University of Tartu
School of Economics and Business Administration

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

Yuliia Puzanova, M. Hakan Eratalay

Tartu 2020

ISSN-L 1406-5967
ISSN 1736-8995
ISBN 978-9985-4-1204-6 (pdf)
The University of Tartu FEBA
https://majandus.ut.ee/en/research/workingpapers

Effect of real estate news sentiments on the stock returns of Swedbank and SEB Bank

Yuliia Puzanova[1], M. Hakan Eratalay[2]

Abstract

This paper analyses the effect of real estate news sentiment on the stock returns of Swedbank and SEB Bank, which are leading banks in Sweden and the Baltic region. For this purpose, we have selected sentiments from news about real estate in the markets of these banks in Sweden, Estonia, Latvia, and Lithuania between 4 January 2016 and 19 February 2019. Estimation results showed that sentiments about the housing market affect stock returns for the banks, and showed the presence of the asymmetric effects of positive and negative news. We also found that there is a difference in the stock returns of these banks in terms of when and to what extent they react to news coming from the Baltic States and Sweden. Moreover, we found that the number of negative news affects the stock returns of the banks more than the strength of the news. We also apply several GARCH specifications to show that negative and positive news explain the asymmetric effects in the volatility processes to some extent. The asymmetric effects in the volatilities are captured much better by the GJR-GARCH and NA-GARCH models, implying that these effects are generated by idiosyncratic shocks rather than the sentiments in the news.

JEL Classification: C320, C520, C580

Keywords: sentiment analysis, real estate market, Sweden, Baltics, Latvia, Lithuania, Estonia.

1 Introduction

The Nordic and Baltic banking sectors are closely connected. Two banks that are systematically important to the Baltic financial system, namely Swedbank and SEB, 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 – in the Baltic states. Real estate news appearing in all markets can be influential for the banks' activities related to lending services, and

[1]Group Business Intelligence, Swedbank AS, Tartu, Estonia. Email: puzanova.yuliia@gmail.com

[2]Corresponding author. Department of Economics, University of Tartu, Tartu, Estonia. Email: hakan.eratalay@ut.ee

This paper is developed from the Master's thesis of Yuliia Puzanova at the University of Tartu, which was supervised by M. Hakan Eratalay. We are grateful for the comments of Luca Alfieri and two anonymous referees which contributed to improving of this paper. All remaining errors are ours.

impact mortgage volume growth. As a result, we can expect an impact on the profitability of the banks and their stock price returns.

News regarding a housing shock and future real estate crash is attracting increasingly 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 is characterised as exhibiting fast growth, declining unemployment, population increase, and low-interest rates, but we can also observe the 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 in news articles. In addition, even after the Swedish Finansinspektionen has introduced measures to handle the risks associated with the rise in household debt, the tension in society and among investors remains.

Baltic news channels also raise the topic of real estate market vulnerabilities. However, factors influencing such discussions differ from that of the Swedish economy. Solid economic growth in Baltic countries is followed by increased activity on the housing market as capital cities, large cities, and resorts are becoming hot locations for real estate objects. The Estonian economy does not show any signs of overheating, as there is no structural economic imbalance. The Lithuanian market is actively boosted due to rising investments. And in the case of Latvia, the residential market is growing due to an increase in the average salary (Ober Haus Report, 2018). Nevertheless, there is concern in the media about a repeat of the housing bubble that happened in the years 2005–2010 in the Baltic states. Therefore, the aim of this paper is to study the extent to which this tension in the media over real estate markets in Sweden and the Baltics can influence the stock returns of the Swedish banks. To that end, we enhance traditional econometric models using a new component – news sentiment. Such a simultaneous use of text analysis and econometric analysis can potentially benefit the financial world and draw attention to the importance of tracking the market mood.

In this research, we use the autoregressive moving average (ARMA) model to estimate the conditional mean for stock returns. The ARMA model makes it possible to identify the best model for a conditional mean equation, and its residuals can be used further to estimate GARCH specifications.

To gather sentiment data, we use open-source tools and libraries, such as the Python library "Beautiful Soup" to conduct web scraping of news pages, and the VADER model for sentiment analysis (Hutto and Gilbert, 2014).

This study contributes to the literature by investigating real estate news sentiment for a period from 4 January 2016 to 19 February 2019. We take into consideration news related to the real estate market in countries which are the home markets of the banks Swedbank and SEB, and its effect on the returns for these banks. In addition, we consider asymmetrical effects in our model. There do not seem to be any other papers investigating non-financial sentiments related to Swedish and Baltic markets, and neither did we find any studies related to investigations of the effect of housing market sentiments on bank stock returns.

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

Moreover, to obtain sentiment data, we wrote an algorithm using open-source libraries and tools, instead of using the news analytic and data gathered by commercial providers, as was done for instance by Sidorov et al. (2014); Verma and Soydemir (2009); Yu (2014).

This paper finds empirical evidence of a positive and statistically significant relationship between an increase in positive real estate news sentiment and upward movements in bank stock returns. Swedish news has a greater impact on Swedbank's stock returns in comparison with other news providers, while for SEB Bank, the Baltic news channels are more influential. When considering negative sentiments and their influence on stocks, we see that extremely negative news related to Baltic real estate markets has a greater influence than news about the Swedish market for both banks in this study. When we have merged news from different sources, we see that the magnitude of negative news available does not have a considerable impact on returns, while the number of negative news has. For positive news, it was found to be working conversely. So, this means that if there is news of possibly large size from only one source, the effect on the stock returns is small. On the other hand, if there is news of even modest size, but published in many sources, the effect on the stock returns is very large.

The paper is structured in the following way: a review and discussion of the literature on the impact of news sentiment on stock prices, the linkages between banks and the real estate market, and approaches to stock price modelling are provided in Section 2. A description of the methods used for data collection, the procedure for news sentiment estimation and an overview of the econometric models used can be found in Section 3. Section 4 presents the data used for modelling and provides its main characteristics considered later in the modelling part. Section 5 reveals the results and interpretations. Finally, section 6 concludes the paper.

2 Literature Review

2.1 News sentiment and its impact on stock prices

Nowadays, there are huge volumes of publicly accessible and quickly distributed news, and this amount is increasing. This makes news one of the main sources of information that helps form opinions and support decision-making. The news is also 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 understanding the behaviour of market players (Arik, 2011). News analytics is a highly popular research topic due to its effective application in predictions in regard to market volatility, prices and trading volumes predictions (Sidorov et al., 2014). In finance, news sentiment is considered an event and a quantitative reflection of information. Simply saying it measures the emotional tone of available 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 analyse the relationship between news sentiment and stock returns. Evidence was found that how positive and negative sentiment influences market volatility differs, and this was proved to be substantial in the case of negative sentiment (Engle and Ng, 1993; Tetlock, 2007). But if we talk about the use of news sentiment as a factor, Tetlock (2007) was the first to prove its significance in a predictive model. Later, Tetlock et al. (2008) obtained better prediction results in comparison with forecasts prepared by analysts by applying the "Bag-Off-Words" model for news sentiment analysis (Harris, 1954).

News sentiment in predictions of stock prices is not new but still is an elusive concept (Yu, 2014). The classical theory about stock price formation is that asset price reflects all available market 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); Engelberg and Parsons (2011). Today, news sources not only include official news agencies or television outlets, but also published company reports, publicly available statistics, Security Exchange Commission reports - all of these are known as "pre-news" and are the first to influence the public mood (Sidorov et al., 2014). Moreover, more power, in the sense of the number of people covered, is getting wielded by social media (e.g. social network posts, blogs, tweets, etc.) For example, anonymised Facebook data is commonly used to evaluate people's beliefs, as was done by Bailey et al. (2017) to investigate home buyers' beliefs regarding future price movements and how this could impact decisions regarding mortgage leverage.

Official news channels such as Thomson Reuters, Bloomberg, Dow Jones, and the Wall Street Journal provide reliable information which is used by investors in forming their opinion about future stock price movement. But blogs, forums, and social media form the opinions of the general public, and their popularity is growing (Yu, 2014). In many cases, these resources create informational topics for mentioned 'reliable' sources. The impact and predictive power of blogging platforms, like Twitter, was studied by Bollen et al. (2011); Zhang et al. (2011); Ranco et al. (2015) and others. But we should take into account that sentiments taken from social media differ in some sense from the overall news sentiments, 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 has been proved using 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 relationships using a Value at Risk model. But, it has been proved that GARCH models better explain the financial fat-tailed data with excess kurtosis. Furthermore, Sidorov et al. (2014) considered the 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).

2.2 Real estate market and bank stock returns

Banks are highly exposed to the real estate market (Igan and Pinheiro, 2010; Martins et al., 2016). This functions via the next scenario: when the housing market has a downtrend, banks have less capital means and their expansion will be shortened, and the most significant changes for the population might be a credit reduction.

The role of the real estate market in the pricing of bank stocks was studied by different authors. Real estate market risk and its influence on US bank stocks were estimated by Carmichael and Coën (2018). Also, the high sensitivity of stock returns to changes in real estate returns was shown by He et al. (1996). Moreover, the vast majority of studies have explored the effects of the real estate crisis on stock prices dynamics. The main idea behind their study results is that financial institution stock price movements and the level of its real estate exposure are significantly connected (Ghosh et al., 1997; Martins et al., 2016; Igan and Pinheiro, 2010). But it is important to mention that the size of bank matters. As was shown in papers Mei and Lee (1994) and (Mei and Saunders, 1995), greater sensitivity to changes in the housing market is mostly experienced by small banks.

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

2.3 Modelling of stock returns

Stock prices and returns are hardly predictable but, at the same time, their forecasts are very desirable. Investors and researchers like to develop algorithms for the prediction of stock movements in order to make better decisions and optimise 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. Primarily, there are two assumptions regarding stock prices predictions: 1) stock price behavior depends on past values and can be modelled, 2) stock prices are not dependent on 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). They are known to be quite efficient for short-term predictions. Such a combination of AR and MA terms was first proposed by Whitle (1951). Thereafter, Box et al. (1970) introduced rARMA(p,q) modeling approach which is still the main strategy for selecting orders of AR and MA polynomials.

The variation in stock returns are commonly modelled using the generalized autoregressive conditional heteroskedasticity (GARCH) method. It is difficult to predict the volatility because it is not stable in time, and 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. Instead of taking it as a problem, conditional volatility models consider it as a variance to be modeled Engle et al. (2012). The commonly used models in financial time series are the autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982), the generalized autoregressive conditional heteroskedasticity (GARCH) model presented by Bollerslev (1986) and their numerous variants. The ARCH model allows the conditional variance to vary in time as a function of past errors, while assuming the unconditional variance constant. The GARCH model allows a more flexible lag structure. As Bollerslev (1986) explained, the main difference is that conditional variance in the ARCH model only has a view of the linear function of past sample variances, while the GARCH process allows us to specify 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 forecast similar behaviour on the next time horizon. The strong evidence in favor of GARCH models with non-normal distributions usage is provided by Liu and Morley (2009). Results presented by Kosapattarapim et al. (2012) also 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 our estimations. Applying the suitable ARMA(p,q) model helps us to solve the autocorrelation problem in the residuals and to see dependencies and effects of sentiments on the stock returns. Thereafter, in follow up studies, the received residuals can be used to estimate the conditional volatility us the best suited GARCH(p,q) model. GARCH(p,q) models that can be used further are shown section 3.

3 Methods and analysis

3.1 Data collection and aggregation

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

Every web page is compiled using "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 to other parts of the article, such as date of publishing or heading. Web scrapers navigate through provided web URLs, looking for the requested information using HTML tags and then download the text on the user's request. We use the Python "Beautiful Soup" (Richardson) library for parsing web pages. Its functions such as "soup.find", "soup.title" or "soup.get_text" are used to extract text from the web page.

One of the difficulties when we are dealing with web scrapping, is that using the same markup language for web pages does not require that the same tags for formatting are used. This creates some inconveniences in writing a universal algorithm to recognize the necessary tags and download the text. For example, after inspecting the web pages' structure, we found that for The Local SE web pages, the body of the article is contained under the tag "div id ="article-body", while for Reuters, this tag is "div class ="StandardArticleBody_body". That is why we have decided to create an algorithm which only requires a couple of tags as input pointing out the part of the web page containing the necessary text.

The resulting algorithm is capable of extracting the main part of the article, its heading and publishing date from different web sites. The only manual step is to inspect the web page and identify tags which lead to the necessary part of the page. Moreover, this algorithm cleans the text of unnecessary tags; for example, those used to format the text, merge the heading and the main part of the article to the one data object and save it 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 the text corpus.

3.2 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 the text. For computers, reading and understanding the language is a highly complex process, involving complicated algorithms with thousands of lines of code.

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

Most of the sentiment analysers are based on sentiment lexicon, list of words labeled either positive, negative or neutral. There are several widely used lexicons, such as LIWC[3] (Linguistic Inquiry and Word Count), mostly used for social media texts, where it is possible to estimate the intensity of words. Hu and Liu (2004) describes the sentiment of text at a high level but has no ability to recognise 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[4] also contains estimated sentiment polarities using the range from -1 to 1. The latter is used for the VADER model (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 and a rule-based model used for general sentiment analysis. The lexicon and rules used by VADER are open and easily accessible, which makes its use for research purposes advantageous (Hutto and Gilbert, 2014).

[3] http://liwc.wpengine.com/
[4] https://sentic.net/

The main task for the evaluation of text sentiments is to calculate sentiment polarity. Before applying the sentiment analyser, in most cases, data pre-processing is needed. In our case, this was carried out after web scraping. Afterward, the "polarity_scores" method is used to obtain polarity indices for the text data.

Sentiment values for every word used in the text are calculated using the "polarity_scores" function in combination with the lexicon of the VADER model (Hutto et al., 2019). The result of the primary calculations is that every word has precalculated values - polarity scores with the format: [x,y,z], where x, y, and z are negative, positive and neutral sentiments, respectively. The compound value is calculated via equation 1 using the sum of all sentiments and the normalisation function, which places the value in the range [-1,1] where a positive sentiment corresponds to a compound score greater than or equal to 0.05, a neutral sentiment is located in a range (-0.05, 0.05), and a negative sentiment corresponds to values less than or equal to -0.05. This indicator is used further in econometric models. The function has the form:

$$C = \frac{S}{\sqrt{S^2 + 15}} \tag{1}$$

where C is the compound sentiment score, and S the 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 the expressed emotion, the capitalization of words, conjunctions 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 implemented in Python. Therefore, this means that as soon as the 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 the econometric models. The VADER model and the calculated sentiments are saved to time series for further modeling. The algorithm can be visualized by the flowchart (Figure 1).

3.3 Overview of models used

The 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;

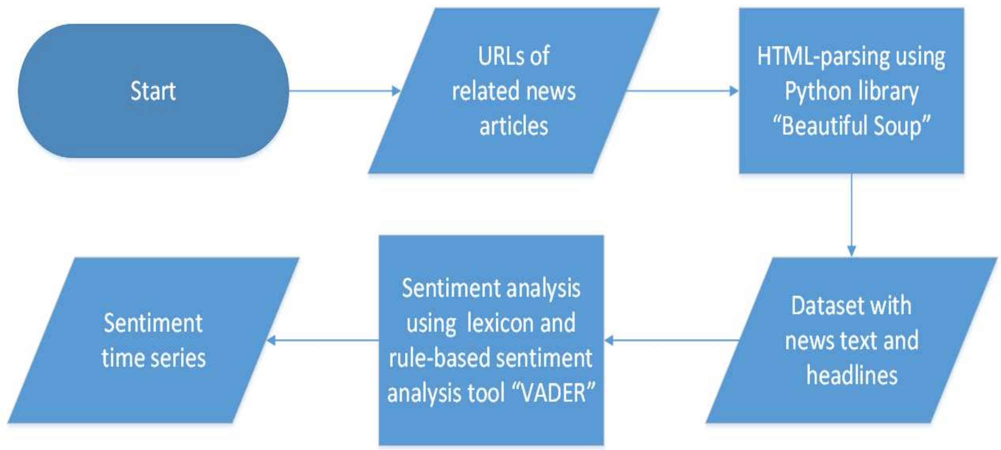


Figure 1: The algorithm used to extract news and estimate its sentiment

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 with the date the article was published. Source: authors' depiction of the steps in the algorithm.

- k - number of exogenous variables.

AR part of the model given in equation 2 describes predicted stock return as a value dependent on the values of previous p periods and random terms, r_t is expressed as its deviation from the mean value:

$$r_t = \mu + \sum_{i=1}^{p} \beta_i r_{t-i} + \varepsilon_t \tag{2}$$

The moving average model given in equation 3 is a linear combination of random unexpected shocks impacting on the stock returns. MA(q) process can be described as:

$$r_t = \mu + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{3}$$

The ARMA(p,q) model is given in equation 4 and is able to describe time series with the characteristics of both the AR(p) and MA(q) processes via their combination in ARMA model (Gujarati, 2003).

Model 1.0. ARMA(p,q) model, our base-line model

$$r_t = \mu + \sum_{i=1}^{p} \beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{4}$$

The model in equation 4 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.

All p-values for the likelihood ratio tests in the ARMA model variants are calculated in comparison with the base model: ARMA(p,q). The choice of orders p and q is discussed in Section 5.

Model 1.1. ARMA(p,q) model with news component

The news sources we consider are from Sweden and the Baltic States. Moreover, we include news from Reuters to take into account international news. News travels very fast in the financial world, and therefore we wouldn't expect many lags for the news variables. Another restriction is that we consider many news variables, and therefore the number of lags should be low for a parsimonious model. The preliminary estimations with two lags for news variables gave the most sensible model. [5]

$$\begin{aligned} r_t = \mu &+ \sum_{i=1}^{p} \beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \\ &+ \delta_{10} Reuters_t + \delta_{20} TheLocal_t + \delta_{30} Others_t + \delta_{40} Baltics_t \\ &+ \delta_{11} Reuters_{t-1} + \delta_{21} TheLocal_{t-1} + \delta_{31} Others_{t-1} + \delta_{41} Baltics_{t-1} \\ &+ \delta_{12} Reuters_{t-2} + \delta_{22} TheLocal_{t-2} + \delta_{32} Others_{t-2} + \delta_{42} Baltics_{t-2} \end{aligned} \tag{5}$$

In equation 5, news variables are defined with the names of their sources and are by construction in the range $[-1, 1]$, where a positive news is higher ranked than negative news.

[5] We estimated the ARMA model with news with no lags and with one lag for the news. The results suggested that the coefficients for the Baltic news are insignificant. This clearly indicated that there are some omitted variables/lags of variables. Therefore, we increased the number of lags to 2 which gave sensible estimation results.

Therefore any increase in the news variable indicate less bad news or more good news; hence, should increase the returns in the next period. Therefore, the restrictions on the news sentiment coefficients are:

1. $\delta_{ji} \geq 0$ for all $i = 1, ..., 4, j = 0, 1, 2$.

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

Model 1.2. ARMA(p,q) model with asymmetric effect of news, neutral news discarded

$$
\begin{aligned}
r_t = {} & \mu + \sum_{i=1}^{p} \beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \\
& + \delta_{10}^{-} Reuters_t^{-} + \delta_{20}^{-} TheLocal_t^{-} + \delta_{30}^{-} Others_t^{-} + \delta_{40}^{-} Baltics_t^{-} \\
& + \delta_{11}^{-} Reuters_{t-1}^{-} + \delta_{21}^{-} TheLocal_{t-1}^{-} + \delta_{31}^{-} Others_{t-1}^{-} + \delta_{41}^{-} Baltics_{t-1}^{-} \\
& + \delta_{12}^{-} Reuters_{t-2}^{-} + \delta_{22}^{-} TheLocal_{t-2}^{-} + \delta_{32}^{-} Others_{t-2}^{-} + \delta_{42}^{-} Baltics_{t-2}^{-} \\
& + \delta_{10}^{+} Reuters_t^{+} + \delta_{20}^{+} TheLocal_t^{+} + \delta_{30}^{-} Others_t^{+} + \delta_{40}^{-} Baltics_t^{+} \\
& + \delta_{11}^{+} Reuters_{t-1}^{+} + \delta_{21}^{+} TheLocal_{t-1}^{+} + \delta_{31}^{-} Others_{t-1}^{+} + \delta_{41}^{-} Baltics_{t-1}^{+} \\
& + \delta_{12}^{+} Reuters_{t-2}^{+} + \delta_{22}^{+} TheLocal_{t-2}^{+} + \delta_{32}^{-} Others_{t-2}^{+} + \delta_{42}^{-} Baltics_{t-2}^{+}
\end{aligned}
\tag{6}
$$

In equation 6, 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.[6] Otherwise, the news sentiment is neutral. Neutral news is discarded because *TheLocal* and *Others* news sources do not have neutral news. It is trivial that positive news should increase returns in the next period. Furthermore, if the size of the negative news falls, then the variable $news^{-}$ increases; therefore, this should increase returns in the next period. Therefore, the restrictions on the news sentiment coefficients are:

1. $\delta_{ji}^{-}, \delta_{ji}^{+} \geq 0$ for all $i = 1, ..., 4, j = 0, 1, 2$.

Model 1.3. ARMA(p,q) model with a naive threshold for extreme news

[6]The values 0.05 and -0.05 are given by the construction of the news sentiment indices.

$$\begin{aligned} r_t = \mu + \sum_{i=1}^{p} \beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \\ &+ \delta_{10} Reuters_t + \delta_{20} TheLocal_t + \delta_{30} Others_t + \delta_{40} Baltics_t \\ &+ \delta_{11} Reuters_{t-1} + \delta_{21} TheLocal_{t-1} + \delta_{31} Others_{t-1} + \delta_{41} Baltics_{t-1} \\ &+ \delta_{12} Reuters_{t-2} + \delta_{22} TheLocal_{t-2} + \delta_{32} Others_{t-2} + \delta_{42} Baltics_{t-2} \\ &+ \delta_{10}^{-} I_t^{Reuters,-e} + \delta_{20}^{-} I_t^{TheLocal,-e} + \delta_{30}^{-} I_t^{Others,-e} + \delta_{40}^{-} I_t^{Baltics,-e} \\ &+ \delta_{11}^{-} I_{t-1}^{Reuters,-e} + \delta_{21}^{-} I_{t-1}^{TheLocal,-e} + \delta_{31}^{-} I_{t-1}^{Others,-e} + \delta_{41}^{-} I_{t-1}^{Baltics,-e} \\ &+ \delta_{12}^{-} I_{t-2}^{Reuters,-e} + \delta_{22}^{-} I_{t-2}^{TheLocal,-e} + \delta_{32}^{-} I_{t-2}^{Others,-e} + \delta_{42}^{-} I_{t-2}^{Baltics,-e} \\ &+ \delta_{10}^{+} I_t^{Reuters,+e} + \delta_{20}^{+} I_t^{TheLocal,+e} + \delta_{30}^{+} I_t^{Others,+e} + \delta_{40}^{+} I_t^{Baltics,+e} \\ &+ \delta_{11}^{+} I_{t-1}^{Reuters,+e} + \delta_{21}^{+} I_{t-1}^{TheLocal,+e} + \delta_{31}^{+} I_{t-1}^{Others,+e} + \delta_{41}^{+} I_{t-1}^{Baltics,+e} \\ &+ \delta_{12}^{+} I_{t-2}^{Reuters,+e} + \delta_{22}^{+} I_{t-2}^{TheLocal,+e} + \delta_{32}^{+} I_{t-2}^{Others,+e} + \delta_{42}^{+} I_{t-2}^{Baltics,+e} \end{aligned} \tag{7}$$

In equation 7, I is a dummy variable for extreme news. A news sentiment is considered to be positive extreme (superscript "+e") if its value is 3 standard deviations higher than its mean, and negative extreme (superscript "-e") if its value is 3 standard deviations less than its mean.

The existence of a positive (negative) extreme news should impact the returns positively (negatively) in the next period. Hence the restrictions on the news sentiment coefficients are:

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

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

Model 1.4. ARMA(p,q) model with merged news data (positive and negative news distinguished)

$$\begin{aligned} r_t = \mu + \sum_{i=1}^{p} \beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \\ &+ \delta_{10}^{+} News_t^{pos} + \delta_{11}^{+} News_{t-1}^{pos} + \delta_{12}^{+} News_{t-2}^{pos} \\ &+ \delta_{20}^{-} News_t^{neg} + \delta_{21}^{-} News_{t-1}^{neg} + \delta_{22}^{-} News_{t-2}^{neg} \\ &+ \delta_{30}^{+} N_t^{pos} + \delta_{31}^{+} N_{t-1}^{pos} + \delta_{32}^{+} N_{t-2}^{pos} \\ &+ \delta_{40}^{-} N_t^{neg} + \delta_{41}^{-} N_{t-1}^{neg} + \delta_{42}^{-} N_{t-2}^{neg} \end{aligned} \tag{8}$$

In equation 8, the $News_{t-1}$ 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 and negative news at time t with the variables N_t^{pos} and N_t^{neg}, respectively.

The amount of positive (negative) news impacts the next period returns positively (negatively); therefore, the restrictions on the news sentiment coefficients are:

1. $\delta_{1i}^{+}, \delta_{2i}^{-}, \delta_{3i}^{+} \geq 0$.

2. $\delta_{4i} \leq 0$.

where $i = 0, 1, 2$.

After estimating these models, the best model for the conditional mean equation was chosen based on the sum of squared errors (SSE). The Akaike (AIC) and Bayesian (BIC) criteria indicated that the base model with no news is the best. However, this model does not give the smallest SSE and does not include the news variables. In fact, AIC and BIC criteria punishes the models for including too many parameters; therefore, they favor the most parsimonious model.

From the chosen conditional mean model, we receive the estimated residuals, $\hat{\varepsilon}_t$. We use this vector of estimated residuals to estimate the following GARCH specifications. For simplicity, we assume the GARCH(1,1) order and keep the focus on the extensions of GARCH models.[7]

When modelling the volatility, we specifically focus on the asymmetric effects. In the literature, asymmetric effects is a stylised fact observed in the volatilities of financial returns, which suggests that volatility increases much more when a negative shock appears compared to a positive shock of the same magnitude. (Ghysels *et al.* 1996, Asai *et al.* 2006) In the base ARMA-GARCH model, news sentiments are not included in the equations; therefore, they appear as shocks to the error term and hence contribute to asymmetric effects through the errors. When we add the news sentiments variables, we can directly control for and estimate the asymmetric effect of news on the volatility of the returns. Therefore, we refer to the asymmetric effect of news as "asymmetric effects" as in the stylized facts, although this may not be perfectly in line with the definition in Asai *et al.* (2006).

[7] There are two reasons for this: the first is the same as in the conditional mean model, we chose a parsimonious model to focus on the effect of news variables. In fact, we estimated higher orders of GARCH model and GARCH(1,1) was outperforming GARCH(p,q) models based on the Bayesian criterion, which favors parsimonious models. The second reason is that we are motivated by the paper of Hansen and Lunde (2005), who found out that for the IBM stock return data the higher orders of the GARCH model cannot outperform the GARCH (1,1). The latter was only outperformed by GARCH models which allowed for asymmetric effects. In our paper we use such models as well: GJR-GARCH and NAGARCH.

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, Engle and Shephard 2001, Carnero and Eratalay 2014).

In the following models, the notation is as follows: $h_t = Var(r_t|F_{t-1}) = Var(\varepsilon_t|F_{t-1}) = E(\varepsilon_t^2|F_{t-1})$, where F_{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. Finally, a negative news would increase volatility because it increases uncertainty, but a positive news may not increase volatility as much or may even decrease volatility.

All p-values for the likelihood ratio tests are calculated in comparison with base model: GARCH(1,1), which is given in equation 9.

Model 2.0. GARCH(1,1) model (base line model)

$$h_t = \gamma + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} \tag{9}$$

In equation 9, the volatilities h_t are positive if $\gamma > 0$; α_1, $\alpha_2 \geq 0$ and stationary if $\alpha_1 + \alpha_2 < 1$.

Model 2.1. GARCH(1,1) model with asymmetric news, neutral news discarded

$$\begin{aligned}
h_t &= \gamma + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} \\
&+ \delta_{10}^- Reuters_t^- + \delta_{20}^- TheLocal_t^- + \delta_{30}^- Others_t^- + \delta_{40}^- Baltics_t^- \\
&+ \delta_{11}^- Reuters_{t-1}^- + \delta_{21}^- TheLocal_{t-1}^- + \delta_{31}^- Others_{t-1}^- + \delta_{41}^- Baltics_{t-1}^- \\
&+ \delta_{12}^- Reuters_{t-2}^- + \delta_{22}^- TheLocal_{t-2}^- + \delta_{32}^- Others_{t-2}^- + \delta_{42}^- Baltics_{t-2}^- \\
&+ \delta_{10}^+ Reuters_t^+ + \delta_{20}^+ TheLocal_t^+ + \delta_{30} Others_t^+ + \delta_{40} Baltics_t^+ \\
&+ \delta_{11}^+ Reuters_{t-1}^+ + \delta_{21}^+ TheLocal_{t-1}^+ + \delta_{31} Others_{t-1}^+ + \delta_{41} Baltics_{t-1}^+ \\
&+ \delta_{12}^+ Reuters_{t-2}^+ + \delta_{22}^+ TheLocal_{t-2}^+ + \delta_{32} Others_{t-2}^+ + \delta_{42} Baltics_{t-2}^+
\end{aligned} \tag{10}$$

Similar to equation 6, in equation 10 the volatilities h_t are positive if $\gamma > 0$, $\alpha_1, \alpha_2, \delta_{ij} \geq 0$ and stationary if $\alpha_1 + \alpha_2 < 1$.

Model 2.2. GJR-GARCH(1,1) model

$$h_t = \gamma + (\alpha_1 + \theta_1 I_{t-1})\, \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} \tag{11}$$

In equation 11, the dummy I_{t-1} takes value 1 if $\varepsilon_{t-1} < 0$. The volatilities h_t are positive if $\gamma > 0$, $\alpha_1, \alpha_2 \geq 0$, and stationary if $\alpha_1 + \theta_1 + \alpha_2 < 1$. Asymmetric effects would occur if $\theta_1 > 0$, because negative errors increase volatility more than positive errors.

Model 2.3. NAGARCH(1,1) model

$$h_t = \gamma + \alpha_1 \left(\varepsilon_{t-1} + \theta h_{t-1}^{1/2}\right)^2 + \alpha_2 h_{t-1} \tag{12}$$

In equation 12, the volatilities h_t are positive if $\gamma > 0$, $\alpha_1, \alpha_2 \geq 0$, and stationary if $\alpha_1 (1 + \theta^2) + \alpha_2 < 1$. Given that:

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

The non-linear asymmetry in equation 12 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 $\theta < 0$.

4 Data

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

4.1 News sentiments data

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

For this purpose, two main sources were chosen to search for news related to the Swedish real estate market: "The Local SE" – a portal that posts Swedish news in English, and

"Reuters" – an international news provider. Posts related to the real estate market in Sweden appear more frequently in The local news portal, and they are necessary to capture possible interactions between daily stock price changes and news. Messages in international news portals about real estate appear more rarely and only in cases of high target-reader interest; or, in other words, when this news is highly important for international audiences. Such news can be a signal for the market about upcoming up or downturns. In addition, this type of news might influence market stability to a greater degree.

To cover those periods when the main selected sources do not provide any articles, we refer to other sources and look for them using the news aggregator, "Google News" hereinafter referred to as "Other Sweden". To these sources, we also include such news providers as "Financial Times", "Business Insider Nordic", "Bloomberg", "The Wall Street Journal" and others [8]. It was highly desirable for further modeling to decrease the number of gaps in news sentiments data set, or simply, to have as many days as possible with at least one news sentiment value.

Regarding the news from the Baltic countries, the main sources of information are "ERR News" – Estonian Public Broadcasting service in English, English-language monthly newspaper "The Baltic Times", "Baltic News Network", and others hereinafter referred to as "Other Baltics". To find these, we refer to the previously mentioned news aggregator "Google News".

In order to find relevant content, we had to select the most precise keywords to direct search engines towards news that might contain content useful for this analysis. We use the following combinations to find articles about the 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 "Baltic property prices".

The best way to see what data we have is to plot it. Figure 2, presents the values for news sentiments by country: Sweden, Latvia, Lithuania, and Estonia, and also news published about the Baltic market. Two horizontal lines (y = 0.05 and y = -0.05) are added to show how many news sentiment values are positive (above y = 0.5 line), negative (below y = -0.05 line), and how many are neutral (between these two horizontal lines). A more detailed explanation about sentiment scores is provided in section 3.

More positive Swedish news is provided by news portals other than "Reuters" or "The Local SE," but the difference is not considerable – all positive news are distributed almost equally. More negative Swedish news is provided by "Reuters", and the distribution is not as equal as for the positive news.

For Baltic news, we cannot divide the results by news portals due to the difficulty in

[8] See Appendix 1 for a list of all sources used

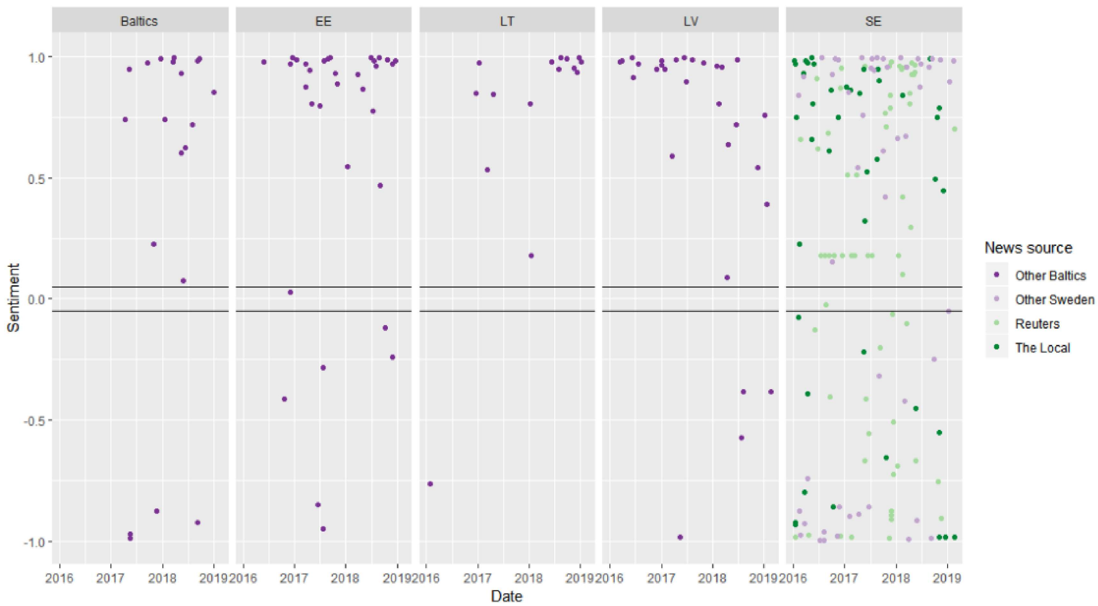


Figure 2: Real estate news sentiments by sources and countries

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. Source: authors' calculations.

finding such resources with sufficient news about the real estate market in 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 the Baltic countries in general ("Baltics" in Figure 2), and also about Estonia and Latvia.

It is also interesting to observe the difference in sentiments' distribution of Swedish news, and also for all Baltic countries news without dividing by country (Figure 3). 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 one reason why neutral sentiments were discarded for the estimation models, and we only considered negative and positive sentiments.

From the histogram of news sentiments (Figure 4), we can see how sentiments are distributed overall. It is clear that positive sentiments prevail, most of the negative sentiments are in the range: $(-0.8, -1)$, most of the positives are in the range: $(0.9, 1)$.

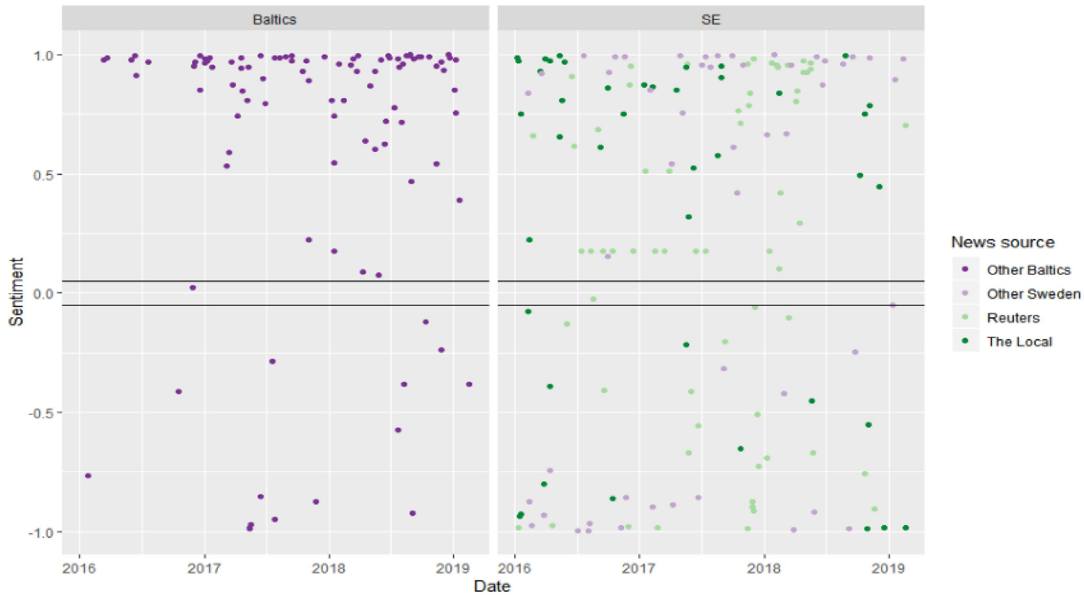


Figure 3: Real estate news sentiments by source and countries

Notes: This figure plots real estate news sentiments by news source and country. 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, SE - Sweden. Period: 04.01.2016 - 19.02.2019. Source: authors' calculations.

To have a better overview of the data we scraped, we provide its descriptive analysis in the Table 1. Statistical values confirm the conclusions made during the visual analysis conducted previously.

The extreme values of the sentiments are very close to the maximum possible values of the positive and negative sentiments. The mean value for news sentiments varies from 0.43 to 0.75 for news in the Baltic states. For Sweden, this value is much less indicating that the Baltic region is more optimistic about their real estate market. But we should take into account that the standard deviation is big enough for all countries and represents that sentiment fluctuations are high, and it is not a stable value. If we will look at the skewness of the sentiment data for all states and apply the rule of thumb, we see that Baltic data is highly negatively skewed, which means that most of the sentiments presented in the data set for these countries are above the average value - mostly positive. Only for Swedish data, can we state that the data is fairly symmetrical, as the skewness is closer to zero (-0.53), which suggests that data is equally distributed. The second value to pay attention to is kurtosis

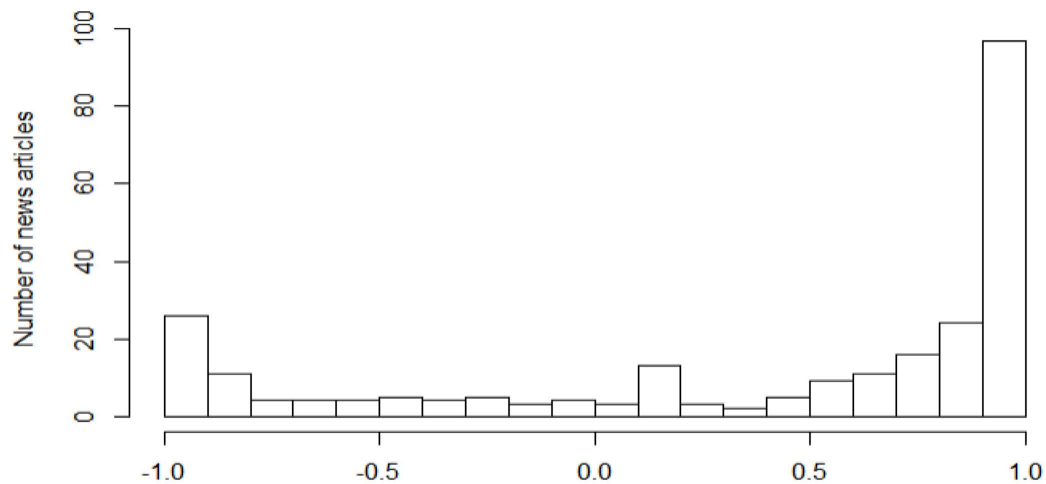


Figure 4: Histogram of news sentiments

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

- a way to notice the outliers in the distribution of data. We can observe a leptokurtic distribution only for Lithuanian news. Any value above 3 (4.09 for Lithuania) indicates that we have to deal with heavy-tailed data, and we might have a large number of outliers. Regarding other kurtosis values, all of them are platykurtic or simply, the data is light-tailed, and the outliers, which also can be present in data set are smaller than those of the normal distribution.

When reviewing the 10 most positive news listed in the Table 2, it can be seen that half of them were published during 2018. And the amount of news describing the Baltic and Swedish real estate market is the same. But taking into account that the data set of Swedish news sentiment is much larger in comparison with Baltic countries, we can state that the news in Baltic countries is more positive in comparison to Sweden. The same can be said when taking into account the mean values of sentiments, the Swedish mean is the lowest among all listed countries (Table 1).

The most negative sentiments are observable mostly for Sweden for different years (Table 3). Furthermore, few Baltic news is also placed at the top of the negatives, but among them, we see only news related to the Latvian or the Baltic market in general.

As a visual representation of the most frequent words in the news, and an overall understanding of the main topics covered in positive and negative news, we have used the word

Table 1: Descriptive statistics of news sentiments 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. Source: authors' calculations.

Table 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. It also shows the name of the country (*States* column) and the region to which the particular news refers. Source: authors' calculations.

Table 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. It also shows the name of the country (*States* column) and the region to what particular news refers. Source: authors' calculations.

clouds shown in Figure 5. The most frequently used words have larger font size and are placed closer to the middle part of the word cloud[9]. It appears that the most positive news contained keywords related to housing, shopping, city, center and macroeconomic terms such as growth, investment and prices, while the most negative news contained keywords such as housing, banks, Sweden, Swedish and prices. This word cloud could suggest that negative news in the Baltics and Sweden concentrates on real estate and banking in Sweden.

4.2 Stocks prices data

To analyse the stock price time series, we use a daily series of adjusted closing stock prices from Swedbank and SEB banks from 04.01.2016 to 19.02.2019. The adjusted closing stock price is the price of last stock traded on a particular day which was adjusted according to the relevant split and dividend paid out to investors (Balasubramaniam, 2018). Historical stock prices data is downloaded directly from "Yahoo! Finance", and the quality of data is assured because it is not stored locally but loaded directly from the web source to the R software[10].

To see how stock prices vary in time, we can look at Figure 6, which presents the adjusted daily closing stock prices for Swedbank and SEB banks for the period: 04.01.2016 -

[9]Word clouds are not used for analysis and are illustrated only for visualization purposes
[10]https://www.r-project.org/

(a) 10 most positive news

(b) 10 most negative news

Figure 5: 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. The word clouds are illustrated only for visualization purposes. Source: authors' calculations.

19.02.2019, and shows that there are some similarities in the behavior of bank stock prices.

An uptrend in Swedbank stock prices is observable after a considerable downturn at the beginning of 2016, albeit with little certainty. For SEB Bank, a decline was observed mostly from March 2016. Similarly for both banks, from July 2016 till March 2017, prices climbed steadily, and remained almost the same with moderate fluctuations till October 2017. After which, Swedbank stock prices fell to a level slightly above 160 SEK per share and remained within the range of 160–175 SEK till July 2018. For SEB Bank, the situation was less positive – falls were present from autumn 2017 till the beginning of June 2018. However, the stocks for both banks jumped considerably around summer and mid-autumn 2018. After which, ups and downs were observable for both financial market leaders.

We calculate daily returns of selected stocks to obtain returns series for the selected stocks. Returns can be defined by

$$r_t = ln\Big(\frac{P_t}{P_{t-1}}\Big) \times 100\% \tag{13}$$

Figure 7 shows the daily stock price returns for 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

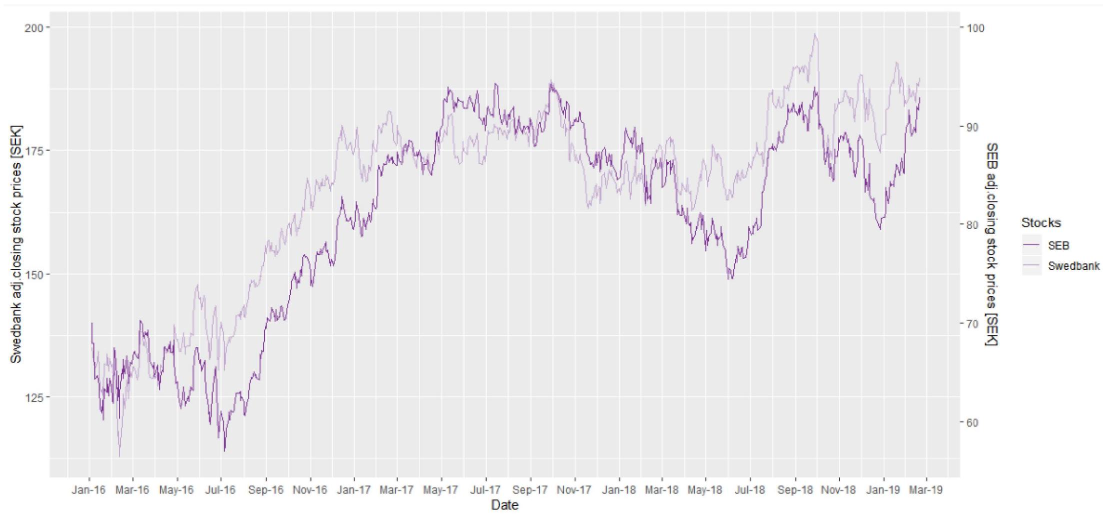


Figure 6: Adjusted daily closing stock prices

Notes: This figure plots adjusted daily closing stock prices for Swedbank and SEB banks measured in SEK, for the period 04.01.2016 - 19.02.2019. Source: "Yahoo! Finance".

because volatility clustering is visible. And, this causes us to 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.

From the skewness and kurtosis, we see that the distribution is not close to normal (skewness is -0.67 and -0.68 and kurtosis is 6.42 and 9.49 for Swedbank and SEB Bank, respectively). Returns data is positively skewed: the right tail is larger than the left. And due to high kurtosis, the distribution is fat-tailed.

Table 4: Descriptive statistics of daily stock returns

Stocks	N	Mean	St.dev	Median	Min	Max	Skew	Kur	St.err
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 for 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, Kur - kurtosis of the data, St.err - standard error. Source: authors' calculations and "Yahoo! Finance".

Figure 7: 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, as a %. Source: authors' calculations and "Yahoo! Finance"

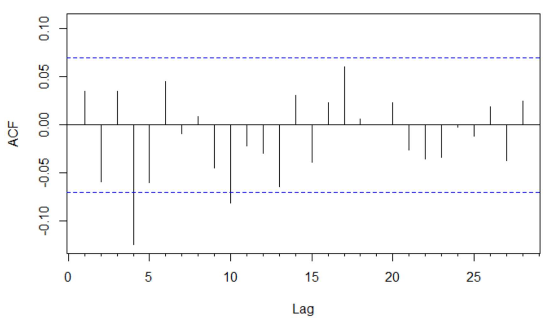
In order to examine the time series and prepare for the modeling, we check for stationarity and make assumptions regarding possible ARMA (p,q) model order. The first test for stationarity is the augmented Dickey-Fuller (ADF) test. From Table 5, based on the null hypothesis, we state that the series is not stationary. For Swedbank and SEB Bank data, the null hypothesis that the data is not stationary can be rejected. Also, from the Figure 8, we can see the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, which allow us to determine the order of the MA(q) and AR(p) terms, respectively.

From Figure 8a, we see that there are significant autocorrelations at lags 4 and 10 for Swedbank. From the partial correlation plot in Figure 8b, we see an increase at lags 4, 10 and 13. For the SEB Bank data, we see significant autocorrelations at lags 1, 4, 10 and 16 (Figure 9a). From Figure 9b, we see an increase at lags 1, 4, 10 and 16. When choosing the ARMA orders for Swedbank and SEB Bank returns, we took into account these lags.

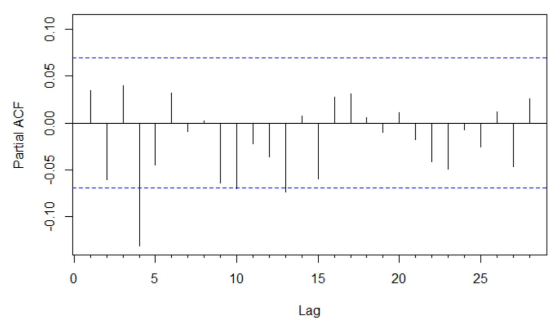
Table 5: Augmented Dickey-Fuller Test for stationarity

	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 Bank 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 resulting p-value, we conclude that null-hypothesis about non-stationary time series can be rejected at a significance level of 0.05. Source: authors' calculations.



(a) ACF plot



(b) PACF plot

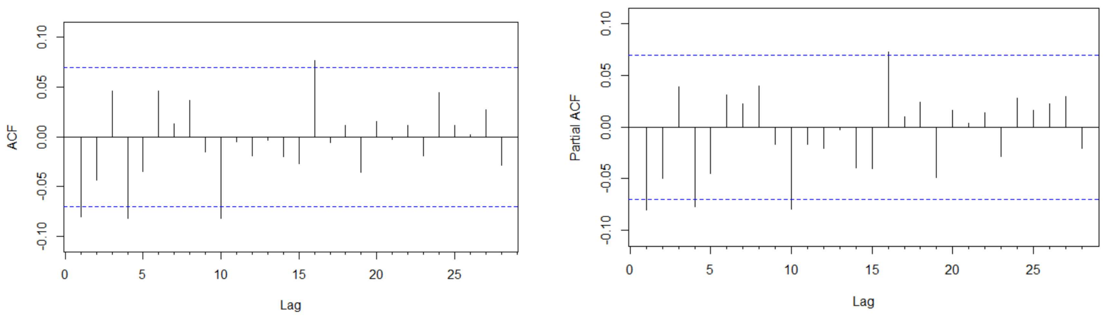
Figure 8: ACF and PACF plots of Swedbank's daily stock returns data

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

5 Results

The news vectors contain many zeros and relatively less non-zero values. Moreover, they are not autocorrelated, [11] and therefore cannot explain the autocorrelation structure in the return series. Based on the AIC criteria from the models which yielded serially uncorrelated

[11] Ljung-Box test was applied to the news vectors and there were no significant autocorrelations for any of the news vectors.



(a) ACF plot (b) PACF plot

Figure 9: 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 for SEB Bank's daily stock returns data. From the ACF, we can conclude that the returns are not highly correlated with its lagged values because most of the spikes are not statistically significant. The PACF plot shows that there is no considerable correlation between residuals and the next lag values. Source: authors' calculations.

residuals [12], we found that the AR(4) model is the most adequate for both Swedbank and SEB Bank returns. The other models with news sentiment variables are built on the AR(4) model. In order to avoid the trap of local optima during the estimations, we considered many starting values for the optimisation procedure. In what follows, the estimation results are reported in equation format where the parameter estimates are given as coefficients, the standard errors are given in parentheses and the p-values are given in brackets.

5.1 Estimation results for Swedbank's stock returns

Model 0. AR(4) model (base-line model)

The AR(4) model, given in equation 2 with p=4, presents the effect of stock return values for the past four periods on stock returns in the current period t.

The estimated model for the AR(4) model is given in equation 14.

$$\hat{r}_t = \underset{\substack{(0.0035)\\ [0.0000]}}{0.0489} + \underset{\substack{(0.0107)\\ [0.0001]}}{0.0422}\, r_{t-1} - \underset{\substack{(0.0213)\\ [0.0011]}}{0.0692}\, r_{t-2} + \underset{\substack{(0.0127)\\ [0.0002]}}{0.0475}\, r_{t-3} - \underset{\substack{(0.0272)\\ [0.0000]}}{0.1299}\, r_{t-4} \tag{14}$$

$$LL = -1277.52,\ AIC = 2565,04,\ BIC = 2588.41,\ SSE = 1174.24$$

From p-values of the parameter estimates, we can conclude that AR components are mostly 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.

[12] The residuals of each model was tested for autocorrelation using the Ljung-Box test.

For the following models which extend the base model by including news sentiment variables, we perform likelihood ratio tests comparing them with the base model. Although the results suggest that the bigger model did not improve the fit significantly, the individual coefficients of interest are found to be significant.

Model 1. AR(4) model model with news component

The model AR(4) that contains news sentiment variables follows the equation 5, and the estimation results are given in equation 15:

$$\begin{aligned}
\hat{r}_t = {} & \underset{\substack{(0.0032)\\ [0.0062]}}{0.0087} + \underset{\substack{(0.0026)\\ [0.0000]}}{0.0350}\, r_{t-1} - \underset{\substack{(0.0133)\\ [0.0000]}}{0.0677}\, r_{t-2} + \underset{\substack{(0.0154)\\ [0.0010]}}{0.0509}\, r_{t-3} - \underset{\substack{(0.0283)\\ [0.0000]}}{0.1250}\, r_{t-4} \\
& + \underset{\substack{(0.0000)\\ [0.0643]}}{0.0000}\, Reuters_t + \underset{\substack{(0.3809)\\ [0.0954]}}{0.6350}\, TheLocal_t + \underset{\substack{(0.1176)\\ [0.0790]}}{0.2066}\, Others_t + \underset{\substack{(0.0000)\\ [0.0045]}}{0.0000}\, Baltics_t \\
& + \underset{\substack{(0.1510)\\ [0.0101]}}{0.3883}\, Reuters_{t-1} + \underset{\substack{(0.0000)\\ [0.0237]}}{0.0000}\, TheLocal_{t-1} + \underset{\substack{(0.0000)\\ [0.0007]}}{0.0000}\, Others_{t-1} + \underset{\substack{(0.0291)\\ [0.0133]}}{0.0720}\, Baltics_{t-1} \\
& + \underset{\substack{(0.1581)\\ [0.0527]}}{0.3064}\, Reuters_{t-2} + \underset{\substack{(0.1327)\\ [0.0626]}}{0.2471}\, TheLocal_{t-2} + \underset{\substack{(0.0000)\\ [0.0073]}}{0.0000}\, Others_{t-2} + \underset{\substack{(0.0505)\\ [0.0400]}}{0.1038}\, Baltics_{t-2}
\end{aligned} \tag{15}$$

$$LL = -1270.30,\ AIC = 2574.59,\ BIC = 2654.06,\ SSE = 1152.97,\ Pval\_LLratio = 0.2732$$

In equation 15, we can see that although the autoregressive parameter estimates did not change much, the intercept is smaller. The coefficients of news sentiment variables are significant. The coefficient estimates suggest that the location of news matter. The Local and Others are news sources from Sweden and they seem to affect the returns on the same day, while the effect of the former continues two days later. Reuters and Baltics news affect the returns the next day and the following day. These coefficients are significant at least at a significance level of 0.1.

Model 2. AR(4) model with asymmetric effect of news, neutral news discarded

The AR(4) model with asymmetric effect of news follows from equation 6 and the estimation results are given in equation 16:

$$
\begin{aligned}
\hat{r}_t = \; & \underset{\substack{(0.0016)\\[0.0000]}}{0.0343} + \underset{\substack{(0.0050)\\[0.0000]}}{0.0342}\; r_{t-1} - \underset{\substack{(0.0088)\\[0.0000]}}{0.0699}\; r_{t-2} + \underset{\substack{(0.0075)\\[0.0000]}}{0.0530}\; r_{t-3} - \underset{\substack{(0.0170)\\[0.0000]}}{0.1221}\; r_{t-4} \\
+ \; & \underset{\substack{(0.0000)\\[0.0006]}}{0.0000}\; Reuters_t^- + \underset{\substack{(0.1604)\\[0.0000]}}{0.8392}\; TheLocal_t^- + \underset{\substack{(0.1764)\\[0.0014]}}{0,5646}\; Others_t^- + \underset{\substack{(0.0230)\\[0.0000]}}{0.1399}\; Baltics_t^- \\
+ \; & \underset{\substack{(0.1441)\\[0.0692]}}{0.2618}\; Reuters_{t-1}^- + \underset{\substack{(0.1197)\\[0.0007]}}{0.4045}\; TheLocal_{t-1}^- + \underset{\substack{(0.0000)\\[0.0219]}}{0.0000}\; Others_{t-1}^- + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\; Baltics_{t-1}^- \\
+ \; & \underset{\substack{(0.0841)\\[0.0001]}}{0.3357}\; Reuters_{t-2}^- + \underset{\substack{(0.0040)\\[0.0000]}}{0.0163}\; TheLocal_{t-2}^- + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\; Others_{t-2}^- + \underset{\substack{(0.1379)\\[0.0000]}}{0.7156}\; Baltics_{t-2}^- \\
+ \; & \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\; Reuters_t^+ + \underset{\substack{(0.3061)\\[0.0703]}}{0.5539}\; TheLocal_t^+ + \underset{\substack{(0.0046)\\[0.0193]}}{0.0107}\; Others_t^+ + \underset{\substack{(0.0000)\\[0.0002]}}{0.0000}\; Baltics_t^+ \\
+ \; & \underset{\substack{(0.3732)\\[0.2155]}}{0.4622}\; Reuters_{t-1}^+ + \underset{\substack{(0.0000)\\[0.0122]}}{0.0000}\; TheLocal_{t-1}^+ + \underset{\substack{(0.0000)\\[0.0096]}}{0.0000}\; Others_{t-1}^+ + \underset{\substack{(0.0378)\\[0.0000]}}{0.1534}\; Baltics_{t-1}^+ \\
+ \; & \underset{\substack{(0.1393)\\[0.0381]}}{0.2888}\; Reuters_{t-2}^+ + \underset{\substack{(0.1638)\\[0.1006]}}{0.2689}\; TheLocal_{t-2}^+ + \underset{\substack{(0.0000)\\[0.0560]}}{0.0000}\; Others_{t-2}^+ + \underset{\substack{(0.0001)\\[0.0000]}}{0.0005}\; Baltics_{t-2}^+
\end{aligned}
\tag{16}
$$

$$LL = -1267.25,\; AIC = 2592.51,\; BIC = 2728.07,\; SSE = 1144.12,\; Pval\_LLratio = 0.6662$$

In equation 16, we can see that when the news are distinguished for negative and positive sentiments, interesting results emerge. It is observed that negative news from the Baltics affect the returns on the same day but the main influence comes with a two-day delay. This is a result that Model 1 could not capture. The negative news from Reuters affects returns with a delay of one and two days. The biggest effect of negative news comes from The Local and Others and the impact continues two days more in the case of The Local. In the case of The Local, Others and Baltics news, the total affect of negative returns is higher than that of the positive returns. Finally, the impact of positive news from The Local and from Reuters sources are relatively high, compared to the other news sources.

Model 3. AR(4) model with a naive threshold for extreme news

This model follows the equation 7 and the estimation results are given in equation 17:

$$
\begin{aligned}
\hat{r}_t = &- \underset{\substack{(0.0015)\\[0.0000]}}{0.0206} + \underset{\substack{(0.0028)\\[0.0000]}}{0.0369}\ r_{t-1} - \underset{\substack{(0.0044)\\[0.0000]}}{0.0628}\ r_{t-2} + \underset{\substack{(0.0040)\\[0.0000]}}{0.0521}\ r_{t-3} - \underset{\substack{(0.0102)\\[0.0000]}}{0.1229}\ r_{t-4} \\
&+ \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ Reuters_t + \underset{\substack{(0.1234)\\[0.0000]}}{0.6337}\ TheLocal_t + \underset{\substack{(0.0606)\\[0.0004]}}{0.2142}\ Others_t + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ Baltics_t \\
&+ \underset{\substack{(0.1101)\\[0.0000]}}{0.5020}\ Reuters_{t-1} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ TheLocal_{t-1} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ Others_{t-1} + \underset{\substack{(0.0093)\\[0.0000]}}{0.0936}\ Baltics_{t-1} \\
&+ \underset{\substack{(0.0987)\\[0.0010]}}{0.3257}\ Reuters_{t-2} + \underset{\substack{(0.0606)\\[0.0000]}}{0.3398}\ TheLocal_{t-2} + \underset{\substack{(0.0063)\\[0.0188]}}{0.0148}\ Others_{t-2} + \underset{\substack{(0.0225)\\[0.0000]}}{0.1044}\ Baltics_{t-2} \\
&+ \underset{\substack{(0.0078)\\[0.0000]}}{0.0344}\ I_t^{Reuters,-e} + \underset{\substack{(0.0000)\\[0.0003]}}{0.0000}\ I_t^{TheLocal,-e} + \underset{\substack{(0.0000)\\[0.0009]}}{0.0000}\ I_t^{Others,-e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ I_t^{Baltics,-e} \\
&+ \underset{\substack{(0.0856)\\[0.0013]}}{0.2758}\ I_{t-1}^{Reuters,-e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ I_{t-1}^{TheLocal,-e} + \underset{\substack{(0.0328)\\[0.0000]}}{0.1435}\ I_{t-1}^{Others,-e} + \underset{\substack{(0.0500)\\[0.0006]}}{0.1709}\ I_{t-1}^{Baltics,-e} \\
&+ \underset{\substack{(0.0147)\\[0.0928]}}{0.0247}\ I_{t-2}^{Reuters,-e} + \underset{\substack{(0.1514)\\[0.0677]}}{0.2765}\ I_{t-2}^{TheLocal,-e} + \underset{\substack{(0.1330)\\[0.0001]}}{0.5177}\ I_{t-2}^{Others,-e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ I_{t-2}^{Baltics,-e} \\
&+ \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ I_t^{Reuters,+e} + \underset{\substack{(0.0000)\\[0.0023]}}{0.0000}\ I_t^{TheLocal,+e} + \underset{\substack{(0.0000)\\[0.0001]}}{0.0000}\ I_t^{Others,+e} + \underset{\substack{(0.0038)\\[0.0001]}}{0.0148}\ I_t^{Baltics,+e} \\
&+ \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ I_{t-1}^{Reuters,+e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ I_{t-1}^{TheLocal,+e} + \underset{\substack{(0.0000)\\[0.0037]}}{0.0000}\ I_{t-1}^{Others,+e} + \underset{\substack{(0.0000)\\[0.0008]}}{0.0000}\ I_{t-1}^{Baltics,+e} \\
&+ \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ I_{t-2}^{Reuters,+e} + \underset{\substack{(0.0000)\\[0.0014]}}{0.0000}\ I_{t-2}^{TheLocal,+e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\ I_{t-2}^{Others,+e} + \underset{\substack{(0.0000)\\[0.0038]}}{0.0000}\ I_{t-2}^{Baltics,+e}
\end{aligned} \tag{17}
$$

$$LL = -1268.51,\ AIC = 2619.03,\ BIC = 2810.68,\ SSE = 1147.78,\ Pval\_LLratio = 0.9946$$

We can see from the results in equation 17 that although the coefficient estimates for the news variables changed, the results are similar to Model 1. It is observed there are many coefficients whose values are very close to zero. The results suggest that negative extreme returns from Reuters affect the returns the same day but more strongly in the next period. Surprisingly, the negative extreme returns from The Local and Others affect the returns with a delay of one or two days. Negative extreme news from the Baltics affects the returns with one day delay. The only significant coefficient for the positive extreme news belonged to the Baltic news source. It is interesting that the returns process incorporates the effect of negative extreme news with a delay although negative news affects the returns on the same day.

Model 4. AR(4) model with merged news data (positive and negative news distinguished)

The AR(4) model with merged news comes from equation 8 and the estimation results are given in equation 18:

$$
\begin{aligned}
\hat{r}_t = \underset{\substack{(0.0064)\\ [0.0000]}}{0.0500} &+ \underset{\substack{(0.0015)\\ [0.0000]}}{0.0386}\ r_{t-1} - \underset{\substack{(0.0051)\\ [0.0000]}}{0.0657}\ r_{t-2} + \underset{\substack{(0.0045)\\ [0.0000]}}{0.0461}\ r_{t-3} - \underset{\substack{(0.0042)\\ [0.0000]}}{0.1273}\ r_{t-4} \\
&+ \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000}\ News_t^{pos} + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000}\ News_{t-1}^{pos} + \underset{\substack{(0.0000)\\ [0.0004]}}{0.0000}\ News_{t-2}^{pos} \\
&+ \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000}\ News_t^{neg} + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000}\ News_{t-1}^{neg} + \underset{\substack{(0.0337)\\ [0.0011]}}{0.1102}\ News_{t-2}^{neg} \\
&+ \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000}\ N_t^{pos} + \underset{\substack{(0.0084)\\ [0.0000]}}{0.0865}\ N_{t-1}^{pos} + \underset{\substack{(0.0369)\\ [0.0093]}}{0.0959}\ N_{t-2}^{pos} \\
&- \underset{\substack{(0.1930)\\ [0.0453]}}{0.3865}\ N_t^{neg} - \underset{\substack{(0.0200)\\ [0.0629]}}{0.0372}\ N_{t-1}^{neg} - \underset{\substack{(0.0000)\\ [0.0423]}}{0.0000}\ N_{t-2}^{neg}
\end{aligned}
\tag{18}
$$

$$LL = -1272.95,\ AIC = 2579.89,\ BIC = 2659.36,\ SSE = 1160.73, Pval\_LLratio = 0.6907$$

The results in equation 18 interestingly suggest that in this model the content of the news (positive or negative) matters only in the case of negative news and with a delay of two days. It seems that the return process is more concerned with the number of news than the content of news. The impact of one more negative news, even if its size is small, affects the returns largely on the same day and to some extent on the next day. This implies that when there is large amount of negative news from different sources on the same day, the returns are pulled down to a large extent. In contrast, if there is only one source providing news whose size might be large (such as speculative or fake news), the effect on the returns is relatively smaller. Typically, speculative or fake news do not appear in many sources on the same day. In this model, it seems that the return process takes into account the probability that the news might be fake. Finally, the effect of the number of positive news affect the returns with delays of one and two days.

Model 0. GARCH(1,1) model (base line model)

The model follows from equation 9 and estimation results are given in 19:

$$\hat{h}_t = \underset{\substack{(0.0149)\\ [0.0000]}}{0.0791} + \underset{\substack{(0.0170)\\ [0.0000]}}{0.1087}\ \varepsilon_{t-1}^2 + \underset{\substack{(0.0174)\\ [0.0000]}}{0.8376}\ h_{t-1} \tag{19}$$

$$LL = -1211.87,\ AIC = 2429.73,\ BIC = 2443.75$$

The results in equation 19 suggest typical coefficients for the GARCH(1,1) processes of stock returns. The persistence (*i.e.* $\alpha_1 + \alpha_2$) is quite high *approx.*0.94, indicating that the effect of large shocks on the volatility lasts.

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

The model follows the equation 10 and the estimation results are given in equation 20:

$$
\begin{aligned}
\hat{h}_t = {} & \underset{[0.0000]}{\underset{(0.0115)}{0.0788}} + \underset{[0.0000]}{\underset{(0.0198)}{0.1196}}\,\varepsilon^2_{t-1} + \underset{[0.0000]}{\underset{(0.0186)}{0.8045}}\,h_{t-1} \\
& + \underset{[0.0488]}{\underset{(0.0000)}{0.0000}}\,Reuters^-_t + \underset{[0.0000]}{\underset{(0.0000)}{0.0000}}\,TheLocal^-_t + \underset{[0.0000]}{\underset{(0.0175)}{0.1649}}\,Others^-_t + \underset{[0.0463]}{\underset{(0.0000)}{0.0000}}\,Baltics^-_t \\
& + \underset{[0.2335]}{\underset{(0.0000)}{0.0000}}\,Reuters^-_{t-1} + \underset{[0.0000]}{\underset{(0.0501)}{0.1683}}\,TheLocal^-_{t-1} + \underset{[0.0000]}{\underset{(0.0121)}{0.0868}}\,Others^-_{t-1} + \underset{[0.0000]}{\underset{(0.0344)}{0.5988}}\,Baltics^-_{t-1} \\
& + \underset{[0.0063]}{\underset{(0.0000)}{0.0000}}\,Reuters^-_{t-2} + \underset{[0.0060]}{\underset{(0.0000)}{0.0000}}\,TheLocal^-_{t-2} + \underset{[0.0000]}{\underset{(0.0087)}{0.0559}}\,Others^-_{t-2} + \underset{[0.0068]}{\underset{(0.0000)}{0.0000}}\,Baltics^-_{t-2} \\
& + \underset{[0.0064]}{\underset{(0.0000)}{0.0000}}\,Reuters^+_t + \underset{[0.0000]}{\underset{(0.0018)}{0.0344}}\,TheLocal^+_t + \underset{[0.0000]}{\underset{(0.0000)}{0.0000}}\,Others^+_t + \underset{[0.0000]}{\underset{(0.0074)}{0.1162}}\,Baltics^+_t \\
& + \underset{[0.0000]}{\underset{(0.0614)}{0.4261}}\,Reuters^+_{t-1} + \underset{[0.0000]}{\underset{(0.0536)}{0.4408}}\,TheLocal^+_{t-1} + \underset{[0.0000]}{\underset{(0.0000)}{0.0000}}\,Others^+_{t-1} + \underset{[0.0000]}{\underset{(0.0027)}{0.0283}}\,Baltics^+_{t-1} \\
& + \underset{[0.0001]}{\underset{(0.0000)}{0.0000}}\,Reuters^+_{t-2} + \underset{[0.0000]}{\underset{(0.0000)}{0.0000}}\,TheLocal^+_{t-2} + \underset{[0.0000]}{\underset{(0.0021)}{0.0329}}\,Others^+_{t-2} + \underset{[0.0000]}{\underset{(0.0071)}{0.0965}}\,Baltics^+_{t-2}
\end{aligned}
\tag{20}
$$

$$LL = -1206.07,\ AIC = 2466.15,\ BIC = 2592.36,\ Pval\_LLratio = 0.9842$$

The results in equation 20 suggest that negative news from Others affect the conditional volatility on the same day and over the next two days. The Local and Baltic negative news affect the conditional volatility with a delay of one day and the affect of Baltic news is quite large. There is no effect of negative news from Reuters on volatility. When we look at the positive news, the story is different. Reuters and The Local positive news affect volatility largely on the following day. Positive news from Baltic sources affects on the same day and the next two days.

Model 2. GJR-GARCH(1,1) model

The model follows the equation 12 and the results are given in equation 21:

$$
\hat{h}_t = \underset{[0.0000]}{\underset{(0.0134)}{0.0988}} + \Big(\underset{[0.0000]}{\underset{(0.0016)}{0.0312}} + \underset{[0.0000]}{\underset{(0.0144)}{0.1425}}\,I_{t-1} \Big)\varepsilon^2_{t-1} + \underset{[0.0000]}{\underset{(0.0160)}{0.8263}}\,h_{t-1}
\tag{21}
$$

$$LL = -1207.14,\ AIC = 2422.28,\ BIC = 2440.98,\ Pval\_LLratio = 0.0021$$

Model 3. NAGARCH(1,1) model

The model follows the equation 12 and the estimation results are given in equation 22:

$$\hat{h}_t = \underset{\substack{(0.0245)\\ [0.0017]}}{0.0766} + \underset{\substack{(0.0124)\\ [0.0000]}}{0.1036} \left(\varepsilon_{t-1} - \underset{\substack{(0.0765)\\ [0.0000]}}{0.4784}\ h_{t-1}^{1/2}\right)^2 + \underset{\substack{(0.0166)\\ [0.0000]}}{0.8226}\ h_{t-1} \tag{22}$$

$$LL = -1207.92,\ AIC = 2423.84,\ BIC = 2442.54,\ Pval\_LLratio = 0.0050$$

Although the results in equation 20 suggest that negative and positive news explain the asymmetric effects[13] in the volatility processes to some extent, the GJR-GARCH and NA-GARCH models in equations 21 and 22 capture the asymmetry with only one additional parameter and improve the estimation greatly. Potentially, this could mean that the asymmetric effects are mostly generated by the idiosyncratic shocks to the Swedbank returns though the error term, rather than the news sentiments.

5.2 Estimation results for SEB Bank's stock returns.

Model 0. AR(4) model (base-line model)

Similar to the case of Swedbank stock returns, the model equation for SEB Bank stock returns follow an AR(4), which presents the effect of stock return values of past four periods on stock return in the current period t.

The estimation results are given in equation 23:

$$\hat{r}_t = \underset{\substack{(0.0025)\\ [0.0000]}}{0.0547} - \underset{\substack{(0.0158)\\ [0.0000]}}{0.0811}\ r_{t-1} - \underset{\substack{(0.0148)\\ [0.0000]}}{0.0637}\ r_{t-2} + \underset{\substack{(0.0013)\\ [0.0000]}}{0.0295}\ r_{t-3} - \underset{\substack{(0.0028)\\ [0.0000]}}{0.0793}\ r_{t-4} \tag{23}$$

$$LL = -1354.36,\ AIC = 2718.72,\ BIC = 2742.09,\ SSE = 1426.41$$

From the p-values of the parameter estimates, we can conclude that AR components are mostly 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.

Model 1. AR(4) model with news component

The AR(4) model with news variables is given in equation 5 and the estimation results are given in equation 24:

[13]See the discussion in Section 3.3. for how news could explain asymmetric effects.

$$\begin{aligned}
\hat{r}_t = & \underset{\substack{(0.0025)\\ [0.0000]}}{0.0115} - \underset{\substack{(0.0202)\\ [0.0000]}}{0.0811} \; r_{t-1} - \underset{\substack{(0.0063)\\ [0.0000]}}{0.0623} \; r_{t-2} + \underset{\substack{(0.0028)\\ [0.0000]}}{0.0297} \; r_{t-3} - \underset{\substack{(0.0119)\\ [0.0000]}}{0.0798} \; r_{t-4} \\
& + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000} \; Reuters_t + \underset{\substack{(0.1331)\\ [0.0000]}}{0.5736} \; TheLocal_t + \underset{\substack{(0.1131)\\ [0.0000]}}{0.5816} \; Others_t + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000} \; Baltics_t \\
& + \underset{\substack{(0.1239)\\ [0.0316]}}{0.2663} \; Reuters_{t-1} + \underset{\substack{(0.0334)\\ [0.0320]}}{0.0716} \; TheLocal_{t-1} + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000} \; Others_{t-1} + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000} \; Baltics_{t-1} \\
& + \underset{\substack{(0.1119)\\ [0.2293]}}{0.1345} \; Reuters_{t-2} + \underset{\substack{(0.0136)\\ [0.0001]}}{0.0543} \; TheLocal_{t-2} + \underset{\substack{(0.0008)\\ [0.0702]}}{0.0015} \; Others_{t-2} + \underset{\substack{(0.0401)\\ [0.0000]}}{0.2323} \; Baltics_{t-2} \\
& LL = -1346.48,\; AIC = 2726.97,\; BIC = 2806.43,\; SSE = 1398.25,\; Pval_LLratio = 0.2028
\end{aligned} \tag{24}$$

Equation 24 suggests that the Swedish news from The Local and Others affect the returns the same day and the effect of The Local continues two more days. On the other hand, Reuters news affect the returns in the next day and Baltic news affect the returns in two days. The differences compared to the estimations with Swedbank returns are that (1) SEB Bank returns reacts slower to Baltic news: two days later, and (2) the SEB Bank returns seem to depend comparatively more on the Others news sources.

Model 2. AR(4) model with asymmetric effect of news, neutral news discarded

The model equation is provided in equation 11 and the estimation results are given in equation 25:

$$
\begin{aligned}
\hat{r}_t = & \underset{\substack{(0.0084)\\[0.0000]}}{0.0385} - \underset{\substack{(0.0059)\\[0.0000]}}{0.0765}\, r_{t-1} - \underset{\substack{(0.0053)\\[0.0000]}}{0.0659}\, r_{t-2} + \underset{\substack{(0.0053)\\[0.0000]}}{0.0333}\, r_{t-3} - \underset{\substack{(0.0090)\\[0.0000]}}{0.0799}\, r_{t-4} \\
& + \underset{\substack{(0.0730)\\[0.0185]}}{0.1720}\, Reuters_t^- + \underset{\substack{(0.2124)\\[0.0215]}}{0.4885}\, TheLocal_t^- + \underset{\substack{(0.2637)\\[0.0001]}}{1.0128}\, Others_t^- + \underset{\substack{(0.0071)\\[0.0000]}}{0.0524}\, Baltics_t^- \\
& + \underset{\substack{(0.0039)\\[0.0403]}}{0.0080}\, Reuters_{t-1}^- + \underset{\substack{(0.2538)\\[0.0077]}}{0.6765}\, TheLocal_{t-1}^- + \underset{\substack{(0.0000)\\[0.0003]}}{0.0000}\, Others_{t-1}^- + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, Baltics_{t-1}^- \\
& + \underset{\substack{(0.1160)\\[0.1074]}}{0.1868}\, Reuters_{t-2}^- + \underset{\substack{(0.0329)\\[0.2076]}}{0.0414}\, TheLocal_{t-2}^- + \underset{\substack{(0.0491)\\[0.0018]}}{0.1533}\, Others_{t-2}^- + \underset{\substack{(0.3414)\\[0.0174]}}{0.8116}\, Baltics_{t-2}^- \\
& + \underset{\substack{(0.0000)\\[0.0001]}}{0.0000}\, Reuters_t^+ + \underset{\substack{(0.2742)\\[0.0235]}}{0.6210}\, TheLocal_t^+ + \underset{\substack{(0.1835)\\[0.0526]}}{0.3555}\, Others_t^+ + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, Baltics_t^+ \\
& + \underset{\substack{(0.1743)\\[0.0327]}}{0.3723}\, Reuters_{t-1}^+ + \underset{\substack{(0.0000)\\[0.0033]}}{0.0000}\, TheLocal_{t-1}^+ + \underset{\substack{(0.0451)\\[0.1209]}}{0.0700}\, Others_{t-1}^+ + \underset{\substack{(0.0573)\\[0.0577]}}{0.1087}\, Baltics_{t-1}^+ \\
& + \underset{\substack{(0.0287)\\[0.0001]}}{0.1101}\, Reuters_{t-2}^+ + \underset{\substack{(0.0000)\\[0.0021]}}{0.0000}\, TheLocal_{t-2}^+ + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, Others_{t-2}^+ + \underset{\substack{(0.0738)\\[0.0722]}}{0.1326}\, Baltics_{t-2}^+
\end{aligned}
\tag{25}
$$

$LL = -1343.27,\ AIC = 2744.53,\ BIC = 2880.09,\ SSE = 1386.91,\ Pval\_LLratio = 0.5681$

We can see in equation 25 that When positive and negative news are considered separately, we can see that negative news from Reuters and Baltics sources affect the returns significantly in the same day, while the large effect of the latter comes with two days lag. Interestingly the SEB Bank returns are influenced by the negative news from Reuters on the same day, which is not the case of Swedbank returns. Both negative and positive news from the Others news sources affect the SEB Bank returns more compared to the Swedbank returns.

Model 3. AR(4) model with a naive threshold for extreme news

The AR(4) model with a threshold for extreme news is given in equation 7 and the estimation results are provided in equation 26:

$$
\begin{aligned}
\hat{r}_t = {} & \underset{\substack{(0.0013)\\[0.0000]}}{-0.0194} - \underset{\substack{(0.0062)\\[0.0000]}}{0.0765}\, r_{t-1} - \underset{\substack{(0.0033)\\[0.0000]}}{0.0617}\, r_{t-2} + \underset{\substack{(0.0026)\\[0.0000]}}{0.0302}\, r_{t-3} - \underset{\substack{(0.0074)\\[0.0000]}}{0.0804}\, r_{t-4} \\
& + \underset{\substack{(0.0000)\\[0.0012]}}{0.0000}\, Reuters_t + \underset{\substack{(0.0983)\\[0.0000]}}{0.4819}\, TheLocal_t + \underset{\substack{(0.2666)\\[0.0254]}}{0.5958}\, Others_t + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, Baltics_t \\
& + \underset{\substack{(0.2442)\\[0.0831]}}{0.4232}\, Reuters_{t-1} + \underset{\substack{(0.0677)\\[0.1358]}}{0.1009}\, TheLocal_{t-1} + \underset{\substack{(0.0000)\\[0.0002]}}{0.0000}\, Others_{t-1} + \underset{\substack{(0.0100)\\[0.0000]}}{0.0459}\, Baltics_{t-1} \\
& + \underset{\substack{(0.0417)\\[0.0009]}}{0.1383}\, Reuters_{t-2} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, TheLocal_{t-2} + \underset{\substack{(0.0056)\\[0.1632]}}{0.0079}\, Others_{t-2} + \underset{\substack{(0.0697)\\[0.0018]}}{0.2181}\, Baltics_{t-2} \\
& + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, I_t^{Reuters,-e} + \underset{\substack{(0.0000)\\[0.3170]}}{0.0000}\, I_t^{TheLocal,-e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, I_t^{Others,-e} + \underset{\substack{(0.0000)\\[0.0249]}}{0.0000}\, I_t^{Baltics,-e} \\
& + \underset{\substack{(0.2353)\\[0.0876]}}{0.4019}\, I_{t-1}^{Reuters,-e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, I_{t-1}^{TheLocal,-e} + \underset{\substack{(0.0616)\\[0.0000]}}{0.3058}\, I_{t-1}^{Others,-e} + \underset{\substack{(0.0686)\\[0.0000]}}{0.5375}\, I_{t-1}^{Baltics,-e} \\
& + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, I_{t-2}^{Reuters,-e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, I_{t-2}^{TheLocal,-e} + \underset{\substack{(0.0000)\\[0.1264]}}{0.0000}\, I_{t-2}^{Others,-e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, I_{t-2}^{Baltics,-e} \\
& + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, I_t^{Reuters,+e} + \underset{\substack{(0.0149)\\[0.0000]}}{0.0856}\, I_t^{TheLocal,+e} + \underset{\substack{(0.0000)\\[0.1192]}}{0.0000}\, I_t^{Others,+e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, I_t^{Baltics,+e} \\
& + \underset{\substack{(0.0000)\\[0.2786]}}{0.0000}\, I_{t-1}^{Reuters,+e} + \underset{\substack{(0.0000)\\[0.0002]}}{0.0000}\, I_{t-1}^{TheLocal,+e} + \underset{\substack{(0.0459)\\[0.0001]}}{0.1768}\, I_{t-1}^{Others,+e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, I_{t-1}^{Baltics,+e} \\
& + \underset{\substack{(0.0000)\\[0.0002]}}{0.0000}\, I_{t-2}^{Reuters,+e} + \underset{\substack{(0.0223)\\[0.0036]}}{0.0649}\, I_{t-2}^{TheLocal,+e} + \underset{\substack{(0.0000)\\[0.0000]}}{0.0000}\, I_{t-2}^{Others,+e} + \underset{\substack{(0.0000)\\[0.0001]}}{0.0000}\, I_{t-2}^{Baltics,+e}
\end{aligned}
\tag{26}
$$

$$LL = -1345.07,\ AIC = 2772.14,\ BIC = 2963.80,\ SSE = 1393.25,\ Pval\_LLratio = 0.9928$$

The estimation results in equation 26 suggest that the negative extreme news of Reuters and Baltics affect the SEB Bank returns in the following day while positive extreme news from the same sources do not have an effect. The extreme news from Others affect the SEB Bank returns with one day delay. These findings are similar to those of Swedbank's stock returns.

Model 4. AR(4) model with merged news data (positive and negative news distinguished)

The model follows the equation 8 and the estimation results are provided in equation 27:

$$
\begin{aligned}
\hat{r}_t = &\underset{\substack{(0.0311)\\ [0.0189]}}{0.0730} - \underset{\substack{(0.0040)\\ [0.0000]}}{0.0804}\, r_{t-1} - \underset{\substack{(0.0208)\\ [0.0024]}}{0.0631}\, r_{t-2} + \underset{\substack{(0.0111)\\ [0.0152]}}{0.0270}\, r_{t-3} - \underset{\substack{(0.0044)\\ [0.0000]}}{0.0809}\, r_{t-4} \\
&+ \underset{\substack{(0.0000)\\ [0.0019]}}{0.0000}\, News_t^{pos} + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000}\, News_{t-1}^{pos} + \underset{\substack{(0.0000)\\ [0.0099]}}{0.0000}\, News_{t-2}^{pos} \\
&+ \underset{\substack{(0.0000)\\ [0.0629]}}{0.0000}\, News_t^{neg} + \underset{\substack{(0.0000)\\ [0.0935]}}{0.0000}\, News_{t-1}^{neg} + \underset{\substack{(0.1628)\\ [0.0549]}}{0.3124}\, News_{t-2}^{neg} \\
&+ \underset{\substack{(0.0101)\\ [0.4083]}}{0.0083}\, N_t^{pos} + \underset{\substack{(0.0501)\\ [0.0109]}}{0.1278}\, N_{t-1}^{pos} + \underset{\substack{(0.0096)\\ [0.0000]}}{0.0573}\, N_{t-2}^{pos} \\
&- \underset{\substack{(0.2198)\\ [0.0199]}}{0.5116}\, N_t^{neg} - \underset{\substack{(0.0000)\\ [0.0109]}}{0.0000}\, N_{t-1}^{neg} - \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000}\, N_{t-2}^{neg}
\end{aligned}
\tag{27}
$$

$$LL = -1347.52,\ AIC = 2729.03,\ BIC = 2808.50,\ SSE = 1401.90, Pval\_LLratio = 0.3209$$

Similar to the case of Swedbank returns, the estimation results in equation 27 suggest that the size of the negative news affect the returns only two days later, however, the number of negative news in the media affect the returns on the same day. The effect of the number of positive news is distributed over the lags and not as large as that for negative news.

Model 0. GARCH(1,1) model (base-line model)

The model equation is given in equation 9 and the estimation results are provided in equation 28:

$$
\hat{h}_t = \underset{\substack{(0.0059)\\ [0.0000]}}{0.0354} + \underset{\substack{(0.0189)\\ [0.0000]}}{0.0924}\, \varepsilon_{t-1}^2 + \underset{\substack{(0.0109)\\ [0.0000]}}{0.8918}\, h_{t-1}
\tag{28}
$$

$$LL = -1271.18,\ AIC = 2548.35,\ BIC = 2562.38$$

The results in equation 28 suggest typical coefficients for the GARCH(1,1) processes of stock returns. The persistence, $\alpha_1 + \alpha_2$, is quite high *approx.* 0.98. This is higher than that of the Swedbank volatility equation estimate, indicating that the effect of large shocks on the volatility lasts longer.

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

The model follows the equation 10 and the estimates are given in equation 29:

$$\begin{aligned}
\hat{h}_t = \; & \underset{\substack{(0.0066)\\ [0.0008]}}{0.0219} + \underset{\substack{(0.0230)\\ [0.0002]}}{0.0859} \varepsilon^2_{t-1} + \underset{\substack{(0.0165)\\ [0.0000]}}{0.8974} h_{t-1} \\
& + \underset{\substack{(0.0000)\\ [0.0931]}}{0.0000} Reuters^-_t + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000} TheLocal^-_t + \underset{\substack{(0.0000)\\ [0.0003]}}{0.0000} Others^-_t + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000} Baltics^-_t \\
& + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000} Reuters^-_{t-1} + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000} TheLocal^-_{t-1} + \underset{\substack{(0.0444)\\ [0.0011]}}{0.1448} Others^-_{t-1} + \underset{\substack{(0.0000)\\ [0.2135]}}{0.0000} Baltics^-_{t-1} \\
& + \underset{\substack{(0.0011)\\ [0.0000]}}{0.0173} Reuters^-_{t-2} + \underset{\substack{(0.0000)\\ [0.0319]}}{0.0000} TheLocal^-_{t-2} + \underset{\substack{(0.0000)\\ [0.0165]}}{0.0000} Others^-_{t-2} + \underset{\substack{(0.0886)\\ [0.0000]}}{0.4605} Baltics^-_{t-2} \\
& + \underset{\substack{(0.0077)\\ [0.0000]}}{0.0884} Reuters^+_t + \underset{\substack{(0.0000)\\ [0.3392]}}{0.0000} TheLocal^+_t + \underset{\substack{(0.0000)\\ [0.0000]}}{0.0000} Others^+_t + \underset{\substack{(0.0458)\\ [0.0597]}}{0.0863} Baltics^+_t \\
& + \underset{\substack{(0.0048)\\ [0.0000]}}{0.0231} Reuters^+_{t-1} + \underset{\substack{(0.0000)\\ [0.0495]}}{0.0000} TheLocal^+_{t-1} + \underset{\substack{(0.0000)\\ [0.2252]}}{0.0000} Others^+_{t-1} + \underset{\substack{(0.0037)\\ [0.0006]}}{0.0125} Baltics^+_{t-1} \\
& + \underset{\substack{(0.0000)\\ [0.0808]}}{0.0000} Reuters^+_{t-2} + \underset{\substack{(0.0059)\\ [0.0000]}}{0.0448} TheLocal^+_{t-2} + \underset{\substack{(0.0000)\\ [0.0005]}}{0.0000} Others^+_{t-2} + \underset{\substack{(0.0113)\\ [0.0000]}}{0.1031} Baltics^+_{t-2} \\
& LL = -1263.82,\; AIC = 2581.64,\; BIC = 2707.86,\; Pval_LLratio = 0.9289
\end{aligned} \tag{29}$$

The results in equation 29 are similar to that of Swedbank in equation 20 except that the effect of negative news from Baltics on volatility is large but appears with a delay of two days. A similar result was found for SEB Bank returns, *i.e.* Model 2 results. The second largest effect belongs to negative news from the Others source with one day delay.

Model 2. GJR-GARCH(1,1) model

The GJR-GARCH model follows the equation 11 and the estimation results are in equation 30:

$$\hat{h}_t = \underset{\substack{(0.0049)\\ [0.0000]}}{0.0518} + \Big(\underset{\substack{(0.0172)\\ [0.0035]}}{0.0502} + \underset{\substack{(0.0192)\\ [0.0005]}}{0.0666} I_{t-1} \Big) \varepsilon^2_{t-1} + \underset{\substack{(0.0247)\\ [0.0000]}}{0.8832} h_{t-1} \tag{30}$$
$$LL = -1267.08,\; AIC = 2542.16,\; BIC = 2560.86,\; Pval\_LLratio = 0.0042$$

Model 3. NAGARCH(1,1) model

The model follows equation 12 and the estimation results are provided in equation 31

$$\hat{h}_t = \underset{\substack{(0.0042)\\ [0.0008]}}{0.0140} + \underset{\substack{(0.0131)\\ [0.0001]}}{0.0514} \left(\varepsilon_{t-1} - \underset{\substack{(0.2585)\\ [0.0013]}}{0.8319} \; h_{t-1}^{1/2}\right)^2 + \underset{\substack{(0.0348)\\ [0.0000]}}{0.9075} \; h_{t-1} \tag{31}$$

$$LL = -1263.23, \; AIC = 2534.47, \; BIC = 2553.17, \; Pval\_LLratio = 0.0001$$

Similar to the results with Swedbank returns in equations 21 and 22, it seems from 30 and 31 that although negative and positive news explain the asymmetric effects[14] in the volatility processes to some extent, the GJR-GARCH and NA-GARCH models capture the asymmetry with only one additional parameter and improve the estimation greatly. Although the GJR-GARCH asymmetry parameter estimate is larger for Swedbank returns, the NA-GARCH asymmetry parameter estimate is larger for SEB Bank returns. As noted earlier, this could mean that the asymmetric effects are mostly generated by the idiosyncratic shocks to the returns through the error term, rather than the news sentiments.

To aid with summarizing the findings, we present Figures 10 and 11. In these figures, the coefficient estimates in the models with news variables are presented in a grouped manner. The number of lags are given on the x-axes and the coefficient value is given on the y-axes. The legends are placed under the figures. The left column refers to the results of Swedbank, while the right column refers SEB Bank. The model coefficient estimates presented here are for:

- Model 1: news variables with lags
- Model 2: (a) negative news variables with lags, (b) positive news variables with lags
- Model 3: (a) indicator variables for negative extreme news with lags, (b) indicator variables for positive extreme news with lags
- Model 4: (a) negative and positive merged news variable, (b) negative and positive number of news variables with lags
- Model 1G: GARCH model where (a) negative news variables with lags, (b) positive news variables with lags.

For both banks, the local news from The Local and Others influence the returns mainly in the same day, while the international news from Reuters and Baltics has a delayed effect. The negative news from local sources affect the returns of both banks in the same day and next day, while the negative news from Reuters and Baltics affect the returns with one or two days lag. It is interesting that the effect of Baltics news comes in two days. When

[14]See the discussion in Section 3.3. for how news could explain asymmetric effects.

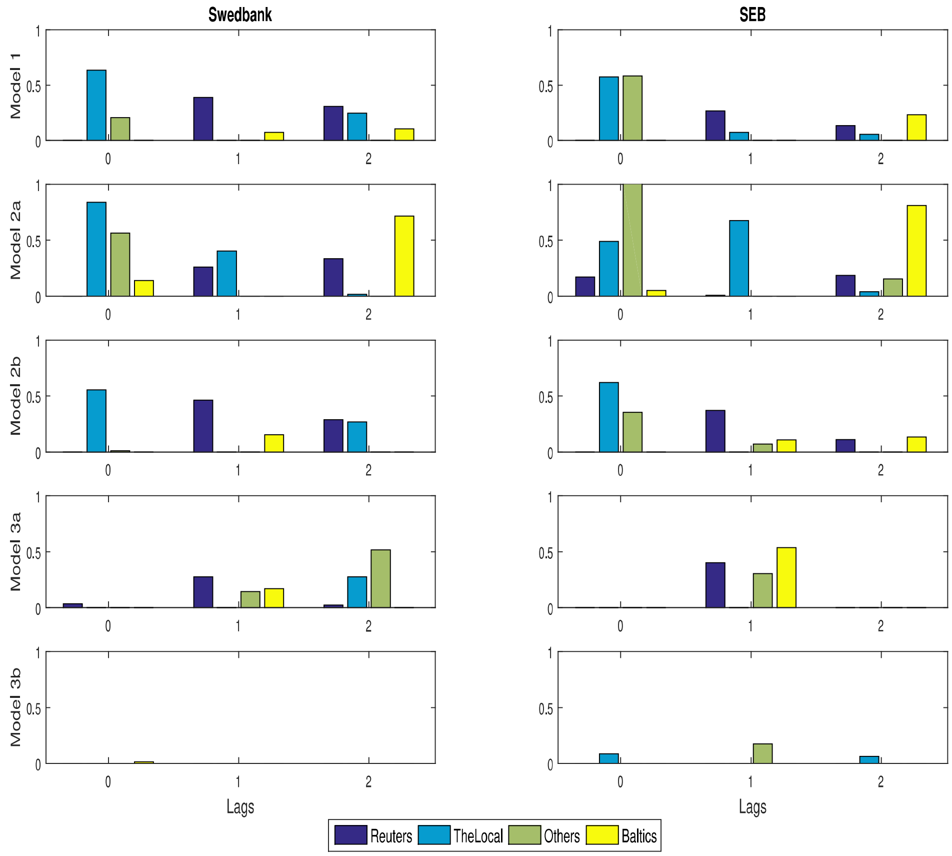


Figure 10: Summary of the coefficients of the news variables in model estimates

Notes: This figure shows the coefficient estimates in a bar chart. Model 2a refers to the coefficients of negative news, while Model 2b refers to the coefficients of positive news. Model 3a plots the coefficients of the indicator variables for negative extreme news, while Model 3B does the same for positive extreme news. Source: authors' calculations.

looking at the effect of positive news, it seems that Swedbank takes into account only The Local, while SEB Bank also takes into account the Others. While the negative extreme news affect Swedbank and SEB Bank returns in one or two days, there is very little effect from positive extreme news. When looking at the merged news data, we can say that the positive news do not influence the returns but negative news has some delayed effect. It turns out that the number of news has more effect than the strength of the news, The number of negative news hits the returns of both banks in the same day in a negative way. The number of positive news has some positive but delayed effect on the returns. When we look at the volatility estimation results, we can say that the effect of negative or positive news is large for Swedbank, but not as much for SEB Bank. For Swedbank negative news from Baltics or positive news from the Local and Reuters have a strong effect on volatility.

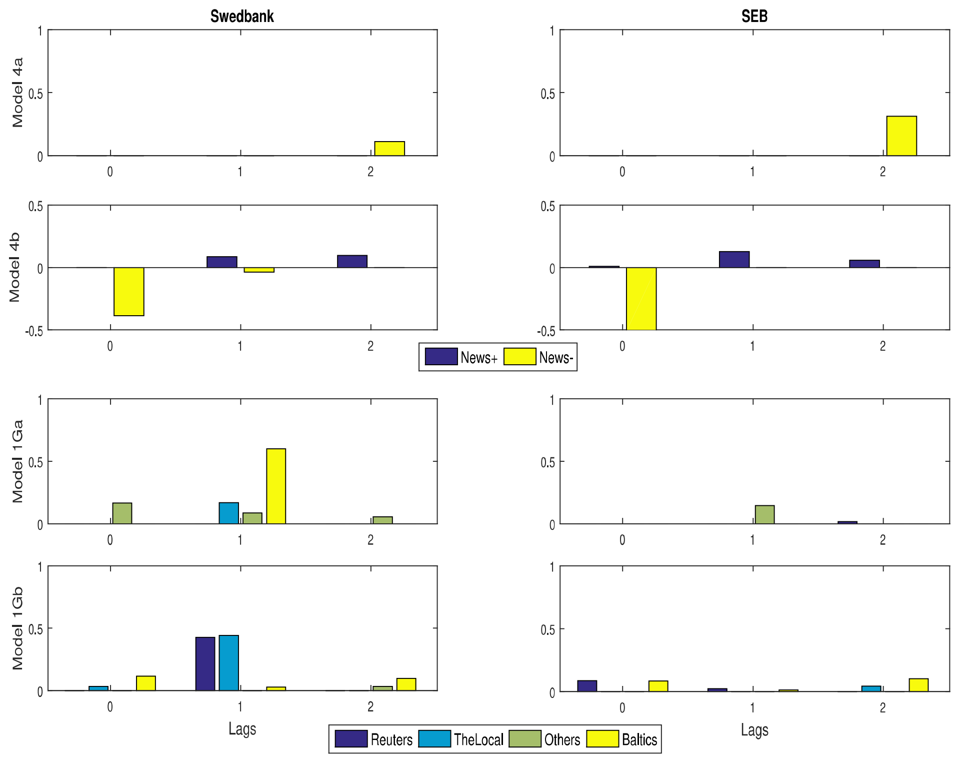


Figure 11: Summary of the coefficients of the news variables in model estimates

Notes: This figure shows the coefficient estimates in a bar chart. Model 4a plots the coefficients of merged positive and negative news, while Model 4B plots the coefficients of the variables: number of positive and negative news. Model 1Ga refers to the coefficients of negative news in the GARCH model, while Model 1Gb does that for the positive news. Source: authors' calculations.

6 Conclusions

In this paper, we have analysed the effect of the sentiment of news related to the real estate market in Sweden, Estonia, Latvia, and Lithuania, on the stock returns for Swedbank and SEB Bank. For this analysis, we considered the period from 04.01.2016 to 19.02.2019. First, we applied the Python open-source tools and libraries for web scraping to obtain text from news web pages, then used a rule-based sentiment analysis tool - VADER model - to compile a news sentiment time series. Subsequently, we used this data to estimate four ARMA models in which we consider several aspects such as extreme news, asymmetric effect of the news and content versus the number of news. We also investigated the effect of the sentiment of news on the volatilities to see if positive and negative news can generate different effects on volatility.

The main finding is that positive and negative news affect the returns and volatilities with different lags. For both banks, local Swedish news has more effect than the Baltic and Reuters news. The size of the negative news only affects the returns two days later; however, the number of negative news affects the returns on the same day. One possible interpretation for this is that the bank stock returns may not react as much to one big negative news but may react much more to many small negative news. We also took into consideration the time lags, and we found that there is a difference between how the two banks react to the news coming from the Baltics and Sweden. SEB Bank stock returns respond in general slower to Baltic news than Swedbank stock returns. On the other hand, SEB Bank stock returns are affected by Reuters news on the same day, while Swedbank stock returns are affected by the Reuters news with a delay. Finally, the volatility of Swedbank stock returns is much more sensitive to local Swedish news and also to the international news, compared to the volatility of SEB Bank stock returns.

This paper can be extended further in several ways. It would be interesting to see how the Swedbank and SEB Bank stock returns and volatilities react to fake or speculative news. It could be that the investors are sensitive to such news, and therefore some volatility behavior can be explained by this. Another extension could be to analyze the posts shared in social media such as Facebook, VK and Twitter to see the mood of the investors. Although these posts could be more emotional compared to actual news, it is a fact that less well-informed investors are influenced by these posts. It is also possible to incorporate news in local languages, *i.e.* Swedish, Latvian, Lithuanian, Estonian, to the sentiment analysis. This would be a challenging task because one needs to construct related libraries in Python for this analysis. However, the news sentiment data extracted from these local news would be highly valuable for this analysis.

7 APPENDICES

Appendix 1. 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

Bibliography

Werner Antweiler and Murray Z Frank. Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance*, 59(3):1259–1294, 2004.

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

Michael Bailey, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel. House price beliefs and mortgage leverage choice. 2017.

Kesavan Balasubramaniam. How to calculate a stock's adjusted closing price, 2018. URL `https://www.investopedia.com`.

Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8, 2011.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

George EP Box, Gwilym M Jenkins, and G Reinsel. Time series analysis: forecasting and control holden-day san francisco. *BoxTime Series Analysis: Forecasting and Control Holden Day1970*, 1970.

Margaret M Bradley and Peter J Lang. Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Citeseer, 1999.

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

Benoît Carmichael and Alain Coën. Real estate as a common risk factor in bank stock returns. *Journal of Banking & Finance*, 94:118–130, 2018.

J Bradford De Long, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann. Noise trader risk in financial markets. *Journal of political Economy*, 98(4):703–738, 1990.

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

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

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

Robert F Engle and Victor K Ng. Measuring and testing the impact of news on volatility. *The journal of finance*, 48(5):1749–1778, 1993.

Robert F Engle, Sergio M Focardi, and Frank J Fabozzi. Arch/garch models in applied financial econometrics. *Encyclopedia of Financial Models*, 2012.

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

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

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

Axel Groß-Klußmann and Nikolaus Hautsch. 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, 2011.

Damodar N Gujarati. Basic econometrics: fourth edition mcgraw-hill. *New York*, 2003.

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

Ling T He, FC Neil Myer, and James R Webb. The sensitivity of bank stock returns to real estate. *The Journal of Real Estate Finance and Economics*, 12(2):203–220, 1996.

C.J. Hutto, Ewan Klein, Pierpaolo Pantone, and George Berry. Vader lexicon, 2019. URL `https://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/vader`$_{l}exicon.txt$.

Clayton J Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*, 2014.

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

Stephen Kelly. *News, Sentiment and Financial Markets: A Computational System to Evaluate the Influence of Text Sentiment on Financial Assets*. PhD thesis, Trinity College Dublin, 2016.

Chaiwat Kosapattarapim, Yan-Xia Lin, and Michael McCrae. Evaluating the volatility forecasting performance of best fitting garch models in emerging asian stock markets. 2012.

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

Qing Li, TieJun Wang, Ping Li, Ling Liu, Qixu Gong, and Yuanzhu Chen. The effect of news and public mood on stock movements. *Information Sciences*, 278:826–840, 2014.

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

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

Antonio Miguel Martins, Ana Paula Serra, and Francisco Vitorino Martins. Real estate market risk in bank stock returns: evidence for 15 european countries. *International Journal of Strategic Property Management*, 20(2):142–155, 2016.

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

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

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

NumFOCUS. Python data analysis library. URL `https://pandas.pydata.org/`.

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

Gabriele Ranco, Darko Aleksovski, Guido Caldarelli, Miha Grčar, and Igor Mozetič. The effects of twitter sentiment on stock price returns. *PloS one*, 10(9):e0138441, 2015.

Leonard Richardson. Beautiful soup 4.4.0 documentation. URL `https://www.crummy.com/software/BeautifulSoup/bs4/doc/`.

Sergei P Sidorov, Andrey Revutskiy, Alexey Faizliev, Eugene Korobov, and Vladimir Balash. Stock volatility modelling with augmented garch model with jumps. *IAENG International Journal of Applied Mathematics*, 44(4):212–220, 2014.

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

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

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

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

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

Rahul Verma and GökÇe Soydemir. 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, 2009.

Peter Whitle. *Hypothesis testing in time series analysis*, volume 4. Almqvist & Wiksells, 1951.

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

Xue Zhang, Hauke Fuehres, and Peter A Gloor. Predicting stock market indicators through twitter "i hope it is not as bad as i fear". *Procedia-Social and Behavioral Sciences*, 26: 55–62, 2011.

KOKKUVÕTTE

Kinnisvarauudiste mõju Swedbanki ja SEB Panga aktsiate tootlusele

Põhja- ja Baltimaade pangandussektorid on tihedalt omavahel seotud. Baltimaade finantssüsteemis on olulilise tähtsusega kaks panka - Swedbank ja SEB Pank, mida omakorda mõjutavad Rootsi kui koduriigi kinnisvaraturu riskid. Lisanduvad Baltimaade endi finantsturgudest tulenevad riskid. Kinnisvarauudised, mis ilmuvad neil kõigil turgudel, mõjutavad suure tõenäosusega pankade laenuteenuseid ning hüpoteeklaenude mahu kasvu ning seega ka pankade kasumlikkust ning aktsiahindade tootlust.

Nii Balti riikides kui Rootsis on järjest rohkem tähelepanu pälvinud uudised kinnisvaraturgude potsensiaalsest langusest ning ohtudest. Kuigi riigiti erinevad põhjused ja argumendid, võib nende mõju turugudele olla ühesugune. Rootsi majandusele on iseloomulik kiire kasv, vähenev tööpuudus, rahvaarvu suurenemine ja madalad intressimäärad, kuid võlakoormus kasvab kiiremini kui leibkondade sissetulekud. Baltikumi uudistekanalid tõstatavad üsna sageli kinnisvaraturu haavatavuse teema, sest Balti riikide kinnisvaraturud on tihedalt seotud Rootsi omaga. Lisaks on Balti riikide kindlale majanduskasvule järgnenud väga aktiivne eluasemeturg, mil pealinnad, suuremad keskused ja kuurortid on muutunud kinnisvaraobjektide populaarseteks kohtadeks. Kõik need tegurid kombineeritult on tõstatanud aktiivseid arutelusid kinnisvarasektori suurenenud riskide teemal nii sotsiaalmeedias kui ka ajakirjanduses. See, kuivõrd on Swedbank ja SEB Pank mõjutatud seda laadi riskidest Balti riikides ning Rootsis, on vaatluse all käesolevas töös.

Antud uurimuses analüüsitakse Rootsi, Eesti, Läti ja Leedu kinnisvaraturuga seotud uudiste mõju Swedbanki ja SEB Panga aktsiate tootlusele ning volatiilsusele ajavahemikus 04.01.2016 kuni 19.02.2019. Analüüsi käigus rakendatakse esmalt Pythoni avatud lähtekoodiga tööriistu ja teeke veebi kraapimiseks, et saada uudiste veebilehtedelt analüüsiks vajaminev tekst. Seejärel kasutatakse reeglipõhist tundmusanalüüsi mudelit VADER uudiste meelestatuse või hoiakute aegridade koostamiseks. Edasiselt kasutatakse neid andmeid nelja ARMA mudeli hindamiseks käsitledes erinevaid aspekte - ekstreemsed uudised, uudiste asümmeetriline mõju, sisu versus uudiste arv. Samuti uuritakse uudiste meelsuse mõju volatiilsusele, et teada saada, kas positiivsed ja negatiivsed uudised võivad volatiilsusele erinevat mõju avaldada.

Töö peamine järeldus on, et positiivsed ja negatiivsed uudised mõjutavad tootlust ja volatiilsust erineva viiteajaga. Swedbanki ja SEB Panga osas on kohalikel Rootsi uudistel rohkem mõju kui Balti ja Reutersi uudistel. Suure tundmusega negatiivsed uudised mõjutavad tootlust alles kaks päeva hiljem; negatiivsete uudiste arv mõjutab aga sama päeva tootlust. Selle üks võimalik tõlgendus on, et pankade aktsiate tootlus ei pruugi reageerida nii palju ühele suurele negatiivsele uudisele, kuid võib reageerida paljudele väikestele negatiivsetele uudistele. Vaadeldes viiteaegu ilmnes erinevus ka selles, kuidas kaks panka reageerivad Baltikumist ja Rootsist saabuvatele uudistele: SEB Panga aktsiate tootlus reageerib Balti uudistele üldiselt aeglasemalt kui Swedbanki aktsiate tootlus. Teisalt mõjutavad SEB Panga aktsiate tootlust samal päeval Reutersi uudised, Swedbanki aktsiate tootlust mõjutavad Reutersi uudised aga viivitusega. Lisaks on Swedbanki aktsiate tootluse volatiilsus võrreldes SEB Panga aktsiatega märgatavalt tundlikum nii kohalike Rootsi uudiste kui ka rahvusvaheliste uudiste suhtes.