UNIVERSITY OF TARTU
Faculty of Social Sciences
School of Economics and Business Administration

Master's thesis

EFFECT OF REAL ESTATE NEWS SENTIMENT ON STOCK RETURNS OF SWEDBANK AND SEB BANK

Yuliia Puzanova

Supervisor: Mustafa Hakan Eratalay (Ph D)

Tartu 2019

Name and signature of supervisor . . . .

Allowed for defence on . . . .
(date)

I have written this master's thesis independently. All viewpoints of other authors, literary sources and data from elsewhere used for writing this paper have been referenced.

. . . .
(signature of author)

ABSTRACT

The paper analyses the effect of real estate news sentiment on stock returns of Swedbank and SEB Bank, leading banks in Sweden and the Baltic region. For this purpose, we have selected sentiments of news about real estate in markets of chosen banks: Sweden, Estonia, Latvia, and Lithuania over the period: 04.01.2016 to 19.02.2019. Estimated models showed interesting results: we proved that housing market sentiments affect banks' well-being, and showed a presence of asymmetry of positive and negative news. Aggregated news without separation by sentiments effects stock returns differently: Swedish news influences the Swedbank's stock returns when for SEB, news about Baltic real estate is more significant. When separating negative news, Baltic real estate news shows a considerable effect on stock returns of both banks. Furthermore, the aggregated number of negative news available does not have a considerable impact on returns, while the sentiment of negative news has. For positive news, it was proved to be working conversely.

Keywords: *econometric models, sentiment analysis, real estate market, Sweden, Baltics.*

ACKNOWLEDGEMENTS

First of all, I thank my supervisor Mustafa Hakan Eratalay for his contribution, and encouragement. I am grateful to all lecturers for their contribution to my luggage of knowledge and organizing the study process during these 2 outstanding years.

I am grateful for the financial support provided to me by the Republic of Estonia and the University of Tartu. Thank you for offering an opportunity to study in this wonderful environment, and also for a chance to meet wonderful people and find amazing friends.

Also, I want to express gratitude to my colleagues for letting me spend as much time as I needed for working on this paper, and for a pleasant atmosphere and great mood.

And, most importantly, I thank a lot to my family, my Mum, Dad, and brother Vanya for unlimited love and support in my decisions which follows me distantly across borders and time.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Nordic and the Baltic banking sectors are closely connected. Two systematically important to the Baltic financial system banks: Swedbank and SEB banks are highly exposed to risks in their real estate home market in Sweden. Moreover, we cannot put aside the risk coming from other home markets - Baltic states. Real estate news appeared in all markets can be influential for the banks' activities related to lending services, and impact the mortgage volume growth. As a result, we can expect an impact on banks' profitability and their stock prices returns.

The news regarding the housing shock and future real estate crash is getting more and more attention. This is happening both in Sweden and in the Baltics. The grounds for such discussions about possible market exposure are different for all countries, although the result might have the same effect.

The Swedish economy characterized as a fast-growing, with declining unemployment, population increase, and low-interest rates, but also there we can observe debt burden rising faster than household incomes (Statistics Sweden, 2019). These factors in combination are the reason for active discussions in social media and loud headlines for news articles. And even after introducing various measures by Swedish Finansinspektionen to handle the household debt rise risks, the tension in society and among investors remains.

Baltic news also raises the topic regarding vulnerabilities on the real estate market. But factors influencing such discussions differ from the Swedish. The solid economic growth of Baltic countries is followed by an increase of activity in the housing market: capitals, largest cities, and resorts are becoming hot locations for real estate objects. Estonian economy does not show any signals of overheating because of no structural economic imbalances. Lithuanian market is actively boosted due to investments rise. And in the case of Latvia, the residential market is growing due to an increase in the average salary (Ober Haus Report, 2018). But, nevertheless, there is a concern in the media of repeating the Baltic states housing bubble happened during 2005-2010 years. That is why the following question is coming: to what extent this news media tension about real estate in Sweden and Baltics can influence the Swedish banks' stock returns. For that, we enhance traditional econometric models with a new component - news sentiment. Such collaboration of text analysis and econometrics can be potentially beneficial for the financial world and bring attention to the importance of tracking the market mood.

In this research, we use the autoregressive moving average (ARMA) model for estimation of the conditional mean for stock returns. ARMA model allows identifying the best model for conditional mean equation and finding estimated residuals can be used further to estimate next GARCH specifications.

For sentiment data gathering, we use open source tools and libraries, such as Python library "Beautiful Soup" for news pages scraping, and VADER model (C. J. Hutto and Gilbert, 2014) for sentiment analysis.

This study contributes to the literature by an investigation of real estate news sentiment for a period from 04.01.2016 to 19.02.2019. We take into consideration news related to the real estate market of countries which are the home markets of considered Swedbank and SEB banks, and its effect on Swedish banks returns. Also, in our model, we take an asymmetrical effect into consideration. There do not seem to be any other papers related to an investigation of non-financial sentiments related to Swedish and Baltic markets, and also we did not find any studies related to the investigation of housing market sentiments effect on banks' stock returns.

Most studies consider company news and other generalized financial news columns and investors' sentiments as explanatory factors for stock price movements (Arik, 2011; Groß-Klußmann and Hautsch, 2011; Q. Li et al., 2014; Tumarkin and Whitelaw, 2001). Housing market sentiments are mostly considered as a factor for real estate prices prediction (Soo, 2018).

Moreover, to get sentiment data, we wrote an algorithm using open-source libraries and tools, instead of using the news analytics and data gathered by commercial providers, like it was done for instance by Sidorov, Revutskiy, Faizliev, Korobov, and Balash (2014), Verma and Soydemir (2009), or Yu (2014).

This paper shows the proofs of the positive and statistically significant relationship between the increase in positive real estate news sentiments and upward movements in banks' stock returns. Swedish news has a bigger impact on Swedbank's stock returns in comparison with other news providers when for SEB bank, Baltic news are more influential. When considering negative sentiments and their influence on stocks, we see that extremely negative news related to Baltic real estate market has a bigger influence than news about the Swedish market for both considered banks. When we have merged news, we see that the number of all negative news available does not have a considerable impact on returns, while the magnitude of negative news has. For positive news, it was proved to be working conversely. So, it means that only the fact of existing negative news can have a considerable impact and it does not matter how much such news was published.

The paper is structured in the following way: review and discussion of the literature about news sentiment impact on stock prices, linkage between banks and real estate market, and approaches to the stock prices modeling are provided in Chapter 2. The description of methods used for data collection, a procedure for news sentiment estimation and overview of econometric models used can be found in Chapter

3. Chapter 4 presents the data used for modeling and provides its main characteristics considered later in the modeling part. Chapter 5 shows the results and interpretations, which followed by Chapter 6 providing brief discussions with a proposal for further investigations. Conclusions of this work are shown in Chapter 7.

CHAPTER 2

Literature Review

News sentiment and its impact on stock prices

Nowadays, there are huge volumes of publicly accessible and quickly distributed news, and its amount is increasing. This makes news one of the main information providers to form opinions and support the decision-making. The news is one of the main sources which forms and reflects the market behavior at the same time. And that is why behavioral economists believe that understanding of the whole market lies in the understanding of the market players' behavior (Arik, 2011). News analytics is highly popular research topic due to its effective application in volatility, prices and trading volumes predictions (Sidorov et al., 2014). In finance, news sentiment is considered as an event and quantitative reflection of information. Simply saying, it measures the emotional tone of delivered news, and its possible values can be: positive, negative or neutral. News sentiments expressed numerically can be used as a component for mathematical and statistical models (Sidorov et al., 2014).

Different authors, such as Kothari and Shanken (1997), De Long, Shleifer, Summers, and Waldmann (1990) are analyzing the relationship between news sentiment and stock returns. Evidence that the influence of positive and negative sentiment on volatility differs was found, and this impact was proved to be substantial in case of negative sentiment (Engle and Ng, 1993; Tetlock, 2007). But if we talk about the usage of news sentiment as a factor, Tetlock (2007) was a first who proved its significance in the predictive model. Later on, Tetlock, Saar-Tsechansky, and Macskassy (2008) got better prediction results in comparison with forecasts prepared by analysts by applying the "Bag-Off-Words" model (Harris, 1954) for news sentiment analysis.

News sentiment topic in stock prices prediction is not new but still is an elusive concept (Yu, 2014). The classical theory about stock price formation is that asset price reflects all market available information (Fama, 1965). However, Fisher and Statman (2000) argued and proved that sentiment is a considerable part of asset price formation.

Before the active spread of the Internet in the World, there were quite many studies about the influence of macroeconomic news (Ederington and Lee, 1993) and also the impact of messages coming from the stock market (Mitchell and Mulherin, 1994). After an increase in the Internet connection coverage, information spread in the World Wide Web became to be actively used for explaining the stock price changes, for instance, by Antweiler and Frank (2004), Tetlock (2007), and Engelberg and Parsons (2011). Now, news sources are not only official news agencies or television, but it is also published companies' reports, publicly available statistics, Security Exchange Commission reports - all of them are so-called "pre-news" and are the first influencers on the public mood (Sidorov et al., 2014). Moreover, more power, in a sense of the number of people covered, is getting by social media, e.g. social network posts, blogs, tweets, etc. For example, anonymized Facebook data is commonly used for evaluation of people beliefs like it was done by Bailey, Dávila, Kuchler, and Stroebel (2017) to investigate homebuyers' beliefs regarding future price movements and how it can impact the decision regarding mortgage leverage.

Official news channels such as Thomson Reuters[1], Bloomberg[2], Dow Jones[3], and Wall Street Journal[4] provide reliable information which is used by investors for formation their opinion about future stock price movement. But blogs, forums, and social media form the general public opinion, and their popularity is growing (Yu,

[1] https://www.thomsonreuters.com
[2] https://www.bloomberg.com
[3] https://www.dowjones.com/
[4] https://www.wsj.com/

2014). In many cases, these resources create informational topics for mentioned 'reliable' sources. Impact and predictive power of blogging platform, like Twitter[5], was studied by Bollen, Mao, and Zeng (2011), Zhang, Fuehres, and Gloor (2011), Ranco, Aleksovski, Caldarelli, Grčar, and Mozetič (2015), and others. But we should take into account that sentiments taken from social media differ in some sense from the overall news sentiment, it can be described more precisely as public mood, but it is still a valuable factor for daily stock price movement prediction (Yu, 2014).

The impact of news sentiment proved via different types of econometric models. Arik (2011), using GARCH-in-mean models with a combination of 17 external variables in the mean equation, found a positive and significant relationship between changes in sentiment and S&P 500 excess returns. Earlier, Verma and Soydemir (2009) investigated stock market returns and investor sentiment relationship using a Value at Risk model. But, it was proved that GARCH models explain better the financial fat-tailed data with excess kurtosis. Furthermore, Sidorov et al. (2014) considered GARCH-Jumps model augmented with news intensity and proved that it has a better performance in comparison with traditional GARCH model with autoregressive conditional jump intensity described by Maheu and McCurdy (2004).

Real estate market and bank stock returns

Banks are highly exposed to the real estate market (Igan and Pinheiro, 2010; A. M. Martins, Serra, and Martins, 2016). It works via the next scenario: in case of housing market downtrend, banks have less capital what means that their expansion will be shortened, and the most significant changes for the population might be the credit reduction.

The role of the real estate market in the pricing of bank stocks was studied by

[5]https://twitter.com/

different authors. Real estate market risk and its influence on US banks stock were estimated by Carmichael and Coën (2018). Also, hight sensitivity of stock returns to changes in real estate returns was shown by He, Myer, and Webb (1996). Moreover, the vast majority of studies look into the effects of the crisis in real estate on stock prices dynamics. The main idea behind their study results is that financial institutions' stock price movements and the level of its real estate exposure are significantly connected (Ghosh, Guttery, and Sirmans, 1997; Igan and Pinheiro, 2010; A. M. Martins et al., 2016). But it is important to mention that the size of bank matters. As it was shown in papers of Mei and Lee (1994) and Mei and Saunders (1995), bigger sensitivity to changes in the housing market is applied mostly to small banks.

Numerous studies have been carried out on understanding the influence of financial news on stock returns, and, the main conclusions are about the need to include sentiment factor to prediction models (Kelly, 2016). The same conclusion we made in our research and used the news sentiment as an explanatory factor for banks' stock returns.

Stock prices modeling

Stock prices and returns are hardly predictable but, at the same time, their forecasts are very desirable. Investors together with researchers tend to develop plenty of algorithms for prediction of stock movements in order to make better decisions and optimize their portfolios. However, the irrational nature of the decision-making process makes price movements complicated with numbers of dependencies and a variety of possible outcomes. Mainly, there are two assumptions regarding stock prices predictions: 1) stock price behavior depends on its past values and can be modeled, 2) stock prices are not dependent from their historical values and are identically distributed random variables (Fama, 1965).

The autoregressive moving average (ARMA) models are linear models widely used to model the mean of processes, e.g. stock returns. And they are known as quite efficient for short-term predictions. Such a combination of AR and MA terms was proposed first by Whitle (1951). Thereafter, Box, Jenkins, and Reinsel (1970) introduced ARMA(p,q) modeling approach which is still the main strategy to select orders of AR and MA-polynomials.

Another class of models used for stock prices predictions is non-linear models. One of the most commonly used models is the generalized autoregressive conditional heteroskedasticity (GARCH) model and often used to estimate stock price volatilities. It is difficult to predict the volatility because it is not stable in time, and the only small part of it can be explained. One of the factors is new information which affects the stock price and makes it highly volatile. Heteroscedasticity describes the volatility changes over the time period. And instead of taking it as a problem, ARCH/GARCH models consider it as a variance to be modeled (Engle, Focardi, and Fabozzi, 2012). The commonly used in financial time series models are autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982), the generalized autoregressive conditional heteroskedasticity (GARCH) model presented by Bollerslev (1986) and numerous their variants.

The ARCH model allows the conditional variance to vary in time as a function of past errors but assuming the unconditional variance constant. The GARCH model allows more flexible lag structure. As Bollerslev explained in (1986), the main difference is that ARCH model's conditional variance has a view of the linear function of past sample variances only, when the GARCH process allows specifying lagged conditional variances as well. It is important to stress that these models are used not only for modeling the volatility of previous periods but also to be able to make a forecast of it for next time horizon. The strong evidence in favor of GARCH models with non-normal distributions usage is provided by W. Liu and Morley (2009). Also, results presented by Kosapattarapim, Lin, and McCrae (2012) showed that a GARCH model with non-normal error distributions suggests a better forecast than the GARCH model with normal error distribution. At the same time, the effectiveness of using GARCH (1,1) in data description and volatility measuring was proved empirically by Taylor (1994), Brook and Burke (2003), Olowe (2009), and others.

We use ARMA models for estimations. Applying the suitable ARMA(p,q) model helps us to solve the autocorrelation problem in the residuals and to see dependencies and effects between sentiments and stock returns. Thereafter, in follow up studies, received residuals can be used to estimate the conditional volatility by the best-fitted GARCH(p,q) model. Possible GARCH(p,q) models which can be used further are shown in Discussions Chapter.

CHAPTER 3

Methods and analysis

Data collection and aggregation

To collect and structure a text, we decided to use Python[1] as the main programming language due to availability of needed libraries and working web scrappers which allow getting the article's text from online sources automatically. All URL[2] of web pages with relevant news was provided in a separate file with .csv extension. In order to read the links of pages, we use the Python Data Analysis Library - "Pandas" (NumFOCUS, 2019), and its function "pandas.read_csv".

Every web page is compiled using the markup language - "Hypertext Markup Language" (HTML), and special tags help to identify the beginning and end of the text placed on the web page and also point other parts of the article, such as date of publishing or heading. Web scrapers navigate through provided URL of web pages, look for requested information using HTML tags and then download text by user's request. We use Python library named "Beautiful Soup"(Richardson, 2015) for parsing web pages. Its functions such as "soup.find", "soup.title" or "soup.get_text" are used to extract a text from the web page.

One of the difficulties, when we are dealing with web scrapping, is that usage of the same markup language for web pages does not require that the same tags for formatting are used. It creates some inconveniences in writing a universal algorithm which can recognize needed tags and download the text. For example, after the inspection of web pages' structure, we found out that for "The Local SE"[3] web pages,

[1]https://www.python.org/
[2]Uniform Resource Locator
[3]https://www.thelocal.se/

the body of the article is contained under the tag “div id =”article-body” when for “Reuters”[4], this tag is named as “div class =”StandardArticleBody_body“. That is why we have decided to create an algorithm which will require as an input only a couple of tags pointing out the part of web page contained needed text.

A created algorithm[5] is capable to get the main part of the article, its heading and publishing date from different web sites. The only manual step is to inspect a web page and identify tags which guide to needed part of the page. Moreover, this algorithm cleans a text from unneeded tags, for example, which are used to format the text, merges the heading, main part of the article to the one data object and save it together with the corresponding date when the article was published. This data will be used to construct the sentiment time series. Such structured and stored news together with dates is called a corpus of the text.

Text analysis and getting the sentiment time series data

Sentiment analysis is one of the areas in the field of Natural Language Processing (NLP) which aims to identify the sentiment of the human text, emotions, and attitude which is delivered by the author via text. For computers, reading and understanding the language is a very complex process, complicated algorithms with thousands of code lines behind.

There are several NLP libraries for Python such as spaCy (Explosion AI, 2019), NLTK (NLTK Project, 2019), TextBlob (Loria, 2018) which provide plenty of useful function to ease the text analysis for researchers and interested parties from the fields other than NLP.

Most of the sentiment analyzers are based on sentiment lexicon, list of words

[4] https://www.reuters.com
[5] Appendix A to the Chapter 3

labeled either positive, negative or neutral. There are several widely used lexicons, such as LIWC[6] (Linguistic Inquiry and Word Count) - mostly used for social media text, and it is able to estimate the intensity of words, Hu-Liu043 (Hu and Liu, 2004) - describes the sentiment of text in high level but has no ability to recognize the emoticons or acronyms, ANEW (Affective Norms for English Words) (Bradley and Lang, 1999) is more advanced because it provides emotional ratings for words in list, SenticNet[7] also contains estimated sentiment polarities using the range from -1 to 1. The last one is used for the VADER model (C. J. Hutto and Gilbert, 2014) which is applied for sentiment analysis in our research.

VADER (Valence Aware Dictionary and Sentiment Reasoner) is an open-source tool, a rule-based model used for general sentiment analysis. Lexicon and rules used by VADER are open and easily accessible which makes its use for research purposes advantageous (C. J. Hutto and Gilbert, 2014).

The main task for the evaluation of text sentiments is to calculate sentiment polarity. Before applying the sentiment analyzer, in most cases, data preprocessing is needed. In our case, it was done after web scraping. Afterward, the method "polarity_scores" is used to get polarity indices for text data.

Sentiment values for every word used in the text are calculated by the function "polarity_scores" in combination with the lexicon of the VADER model (C. Hutto, Klein, Pantone, and Berry, 2019). The result of primary calculations is that every word has precalculated values - polarity scores, they have a format: [x,y,z], where x, y, and z are negative, positive and neutral sentiments correspondingly. The compound value is calculated using the sum of all sentiments and normalization function which makes the value lies in the range [-1,1] where positive sentiment corresponds to compound score greater or equal to 0.05, neutral is located in a range (-0.05, 0.05),

[6] http://liwc.wpengine.com/
[7] https://sentic.net/

and negative sentiment corresponds to values less or equal to -0.05. This indicator is used further in econometric models. The function has the form:

$$C = \frac{S}{\sqrt{S^2 + 15}} \tag{3.1}$$

where C - compound sentiment score, S - a sum of negative (x), positive (y) and neutral (z) scores.

The main advantage of the VADER model is that it takes into account, for example, exclamation marks which show the intensity of expressed emotion, capitalization of used words, conjunctions which are signaling that there is a change in sentiment polarity, and even emojis, slang words and emoticons (e.g. ":D", ":)").

Afterward, the VADER model is included in the algorithm[8] implemented in Python. So, it means, that as soon as text data is extracted from the web page, it is used by VADER to evaluate the sentiment and generate time series used as an input for econometric models. VADER model and the calculated sentiments are saved to time series for further modeling. The algorithm can be visualized by the flowchart (Figure 3.1).

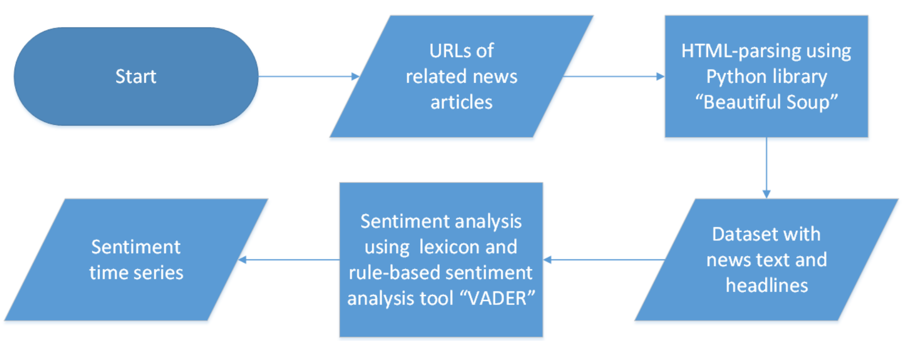


Figure 3.1: The algorithm used to extract news and estimate its sentiment

Notes: This flowchart represents an algorithm used to extract the news from web pages, calculate the sentiment of text and save as a time series together with a date when the article was published

[8]Snippet of the code can be found in Appendix A to the Chapter 3

Overview of models used

ARMA model has two components: autoregressive (AR) and moving average (MA) processes. We use the following common notation for all models introduced further:

- r_t - returns;
- p - AR order;
- q - MA order;
- k - number of exogenous variables.

AR part of the model describes predicted stock return as a value dependent on previous p periods values and random terms, r_t is expressed as its deviation from the mean value:

$$r_t = \mu + \sum_{i=1}^{p} \beta_i r_{t-i} + \varepsilon_t \tag{3.2}$$

The moving average component is a linear combination of random unexpected shocks impacted stock returns. MA(q) process can be described as:

$$r_t = \mu + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{3.3}$$

ARMA(p,q) model is able to describe time series which has characteristics of both AR(p) and MA(q) process via their combination in ARMA model (Gujarati, 2003).

Model 0: ARMA(p,q) model, our base-line model

$$r_t = \mu + \sum_{i=1}^{p} \beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{3.4}$$

The model should satisfy the stationarity and invertibility restrictions:

1. for stationarity, all z that solves $1-\beta_1 z-\beta_2 z^2-...-\beta_p z^p=0$ should lie outside the unit circle.
2. for invertability, all z that solves $1-\theta_1 z-\theta_2 z^2-...-\theta_q z^q=0$ should lie outside the unit circle.

Model 1: ARMA(p,q) model with news component

$$\begin{aligned}r_t = \mu + \sum_{i=1}^{p}\beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q}\theta_j\varepsilon_{t-j}+ \\ + \delta_1 Reuters_{t-1} + \delta_2 TheLocal_{t-1} + \delta_3 Others_{t-1} + \delta_4 Baltics_{t-1}\end{aligned} \tag{3.5}$$

News sentiments are defined in the range $[-1,1]$, where a positive news is higher ranked than negative news. Therefore any increase in the news variable indicates less bad news or more good news hence should increase the returns in the next period. Therefore, the restrictions for this model are:

1. $\delta_i \geq 0$ for all $i=1,...,4$.

For the next models, we separate negative and positive news in order to take into account an asymmetric effect of positive and negative news.

Model 2: ARMA(p,q) model with asymmetric effect of news, neutral news discarded

$$\begin{aligned}r_t = \mu + \sum_{i=1}^{p}\beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q}\theta_j\varepsilon_{t-j}+ \\ + \delta_1^- Reuters_{t-1}^- + \delta_2^- TheLocal_{t-1}^- + \delta_3^- Others_{t-1}^- + \delta_4^- Baltics_{t-1}^- + \\ + \delta_1^+ Reuters_{t-1}^+ + \delta_2^+ TheLocal_{t-1}^+ + \delta_3 Others_{t-1}^+ + \delta_4 Baltics_{t-1}^+\end{aligned} \tag{3.6}$$

where a news sentiment is considered to be positive (superscript "+") if its value is higher than 0.05, and negative (superscript "-") if its value is less than -0.05. Otherwise, the news sentiment is neutral. Neutral news is discarded because *TheLocal* and *Others* news sources do not have neutral news. The values 0.05 and -0.05 are given by the construction of the news sentiment scores[9]. It is trivial that positive news should increase returns in the next period. Also if the size of the negative news falls, then the $news^-$ increases, therefore this should increase returns in the next period. Therefore, the restrictions are:

1. $\delta_i, \delta_i^-, \delta_i^+ \geq 0$ for all $i = 1, ..., 4$.

Asymmetric effect of the news would appear if the impact of negative news is higher than that of positive news, *i.e.*if $\delta_i^- > \delta_i^+$ for any i. This is not imposed as a restriction.

Model 3: ARMA(p,q) model with sign and size effect of news

$$
\begin{aligned}
r_t = \mu + \sum_{i=1}^{p} \beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \\
+ \delta_1^{abs} |Reuters_{t-1}| + \delta_2^{abs} |TheLocal_{t-1}| + \delta_3^{abs} |Others_{t-1}| + \delta_4^{abs} |Baltics_{t-1}| + \\
+ \delta_1 Reuters_{t-1} + \delta_2 TheLocal_{t-1} + \delta_3 Others_{t-1} + \delta_4 Baltics_{t-1}
\end{aligned} \tag{3.8}
$$

where δ_i^{abs} for the size effect and δ_i is for the sign effect. The impact of an increase in the positive news is $\delta_i^{abs} + \delta_i$ and the impact of an increase in the negative news (that is, the magnitude of the news is decreasing towards 0) is $-\delta_i^{abs} + \delta_i$. Then the restrictions are:

1. $\delta_i \geq 0$ for all $i = 1, ..., 4$.
2. $\delta_i^{abs} + \delta_i \geq 0$ and $-\delta_i^{abs} + \delta_i \geq 0$, as the increase in the news variable would increase the returns.

[9] It is described in the subchapter "Text analysis and getting the sentiment time series data"

Then asymmetric effect of the news would appear if $\delta_i^{abs} + \delta_i > -\delta_i^{abs} + \delta_i$, $i.e.$if $\delta_i^{abs} < 0$ for any i. In this case, if $\delta_i \geq 0$ and $\delta_i^{abs} < 0$, the restriction $-\delta_i^{abs} + \delta_i \geq 0$ is redundant.

Model 4: ARMA(p,q) model with merged news data (positive and negative news distinguished)

$$\begin{aligned} r_t = \mu + \sum_{i=1}^{p} \beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \\ + \delta_1^{+} News_{t-1}^{pos} + \delta_2^{-} News_{t-1}^{neg} + \delta_3^{+} N_{t-1}^{pos} + \delta_4^{-} N_{t-1}^{neg} \end{aligned} \tag{3.9}$$

where $News_{t-1}$ variable is the sum of positive or negative news sentiments at time $t-1$. Given that in some dates there may be more than one news, we also consider the number of positive or negative news at time $t-1$ with the variable N_{t-1}.

The restrictions are:

1. $\delta_1^{+}, \delta_2^{-}, \delta_3^{+} \geq 0$.
2. $\delta_4 \leq 0$.

Then asymmetric effect of the news would appear if $\delta_2^{-} \geq \delta_1^{+}$.

CHAPTER 4

Data

We utilize two main data sets for this research: the first one contains all sentiments of news gathered for selected topics, countries and time when the news was published. The second data set is the daily adjusted closing stock prices of two banks, both operating in Sweden and Baltic countries: Swedbank and SEB banks, we used this data for the modeling part.

News sentiments data

The main difficulty for the data gathering is to choose the most suitable sources which are able to provide reliable and needed information on a certain topic. In our selection criteria, we intend to select Swedish and Baltic countries (Estonia, Latvia, Lithuania) news about the housing market and use this data for sentiment analysis.

For this purpose, two main sources were chosen to search for the news related to Swedish real estate market: "The Local SE"[1]- the portal which posts Swedish news in English, and "Reuters"[2] - the international news provider. Posts related to the real estate market in Sweden appear more frequently in the local news portal, and they are needed to capture possible interactions between daily stock price changes and news. Messages in international news portals about real estate appear rarer and only in case of high interest to it, or, in other words - when this news is highly important for the international audience. Such news can be a signal for the market about upcoming growth or downturn. Also, this type of news might influence the market stability in a bigger manner.

[1]https://www.thelocal.se
[2]https://www.reuters.com

To cover time periods for which main selected sources do not provide any articles, we refer to other sources and look for them using news aggregator - "Google News"[3]. We name them further as "Other Sweden". To these sources, we include such news providers as "Financial Times"[3], "Business Insider Nordic"[3], "Bloomberg"[3], "The Wall Street Journal"[3] and others[3]. It was highly desirable for further modeling to decrease the number of gaps in news sentiments dataset, or simply saying, to have as many days as possible with at least one news sentiment value.

Regarding Baltic countries news, main sources of information are "ERR News" - Estonian Public Broadcasting service in English[3], English-language monthly newspaper "The Baltic Times"[3], "Baltic News Network"[3], and others[3]. Further, we name them as "Other Baltics". To find them, we refer to mentioned previously news aggregator "Google News".

In order to find a content related to the needed topic, we have to use the most precise keywords for a search engine to make a direction for it what keywords the interesting for us news might contain. We use the following combinations to find articles about Swedish market: "Swedish real estate", "Real estate Sweden", "Swedish housing market", "Sweden housing", "Sweden property", "Construction Sweden", "Stockholm real estate", "Stockholm housing", "Stockholm flats", "Real estate bubble Sweden". For the Baltic market, we use "Baltics real estate", "Dwelling prices in Baltic countries", "Housing market Baltics" and "Baltics property prices".

The best way to see what data do we have - to plot it. At the plotted news sentiments[4] we can see values by countries: Sweden, Latvia, Lithuania, Estonia, and also news published about Baltic countries market. Also, two horizontal lines (y = 0.05 and y = -0.05) are added to show how many news sentiment are positive (above y = 0.5 line), negative (below y = -0.05 line), and how many are neutral (between these

[3]See Appendix C for a list of all sources used

[4]See Figure 4.1

two horizontal lines). More detailed explanation about sentiment scores is provided in Chapter 3.

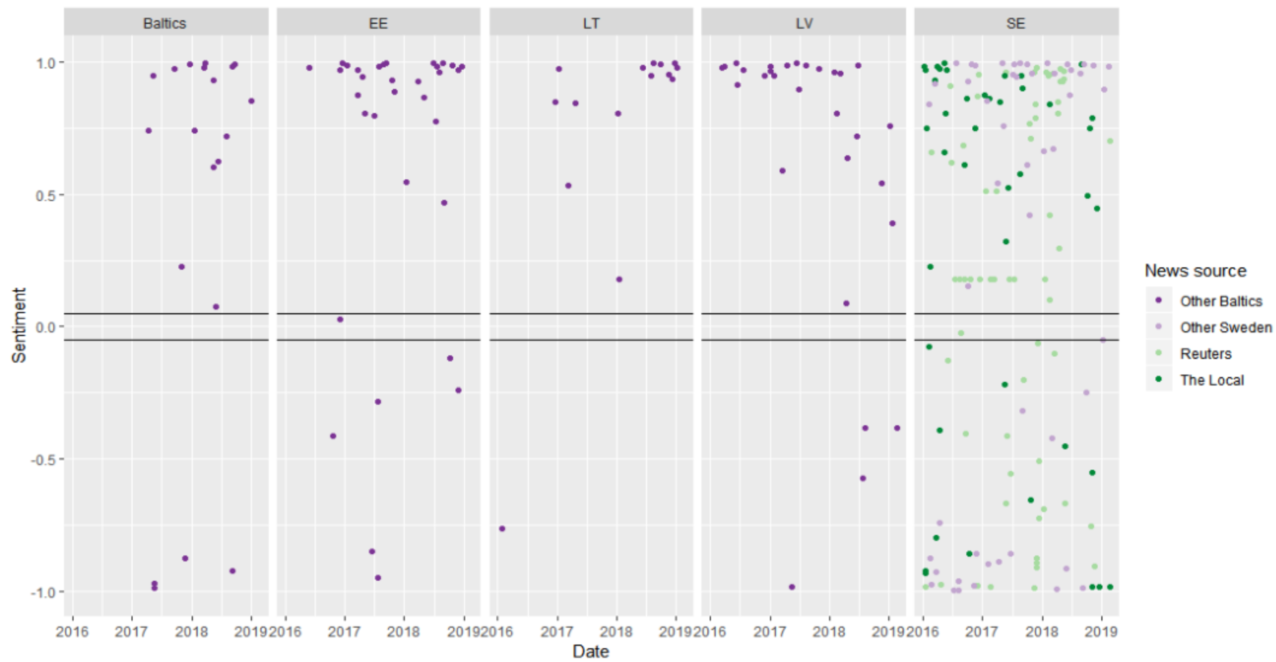


Figure 4.1: Real estate news sentiments by sources and countries

Notes: This figure plots real estate news sentiments by news sources and countries. News sources: *Other Baltics* - all sentiments of news about Baltic countries, *Other Sweden* - all sentiments of Swedish news except news from ”Reuters” and ”The Local SE” portals, *Reuters* - news about real estate in Sweden published by ”Reuters”, *The Local* - news about real estate in Sweden published by ”The Local SE” portal. Countries: *Baltics* - news about Baltic countries overall, *EE* - Estonia, *LV* - Latvia, *LT* - Lithuania. Period: 04.01.2016 - 19.02.2019.

More positive Swedish news is provided by other than “Reuters”[5] or “The Local SE” [6] news portals but the difference is not considerable, all positives news are distributed almost equally. More negative Swedish news is provided by “Reuters”, and the distribution is not already equal as we had for positive news.

For Baltics news, we do not have splitting by news portals due to difficulties to find such resources with sufficient amount of news about the real estate market in

[5]https://www.reuters.com
[6]https://www.thelocal.se/

these countries. But taking into account all the sentiments we have, we can see that they are mostly positive and concentrated near the maximum sentiment score. Negative sentiments are mostly related to real estate news about Baltic countries overall ("Baltics" in Figure 4.1), and also about Estonian and Latvian.

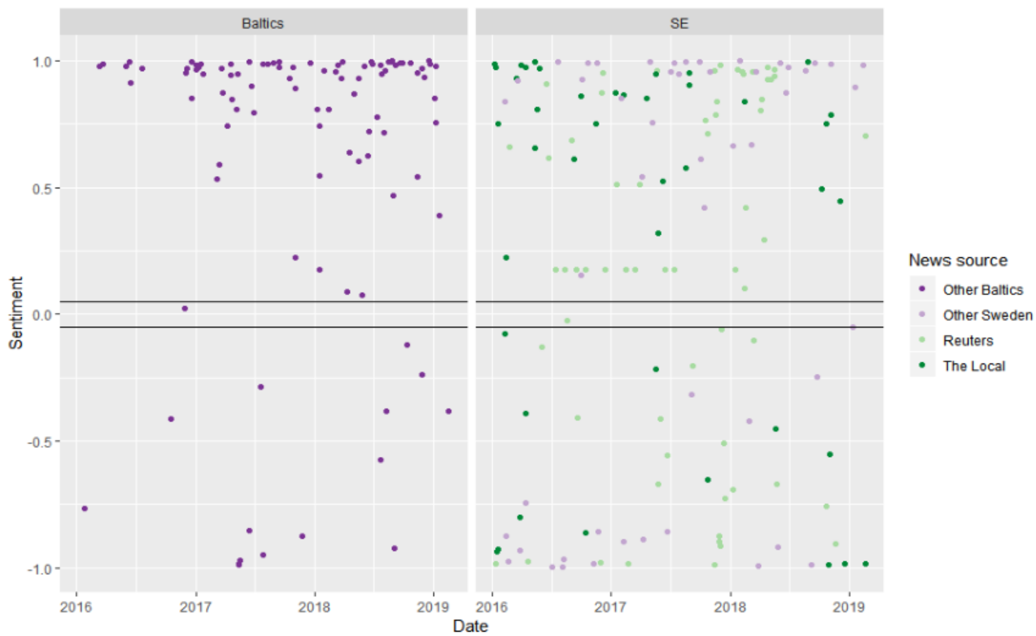


Figure 4.2: Real estate news sentiments by sources and regions

Notes: This figure plots real estate news sentiments by sources and countries. News sources: *Other Baltics* - all sentiments of news about Baltic countries, *Other Sweden* - all sentiments of Swedish news except news from "Reuters" and "The Local SE" portals, *Reuters* - news about real estate in Sweden published by "Reuters", *The Local* - news about real estate in Sweden published by "The Local SE" portal. Countries: Baltics (Estonia, Latvia, and Lithuania), SE (Sweden). Period: 04.01.2016 - 19.02.2019.

It is also interesting to see the difference in sentiments' distribution of Swedish news, and also for all Baltic countries news without splitting by countries[7]. Overall, we can see a considerable difference in the amount of negative news about the real estate market. Furthermore, for all markets, the neutral amount of news is extremely low, and it was a reason why neutral sentiments were discarded for the estimation

[7]See Figure 4.2

models, and we consider only negative and positive sentiments.

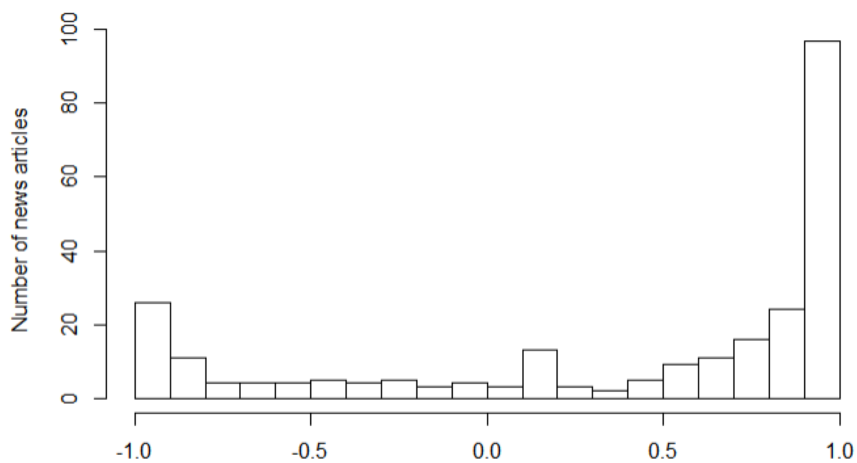


Figure 4.3: Histogram of news sentiments

Notes: This figure plots the distribution of real estate news sentiments for the period: 04.01.2016 - 19.02.2019.

From the histogram of news sentiments[8], we can see how sentiments are distributed overall. It is clear that positive sentiments prevail, most of the negative sentiments are in a range: $(-0.8, -1)$, most of the positives are in a range: $(0.9, 1)$.

To have a better overview of the data we scraped, we show its descriptive analysis - main statistical characteristics[9]. Most of the statistical values confirm the conclusions made during visual analysis.

Extreme values of sentiments are very close to maximum possible positive and maximum negative sentiments. The mean value for news sentiments is varying from 0.43 to 0.75 for Baltic states news. For Sweden, this value is much less indicating that the Baltics region is more optimistic about their real estate market. But we should take into account that standard deviation is big enough for all countries and

[8]See Figure 4.3
[9]See Table 4.1

represents that sentiment fluctuations are high, and it is not a stable value. If we will look at skewness of sentiment data for all states and apply the rule of thumb, we see that Baltic data is highly negatively skewed what means that most of the sentiments presented in data set for these countries are above the average value, so - mostly positive. Only for Swedish data, we can state that the data is fairly symmetrical due to the skewness close to zero (-0.53) what explains that data is equally distributed. The second value to pay our attention to is kurtosis needed to measure our outliers in the distribution of data. The leptokurtic distribution we can observe only for Lithuanian news. The value above 3 (4.09 for Lithuania) indicates that we have to deal with heavy-tailed data, we might have a huge number of outliers. Regarding other kurtosis values - all of them are platykurtic or simply saying, data is light-tailed, and outliers which also can be presented in data set are smaller than values of the normal distribution.

Table 4.1: Descriptive statistics of news sentments data

Variable	N	Mean	St.dev	Median	Min	Max	Skew	Kurtosis	St.er
Baltics									
sentiments	20	0.43	0.75	0.74	-0.99	1	-1.07	-0.58	0.17
Estonia									
sentiments	34	0.64	0.57	0.94	-0.95	1	-1.53	0.96	0.1
Lithuania									
sentiments	15	0.75	0.47	0.95	-0.76	1	-2.22	4.09	0.12
Latvia									
sentiments	29	0.64	0.55	0.95	-0.98	1	-1.63	1.43	0.1
Sweden									
sentiments	155	0.24	0.75	0.54	-1	1	-0.53	-1.35	0.06

Notes: This table shows descriptive statistics of news sentiments data by states for the period: 04.01.2016 - 19.02.2019. N - number of news sentiments available in data set, St.dev - standard deviation of sentiments, Min/Max - minimum and maximum sentiment score, Skew - skewness of the data, St.er - standard error.

When selecting the 10 most positive news[10], it can be seen that half of them

[10] See Table 4.2

were published during the 2018 year. And the amount of news described Baltics and Swedish real estate market is the same. But taking into account that data set of Swedish news sentiment is much bigger in comparison with Baltics, we can state that Baltic countries news is more positive in comparison to Swedish. Also, the same statement can be formulated by taking into account the mean values of sentiments, Swedish mean is the lowest among all listed countries[11].

Table 4.2: The most positive news sentiments

Date (Y-M-D)	sentiments	States	Region
2018-08-27	0.9990	Estonia	Baltics
2018-12-19	0.9990	Lithuania	Baltics
2018-02-01	0.9989	Sweden	SE
2016-05-11	0.9984	Sweden	SE
2016-07-20	0.9981	Sweden	SE
2018-08-28	0.9973	Estonia	Baltics
2017-04-28	0.9969	Sweden	SE
2018-03-27	0.9965	Baltics	Baltics
2016-06-10	0.9961	Latvia	Baltics
2017-08-17	0.9961	Sweden	SE

Notes: This table shows the 10 highest sentiments of news published during the period: 04.01.2016 - 19.02.2019. Also, it shows the name of the country (*States* column) and the region to what particular news refers

It was not a surprise that the most negative sentiments are observable mostly for Sweden for different years[12]. Furthermore, few Baltic news is also placed on top of the negatives, but among them, we see only news related to Latvian and overall Baltic market.

To have a visual representation of the most frequent words in the news, and to have an overall understanding of the main topics covered in positive and negative news, we have decided to use word clouds[13]. The most frequently used words have bigger font size and are placed closer to the middle part of the word cloud[14].

[11]See Table 4.1
[12]See Table 4.3
[13]See Figure 4.4
[14]Word clouds are not used for analysis and are illustrated only for visualization purposes

Table 4.3: The most negative news sentiments

Date (Y-M-D)	sentiments	States	Region
2016-07-04	-0.9973	Sweden	SE
2016-08-03	-0.9971	Sweden	SE
2018-03-28	-0.9919	Sweden	SE
2017-05-15	-0.9886	Baltics	Baltics
2018-09-07	-0.9881	Sweden	SE
2017-11-13	-0.9877	Sweden	SE
2018-11-01	-0.9869	Sweden	SE
2019-02-19	-0.9851	Sweden	SE
2018-12-19	-0.9849	Sweden	SE
2017-05-12	-0.9847	Latvia	Baltics

Notes: This table shows the 10 lowest sentiments of news published during the period: 04.01.2016 - 19.02.2019. Also, it shows the name of the country (*States* column) and the region to what particular news refers

(a) 10 most positive news **(b)** 10 most negative news

Figure 4.4: Word clouds

Notes: Each word cloud represents the frequency of words used in 10 most positive (a) and 10 most negative (b) news about real estate; the size of the word indicates its occurring frequency within selected articles. Word clouds are illustrated only for visualization purposes

Stocks prices data

For stock prices time series analysis, we use a daily series of the Swedbank and SEB banks adjusted closing stock prices from 04.01.2016 to 19.02.2019. The adjusted closing stock price is a price of last stock traded for a particular day which was adjusted by relevant split and dividend paid out to investors (Balasubramaniam, 2018). Historical stock prices data is downloaded directly from “Yahoo! Finance” (Yahoo, 2019), and the quality of data is assured because it is not stored locally but loaded directly from the web source to the R software[15].

To see how stocks' prices are varying in time, we can look at Figure 4.5. It presents the adjusted daily closing stock prices of Swedbank and SEB banks for the period: 04.01.2016 - 19.02.2019, and shows that there are some similarities in the behavior of banks' stock prices.

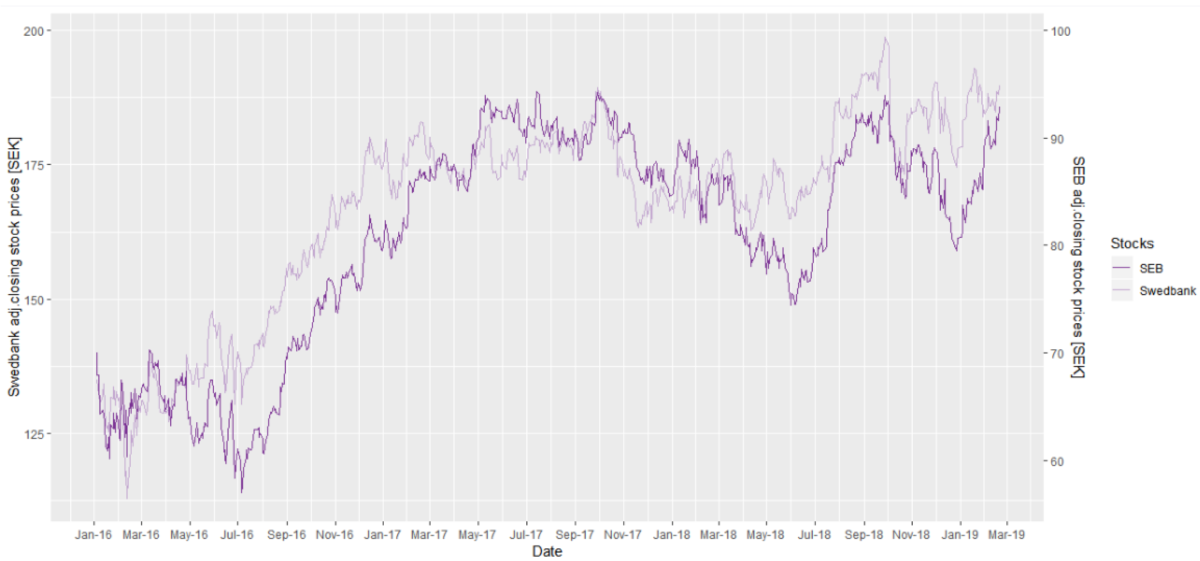


Figure 4.5: Adjusted daily closing stock prices
Notes: This figure plots adjusted daily closing stock prices of Swedbank and SEB banks measured in SEK, for the period: 04.01.2016 - 19.02.2019. Data is downloaded from "Yahoo! Finance"

Swedbank stock prices uptrend is observable after a considerable downturn at the beginning of 2016, albeit uncertainly. For SEB bank, the decline was observed mostly from March 2016. Similarly for both banks, starting from July till March 2017, prices were climbing up steadily, and they were remaining almost the same with moderate fluctuations till October 2017. Afterward, it was a decline in Swedbank stock prices to a level slightly above 160 SEK per share and it was remaining in

[15] https://www.r-project.org/

a range of 160-175 SEK till July 2018. For SEB bank, the situation was less pleasant: lowering was present from autumn 2017 till the beginning of June 2018. However, both banks' stocks had a considerable jump during the summer - mid of autumn 2018. Afterward, ups and downs were observable for both financial market leaders.

Table 4.4: Descriptive statistics of daily adjusted closing stock prices

Stocks	N	Mean	St.dev	Median	Min	Max	Skew	Kurtosis	St.er
Swedbank									
	793	167.3	18.31	172.87	112.87	198.83	-0.95	-0.12	0.65
SEB									
	793	80.8	10.10	84.64	57.02	94.34	-0.68	-0.68	0.36

Notes: This table shows descriptive statistics of daily adjusted closing stock prices of Swedbank and SEB banks for the period: 04.01.2016 - 19.02.2019. N - number of news sentiments available in data set, St.dev - standard deviation of sentiments, Min/Max - minimum and maximum sentiment score, Skew - skewness of the data, St.er - standard error.

Table 4.4 shows summary statistics of adjusted closing stock prices for a given period. We can see that both types of stocks data is moderately negatively skewed. Regarding "peakedness", distribution has thinner tails and fewer outliers in comparison with normally distributed data. It can tell us that data does not have rapid fluctuations and investments might be less risky. Also, we should also consider that variance for both stocks is very hight, stocks' movements are quite erratic.

But for modeling purposes, we are more interested in stock price relative changes than in its price. We calculate daily geometric returns of selected stocks to get returns series for selected stocks. Returns can be defined by

$$r_{[0,t]} = ln\left(\frac{P_t}{P_0}\right) \times 100\% \tag{4.1}$$

Figure 4.6 shows the daily stock price returns of Swedbank and SEB banks. We see that returns fluctuate near the mean value (close to zero), and fluctuations are different for time periods. Moreover, it seems that the returns series have conditional heteroskedasticity because volatility clustering is visible. And it makes us assume that returns volatility might be described by an autoregressive process. A statistical overview of log returns for both stocks is provided in Table 4.5.

Figure 4.6: Daily stock returns of (A) Swedbank, and (B) SEB banks

Notes: This figure shows the daily stock returns of (A) Swedbank, and (B) SEB banks for the period: 04.01.2016 - 19.02.2019, evaluated in % and calculated using the formula (4.1)

From skewness and kurtosis, we see that the distribution is not closed to the normal (skewness is -0.67 and -0.68 for Swedbank and SEB bank correspondingly, instead of 0 for normally distributed data, kurtosis is 6.42 and 9.49, for normal distribution it should be 3). Returns data is positively skewed: the right tail is larger than the left. And due to hight kurtosis, the distribution is fat-tailed.

In order to examine time-series and prepare for the modeling, we would check it for stationarity and make assumptions regarding possible ARMA (p,q) model's order. The first test for stationarity performed is the augmented Dickey-Fuller (ADF) test. The null hypothesis states that the series is not-stationary[16]. For Swedbank and SEB

[16] See Table 4.6

Table 4.5: Descriptive statistics of daily adjusted closing stock returns

Stocks	N	Mean	St.dev	Median	Min	Max	Skew	Kurtosis	St.er
Swedbank									
	792	0.04	1.24	0.01	-9.48	5.14	-0.67	6.42	0.04
SEB									
	792	0.04	1.37	0.00	-11.63	8.66	-0.68	9.49	0.05

Notes: This table shows descriptive statistics of daily adjusted closing stock returns of Swedbank and SEB banks for the period: 04.01.2016 - 19.02.2019. N - number of news sentiments available in data set, St.dev - standard deviation of sentiments, Min/Max - minimum and maximum sentiment score, Skew - skewness of the data, St.er - standard error.

Table 4.6: Augmented Dickey-Fuller Test for stationarity

	Swedbank	SEB
Dickey-Fuller	-10.664	-9.8816
Lag order	9	9
p-value	0.01	0.01
alternative hypothesis:	stationary	

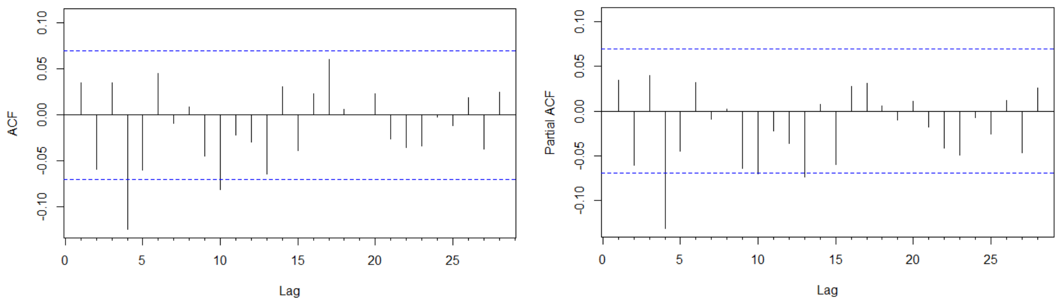
Notes: Augmented Dickey-Fuller test for stationarity of daily stock returns of Swedbank and SEB banks for the period: 04.01.2016 - 19.02.2019. Dickey-Fuller - Augmented Dickey-Fuller test statistic, lag - number of lagged terms, the alternative hypothesis is that time series is stationary. From the resulted p-value, we conclude that null-hypothesis about non-stationary time series can be rejected at a significance level of 0.05.

banks data, the null hypothesis about non-stationarity of the data can be rejected. Also, we refer to autocorrelation function (ACF) and partial autocorrelation function (PACF) plots which allow determining the order of MA(q) and AR(p) terms respectively [17].

From Figure 4.7a, we see that there are significant autocorrelations at lags 4 and 10. From the partial correlation plot (Figure 4.7b), we see an increase at lags 4, 10 and 13. It means that we should try models with AR and MA terms a least for the orders 4, 10, 13.

For the SEB bank data, we see significant autocorrelations at lags 1, 4, 10 and 16 (Figure 4.8a). From Figure 4.8b, we see an increase at lags 1, 4, 10 and 16. It

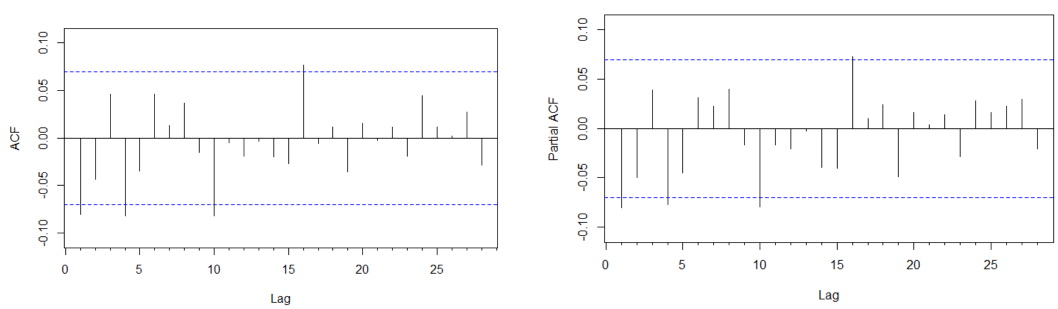
[17] See Figure 4.7 and Figure 4.8



(a) ACF plot **(b)** PACF plot

Figure 4.7: ACF and PACF plots of Swedbank's daily stock returns data

Notes: This figure shows (A) autocorrelation function (ACF) and (B) partial autocorrelation function (PACF) plots of Swedbank's daily stock returns data. From ACF, we can conclude that the returns are not highly correlated with its lagged values because most of the spikes are not statistically significant. PACF plot shows that there is no correlation between residuals and next lag values.



(a) ACF plot **(b)** PACF plot

Figure 4.8: ACF and PACF plots of SEB bank's daily stock returns data

Notes: This figure shows (A) autocorrelation function (ACF) and (B) partial autocorrelation function (PACF) plots of SEB bank's daily stock returns data. From ACF, we can conclude that the returns are not highly correlated with its lagged values because most of the spikes are not statistically significant. PACF plot shows that there is no considerable correlation between residuals and next lag values.

means that we should most probably refer to the models with AR and MA terms a least for the orders 1, 4, 10, 16.

CHAPTER 5

Results

The news vectors contain many zeros and relatively less non-zero values. The news is not autocorrelated and therefore cannot explain the autocorrelation structure in the return series. Therefore AR(4) model is chosen based on AIC criteria from the models which yielded serially uncorrelated residuals based on the Ljung-Box test. The other models are built on that. When estimating the following models, many starting values are considered for the optimization procedure.

Estimation results for Swedbank's stock returns.

Model 0: ARMA(p,q) model (baseline model)[1]

The model presents the effect of stock return values of past p periods on stock return for the current period t (via AR term). MA part sets a linear combination of error terms observed in the previous q periods.

The estimated model (equation 3.4) is:

$$\begin{array}{rclclclclcl} \hat{r}_t & = & \underset{\substack{(0.0234)\\ [0.0286]}}{0.0513} & + & \underset{\substack{(0.0063)\\ [0.0000]}}{0.0430} \; r_{t-1} & - & \underset{\substack{(0.0149)\\ [0.0000]}}{0.0694} \; r_{t-2} & + & \underset{\substack{(0.0268)\\ [0.0910]}}{0.0453} \; r_{t-3} & - & \underset{\substack{(0.0601)\\ [0.0278]}}{0.1319} \; r_{t-4} \end{array}$$

$$LL = -1280.31,\ AIC = 2570.63,\ BIC = 2593.99,\ SSR = 1175.91 \tag{5.1}$$

From p-values of the parameter estimates, we can conclude that AR components are significant at a significant level of 0.05 (only r_{t-3} is significant at 0.1 level) and stock returns are correlated with returns for the last 4 periods.

[1] In parentheses are the standard errors and in brackets are the p-values.

Model 1: ARMA(p,q) model with news[2]

The estimated model (equation 3.5) is:

$$
\begin{aligned}
\hat{r}_t &= \underset{\substack{(0.0069)\\ [0.0000]}}{0.0340} + \underset{\substack{(0.0107)\\ [0.0003]}}{0.0387}\ r_{t-1} - \underset{\substack{(0.0197)\\ [0.0005]}}{0.0684}\ r_{t-2} + \underset{\substack{(0.0240)\\ [0.0626]}}{0.0447}\ r_{t-3} - \underset{\substack{(0.0355)\\ [0.0002]}}{0.1328}\ r_{t-4} \\
&\quad + \underset{\substack{(0.1602)\\ [0.0392]}}{0.3303}\ Reuters_{t-1} + \underset{\substack{(0.0477)\\ [0.0001]}}{0.1922}\ TheLocal_{t-1} + \underset{\substack{(0.0000)\\ [0.1086]}}{0.0000}\ Others_{t-1} + \underset{\substack{(0.0563)\\ [0.0379]}}{0.1169}\ Baltics_{t-1} \\
LL &= -1277.29,\ AIC = 2572.58,\ BIC = 2614.65,\ SSR = 1170.26,\ Pval\_LLratio = 0.1957
\end{aligned}
\tag{5.2}
$$

The likelihood ratio test is performed compared to the base model $AR(4)$ and suggests that the base model gives better results.

For this model, all AR terms are significant at what refers to the presence of correlation between the stock return of period t and previous periods' returns.

All news components except $Others$ are significant at 5% significant level. From parameters values, we can conclude that an increase in these news sentiments is followed by an increase in stock returns. Moreover, we can notice that effect from news published by “Reuters” is considerably bigger in comparison with other news, so, it means that global news provider impacts stock returns of Swedbank more considerably. And, in general, the biggest part of stock returns is explained by sentiments of news about the Swedish real estate market.

Model 2: ARMA(p,q) model with asymmetric effect of news, neutral news discarded[2]

The estimated model (equation 3.6) is:

[2] In parentheses are the standard errors and in brackets are the p-values.

$$
\begin{aligned}
\hat{r}_t &= \underset{\substack{(0.0473)\\[0.3022]}}{0.0488} + \underset{\substack{(0.0180)\\[0.0227]}}{0.0410}\ r_{t-1} - \underset{\substack{(0.0294)\\[0.0199]}}{0.0685}\ r_{t-2} + \underset{\substack{(0.0135)\\[0.0006]}}{0.0460}\ r_{t-3} - \underset{\substack{(0.0938)\\[0.1610]}}{0.1315}\ r_{t-4} \\
&\quad + \underset{\substack{(0.2181)\\[0.0847]}}{0.3760}\ Reuters^{-}_{t-1} + \underset{\substack{(0.0407)\\[0.0046]}}{0.1154}\ TheLocal^{-}_{t-1} + \underset{\substack{(0.0000)\\[0.0620]}}{0.0000}\ Others^{-}_{t-1} + \underset{\substack{(0.2497)\\[0.0045]}}{0.7090}\ Baltics^{-}_{t-1} \\
&\quad + \underset{\substack{(0.1760)\\[0.0834]}}{0.3048}\ Reuters^{+}_{t-1} + \underset{\substack{(0.0906)\\[0.0231]}}{0.2058}\ TheLocal^{+}_{t-1} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ Others^{+}_{t-1} + \underset{\substack{(0.0330)\\[0.3210]}}{0.0327}\ Baltics^{+}_{t-1} \\
LL &= -1276.29,\ AIC = 2578.57,\ BIC = 2639.34,\ SSR = 1167.29,\ Pval\_LLratio = 0.4285
\end{aligned}
$$

All news sentiments are significant at 0.9 significant level except $Baltics^{+}$, and it is trivial that increase in sentiments is followed by an increase in stock returns. It is more important for us to look at the difference in estimated parameters for every source of news.

When separating the effects of negative and positive news, we can see one interesting result: Baltic negative news has the greatest effect on stock returns when positive is not significant at all. *Reuters* sentiments are the most influential only among positive sentiment scores. But the comparison of estimates for negatives shows that the difference between *Reuters* and *TheLocal* in case of negative sentiments is much bigger. It proofs that for Swedish news, negative real estate news published be global news provider impacts the stock returns in a bigger extent than a local provider. But still, Baltic negative news has a bigger impact on Swedbank's returns than all Swedish news. Estimates for *Other* news are consistent with the previous model because it is similarly non-significant.

Model 3: ARMA(p,q) model with sign and size effect of news[3]

[3] In parentheses are the standard errors and in brackets are the p-values.

The estimated model (equation 3.7) is:

$$\begin{aligned}
\hat{r}_t \;=\;& \underset{\substack{(0.0096)\\[0.0003]}}{0.0343} + \underset{\substack{(0.0123)\\[0.0015]}}{0.0388}\; r_{t-1} - \underset{\substack{(0.0438)\\[0.1183]}}{0.0684}\; r_{t-2} + \underset{\substack{(0.0177)\\[0.0114]}}{0.0447}\; r_{t-3} - \underset{\substack{(0.0527)\\[0.0118]}}{0.1327}\; r_{t-4} \\
& - \underset{\substack{(0.0077)\\[0.0000]}}{0.0585}\; |Reuters_{t-1}| - \underset{\substack{(0.0066)\\[0.0002]}}{0.0244}\; |TheLocal_{t-1}| - \underset{\substack{(0.0000)\\[0.1256]}}{0.0000}\; |Others_{t-1}| - \underset{\substack{(0.0177)\\[0.0074]}}{0.0473}\; |Baltics_{t-1}| \\
& + \underset{\substack{(0.0377)\\[0.0001]}}{0.3328}\; Reuters_{t-1} + \underset{\substack{(0.0513)\\[0.0002]}}{0.1932}\; TheLocal_{t-1} + \underset{\substack{(0.0000)\\[0.0686]}}{0.0000}\; Others_{t-1} + \underset{\substack{(0.0534)\\[0.0093]}}{0.1387}\; Baltics_{t-1} \\
LL \;=\;& -1277.07,\; AIC = 2580.14,\; BIC = 2640.91,\; SSR = 1169.61,\; Pval_LLratio = 0.5926
\end{aligned}$$

We can observe a significant size effect (0.05 significant level) of all news published except Swedish news published by $Others$, same for the sign effect ($Others$ is significant only at 0.1 level). But as we noticed, there is no effect of $Others$ still.

$\delta_i^{abs} < 0$ for all i what means that we have an asymmetric effect of the news. And Swedish news shows the greatest impact on stock returns.

Model 4: ARMA(p,q) model with merged news data (positive and negative news distinguished)[4]

Model's (equation 3.8) estimates are:

$$\begin{aligned}
\hat{r}_t \;=\;& \underset{\substack{(0.0099)\\[0.0006]}}{0.0339} + \underset{\substack{(0.0138)\\[0.0029]}}{0.0411}\; r_{t-1} - \underset{\substack{(0.0373)\\[0.0618]}}{0.0697}\; r_{t-2} + \underset{\substack{(0.0296)\\[0.1142]}}{0.0467}\; r_{t-3} - \underset{\substack{(0.0490)\\[0.0065]}}{0.1332}\; r_{t-4} \\
& + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\; News_{t-1}^{pos} + \underset{\substack{(0.0650)\\[0.0791]}}{0.1141}\; News_{t-1}^{neg} + \underset{\substack{(0.0355)\\[0.0041]}}{0.1018}\; N_{t-1}^{pos} - \underset{\substack{(0.0000)\\[0.2193]}}{0.0000}\; N_{t-1}^{neg} \\
LL \;=\;& -1278.27,\; AIC = 2574.54,\; BIC = 2616.61,\; SSR = 1171.16,\; Pval_LLratio = 0.3944
\end{aligned}$$

[4] In parentheses are the standard errors and in brackets are the p-values.

This model includes only aggregated news without separation by news sources. From the estimates and their p-values, we can see the difference in results for positive and negative news. The magnitude of negative news has an effect on stock returns while the number of news does not. At the same time, for positive news it is vice versa: the number of positive news affects the returns while the magnitude has no influence on stock prices movements.

All considered models' estimates are summarized in Table 5.1.

Estimation results for SEB bank's stock returns.

Model 0: ARMA(p,q) model (baseline model)[5]

The model (equation 3.4) presents the effect of stock return values of past p periods on stock return for the current period t (via AR term). MA part sets a linear combination of error terms observed in the previous q periods.

The model's estimates are:

$$\begin{array}{cccccccccccc} \hat{r}_t & = & \underset{\substack{(0.0247)\\[0.0283]}}{0.0541} & - & \underset{\substack{(0.0264)\\[0.0021]}}{0.0811} & r_{t-1} - & \underset{\substack{(0.0180)\\[0.0004]}}{0.0636} & r_{t-2} + & \underset{\substack{(0.0282)\\[0.2692]}}{0.0312} & r_{t-3} - & \underset{\substack{(0.0185)\\[0.0000]}}{0.0792} & r_{t-4} \end{array}$$
$$LL = -1355.63,\ AIC = 2721.26,\ BIC = 2744.63,\ SSR = 1426.61 \quad (5.3)$$

AR(4) model based on AIC from the models which yielded serially uncorrelated residuals based on the Ljung-Box test.

From p-values of the parameter estimates, we can conclude that AR components are significant at a significant level of 0.05 (only r_{t-3} is not significant) and stock returns are correlated with returns at periods 1, 2 and 4.

[5] In parentheses are the standard errors and in brackets are the p-values.

Table 5.1: Estimation results for Swedbank's stock returns.

	Model 0	Model 1	Model 2	Model 3	Model 4
Estimates of AR(4) model					
μ	0.513	0.0340	0.0488	0.0343	0.0339
	[0.0286]	[0.0000]	[0.3022]	[0.0003]	[0.0006]
r_{t-1}	0.0430	0.0387	0.0410	0.0388	0.0411
	[0.0000]	[0.0003]	[0.0227]	[0.0015]	[0.0029]
r_{t-2}	-0.0694	-0.0684	-0.0685	-0.0684	-0.0697
	[0.0000]	[0.0005]	[0.0199]	[0.1183]	[0.0618]
r_{t-3}	0.0453	0.0447	0.0460	0.0447	0.0467
	[0.0910]	[0.0626]	[0.0006]	[0.0114]	[0.1142]
r_{t-4}	-0.1319	-0.1328	-0.1315	-0.1327	-0.1332
	[0.0278]	[0.0002]	[0.1610]	[0.0118]	[0.0065]
Estimates of news components					
$Reuters_{t-1}$		0.3303		0.3328	
		[0.0392]		[0.0001]	
$TheLocal_{t-1}$		0.1922		0.1932	
		[0.0001]		[0.0002]	
$Others_{t-1}$		0.0000		0.0000	
		[0.1086]		[0.0686]	
$Baltics_{t-1}$		0.1169		0.1387	
		[0.0379]		[0.0093]	
$Reuters^{-}_{t-1}$			0.3760		
			[0.0847]		
$TheLocal^{-}_{t-1}$			0.1154		
			[0.0046]		
$Others^{-}_{t-1}$			0.0000		
			[0.0620]		
$Baltics^{-}_{t-1}$			0.7090		
			[0.0045]		
$Reuters^{+}_{t-1}$			0.3048		
			[0.0834]		
$TheLocal^{+}_{t-1}$			0.2058		
			[0.0231]		
$Others^{+}_{t-1}$			0.0000		
			[0.0000]		
$Baltics^{+}_{t-1}$			0.0327		
			[0.3210]		
$\|Reuters_{t-1}\|$				-0.0585	
				[0.0000]	
$\|TheLocal_{t-1}\|$				-0.0244	
				[0.0002]	
$\|Others_{t-1}\|$				-0.0000	
				[0.1256]	
$\|Baltics_{t-1}\|$				-0.0473	
				[0.0074]	
$News^{pos}_{t-1}$					0.0000
					[0.0000]
$News^{neg}_{t-1}$					0.1141
					[0.0791]
N^{pos}_{t-1}					0.1018
					[0.0041]
N^{neg}_{t-1}					-0.0000
					[0.2193]

Notes: Model 0: ARMA(p,q); Model 1: ARMA(p,q) model with news; Model 2: ARMA(p,q) model with asymmetric effect of news, neutral news discarded; Model 3: ARMA(p,q) model with sign and size effect of news; Model 4: ARMA(p,q) model with merged news data (positive and negative news distinguished). In brackets are the p-values.

Model 1: ARMA(p,q) model with news[6]

The estimated model (equation 3.5) is:

$$
\begin{aligned}
\hat{r}_t &= \underset{\substack{(0.0274)\\ [0.2378]}}{0.0323} - \underset{\substack{(0.0109)\\ [0.0000]}}{0.0821}\; r_{t-1} - \underset{\substack{(0.0185)\\ [0.0009]}}{0.0615}\; r_{t-2} + \underset{\substack{(0.0116)\\ [0.0048]}}{0.0327}\; r_{t-3} - \underset{\substack{(0.0337)\\ [0.0147]}}{0.0821}\; r_{t-4} \\
&\quad + \underset{\substack{(0.0789)\\ [0.0485]}}{0.1556}\; Reuters_{t-1} + \underset{\substack{(0.0004)\\ [0.0090]}}{0.0009}\; TheLocal_{t-1} + \underset{\substack{(0.0000)\\ [0.2288]}}{0.0000}\; Others_{t-1} + \underset{\substack{(0.0976)\\ [0.0090]}}{0.2550}\; Baltics_{t-1} \\
LL &= -1354.19,\ AIC = 2726.39,\ BIC = 2768.46,\ SSR = 1421.44,\ Pval\_LLratio = 0.5792
\end{aligned}
\tag{5.4}
$$

All news components except *Others* are significant at 5% significant level. From parameters values, we can conclude that an increase in these news sentiments is followed by an increase in stock returns. Moreover, we can notice that effect from Baltic news is considerably bigger in comparison with news about the Swedish real estate market.

Model 2: ARMA(p,q) model with asymmetric effect of news, neutral news discarded[7]

The estimated model (equation 3.6) is:

[6] In parentheses are the standard errors and in brackets are the p-values.

$$
\begin{aligned}
\hat{r}_t &= \underset{\substack{(0.0058)\\[0.0000]}}{0.0545} - \underset{\substack{(0.0175)\\[0.0000]}}{0.0792}\ r_{t-1} - \underset{\substack{(0.0135)\\[0.0000]}}{0.0645}\ r_{t-2} + \underset{\substack{(0.0045)\\[0.0000]}}{0.0322}\ r_{t-3} - \underset{\substack{(0.0131)\\[0.0000]}}{0.0823}\ r_{t-4} \\
&\quad + \underset{\substack{(0.1324)\\[0.0730]}}{0.2374}\ Reuters^{-}_{t-1} + \underset{\substack{(0.1039)\\[0.0573]}}{0.1975}\ TheLocal^{-}_{t-1} + \underset{\substack{(0.1062)\\[0.1487]}}{0.1534}\ Others^{-}_{t-1} + \underset{\substack{(0.1215)\\[0.0000]}}{0.8257}\ Baltics^{-}_{t-1} \\
&\quad + \underset{\substack{(0.0341)\\[0.0026]}}{0.1027}\ Reuters^{+}_{t-1} + \underset{\substack{(0.0000)\\[0.0638]}}{0.0000}\ TheLocal^{+}_{t-1} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ Others^{+}_{t-1} + \underset{\substack{(0.0358)\\[0.0000]}}{0.1722}\ Baltics^{+}_{t-1} \\
LL &= -1353.27,\ AIC = 2732.54,\ BIC = 2793.31,\ SSR = 1418.12,\ Pval\_LLratio = 0.7867
\end{aligned}
$$

All news sentiments are significant at 0.9 significant level except $Others^{-}_{t-1}$.

When separating the effects of negative and positive news, we can see that Baltic negative and positive news has the greatest effect on stock returns. However, when looking at positive sentiments, they do not have an extremely high effect on stock returns, only Baltic news and Swedish news from Reuther affects the SEB stock prices movements.

Model 3: ARMA(p,q) model with sign and size effect of news[7]

The estimated model (equation 3.7) is:

[7] In parentheses are the standard errors and in brackets are the p-values.

$$\begin{array}{rcllllllllll}
\hat{r}_t & = & \underset{\substack{(0.0142)\\[0.0003]}}{0.0520} & - & \underset{\substack{(0.0244)\\[0.0010]}}{0.0803} & r_{t-1} - & \underset{\substack{(0.0364)\\[0.0858]}}{0.0626} & r_{t-2} + & \underset{\substack{(0.0054)\\[0.0000]}}{0.0316} & r_{t-3} - & \underset{\substack{(0.0233)\\[0.0004]}}{0.0829} & r_{t-4} \\
& & - \underset{\substack{(0.0186)\\[0.0001]}}{0.0745} & |Reuters_{t-1}| - & \underset{\substack{(0.0125)\\[0.0000]}}{0.1336} & |TheLocal_{t-1}| - & \underset{\substack{(0.0000)\\[0.1208]}}{0.0000} & |Others_{t-1}| - & \underset{\substack{(0.0531)\\[0.0000]}}{0.3196} & |Baltics_{t-1}| & & \\
& & + \underset{\substack{(0.0393)\\[0.0000]}}{0.1887} & Reuters_{t-1} + & \underset{\substack{(0.0358)\\[0.0002]}}{0.1338} & TheLocal_{t-1} + & \underset{\substack{(0.0000)\\[0.3734]}}{0.0000} & Others_{t-1} + & \underset{\substack{(0.1021)\\[0.0000]}}{0.4953} & Baltics_{t-1} & & \\
LL & = & \multicolumn{10}{l}{-1353.36,\ AIC = 2732.72,\ BIC = 2793.49,\ SSR = 1418.44,\ Pval_LLratio = 0.8053}
\end{array}$$

We can observe a significant size effect (0.05 significant level) of all news published except Swedish news published by $Others$, same for the sign effect. $\delta_i^{abs} < 0$ for all i what means that we have an asymmetric effect of the news. We also can observe a consistent result with Model's 2 results: the effect of negative news from news published by "The Local SE" is not zero, but the effect of positive news is around zero.

2.

Model 4: ARMA(p,q) model with merged news data (positive and negative news distinguished)[8]

Model's (equation 3.8) estimates are:

[8] In parentheses are the standard errors and in brackets are the p-values.

$$
\begin{aligned}
\hat{r}_t \quad = \quad & \underset{\substack{(0.0184)\\ [0.0020]}}{0.0571} - \underset{\substack{(0.0090)\\ [0.0000]}}{0.0811}\ r_{t-1} - \underset{\substack{(0.0088)\\ [0.0000]}}{0.0689}\ r_{t-2} + \underset{\substack{(0.0101)\\ [0.0018]}}{0.0314}\ r_{t-3} - \underset{\substack{(0.0192)\\ [0.0000]}}{0.0823}\ r_{t-4} \\
& + \underset{\substack{(0.0000)\\ [0.0012]}}{0.0000}\ News^{pos}_{t-1} + \underset{\substack{(0.1582)\\ [0.0472]}}{0.3139}\ News^{neg}_{t-1} + \underset{\substack{(0.0435)\\ [0.1121]}}{0.0691}\ N^{pos}_{t-1} - \underset{\substack{(0.0000)\\ [0.0001]}}{0.0000}\ N^{neg}_{t-1} \\
LL \quad = \quad & -1354.29,\ AIC = 2726.58,\ BIC = 2768.65,\ SSR = 1421.79,\ Pval\_LLratio = 0.6134
\end{aligned}
$$

This model includes only aggregated news without separation by news sources. We can see interesting results similar to the same model using Swedbank's data. From the estimates and their p-values, we can see the difference in results for positive and negative news. The magnitude of negative news has an effect on stock returns while the number of news does not. At the same time, for positive news it is vice versa: a number of positive news affects the returns (not significantly though) while the magnitude does not influence on stock prices movements.

All considered models' estimates are summarized in Table 5.2.

Summary

From estimations results (Table 5.1 and Table 5.2), we can conclude that news component must be included in econometric models and can be used to explain the mean value of bank stock returns. We have estimated 5 different ARMA (p,q) models for Swedbank and SEB banks' stock returns including the news variables. The main similarity for both banks is that the magnitude of negative news and amount of positive real estate news influence the stock returns when for the magnitude of positive and amount of negative do not have a significant influence. It means that just several extremely negative news might have the same effect as numerous published positive

Table 5.2: Estimation results for SEB's stock returns.

	Model 0	Model 1	Model 2	Model 3	Model 4
Estimates of AR(4) model					
μ	0.0541	0.0323	0.0545	0.0520	0.0571
	[0.0283]	[0.2378]	[0.0000]	[0.0003]	[0.0020]
r_{t-1}	-0.0811	-0.0821	-0.0792	-0.0803	-0.0811
	[0.0021]	[0.0000]	[0.0000]	[0.0010]	[0.0000]
r_{t-2}	-0.0636	-0.0615	-0.0645	-0.0626	-0.0689
	[0.0004]	[0.0009]	[0.0000]	[0.0858]	[0.0000]
r_{t-3}	0.0312	0.0327	0.0322	0.0316	0.0314
	[0.2692]	[0.0048]	[0.0000]	[0.0000]	[0.0018]
r_{t-4}	-0.0792	-0.0821	-0.0823	-0.0829	-0.0823
	[0.0000]	[0.0147]	[0.0000]	[0.0004]	[0.0000]
Estimates of news components					
$Reuters_{t-1}$		0.1556		0.1887	
		[0.0485]		[0.0000]	
$TheLocal_{t-1}$		0.0009		0.1338	
		[0.0090]		[0.0002]	
$Others_{t-1}$		0.0000		0.0000	
		[0.2288]		[0.3734]	
$Baltics_{t-1}$		0.2550		0.4953	
		[0.0090]		[0.0000]	
$Reuters^{-}_{t-1}$			0.2374		
			[0.0730]		
$TheLocal^{-}_{t-1}$			0.1975		
			[0.0573]		
$Others^{-}_{t-1}$			0.1534		
			[0.1487]		
$Baltics^{-}_{t-1}$			0.8257		
			[0.0000]		
$Reuters^{+}_{t-1}$			0.1027		
			[0.0026]		
$TheLocal^{+}_{t-1}$			0.0000		
			[0.0638]		
$Others^{+}_{t-1}$			0.0000		
			[0.0000]		
$Baltics^{+}_{t-1}$			0.1722		
			[0.0000]		
$\|Reuters_{t-1}\|$				-0.0745	
				[0.0001]	
$\|TheLocal_{t-1}\|$				-0.1336	
				[0.0000]	
$\|Others_{t-1}\|$				-0.0000	
				[0.1208]	
$\|Baltics_{t-1}\|$				-0.3196	
				[0.0000]	
$News^{pos}_{t-1}$					0.0000
					[0.0012]
$News^{neg}_{t-1}$					0.3139
					[0.0472]
N^{pos}_{t-1}					0.0691
					[0.1121]
N^{neg}_{t-1}					-0.0000
					[0.0001]

Notes: Model 0: ARMA(p,q); Model 1: ARMA(p,q) model with news; Model 2: ARMA(p,q) model with asymmetric effect of news, neutral news discarded; Model 3: ARMA(p,q) model with sign and size effect of news; Model 4: ARMA(p,q) model with merged news data (positive and negative news distinguished). In brackets are the p-values.

articles.

Among negative sentiments, news about Baltic Real estate has the biggest impact on both banks returns. However, when not separating the negative and positive news, for SEB bank it has still the greatest effect when for Swedbank, the Swedish news published by “Reuters” and “The Local SE” is becoming more significant.

Swedish news published by other than “Reuters” and “The Local SE” is not significant at all meaning that Swedish news published only either by global news provider or by local have the impact on banks stocks. Furthermore, there is an asymmetric effect of news what is shown with estimations of a Model 2 (equation 3.6).

CHAPTER 6

Discussion

The leverage effect cannot be used to explain asymmetries in stock prices (Ghysels, Harvey, and Renault, 1996). That is why, in this research, we considered an asymmetry in our models. For now, we refer only to one type of asymmetry which appears with negative and positive news. But we are aiming to analyze two types of asymmetry coming from news and volatility how it was done by Asai and McAleer (2006), and it can be estimated further using the GARCH models described below (Models 1 - 4).

At the point when we have the best ARMA(p,q) model for the conditional mean equation, we receive the estimated residuals from this model, $\hat{\varepsilon}_t$. We can use this vector of estimated residuals to estimate the GARCH models. For simplicity, we can assume GARCH(1,1) order and keep the focus on the extensions of GARCH models. The estimation of conditional mean and variance structures can be done separately. The resulting estimator is a quasi-maximum likelihood estimator and it is consistent, asymptotically normal but not efficient (Bollerslev and Wooldridge, 1992; Carnero and Eratalay, 2014).

We propose the following models to use for further investigation, and the notation is as follows:

$h_t = Var(r_t|I_{t-1}) = Var(\varepsilon_t|I_{t-1}) = E(\varepsilon_t^2|I_{t-1})$, where I_{t-1} is the information available up to $t-1$ and h_t is referred to as conditional variance or volatility.

Given that news can take values between -1 and 1, ensuring the positivity of the volatilities is not as trivial as setting the coefficients to be positive.

Model 0: GARCH(1,1) model (baseline model)

$$h_t = \gamma + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1}$$

where the volatilities are positive if $\gamma > 0$; $\alpha_1, \alpha_2 \geq 0$ and stationary if $\alpha_1 + \alpha_2 < 1$.

Model 1: GARCH(1,1) model with news

$$\begin{aligned} h_t &= \gamma + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} \\ &\quad + \delta_1 Reuters_{t-1} + \delta_2 TheLocal_{t-1} + \delta_3 Others_{t-1} + \delta_4 Baltics_{t-1} \end{aligned}$$

where the volatilities are positive if $\gamma > 0$; $\alpha_1, \alpha_2 \geq 0$, $\gamma - \sum_{i=1}^4 \delta_i > 0$, and stationary if $\alpha_1 + \alpha_2 < 1$.

Model 2: GARCH(1,1) model with asymmetric news, neutral news discarded

$$\begin{aligned} h_t &= \gamma + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} \\ &+ \delta_1^- Reuters_{t-1}^- + \delta_2^- TheLocal_{t-1}^- + \delta_3^- Others_{t-1}^- + \delta_4^- Baltics_{t-1}^- \\ &+ \delta_1^+ Reuters_{t-1}^+ + \delta_2^+ TheLocal_{t-1}^+ + \delta_3 Others_{t-1}^+ + \delta_4 Baltics_{t-1}^+ \end{aligned}$$

where the volatilities are positive if $\gamma > 0$; $\alpha_1, \alpha_2 \geq 0$; $\gamma + \sum_{i=1}^4 \delta_i^- + \sum_{i=1}^4 \delta_i^+ > 0$ and stationary if $\alpha_1 + \alpha_2 < 1$. Asymmetric effect would occur if $\delta_i^- > \delta_i^+$ for any $i = 1, ..., 4$.

Model 3: GARCH(1,1) model with sign and size effect of news

$$
\begin{aligned}
h_t &= \gamma + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} \\
&+ \delta_1^{abs} |Reuters_{t-1}| + \delta_2^{abs} |TheLocal_{t-1}| + \delta_3^{abs} |Others_{t-1}| + \delta_4^{abs} |Baltics_{t-1}| \\
&+ \delta_1 Reuters_{t-1} + \delta_2 TheLocal_{t-1} + \delta_3 Others_{t-1} + \delta_4 Baltics_{t-1}
\end{aligned}
$$

where the volatilities are positive if $\gamma > 0$; $\alpha_1, \alpha_2 \geq 0$; $\gamma + \sum_{i=1}^{4} |\delta_i^{abs}| + \sum_{i=1}^{4} \delta_i > 0$ and stationary if $\alpha_1 + \alpha_2 < 1$. A positive news of value 1 would increase volatility by $\delta_i^{abs} + \delta_i$ while a negative news of value 1 would increase volatility by $\delta_i^{abs} - \delta_i$. Asymmetric effect would occur if $\delta_i^{abs} + \delta_i < \delta_i^{abs} - \delta_i$ which implies $\delta_i < 0$. It is expected that $\delta_i^{abs} > 0$ as it is the coefficient for the size effect.

Model 4: GJR-GARCH(1,1) model

$$
h_t = \gamma + (\alpha_1 + \theta_1 I_{t-1}) \varepsilon_{t-1}^2 + \alpha_2 h_{t-1}
$$

where the dummy I_{t-1} takes value 1 if $\varepsilon_{t-1} < 0$. The volatilities are positive if $\gamma > 0$; $\alpha_1, \alpha_2 \geq 0$, and stationary if $\alpha_1 + \theta_1 + \alpha_2 < 1$. Asymmetric effect would occur if $\theta_1 > 0$, because negative errors increase volatility more, than positive errors.

Model 5: EGARCH(1,1) model

$$
\begin{aligned}
\log(h_t) &= \gamma + \alpha_1 g(Z_{t-1}) + \alpha_2 (h_{t-1}) \\
g(Z_t) &= \theta Z_t + \lambda (|Z_t| - E(Z_t))
\end{aligned}
$$

where $Z_t = \frac{\varepsilon_t}{\sqrt{h_t}}$ and if $Z_t \sim N(0,1)$, $E(Z_t) = \sqrt{2/\pi}$. The volatilities are always positive by construction and if the error distribution is normal, the volatilities are stationary if $|\alpha_2| < 1$. θZ_t represents the sign effect while $\lambda\left(|Z_t| - E\left(Z_t\right)\right)$ represents the size effect. $Z_t = 1$ would increase volatility by $\theta + \lambda$ while $Z_t = -1$ would increase volatility by $-\theta + \lambda$. Asymmetric effect would occur if $\theta + \lambda < -\theta + \lambda$ which implies that $\theta < 0$. Typically $\lambda > 0$.

Model 6: NAGARCH(1,1) model

$$h_t = \gamma + \alpha_1\left(\varepsilon_{t-1} + \theta h_{t-1}^{1/2}\right)^2 + \alpha_2 h_{t-1}$$

The volatilities are positive if $\gamma > 0$; $\alpha_1, \alpha_2 \geq 0$, and stationary if $\alpha_1\left(1+\theta^2\right) + \alpha_2 < 1$. Given that:

$$\begin{gathered}\left(\varepsilon_{t-1} - \theta h_{t-1}^{1/2}\right)^2 = \\ \varepsilon_{t-1}^2 + \theta^2 h_{t-1} + 2\theta\varepsilon_{t-1}h_{t-1}^{1/2}\end{gathered}$$

the nonlinear asymmetry is generated by $2\theta\varepsilon_{t-1}h_{t-1}^{1/2}$. As negative shocks are expected to increase volatility, asymmetric effect would occur if $\theta < 0$.

CHAPTER 7

Conclusions

In this paper, we have analyzed an effect of the sentiment of news related to the real estate market in Sweden, Estonia, Latvia, and Lithuania, on stock returns of Swedbank and SEB Bank. A considered period is from 04.01.2016 to 19.02.2019. At first, we have applied the Python open source tools and libraries for web scraping to get a text from news web pages, then used rule-based sentiment analysis tool - VADER model to compile a news sentiment time series. This data is used further to estimate four ARMA models for which we consider an asymmetric effect of the news.

The main finding is that there exists an asymmetry of positive and negative news. After separation of negative and positive news, Baltic real estate news shows a considerable effect on stock returns of both banks. Without separation, Swedish news influences the Swedbank's stock returns when for SEB, news about Baltic real estate is more significant.

Furthermore, an important result is that the aggregated number of negative news available does not have a considerable impact on returns, while the sentiment of negative news has. For positive news, it works vice versa.

Also, we have discussed that there is a need to investigate the shocks coming from volatilities in order to compare two types of asymmetry, and for that, the estimates received in this research will be used.

REFERENCES

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

Arik, A. (2011). Modeling market sentiment and conditional distribution of stock index returns under garch process. *Claremont Graduate University*.

Asai, M., and McAleer, M. (2006). Asymmetric multivariate stochastic volatility. *Econometric Reviews*, *25*(2-3), 453–473.

Bailey, M., Dávila, E., Kuchler, T., and Stroebel, J. (2017). *House price beliefs and mortgage leverage choice*. National Bureau of Economic Research.

Balasubramaniam, K. (2018, January 12). How to calculate a stock's adjusted closing price. Retrieved from https://www.investopedia.com

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T., and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, *11*(2), 143–172.

Box, G. E., Jenkins, G. M., and Reinsel, G. (1970). Time series analysis: Forecasting and control holden-day san francisco. *BoxTime Series Analysis: Forecasting and Control Holden Day1970*.

Bradley, M. M., and Lang, P. J. (1999). *Affective norms for english words (anew): Instruction manual and affective ratings*. Citeseer.

Brook, C., and Burke, S. (2003). Information criteria for garch model selection: An application to high frequency data. *European Journal of Finance*, *9*(6), 557–580.

Carmichael, B., and Coën, A. (2018). Real estate as a common risk factor in bank stock returns. *Journal of Banking & Finance*, *94*, 118–130.

Carnero, M. A., and Eratalay, M. H. (2014). Estimating var-mgarch models in multiple steps. *Studies in Nonlinear Dynamics & Econometrics*, *18*(3), 339–365.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738.

Ederington, L. H., and Lee, J. H. (1993). How markets process information: News releases and volatility. *The Journal of Finance*, *48*(4), 1161–1191.

Engelberg, J. E., and Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, *66*(1), 67–97.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F., and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749–1778.

Engle, R. F., Focardi, S. M., and Fabozzi, F. J. (2012). ARCH/garch models in applied financial econometrics. *Encyclopedia of Financial Models*.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, *38*(1), 34–105.

Fisher, K. L., and Statman, M. (2000). Investor sentiment and stock returns. *Financial Analysts Journal*, *56*(2), 16–23.

Ghosh, C., Guttery, R. S., and Sirmans, C. (1997). The effects of the real estate crisis on institutional stock prices. *Real Estate Economics*, *25*(4), 591–614.

Ghysels, E., Harvey, A. C., and Renault, E. (1996). 5 stochastic volatility. *Handbook of Statistics*, *14*, 119–191.

Groß-Klußmann, A., and Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, *18*(2), 321–340.

Gujarati, D. N. (2003). Basic econometrics: Fourth edition mcgraw-hill. *New York*.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2-3), 146–162.

He, L. T., Myer, F. N., and Webb, J. R. (1996). The sensitivity of bank stock returns to real estate. *The Journal of Real Estate Finance and Economics*, *12*(2), 203–220.

Hu, M., and Liu, B. (2004). Mining opinion features in customer reviews. In *AAAI* (Vol. 4, pp. 755–760).

Hutto, C. J., and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international aaai conference on weblogs and social media*.

Hutto, C., Klein, E., Pantone, P., and Berry, G. (2019). Vader lexicon. Retrieved from https://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/

vader_lexicon.txt

Igan, D., and Pinheiro, M. (2010). Exposure to real estate in bank portfolios. *Journal of Real Estate Research*, *32*(1), 47–74.

Kelly, S. (2016). *News, sentiment and financial markets: A computational system to evaluate the influence of text sentiment on financial assets* (PhD thesis). Trinity College Dublin.

Kosapattarapim, C., Lin, Y.-X., and McCrae, M. (2012). Evaluating the volatility forecasting performance of best fitting garch models in emerging asian stock markets.

Kothari, S. P., and Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics*, *44*(2), 169–203.

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., and Chen, Y. (2014). The effect of news and public mood on stock movements. *Information Sciences*, *278*, 826–840.

Liu, W., and Morley, B. (2009). Volatility forecasting in the hang seng index using the garch approach. *Asia-Pacific Financial Markets*, *16*(1), 51–63.

Maheu, J. M., and McCurdy, T. H. (2004). News arrival, jump dynamics, and volatility components for individual stock returns. *The Journal of Finance*, *59*(2), 755–793.

Martins, A. M., Serra, A. P., and Martins, F. V. (2016). Real estate market risk in bank stock returns: Evidence for 15 european countries. *International Journal of Strategic Property Management*, *20*(2), 142–155.

Mei, J., and Lee, A. (1994). Is there a real estate factor premium? *The Journal of Real Estate Finance and Economics*, *9*(2), 113–126.

Mei, J., and Saunders, A. (1995). Bank risk and real estate: An asset pricing perspective. *The Journal of Real Estate Finance and Economics*, *10*(3), 199–224.

Mitchell, M. L., and Mulherin, J. H. (1994). The impact of public information on the stock market. *The Journal of Finance*, *49*(3), 923–950.

NumFOCUS. (2019). Python data analysis library. Retrieved from https://pandas.pydata.org/

Olowe, R. A. (2009). Modelling naira/dollar exchange rate volatility: Application of garch and assymetric models. *International Review of Business Research Papers*, *5*(3), 377–398.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. (2015). The effects of twitter sentiment on stock price returns. *PloS One*, *10*(9), e0138441.

Richardson, L. (2015). Beautiful soup 4.4.0 documentation. Retrieved from https://

www.crummy.com/software/BeautifulSoup/bs4/doc/

Sidorov, S. P., Revutskiy, A., Faizliev, A., Korobov, E., and Balash, V. (2014). Stock volatility modelling with augmented garch model with jumps. *IAENG International Journal of Applied Mathematics*, *44*(4), 212–220.

Soo, C. K. (2018). Quantifying sentiment with news media across local housing markets. *The Review of Financial Studies*, *31*(10), 3689–3719.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, *4*(2), 183–204.

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

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, *63*(3), 1437–1467.

Tumarkin, R., and Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, *57*(3), 41–51.

Verma, R., and Soydemir, G. (2009). The impact of individual and institutional investor sentiment on the market price of risk. *The Quarterly Review of Economics and Finance*, *49*(3), 1129–1145.

Whitle, P. (1951). *Hypothesis testing in time series analysis* (Vol. 4). Almqvist & Wiksells.

Yahoo. (2019). Yahoo! Finance. Retrieved from https://finance.yahoo.com

Yu, X. (2014). *Analysis of new sentiment and its application to finance* (PhD thesis).

Zhang, X., Fuehres, H., and Gloor, P. A. (2011). Predicting stock market indicators through twitter "i hope it is not as bad as i fear". *Procedia-Social and Behavioral Sciences*, *26*, 55–62.

APPENDICES

Appendix A: Snippet of algorithm described in Chapter 3

Sentiment Analyzer needed:

```
analyser = SentimentIntensityAnalyzer()
```

To read 'csv' file contaned URL to news related to the topic for specified period:

```
data = pd.read_csv("URL_reuters.csv", header = 0, sep = ";")
```

To clear text from tags which are not needed:

```
invalid_tags = ['b', 'i', 'u','div','a','strong','span','n']
```

Part of the algorithm used to read web pages, extract the text, evaluate the sentiments and save them as lists with dates when news appeared and sentiments:

```
urls = data['URL'].count()

scores = []
article = []
ldate = []
i = 0

while i < urls:
    try:
        quote_page = data['URL'][i]
        page = requests.get(quote_page)
        soup = BeautifulSoup(page.content, 'html.parser')
        text = soup.find(class_ = "StandardArticleBody_body")

        for tag in invalid_tags:
            for match in text.findAll(tag):
                match.replaceWithChildren()
                main = str(text.get_text())

        main_n = textwrap.dedent(main).rstrip()
        heading = soup.find(class_ = "ArticleHeader_headline")
        heading = heading.get_text()
        heading = textwrap.dedent(heading).strip()

        article_res = heading + ' ' + main_n
        article.append(article_res)

        score = analyser.polarity_scores(article_res)
        scores.append(score)
```

```
        article_date = soup.find(class_ = "ArticleHeader_date")
        article_date = article_date.get_text()
        article_date = textwrap.dedent(article_date).strip()
        match = re.search(r'.*\d{4}\s', article_date)
        article_date = match.group()

        ldate.append(article_date)
        i = i + 1

        delays = [3, 4, 6, 2, 5, 1]
        delay = numpy.random.choice(delays)
        time.sleep(delay)

    except AttributeError:
        print (data['URL'][i] + " was skipped due to AttributeError")
        i = i + 1
```

Appendix B: Snippet of R Code for Chapter 4

Needed libraries: ggplot2. To read file with stock prices and sentiments (file must be saved in the same folder where R file is placed):

```
sentiments <- read.csv(file="Sentiments with stocks.csv",
                       header=TRUE, sep=";")
sentiments$Date <- as.Date(sentiments$Date, format = "%d.%m.%Y")
```

To upload the Swedbank and SEB daily adjusted closing stock prices directly from Yahoo!Finance for further investigation, ‘quantmod’ R package is needed:

```
getSymbols('SWED-A.ST', src = 'yahoo', from = '2016-1-3', to = '2018-2-19',
           format="%Y-%M-%D", auto.assign = T)
adj_close_sw <- `SWED-A.ST`$`SWED-A.ST.Adjusted`

getSymbols('SEB-A.ST', src = 'yahoo', from = '2016-1-3', to = '2018-2-19',
           format="%Y-%M-%D", auto.assign = T)
adj_close_seb <- `SEB-A.ST`$`SEB-A.ST.Adjusted`
```

Appendix C: List of news sources used for the sentiment analysis

1. Baltic news.
   - Eesti Rahvusringhääling: https://news.err.ee/
   - The Baltic Times: https://www.baltictimes.com/
   - Baltic News Network: https://bnn-news.com/
   - Reuters: https://www.reuters.com/
   - EN.DELFI: https://en.delfi.lt/
   - The Baltic Course: http://www.baltic-course.com/
   - LETA: http://www.leta.lv/
   - The New York Times: https://www.nytimes.com
   - REINVEST24: https://blog.reinvest24.com/
   - NEWSEC: http://newsecbaltics.com/
2. Swedish news.
   - Reuters: https://www.reuters.com/
   - The Local SE: https://www.thelocal.se/
   - Quartz: https://qz.com
   - Business Insider: https://www.businessinsider.com
   - Bloomberg: https://www.bloomberg.com/
   - Telegraph: https://www.telegraph.co.uk/
   - Financial TImes: https://www.ft.com
   - The Brampton Guardian: https://www.bramptonguardian.com/
   - Sputnik: https://sputniknews.com/
   - CNBC: https://www.cnbc.com/
   - Politico: https://www.politico.eu/
   - IMF News: https://www.imf.org/en/News/
   - Forbes: https://www.forbes.com/
   - The Construction Index: https://www.theconstructionindex.co.uk/
   - The New York Times: https://www.nytimes.com
   - Euromoney: https://www.euromoney.com/

- BuyAssociation: https://www.buyassociation.co.uk/
- The Wall Street Journal: https://www.wsj.com
- Coindesk: https://www.coindesk.com/
- The Economist: https://www.economist.com
- Expert Investor: https://expertinvestoreurope.com/
- Cision PR Newswire: https://www.prnewswire.com/
- Property Funds World: https://www.propertyfundsworld.com
- GlobeNewswire: https://www.globenewswire.com/
- Global Property Guide: https://www.globalpropertyguide.com/
- Data Centre Dynamics: https://www.datacenterdynamics.com/
- ING THINK: https://think.ing.com

Non-exclusive licence to reproduce thesis and make thesis public

I, Yuliia Puzanova

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,
   Effect of real estate news sentiment on stock returns of Swedbank and SEB Bank,
   supervised by Mustafa Hakan Eratalay.
2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.
4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Yuliia Puzanova

23.05.2019