test for "returns Granger cause sentiment."

Returns cause sentiment	
Company	p-value
Nokia	0.225
Bittium	0.017
Biohit	0.859
Outokumpu	0.653
Ovaro Kiinteistösijoitus	0.718
Valoe	0.122
Sotkamo Silver	0.440
Revenio Group	0.446
Metsä Board	0.535
Nordea	0.081

The results of both Granger causality tests show that there does not seem to be any statistically significant predictive power in the sentiment variable towards returns, or vice versa. Therefore, the sentiment and returns seem to be completely independent from each other.

5.4 Discussion

There can be multiple reasons why Granger causality test did not manage to find any predictive power in the sentiment variable. The obvious one, which should be noted again, is that the poor performance of the classifier did not manage to completely capture the investor sentiment, and thus the results might not reflect the reality of the market. Baker & Wurgler (2007) have even stated that the way investor sentiment is measured has enormous impact on the results of any study studying it, making this a very likely cause. Audrino et al. (2020) state that high number of institutional investors in the market significantly lower the predictive power of investor sentiment. This also happens to be one of the defining characteristics of Finnish market and is strengthened by the low liquidity of the market (Jakobson & Korkeamäki 2014). The markets could also work more efficiently as the small number of retail investors cannot offset the actions of institutional investors who can be assumed to interpret any new information faster and more accurately than amateurs, leading to prices starting to reflect new information before it can be even fully discussed on the internet.

The chosen timeframe also has an effect. According to Ho et al. (2017) the existence of the relationship varies in time, and thus the results also differ based on the time frame and frequency of data chosen. It could be possible that even though there is no relationship in the data from 2019, it could still exist during other years. Other limitations of this study, such as only using the popular stocks, might affect the results. The less known stocks might have some effect that was not studied at all here. Rando et al. (2015) also found that the relationship between sentiment and returns only exists during major events when new information was revealed. Outside of these events the sentiment did not have any statistically significant effect but during them it could be used to make predictions on price direction.

6. SUMMARY AND CONCLUSION

This final chapter consists of a summary and conclusions. The first subchapter gives a quick summary of the data, methodology, and the findings of the study conducted. This is followed by a subchapter presenting the conclusions that could be made from the findings and aims to answer the research question presented in the introduction chapter. Finally, the last subchapter discusses possible future research that could be conducted on this topic.

6.1 Summary

Sentiment analysis allows a computer to automatically assess sentiment of a piece of text. It has been made possible with the advancements in AI research and, more importantly, the availability of large amounts of opinionated data online because of the increased prevalence of social media in modern daily life. Recently, there has been increasing number of attempts to incorporate this sentiment data into models that predict asset prices. However, this research has focused mainly on the English-speaking markets, such as NYSE and there is not much research done on smaller languages and markets, which is a gap this thesis attempted to fulfill.

This thesis attempted to study the relationship between investor sentiment and daily stock prices in Finnish markets with similar methodology that had been used to study the relationship in English-speaking markets. The stocks considered were limited to those traded on OMX Helsinki. The investor sentiment was measured by collecting a large dataset of message board posts regarding the most discussed companies on a Finnish financial newspaper Kauppalehti's discussion board in 2019. The posts were vectorized and classified as either positive (1), negative (-1), or neutral (0) using a Naïve Bayes classifier, which had been trained on a premade Finnish social media corpus. The classified posts were aggregated to create a daily sentiment variable, which was used as regressor in VAR model aiming to predict returns. The predictive power was evaluated by comparing predictive power of the models against similar ones which did not include sentiment using RMSE and correct sign percentage. Finally, a Granger causality analysis was conducted to see whether the sentiment variable had any statistically significant predictive power towards returns, or vice versa.

6.2 Conclusion

The sentiment variable calculated like in study by Anteweiler & Frank (2004) did not have statistically significant predictive power towards stock returns as the Granger causality was rejected on all stocks at 1% significance level. Neither was there any Granger causal relationship from the returns towards the sentiment variable. This goes against the results of research of, for example, Rao & Srivastava (2012), Ho et al. (2017) and Piñeiro-Chousa et al. (2018) who all found significant predictive power using similar methodology in English-speaking context. However, not all past research has been in agreement about the existence of the relationship. Ranco et al. (2015) and McGurk (2020) also did not find statistically significant links between sentiment and stock returns in their studies with similar methodology to this one.

The correct sign percentage improved slightly with the introduction of the sentiment variable. The highest value was 68% which was 10% higher than the percentage achieved by a model without sentiment. However, on average the correct sign percentage remained very close to 50% and did improve a lot after the introduction of sentiment variable. This is in line with research on English and Turkish markets by Nguyen et al. (2015) and Derakhshan & Beigy (2019), respectively, where the average was also close to 50% and the maximum reached was 70%. The 50% correct sign percentage is not that impressive and can arguably be achieved by a random number generator. It could also be as evidence that the prices follow a random walk from investor's point of view.

Overall, the results of this study seem to reject the idea that sentiment could be used to accurately predict stock prices. The research question presented in the introduction chapter can be answered that sentiment cannot be used to predict stock prices of OMX Helsinki. It would seem that any information existing in the sentiment is already reflected by the prices. However, there is a big caveat with the claim as the performance of the classifier used to classify the sentiment of social media posts is far from optimal and is bound to misclassify some of the posts. Thus, it is also possible that the results were caused more by the classifiers poor ability to measure investor sentiment correctly and completely. Therefore, the results cannot be considered reliable and should not be generalized.

6.3 Future research

Even though this study did not find evidence for predictive power in social media sentiment, there still is possibility of the relationship existing and further research into more accurate sentiment analysis models that could perform well on Finnish language is required. In addition, even studies English-speaking sentiment does not completely agree on whether there is any link between social media sentiment and returns. To further study the relationship, it could be beneficial to attempt to use different models in sentiment analysis, for example Artificial Neural Networks, such as Convoluted Neural Networks, have had some success in English literature (Derakhshan & Beigy 2019). Some literature, while using simple classifiers like Naïve Bayes or Support vector machines, also had used more advanced preprocessing to include part-of-speech tagging, negation handling, or intensification words. It could also be beneficial to combine multiple classifiers to achieve higher accuracy similar to the voting system by Das & Chen (2007). As Baker & Wurgler (2007) had assessed, the most challenging and important part is measuring the investor sentiment correctly.

In addition, using other sources for analysis could be beneficial. For example, using other sources, like Twitter, might capture a broader group of people and thus might include more general sentiment among the population, which might affect the valuation of assets more. The combination of multiple sources, such as multiple social media sites and various news articles could improve the accuracy of prediction (Li, Shang & Wang 2019). The sentiment's effect is also time-varying, which leads to possibility to using different timeframe or frequency yielding different results (Ho et al. 2017). Inclusion of news article sentiment could also improve the predictive models.

Finally, the further research could incorporate more variables into the predictive models. For example, the effect of sentiment could be used to predict future implied volatility or trading volume, which has been found to have stronger relationship by Checkley et al. (2017). In addition, number of messages, which was normalized away in this study, could be included as a variable to see whether asset's popularity online can be used to predict its price, volatility, or trading volume.

REFERENCES

Akaike, H. (1974) A new look at the statistical model identification. IEEE transactions on automatic control. 19 (6), 716–723.

Ali, S. R. M., Ahmed, S., & Östermark, R. (2020). Extreme returns and the investor's expectation for future volatility: Evidence from the Finnish stock market. The Quarterly Review of Economics and Finance, 76, 260-269.

Allen, D. M. (1974). The relationship between variable selection and data agumentation and a method for prediction. technometrics, 16(1), 125-127.

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

Audrino, F., Sigrist, F. & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. International Journal of Forecasting, 36 (2), pp. 334-357.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. Journal of economic perspectives, 21(2), 129-152.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. Journal of financial economics, 49(3), 307-343.

Beleites, C., Salzer, R., & Sergo, V. (2013). Validation of soft classification models using partial class memberships: An extended concept of sensitivity & co. applied to grading of astrocytoma tissues. Chemometrics and Intelligent Laboratory Systems, 122, 12-22.

Brill, E., & Mooney, R. J. (1997). An overview of empirical natural language processing. AI magazine, 18(4), 13-13.

Brooks, C. (2014) Introductory econometrics for finance. 3rd ed. Cambridge: Cambridge University Press.

Checkley, M. S., Higón, D. A., & Alles, H. (2017). The hasty wisdom of the mob: How market sentiment predicts stock market behavior. Expert Systems with applications, 77, 256-263.

Daniel, K., & Titman, S. (1999). Market efficiency in an irrational world. Financial Analysts Journal, 55(6), 28-40.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. Management science, 53(9), 1375-1388.

Deng, S., Huang, Z., Sinha, A.P. & Zhao, H. (2018) The Interaction Between Microblog Sentiment and Stock Returns: An Empirical Examination. MIS quarterly, 42(3), pp. 895–918.

Derakhshan, A., & Beigy, H. (2019). Sentiment analysis on stock social media for stock price movement prediction. Engineering Applications of Artificial Intelligence, 85, 569-578.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, 74(366a), 427-431.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. Machine learning, 29(2), 103-130.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the econometric society, 987-1007.

Evans, L., Owda, M., Crockett, K., & Vilas, A. F. (2019). A methodology for the resolution of cashtag collisions on Twitter–A natural language processing & data fusion approach. Expert Systems with Applications, 127, 353-369.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The journal of Finance, 25(2), 383-417.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

Freeman, J. R. (1983). Granger causality and the times series analysis of political relationships. American Journal of Political Science, 327-358.

G. M. Ljung; G. E. P. Box (1978). "On a Measure of a Lack of Fit in Time Series Models". Biometrika. 65 (2): 297–303.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society, 424-438.

Grigaliūnienė, Ž., & Cibulskienė, D. (2010). Investor sentiment effect on stock returns in Scandinavian stock market. Economics and Management, 15, 929-940.

Hamilton, J. D. (1994). Time series analysis. Princeton university press.

Ho, C. S., Damien, P., Gu, B., & Konana, P. (2017). The time-varying nature of social media sentiments in modeling stock returns. Decision Support Systems, 101, 69-81.

Hodge, V.J. & Austin, J. (2004) A survey of outlier detection methodologies. Artificial Intelligence Review, 22 (2). 85-126.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM computing surveys (CSUR), 31(3), 264-323.

Jakobson, U. & Korkeamäki, T. (2014) Omistus, omistajaohjaus ja määräysvalta suurissa suomalaisyrityksissä. Valtionneuvoston kanslian raporttisarja 5/2014. [In Finnish, abstract in Swedish]

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

Karppinen, K., Nieminen, H., & Markkanen, A. L. (2011). High professional ethos in a small, concentrated media market. The media for democracy monitor/J. Trappel, H. Nieminen, LW Nord (Eds.). Göteborg: Nordicom.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics, 54(1-3), 159–178.

Li, X., Shang, W., & Wang, S. (2019). Text-based crude oil price forecasting: A deep learning approach. International Journal of Forecasting, 35(4), 1548-1560.

Lindén, K., Jauhiainen, T., & Hardwick, S. (2020). FinnSentiment--A Finnish Social Media Corpus for Sentiment Polarity Annotation. arXiv preprint arXiv:2012.02613.

Liu, B. (2010). Sentiment analysis and subjectivity. Handbook of natural language processing, 2(2010), 627-666.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. Journal of Accounting Research, 54(4), 1187-1230.

Lütkepohl, H. (2005). Introduction to multiple time series analysis. Springer Science & Business Media.

MacKinnon, J.G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. Journal of Business and Economic Statistics 12, 167-76.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. Journal of economic perspectives, 17(1), 59-82.

Malkiel, B. G. (2005). Reflections on the efficient market hypothesis: 30 years later. Financial review, 40(1), 1-9.

Manning, C.D., Raghavan, P., Schuetze, H. (2008). Introduction to Information Retrieval. Cambridge University Press, pp. 234-265.

Maron, M. E. (1961). Automatic indexing: an experimental inquiry. Journal of the ACM (JACM), 8(3), 404-417.

McGurk, Z., Nowak, A. & Hall, J.C. (2020). Stock returns and investor sentiment: textual analysis and social media. Journal of Economics and Finance 44, 458–485.

Mittal, A., & Goel, A. (2012). Stock prediction using twitter sentiment analysis. Standford University

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the 2nd international conference on Knowledge capture (pp. 70-77).

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems with Applications, 42(24), 9603-9611.

Piñeiro-Chousa, J., López-Cabarcos, M. Á., Pérez-Pico, A. M., & Ribeiro-Navarrete, B. (2018). Does social network sentiment influence the relationship between the S&P 500 and gold returns?. International Review of Financial Analysis, 57, 57-64.

Piryani, R, Madhavi, D., Singh, V. K. (2017). Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. Information processing & management. 53 (1), 122–150.

Poole, D., Mackworth, A., Goebel, R. (1998). Computational Intelligence: A Logical Approach. New York: Oxford University Press

Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. PloS one. 10. e0138441. 10.1371/journal.pone.0138441.

Rao, T., & Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis. Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)

Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. In ICML (Vol. 3, pp. 616-623).

Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika, 71(3), 599-607.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information processing & management, 24(5), 513-523.

Salton, G., Fox, E. A., & Wu, H. (1983). Extended boolean information retrieval. Communications of the ACM, 26(11), 1022-1036.

Schwert, G. W. (2003). Anomalies and market efficiency. Handbook of the Economics of Finance, 1, 939-974.

Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.

Shapiro, S. S. & Wilk, M.B (1965). An analysis of variance test for normality (complete samples), Biometrika, Vol. 52, pp. 591-611.

Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., & Deng, X. (2013). Exploiting topic based twitter sentiment for stock prediction. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 24-29).

Simon, H. A. (1955). A behavioral model of rational choice. The quarterly journal of economics, 69(1), 99-118.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 111-133.

Timmermann, A., & Granger, C. W. (2004). Efficient market hypothesis and forecasting. International Journal of forecasting, 20(1), 15-27.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. science, 185(4157), 1124-1131.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. Psychonomic bulletin & review, 11(1), 192-196.

Wiriyathammabhum, P., Summers-Stay, D., Fermüller, C., & Aloimonos, Y. (2016). Computer vision and natural language processing: recent approaches in multimedia and robotics. ACM Computing Surveys (CSUR), 49(4), 1-44.

Yang, S. Y., Mo, S. Y. K., & Liu, A. (2015). Twitter financial community sentiment and its predictive relationship to stock market movement. Quantitative Finance, 15(10), 1637-1656.

Yu, Y., Duan, W., Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. Decision Support Systems. 55(4). 919-926.

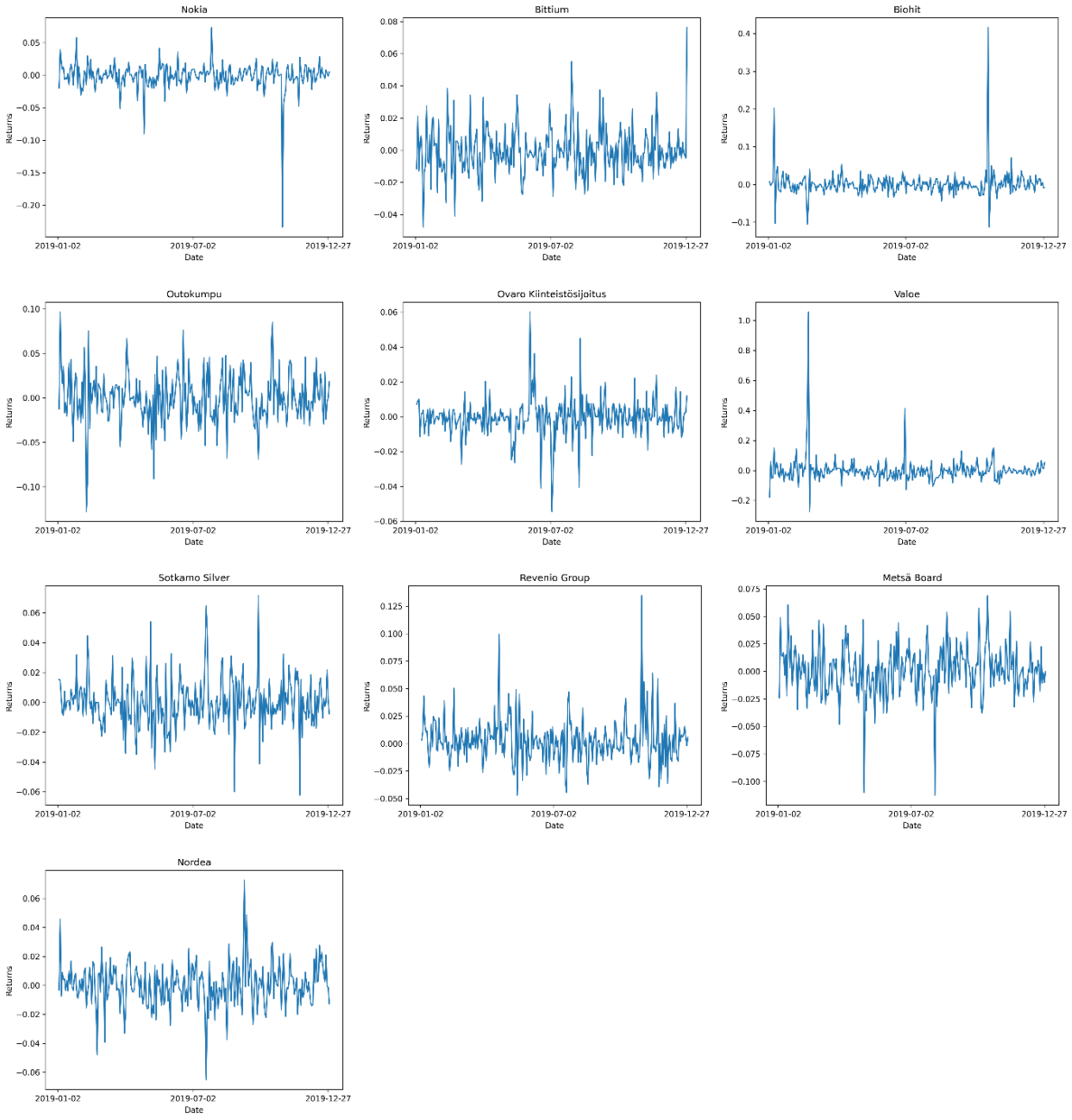
APPENDICES

Appendix 1. Number of messages for OMX Helsinki companies in 2019.

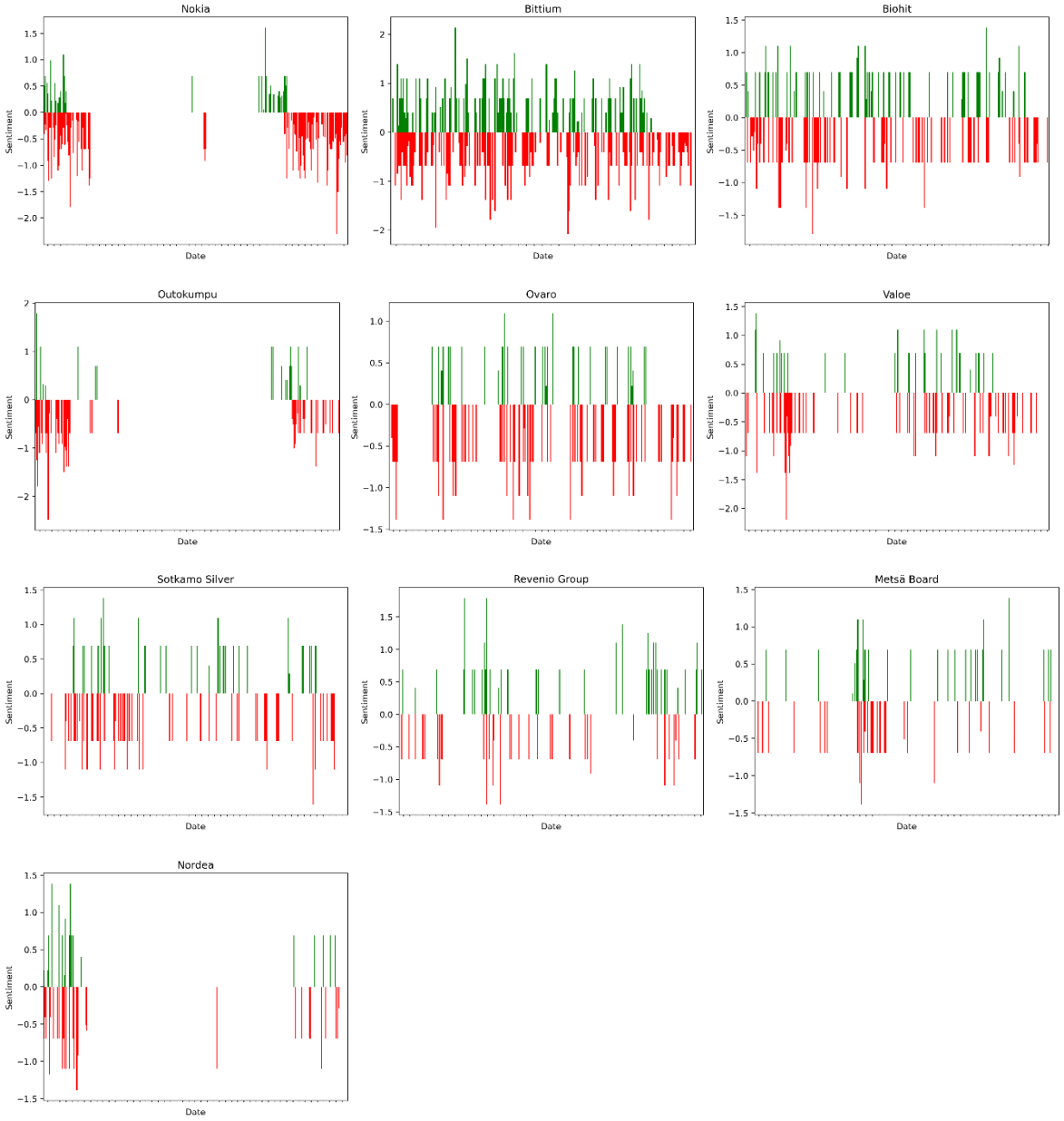
	Company	Posts
1	Nokia	16871
2	Bittium	8400
3	Biohit	3955
4	Outokumpu	3266
5	Ovaro Kiinteistösijoitus	3221
6	Valoe	2894
7	Sotkamo Silver	1725
8	Revenio Group	1665
9	Metsä Board	1230
10	Nordea	1064
11	Stockmann	1035
12	SSH Communications Security	930
13	Afarak Group	915
14	Componenta	900
15	SSAB	810
16	Fortum	780
17	INCAP	716
18	Lehto Group	705
19	Optomed	703
20	Rovio	645
21	Aspocomp Group	600
22	Rapala VMC	570
23	Wärtsilä	570
24	Orion	465
25	Stora Enso	450
26	Glaston	420
27	Tulikivi	420
28	UPM-Kymmene	390
29	Dovre Group	375
30	HKScan	345
31	Neste	334
32	Sampo	300
33	Nokian Renkaat	255
34	YIT	240
35	Ilkka-Yhtymä	226
36	Saga Furs	225
37	Metso Outotec	195
38	Verkkokauppa.com	165
39	Basware	165
40	Exel Composites	150
41	Atria	150
42	Panostaja	150
43	Citycon	150
44	F-Secure	150
45	Endomines	135
46	SRV Yhtiöt	105

	Company	Posts
47	Qt Group	105
48	Solteq	105
49	Teleste	105
50	CapMan	105
51	Raisio	95
52	Harvia	90
53	Kamux	90
54	Oriola	90
55	Valmet	90
56	Apetit	75
57	Fiskars	75
58	Finnair	75
59	Nixu	75
60	Ericsson	60
61	Innofactor	60
62	Talenom	60
63	Robit	45
64	Tokmanni	45
65	Digia	45
66	Marimekko	45
67	Terveystalo	45
68	Trainers House	36
69	Elisa	30
70	Lassila & Tikanoja	30
71	Konecranes	30
72	Kone	30
73	Kemira	30
74	Sievi Capital	30
75	Uponor	30
76	Oma Säästöpankki	30
77	Ponsse	30
78	Investors House	30
79	Telia Company	30
80	Caverion	30
81	eQ	15
82	Tieto EVRY	15
83	Plc Uutechnic Group	15
84	Consti	15
85	Taaleri	15
86	NoHo Partners	15
87	Pihlajalinna	15
88	Aktia	15
89	Honkarakenne	15
90	Scanfil	15
91	Altia	15
92	Elecster	5

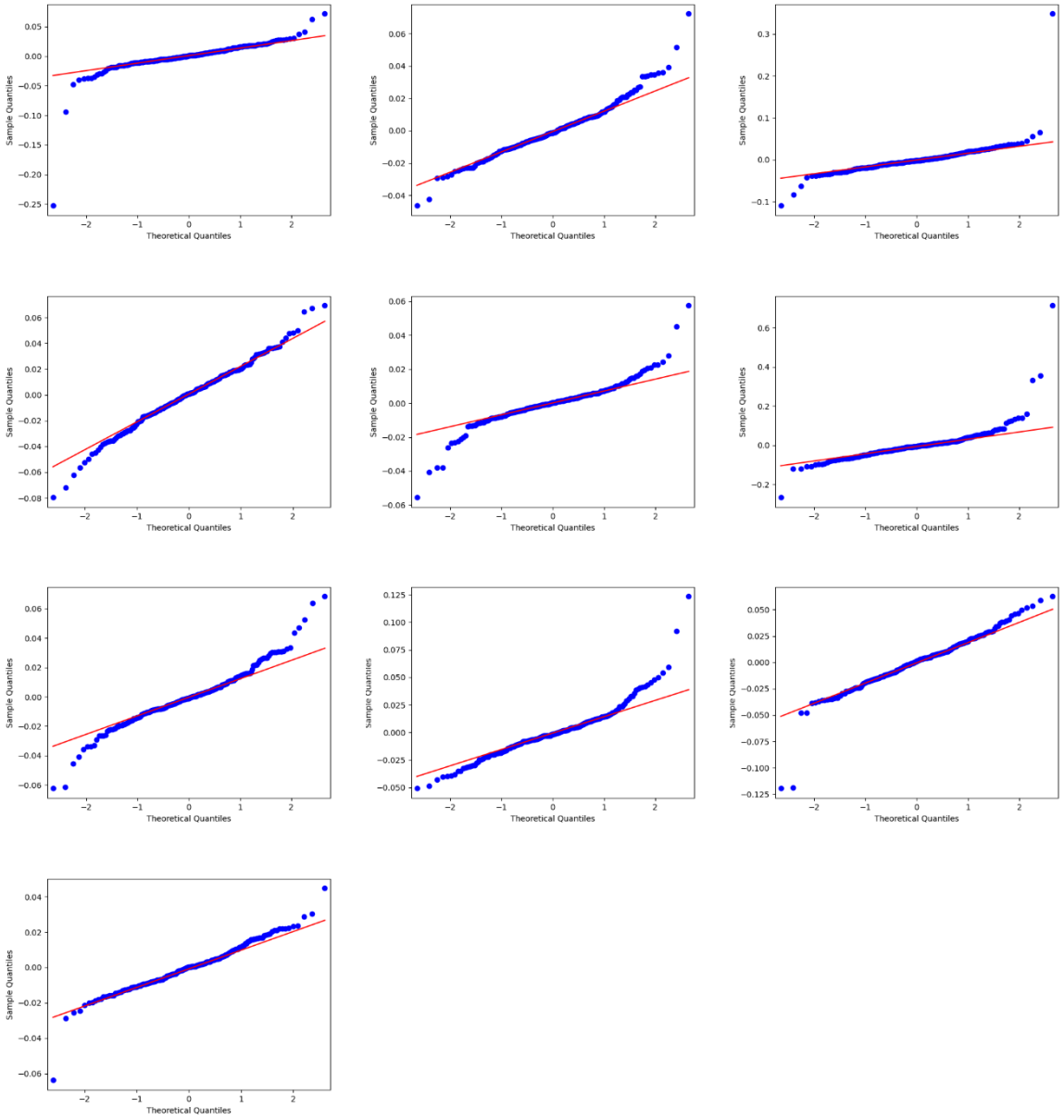
Appendix 2. Plots for daily returns



Appendix 3. Plots for daily sentiment



Appendix 4. Q-Q plots for residuals



Appendix 5. Regression results (VAR)

```
Nokia
VAR
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
---------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                   -9.05726
Nobs:                     243.000    HQIC:                  -9.28046
Log likelihood:           482.263    FPE:                8.02150e-05
AIC:                     -9.43100    Det(Omega_mle):     7.22750e-05
---------------------------------------------------------------------
Results for equation Returns
=============================================================================
                  coefficient       std. error           t-stat            prob
-----------------------------------------------------------------------------
const               -0.001805         0.001883           -0.959           0.338
L1.Returns           0.087829         0.066319            1.324           0.185
L1.Sentiment        -0.003708         0.004206           -0.881           0.378
L2.Returns           0.072065         0.065867            1.094           0.274
L2.Sentiment         0.001042         0.004165            0.250           0.802
L3.Returns           0.008922         0.066024            0.135           0.893
L3.Sentiment         0.009329         0.004174            2.235           0.025
L4.Returns          -0.020056         0.065558           -0.306           0.760
L4.Sentiment         0.000664         0.004158            0.160           0.873
L5.Returns          -0.083543         0.065420           -1.277           0.202
L5.Sentiment        -0.007848         0.004124           -1.903           0.057
L6.Returns          -0.024758         0.064989           -0.381           0.703
L6.Sentiment        -0.000194         0.004224           -0.046           0.963
=============================================================================

Results for equation Sentiment
=============================================================================
                  coefficient       std. error           t-stat            prob
-----------------------------------------------------------------------------
const               -0.062188         0.028830           -2.157           0.031
L1.Returns           1.692836         1.015162            1.668           0.095
L1.Sentiment         0.032514         0.064387            0.505           0.614
L2.Returns           0.037566         1.008236            0.037           0.970
L2.Sentiment         0.143086         0.063761            2.244           0.025
L3.Returns           0.718500         1.010652            0.711           0.477
L3.Sentiment         0.026195         0.063900            0.410           0.682
L4.Returns           1.850259         1.003516            1.844           0.065
L4.Sentiment         0.158298         0.063645            2.487           0.013
L5.Returns          -0.101805         1.001405           -0.102           0.919
L5.Sentiment         0.151383         0.063127            2.398           0.016
L6.Returns          -0.087436         0.994807           -0.088           0.930
L6.Sentiment         0.072190         0.064659            1.116           0.264
=============================================================================

Correlation matrix of residuals
              Returns  Sentiment
Returns      1.000000   0.112372
Sentiment    0.112372   1.000000


Bittium
VAR
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
```

```
---------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                   -8.96546
Nobs:                     248.000    HQIC:                  -9.01624
Log likelihood:           424.464    FPE:                0.000117337
AIC:                     -9.05046    Det(Omega_mle):     0.000114549
---------------------------------------------------------------------
Results for equation Returns
=============================================================================
                  coefficient       std. error           t-stat            prob
-----------------------------------------------------------------------------
const               -0.000387         0.000960           -0.403           0.687
L1.Returns          -0.034048         0.067499           -0.504           0.614
L1.Sentiment         0.003215         0.001348            2.385           0.017
=============================================================================

Results for equation Sentiment
=============================================================================
                  coefficient       std. error           t-stat            prob
-----------------------------------------------------------------------------
const               -0.070123         0.046226           -1.517           0.129
L1.Returns           0.233170         3.251910            0.072           0.943
L1.Sentiment         0.007138         0.064949            0.110           0.912
=============================================================================

Correlation matrix of residuals
              Returns  Sentiment
Returns      1.000000   0.174678
Sentiment    0.174678   1.000000


Biohit
VAR
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
---------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                   -7.92038
Nobs:                     244.000    HQIC:                  -8.10870
Log likelihood:           334.313    FPE:                0.000265055
AIC:                     -8.23570    Det(Omega_mle):     0.000242681
---------------------------------------------------------------------
Results for equation Returns
=============================================================================
                  coefficient       std. error           t-stat            prob
-----------------------------------------------------------------------------
const                0.000012         0.002035            0.006           0.995
L1.Returns          -0.228030         0.061172           -3.728           0.000
L1.Sentiment        -0.001557         0.004011           -0.388           0.698
L2.Returns          -0.048660         0.062362           -0.780           0.435
L2.Sentiment         0.001902         0.004016            0.474           0.636
L3.Returns           0.030894         0.062586            0.494           0.622
L3.Sentiment        -0.000755         0.004013           -0.188           0.851
L4.Returns          -0.065218         0.062841           -1.038           0.299
L4.Sentiment         0.005177         0.003970            1.304           0.192
L5.Returns          -0.018100         0.061541           -0.294           0.769
L5.Sentiment        -0.000509         0.003973           -0.128           0.898
=============================================================================

Results for equation Sentiment
=============================================================================
                  coefficient       std. error           t-stat            prob
-----------------------------------------------------------------------------
const               -0.044577         0.033234           -1.341           0.180
L1.Returns           0.068916         0.999008            0.069           0.945
L1.Sentiment         0.103469         0.065504            1.580           0.114
```

```
L2.Returns           1.257175         1.018447           1.234           0.217
L2.Sentiment        -0.075836         0.065585          -1.156           0.248
L3.Returns          -1.492616         1.022102          -1.460           0.144
L3.Sentiment         0.042079         0.065543           0.642           0.521
L4.Returns          -1.233632         1.026265          -1.202           0.229
L4.Sentiment         0.010934         0.064840           0.169           0.866
L5.Returns          -0.589557         1.005024          -0.587           0.557
L5.Sentiment         0.037437         0.064889           0.577           0.564
================================================================================

Correlation matrix of residuals
              Returns  Sentiment
Returns      1.000000   0.063484
Sentiment    0.063484   1.000000


Outokumpu
VAR
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
---------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                   -7.64904
Nobs:                     222.000    HQIC:                  -8.65434
Log likelihood:           516.182    FPE:                9.01540e-05
AIC:                     -9.33505    Det(Omega_mle):     5.79070e-05
---------------------------------------------------------------------
Results for equation Returns
================================================================================
                    coefficient       std. error           t-stat            prob
--------------------------------------------------------------------------------
const                -0.002247         0.002086          -1.077           0.281
L1.Returns           -0.024413         0.075426          -0.324           0.746
L1.Sentiment         -0.012859         0.007502          -1.714           0.087
L2.Returns            0.093614         0.072057           1.299           0.194
L2.Sentiment         -0.008885         0.007252          -1.225           0.220
L3.Returns            0.096201         0.071497           1.346           0.178
L3.Sentiment         -0.002133         0.007087          -0.301           0.763
L4.Returns            0.033628         0.070451           0.477           0.633
L4.Sentiment         -0.005316         0.007126          -0.746           0.456
L5.Returns           -0.040008         0.069359          -0.577           0.564
L5.Sentiment          0.000308         0.006904           0.045           0.964
L6.Returns           -0.052660         0.068979          -0.763           0.445
L6.Sentiment         -0.001751         0.006917          -0.253           0.800
L7.Returns           -0.086842         0.069296          -1.253           0.210
L7.Sentiment          0.003737         0.006894           0.542           0.588
L8.Returns           -0.090844         0.068966          -1.317           0.188
L8.Sentiment          0.002776         0.007095           0.391           0.696
L9.Returns           -0.034820         0.069654          -0.500           0.617
L9.Sentiment         -0.003083         0.007108          -0.434           0.665
L10.Returns           0.059580         0.069691           0.855           0.393
L10.Sentiment         0.000904         0.007042           0.128           0.898
L11.Returns           0.046153         0.068545           0.673           0.501
L11.Sentiment         0.012229         0.006783           1.803           0.071
L12.Returns          -0.045885         0.068939          -0.666           0.506
L12.Sentiment        -0.008093         0.006986          -1.159           0.247
L13.Returns          -0.154344         0.068908          -2.240           0.025
L13.Sentiment        -0.001032         0.007091          -0.145           0.884
L14.Returns          -0.058896         0.068645          -0.858           0.391
L14.Sentiment         0.004484         0.007119           0.630           0.529
L15.Returns          -0.092236         0.067438          -1.368           0.171
L15.Sentiment         0.000473         0.007176           0.066           0.947
L16.Returns          -0.019663         0.067548          -0.291           0.771
L16.Sentiment        -0.003000         0.007278          -0.412           0.680
L17.Returns           0.009509         0.066242           0.144           0.886
```

L17.Sentiment	-0.006864	0.007117	-0.964	0.335
L18.Returns	0.018618	0.066222	0.281	0.779
L18.Sentiment	-0.001950	0.006652	-0.293	0.769
L19.Returns	0.043508	0.066147	0.658	0.511
L19.Sentiment	0.003144	0.006527	0.482	0.630
L20.Returns	-0.028016	0.065926	-0.425	0.671
L20.Sentiment	-0.007677	0.006380	-1.203	0.229
L21.Returns	-0.032161	0.065355	-0.492	0.623
L21.Sentiment	0.012258	0.006492	1.888	0.059
L22.Returns	-0.109639	0.065404	-1.676	0.094
L22.Sentiment	0.002468	0.006657	0.371	0.711
L23.Returns	-0.136128	0.066334	-2.052	0.040
L23.Sentiment	0.007311	0.005900	1.239	0.215
L24.Returns	0.002057	0.067050	0.031	0.976
L24.Sentiment	-0.006224	0.006180	-1.007	0.314
L25.Returns	-0.048631	0.067138	-0.724	0.469
L25.Sentiment	-0.008058	0.006255	-1.288	0.198
L26.Returns	0.055638	0.067204	0.828	0.408
L26.Sentiment	0.003882	0.006000	0.647	0.518
L27.Returns	-0.044663	0.067457	-0.662	0.508
L27.Sentiment	-0.004599	0.005928	-0.776	0.438

Results for equation Sentiment

	coefficient	std. error	t-stat	prob
const	-0.023362	0.021529	-1.085	0.278
L1.Returns	0.544681	0.778587	0.700	0.484
L1.Sentiment	0.050603	0.077436	0.653	0.513
L2.Returns	0.733963	0.743811	0.987	0.324
L2.Sentiment	0.071087	0.074858	0.950	0.342
L3.Returns	-0.348709	0.738034	-0.472	0.637
L3.Sentiment	-0.108943	0.073153	-1.489	0.136
L4.Returns	-0.201128	0.727235	-0.277	0.782
L4.Sentiment	-0.032500	0.073560	-0.442	0.659
L5.Returns	-0.395326	0.715961	-0.552	0.581
L5.Sentiment	0.113296	0.071266	1.590	0.112
L6.Returns	0.770153	0.712041	1.082	0.279
L6.Sentiment	0.004741	0.071404	0.066	0.947
L7.Returns	-0.148819	0.715310	-0.208	0.835
L7.Sentiment	0.146966	0.071163	2.065	0.039
L8.Returns	-1.300228	0.711905	-1.826	0.068
L8.Sentiment	0.089040	0.073239	1.216	0.224
L9.Returns	-0.402089	0.719013	-0.559	0.576
L9.Sentiment	0.088122	0.073378	1.201	0.230
L10.Returns	0.157643	0.719393	0.219	0.827
L10.Sentiment	-0.023206	0.072694	-0.319	0.750
L11.Returns	-0.965059	0.707555	-1.364	0.173
L11.Sentiment	-0.057220	0.070017	-0.817	0.414
L12.Returns	-0.213593	0.711627	-0.300	0.764
L12.Sentiment	-0.134378	0.072112	-1.863	0.062
L13.Returns	0.009791	0.711310	0.014	0.989
L13.Sentiment	0.051040	0.073201	0.697	0.486
L14.Returns	-0.560781	0.708595	-0.791	0.429
L14.Sentiment	-0.022522	0.073486	-0.306	0.759
L15.Returns	0.539868	0.696128	0.776	0.438
L15.Sentiment	0.053913	0.074071	0.728	0.467
L16.Returns	-0.610511	0.697267	-0.876	0.381
L16.Sentiment	-0.134721	0.075128	-1.793	0.073
L17.Returns	0.533989	0.683790	0.781	0.435
L17.Sentiment	-0.075584	0.073469	-1.029	0.304
L18.Returns	0.495514	0.683576	0.725	0.469
L18.Sentiment	0.007246	0.068668	0.106	0.916
L19.Returns	-0.601578	0.682811	-0.881	0.378
L19.Sentiment	0.025923	0.067377	0.385	0.700

```
L20.Returns          -0.117713         0.680523           -0.173           0.863
L20.Sentiment         0.118638         0.065860            1.801           0.072
L21.Returns           0.362828         0.674627            0.538           0.591
L21.Sentiment         0.209442         0.067010            3.126           0.002
L22.Returns          -0.674005         0.675142           -0.998           0.318
L22.Sentiment         0.063758         0.068718            0.928           0.353
L23.Returns          -0.515222         0.684739           -0.752           0.452
L23.Sentiment        -0.255722         0.060907           -4.199           0.000
L24.Returns           0.597487         0.692128            0.863           0.388
L24.Sentiment        -0.032043         0.063799           -0.502           0.615
L25.Returns          -0.987579         0.693039           -1.425           0.154
L25.Sentiment         0.004510         0.064569            0.070           0.944
L26.Returns           0.524001         0.693713            0.755           0.450
L26.Sentiment        -0.011966         0.061932           -0.193           0.847
L27.Returns          -0.177461         0.696329           -0.255           0.799
L27.Sentiment        -0.002520         0.061192           -0.041           0.967
==============================================================================

Correlation matrix of residuals
              Returns  Sentiment
Returns      1.000000  -0.015273
Sentiment   -0.015273   1.000000


Ovaro Kiinteistösijoitus
VAR
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
----------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                   -10.2474
Nobs:                     245.000    HQIC:                  -10.4010
Log likelihood:           609.533    FPE:                2.74121e-05
AIC:                     -10.5046    Det(Omega_mle):     2.55039e-05
----------------------------------------------------------------------
Results for equation Returns
==============================================================================
                  coefficient       std. error           t-stat            prob
------------------------------------------------------------------------------
const               -0.000619         0.000836           -0.741           0.459
L1.Returns          -0.049235         0.064854           -0.759           0.448
L1.Sentiment        -0.000149         0.001594           -0.094           0.925
L2.Returns           0.102968         0.064629            1.593           0.111
L2.Sentiment        -0.002152         0.001596           -1.348           0.178
L3.Returns           0.122986         0.064939            1.894           0.058
L3.Sentiment         0.000648         0.001578            0.411           0.681
L4.Returns           0.087615         0.065468            1.338           0.181
L4.Sentiment         0.000648         0.001578            0.411           0.681
==============================================================================

Results for equation Sentiment
==============================================================================
                  coefficient       std. error           t-stat            prob
------------------------------------------------------------------------------
const               -0.104976         0.033627           -3.122           0.002
L1.Returns           4.170644         2.608100            1.599           0.110
L1.Sentiment        -0.033874         0.064091           -0.529           0.597
L2.Returns          -5.117031         2.599035           -1.969           0.049
L2.Sentiment         0.036421         0.064191            0.567           0.570
L3.Returns          -2.708276         2.611524           -1.037           0.300
L3.Sentiment         0.002512         0.063479            0.040           0.968
L4.Returns          -0.897971         2.632798           -0.341           0.733
L4.Sentiment         0.124656         0.063441            1.965           0.049
==============================================================================
```

```
Correlation matrix of residuals
              Returns  Sentiment
Returns      1.000000   0.024505
Sentiment    0.024505   1.000000


Valoe
VAR
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
---------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                   -6.58690
Nobs:                     248.000    HQIC:                  -6.63768
Log likelihood:           129.522    FPE:                  0.00126600
AIC:                     -6.67190    Det(Omega_mle):       0.00123591
---------------------------------------------------------------------
Results for equation Returns
=============================================================================
                  coefficient       std. error           t-stat            prob
-----------------------------------------------------------------------------
const               -0.000448         0.004873           -0.092           0.927
L1.Returns          -0.025398         0.064264           -0.395           0.693
L1.Sentiment         0.015451         0.009986            1.547           0.122
=============================================================================

Results for equation Sentiment
=============================================================================
                  coefficient       std. error           t-stat            prob
-----------------------------------------------------------------------------
const               -0.120468         0.031818           -3.786           0.000
L1.Returns          -0.567603         0.419644           -1.353           0.176
L1.Sentiment        -0.023435         0.065209           -0.359           0.719
=============================================================================

Correlation matrix of residuals
              Returns  Sentiment
Returns      1.000000  -0.220692
Sentiment   -0.220692   1.000000


Sotkamo Silver
VAR
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
---------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                   -9.59625
Nobs:                     248.000    HQIC:                  -9.64703
Log likelihood:           502.681    FPE:                6.24436e-05
AIC:                     -9.68125    Det(Omega_mle):     6.09598e-05
---------------------------------------------------------------------
Results for equation Returns
=============================================================================
                  coefficient       std. error           t-stat            prob
-----------------------------------------------------------------------------
const               -0.000372         0.001102           -0.338           0.736
L1.Returns          -0.028991         0.063838           -0.454           0.650
L1.Sentiment        -0.001841         0.002383           -0.773           0.440
=============================================================================

Results for equation Sentiment
=============================================================================
                  coefficient       std. error           t-stat            prob
```

```
------------------------------------------------------------------------------
const               -0.066929         0.029249           -2.288           0.022
L1.Returns           4.108676         1.694702            2.424           0.015
L1.Sentiment        -0.006964         0.063251           -0.110           0.912
==============================================================================

Correlation matrix of residuals
              Returns  Sentiment
Returns      1.000000  -0.060775
Sentiment   -0.060775   1.000000


Revenio Group
VAR
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
---------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                   -9.50117
Nobs:                     248.000    HQIC:                  -9.55196
Log likelihood:           490.892    FPE:                6.86718e-05
AIC:                     -9.58617    Det(Omega_mle):     6.70401e-05
---------------------------------------------------------------------
Results for equation Returns
==============================================================================
                  coefficient       std. error           t-stat            prob
------------------------------------------------------------------------------
const                0.002972         0.001327            2.241           0.025
L1.Returns          -0.003444         0.063978           -0.054           0.957
L1.Sentiment         0.002479         0.003249            0.763           0.445
==============================================================================

Results for equation Sentiment
==============================================================================
                  coefficient       std. error           t-stat            prob
------------------------------------------------------------------------------
const               -0.001494         0.025493           -0.059           0.953
L1.Returns           3.801501         1.229515            3.092           0.002
L1.Sentiment        -0.007284         0.062444           -0.117           0.907
==============================================================================

Correlation matrix of residuals
              Returns  Sentiment
Returns      1.000000  -0.072090
Sentiment   -0.072090   1.000000


Metsä Board
VAR
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
---------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                   -9.48019
Nobs:                     246.000    HQIC:                  -9.59936
Log likelihood:           506.483    FPE:                6.25432e-05
AIC:                     -9.67968    Det(Omega_mle):     5.91302e-05
---------------------------------------------------------------------
Results for equation Returns
==============================================================================
                  coefficient       std. error           t-stat            prob
------------------------------------------------------------------------------
const                0.000620         0.001487            0.417           0.677
L1.Returns           0.071017         0.065523            1.084           0.278
```

```
L1.Sentiment        -0.000094         0.004449           -0.021           0.983
L2.Returns           0.057246         0.064885            0.882           0.378
L2.Sentiment         0.000500         0.004447            0.112           0.911
L3.Returns           0.005126         0.064640            0.079           0.937
L3.Sentiment         0.006563         0.004439            1.478           0.139
================================================================================

Results for equation Sentiment
================================================================================
                  coefficient       std. error           t-stat            prob
--------------------------------------------------------------------------------
const               -0.012403         0.021622           -0.574           0.566
L1.Returns          -0.121240         0.952591           -0.127           0.899
L1.Sentiment        -0.062583         0.064687           -0.967           0.333
L2.Returns           0.258370         0.943309            0.274           0.784
L2.Sentiment         0.005029         0.064653            0.078           0.938
L3.Returns           2.609647         0.939748            2.777           0.005
L3.Sentiment         0.037412         0.064533            0.580           0.562
================================================================================

Correlation matrix of residuals
              Returns  Sentiment
Returns      1.000000   0.186061
Sentiment    0.186061   1.000000


Nordea
VAR
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
---------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                   -9.07731
Nobs:                     214.000    HQIC:                  -10.4083
Log likelihood:           744.950    FPE:                1.28957e-05
AIC:                     -11.3108    Det(Omega_mle):     7.27081e-06
---------------------------------------------------------------------
Results for equation Returns
================================================================================
                  coefficient       std. error           t-stat            prob
--------------------------------------------------------------------------------
const               -0.000925         0.001332           -0.695           0.487
L1.Returns           0.153890         0.082346            1.869           0.062
L1.Sentiment         0.003027         0.006733            0.450           0.653
L2.Returns           0.056137         0.082479            0.681           0.496
L2.Sentiment         0.003500         0.006073            0.576           0.564
L3.Returns           0.018215         0.081369            0.224           0.823
L3.Sentiment         0.002572         0.006132            0.419           0.675
L4.Returns           0.093678         0.081532            1.149           0.251
L4.Sentiment        -0.007879         0.006177           -1.276           0.202
L5.Returns          -0.058165         0.080845           -0.719           0.472
L5.Sentiment        -0.001146         0.006069           -0.189           0.850
L6.Returns          -0.069204         0.081106           -0.853           0.394
L6.Sentiment         0.010324         0.006114            1.689           0.091
L7.Returns           0.155149         0.080169            1.935           0.053
L7.Sentiment        -0.000507         0.005987           -0.085           0.933
L8.Returns          -0.092813         0.082638           -1.123           0.261
L8.Sentiment         0.003005         0.006173            0.487           0.626
L9.Returns          -0.063177         0.083844           -0.753           0.451
L9.Sentiment         0.000981         0.006252            0.157           0.875
L10.Returns          0.033746         0.085291            0.396           0.692
L10.Sentiment       -0.006793         0.006239           -1.089           0.276
L11.Returns          0.103705         0.083899            1.236           0.216
L11.Sentiment        0.006265         0.006283            0.997           0.319
L12.Returns         -0.141487         0.084084           -1.683           0.092
```

L12.Sentiment	-0.001334	0.006208	-0.215	0.830
L13.Returns	0.029162	0.084214	0.346	0.729
L13.Sentiment	-0.016521	0.006072	-2.721	0.007
L14.Returns	-0.046521	0.084062	-0.553	0.580
L14.Sentiment	0.004182	0.006105	0.685	0.493
L15.Returns	0.026224	0.082608	0.317	0.751
L15.Sentiment	0.006032	0.005742	1.051	0.293
L16.Returns	0.060898	0.083573	0.729	0.466
L16.Sentiment	-0.011987	0.005793	-2.069	0.039
L17.Returns	-0.071734	0.083109	-0.863	0.388
L17.Sentiment	-0.007230	0.005601	-1.291	0.197
L18.Returns	-0.177733	0.081022	-2.194	0.028
L18.Sentiment	0.009662	0.005348	1.807	0.071
L19.Returns	0.090511	0.081537	1.110	0.267
L19.Sentiment	-0.003061	0.005668	-0.540	0.589
L20.Returns	0.040792	0.080832	0.505	0.614
L20.Sentiment	-0.002839	0.005615	-0.506	0.613
L21.Returns	-0.066877	0.080570	-0.830	0.407
L21.Sentiment	0.002336	0.005664	0.412	0.680
L22.Returns	-0.027013	0.080769	-0.334	0.738
L22.Sentiment	-0.002533	0.005529	-0.458	0.647
L23.Returns	-0.074527	0.081194	-0.918	0.359
L23.Sentiment	0.005818	0.005431	1.071	0.284
L24.Returns	0.001583	0.081377	0.019	0.984
L24.Sentiment	-0.002540	0.005346	-0.475	0.635
L25.Returns	0.151800	0.081144	1.871	0.061
L25.Sentiment	-0.003424	0.005270	-0.650	0.516
L26.Returns	0.026343	0.081985	0.321	0.748
L26.Sentiment	0.000474	0.005189	0.091	0.927
L27.Returns	-0.049935	0.080690	-0.619	0.536
L27.Sentiment	-0.002457	0.005004	-0.491	0.623
L28.Returns	-0.012995	0.080075	-0.162	0.871
L28.Sentiment	-0.002265	0.005137	-0.441	0.659
L29.Returns	0.007375	0.078610	0.094	0.925
L29.Sentiment	-0.011533	0.005042	-2.287	0.022
L30.Returns	-0.059178	0.079407	-0.745	0.456
L30.Sentiment	-0.002039	0.005026	-0.406	0.685
L31.Returns	-0.013741	0.079433	-0.173	0.863
L31.Sentiment	-0.002472	0.004978	-0.497	0.619
L32.Returns	0.025524	0.079061	0.323	0.747
L32.Sentiment	-0.002103	0.005035	-0.418	0.676
L33.Returns	0.028739	0.078989	0.364	0.716
L33.Sentiment	0.007835	0.004834	1.621	0.105
L34.Returns	0.018870	0.077056	0.245	0.807
L34.Sentiment	-0.005667	0.004760	-1.190	0.234
L35.Returns	-0.200123	0.077729	-2.575	0.010
L35.Sentiment	0.004563	0.004598	0.992	0.321

Results for equation Sentiment

	coefficient	std. error	t-stat	prob
const	-0.028678	0.016470	-1.741	0.082
L1.Returns	-0.718834	1.018441	-0.706	0.480
L1.Sentiment	0.013596	0.083276	0.163	0.870
L2.Returns	-0.315256	1.020088	-0.309	0.757
L2.Sentiment	0.148594	0.075112	1.978	0.048
L3.Returns	-0.749816	1.006358	-0.745	0.456
L3.Sentiment	-0.159638	0.075844	-2.105	0.035
L4.Returns	0.361074	1.008376	0.358	0.720
L4.Sentiment	-0.054031	0.076399	-0.707	0.479
L5.Returns	-1.349233	0.999879	-1.349	0.177
L5.Sentiment	-0.106240	0.075055	-1.415	0.157
L6.Returns	0.185485	1.003104	0.185	0.853
L6.Sentiment	0.033411	0.075611	0.442	0.659

```
L7.Returns          -1.424931        0.991511          -1.437          0.151
L7.Sentiment         0.190652        0.074048           2.575          0.010
L8.Returns           1.935074        1.022049           1.893          0.058
L8.Sentiment        -0.149379        0.076341          -1.957          0.050
L9.Returns           0.403208        1.036972           0.389          0.697
L9.Sentiment         0.092207        0.077324           1.192          0.233
L10.Returns          1.133566        1.054858           1.075          0.283
L10.Sentiment       -0.006883        0.077167          -0.089          0.929
L11.Returns         -1.018212        1.037654          -0.981          0.326
L11.Sentiment        0.094829        0.077709           1.220          0.222
L12.Returns         -0.396867        1.039932          -0.382          0.703
L12.Sentiment       -0.012839        0.076775          -0.167          0.867
L13.Returns          0.995285        1.041545           0.956          0.339
L13.Sentiment        0.018184        0.075104           0.242          0.809
L14.Returns         -0.851539        1.039670          -0.819          0.413
L14.Sentiment        0.142188        0.075504           1.883          0.060
L15.Returns         -1.813871        1.021680          -1.775          0.076
L15.Sentiment        0.062591        0.071017           0.881          0.378
L16.Returns          1.390281        1.033617           1.345          0.179
L16.Sentiment       -0.096279        0.071646          -1.344          0.179
L17.Returns         -0.902025        1.027874          -0.878          0.380
L17.Sentiment        0.102618        0.069273           1.481          0.139
L18.Returns         -0.847971        1.002062          -0.846          0.397
L18.Sentiment       -0.032318        0.066137          -0.489          0.625
L19.Returns         -0.521463        1.008436          -0.517          0.605
L19.Sentiment        0.079606        0.070101           1.136          0.256
L20.Returns         -0.163842        0.999712          -0.164          0.870
L20.Sentiment       -0.148883        0.069441          -2.144          0.032
L21.Returns         -0.716560        0.996481          -0.719          0.472
L21.Sentiment        0.099028        0.070052           1.414          0.157
L22.Returns          1.352287        0.998941           1.354          0.176
L22.Sentiment        0.041232        0.068378           0.603          0.547
L23.Returns         -1.607968        1.004196          -1.601          0.109
L23.Sentiment        0.079860        0.067171           1.189          0.234
L24.Returns         -0.861480        1.006453          -0.856          0.392
L24.Sentiment       -0.088431        0.066120          -1.337          0.181
L25.Returns         -0.048371        1.003573          -0.048          0.962
L25.Sentiment        0.116086        0.065180           1.781          0.075
L26.Returns         -0.147979        1.013977          -0.146          0.884
L26.Sentiment        0.043063        0.064172           0.671          0.502
L27.Returns          0.324068        0.997958           0.325          0.745
L27.Sentiment       -0.067626        0.061889          -1.093          0.275
L28.Returns          0.381466        0.990355           0.385          0.700
L28.Sentiment       -0.022720        0.063530          -0.358          0.721
L29.Returns         -1.097372        0.972238          -1.129          0.259
L29.Sentiment       -0.105420        0.062354          -1.691          0.091
L30.Returns          0.784593        0.982094           0.799          0.424
L30.Sentiment        0.001618        0.062160           0.026          0.979
L31.Returns          1.111564        0.982415           1.131          0.258
L31.Sentiment       -0.117915        0.061566          -1.915          0.055
L32.Returns         -0.115303        0.977814          -0.118          0.906
L32.Sentiment       -0.062358        0.062268          -1.001          0.317
L33.Returns          0.178773        0.976920           0.183          0.855
L33.Sentiment       -0.034559        0.059789          -0.578          0.563
L34.Returns         -1.921186        0.953017          -2.016          0.044
L34.Sentiment       -0.018818        0.058877          -0.320          0.749
L35.Returns         -0.981447        0.961334          -1.021          0.307
L35.Sentiment       -0.009652        0.056868          -0.170          0.865
==============================================================================

Correlation matrix of residuals
              Returns  Sentiment
Returns      1.000000   0.103774
Sentiment    0.103774   1.000000
```