

Effect of Real Estate News Sentiments on the Stock Returns of Swedbank and SEB Bank

Yuliia Puzanova[♣] and Mustafa Hakan Eratalay^{*♠}

[♣]Swedbank AS, Estonia

[♠]University of Tartu, Estonia

Submitted: October 10, 2020 • Accepted: July 8, 2021

ABSTRACT: This paper explores 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 both banks, and the effect is different for 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 explore if negative and positive news affect the volatility processes to some extent. We found out that the volatilities are explained better by the GJR-GARCH and NAGARCH models. Overall, the volatility of SEB stock returns depends more on the news sentiments compared to the volatility of Swedbank stock returns.

JEL classification: C320, C520, C580.

Keywords: Sentiment Analysis, ARMA, GARCH, Real Estate Market, Sweden, Baltics, Latvia, Lithuania, Estonia.

*Corresponding Author. E-mail: hakan.eratalay@ut.ee

1 Introduction

The Nordic and the Baltic banking sectors are closely connected. Two banks which 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 influence for the banks' activities related to lending services, and impact mortgage volume growth. As a result, we can expect an impact on the profitability of these banks and their stock price returns.

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

The Swedish economy is characterised as exhibiting fast growth, declining unemployment, population increase, and low-interest rates. On the other hand, the debt burden rising faster than household incomes is also observed (Statistics Sweden, 2019). The combination of these factors is the reason for active discussions in social media and for (loud) headlines in news articles. In addition, even after the Swedish Finansinspektionen 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 being actively boosted due to rising investments. 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 this 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 particular, we use the autoregressive moving average (ARMA) models with news sentiment variables 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 used 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). Moreover, 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. (2014a); Verma and Soydemir (2009); Yu (2014).

This study contributes to the literature by investigating real estate news sentiment for a period between 4 January 2016 and 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 of the news sentiments in our model. To the best of our knowledge, there are no other papers investigating the real estate news sentiments in Swedish and Baltic markets, and their effect on these banks' 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).

The paper is further divided into five sections. Section 2 contains 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. Section 3 provides description of the methods used for data collection, the procedure for news sentiment estimation and an overview of the econometric models used. 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. In addition, the news is one of the main sources which forms and reflects the market behaviour at the same time. And this is why behavioural economists believe that the 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., 2014a). 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., 2014a).

Different authors, such as Kothari and Shanken (1997), De Long et al. (1990) analysed 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-Of-Words" model for news sentiment analysis (Harris, 1954).

News sentiment in predictions of stock prices is not new but it is still 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 publish 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., 2014a). 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.

The analysis of news releases and their impact on the financial world was particularly important during the euro-zone sovereign debt crisis. Starting with Greece in early 2010, Ireland, Portugal, Spain and Cyprus required financial bailouts or assistance in the next few years. This was followed by the Brexit voting in 2016, which brought some speculations on the EU economy. Afterwards, the news and speculations were further fuelled by the worsened situation of the Italian banks. Acharya et al. (2018) provides a discussion on the time line, causes and effects of the euro-zone debt crisis, while Beetsma et al. (2013, 2017) analyses the effect of the news sentiments on the euro-zone sovereign years during the crisis period.

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. Deng et al. (2018) used a large microblog data set to analyse the effect of positive and negative microblog sentiments on the stock returns. However 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. (2014a) 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; Kwan, 2019). This functions via the following 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 a number of 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 discussed in papers Mei and Lee (1994); Mei and Saunders (1995) and Ding et al. (2017), 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 the prediction models (Kelly, 2016; Gupta and Banerjee, 2019). Soroka (2006) finds strong evidence for the asymmetric responses, stating that negative economic news generate a bigger impact on the individuals' attitudes. Cepoi (2020) considers the asymmetric dependence between the stock market returns and news during the Covid-19 crisis.

Following the previous research, we used the news sentiment as an explanatory factor for banks' stock returns and considered the asymmetric effect of news on the stock return behaviour.

3 Methods and Analysis

3.1 Data Collection and Aggregation

In order to collect and structure a text, we decided to use Python as the main programming language due to the 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 used 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 as well as point to other parts of the article, such as the 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 upon 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 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 is inconvenient when writing a universal algorithm to recognize the necessary tags and download the text. For instance, when inspecting the web pages' structure, we found that in the case of The Local SE web pages the body of the article was contained under the tag "div id ="article-body", while for Reuters, this tag is "div class ="StandardArticle-Body_body". This 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 removes 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, a list of words labelled either positive, negative or neutral. There are several widely used lexicons, such as LIWC (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; SenticNet¹ 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).

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

¹<https://sentic.net/>

in econometric models. The function has the form:

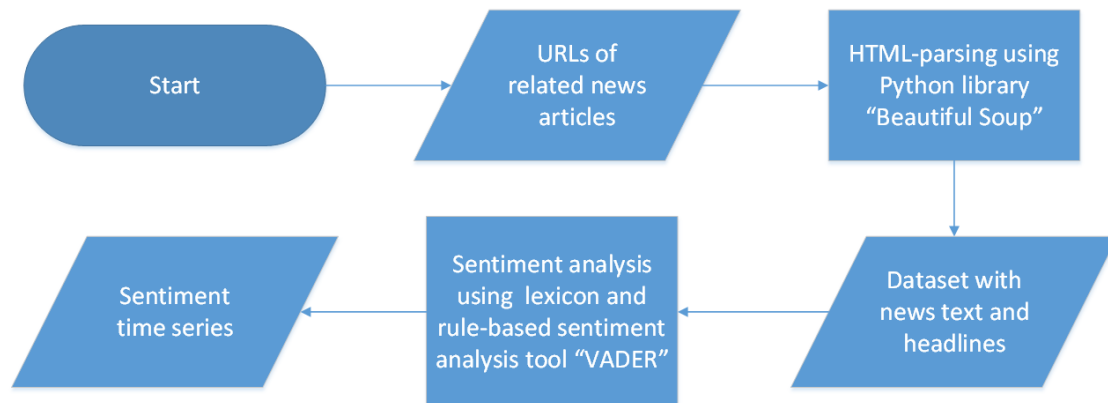
$$C = \frac{S}{\sqrt{S^2 + 15}} \quad (1)$$

where C is the compound sentiment score, and S the sum of negative (x), positive (y) and neutral (z) scores. The value 15 is the default normalization constant mentioned in (Hutto and Gilbert, 2014).

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 signalling 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, 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 modelling. The algorithm can be visualized by the flowchart (Figure 1).

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



Note: 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.

3.3 Overview of Models Used

The base model we consider is an ARMA(p , q) model, where p and q are the autoregressive and moving average orders, respectively. In all the models that follow, the ARMA parameters satisfy the typical stationarity and invertability conditions. We call the base model Model 1.0 and it is nested in Model 1.1. If we set the coefficients of the news components equal to zero in Equation 2 of Model 1.1, we obtain Model 1.0 equation. 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 and other international news sources to take into account the 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. 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.

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

In Equation 2, news variables are defined with the names of their sources (as explained in section 4.1) and are by construction in the range $[-1, 1]$, where a positive news is higher ranked than negative news. Therefore any increase in the news variable indicates less bad news or more good news; hence, should have a positive effect on the returns. This positive effect may come on the same day or with a delay. Therefore, we apply the following restrictions on the news sentiment coefficients: $\delta_{ji} \geq 0$ for all $i = 1, \dots, 4, j = 0, 1, 2$.

For the next models, we distinguish negative and positive news in the equations in order to take into account their asymmetric effects.

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

In Equation 3, 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 -0.05². Otherwise, the news sentiment is neutral. Neutral news is discarded because *TheLocal* and *Others* news sources do not have neutral news. Considering negative and positive sentiments separately is in line with Soroka (2006), where the author pointed out the

²The values 0.05 and -0.05 are given by the construction of the news sentiment indices.

larger impact of negative economic news on individuals' attitudes.

$$\begin{aligned}
r_t = & \mu + \sum_{i=1}^p \beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \\
& + \delta_{10}^- Reuters_t^- + \delta_{20}^- TheLocal_t^- + \delta_{30}^- Others_t^- + \delta_{40}^- Baltics_t^- + \\
& + \delta_{11}^- Reuters_{t-1}^- + \delta_{21}^- TheLocal_{t-1}^- + \delta_{31}^- Others_{t-1}^- + \delta_{41}^- Baltics_{t-1}^- + \\
& + \delta_{12}^- Reuters_{t-2}^- + \delta_{22}^- TheLocal_{t-2}^- + \delta_{32}^- Others_{t-2}^- + \delta_{42}^- Baltics_{t-2}^- + \\
& + \delta_{10}^+ Reuters_t^+ + \delta_{20}^+ TheLocal_t^+ + \delta_{30}^- Others_t^+ + \delta_{40}^- Baltics_t^+ + \\
& + \delta_{11}^+ Reuters_{t-1}^+ + \delta_{21}^+ TheLocal_{t-1}^+ + \delta_{31}^- Others_{t-1}^+ + \delta_{41}^- Baltics_{t-1}^+ + \\
& + \delta_{12}^+ Reuters_{t-2}^+ + \delta_{22}^+ TheLocal_{t-2}^+ + \delta_{32}^- Others_{t-2}^+ + \delta_{42}^- Baltics_{t-2}^+
\end{aligned} \tag{3}$$

As discussed in Model 1.1, an increase in a news variable would mean less bad news or more good news. It is trivial that more good news should increase returns. An increase in the negative news sentiment would mean less bad news and therefore, this should increase returns. The effect on the returns may appear on the same day or with a delay. Therefore, the restrictions on the news sentiment coefficients are: $\delta_{ji}^-, \delta_{ji}^+ \geq 0$ for all $i = 1, \dots, 4, j = 0, 1, 2$.

Model 1.3. ARMA(p, q) model with merged news data (positive and negative news distinguished)

In Equation 4, the $News_t^-$ and $News_t^+$ variables are the sum of negative and positive news at time t , respectively. Given that in some dates there may be more than one piece of news, we also consider the number of negative and positive news at time t with the variables N_t^- and N_t^+ , respectively. A related approach was considered by Sidorov et al. (2014a), where the authors used the number of all relevant news as an exogenous variable for the GARCH model and referred to this number as news intensity.

$$\begin{aligned}
r_t = & \mu + \sum_{i=1}^p \beta_i r_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \\
& + \delta_{10}^- News_t^- + \delta_{11}^- News_{t-1}^- + \delta_{12}^- News_{t-2}^- + \\
& + \delta_{20}^+ News_t^+ + \delta_{21}^+ News_{t-1}^+ + \delta_{22}^+ News_{t-2}^+ + \\
& + \delta_{30}^- N_t^- + \delta_{31}^- N_{t-1}^- + \delta_{32}^- N_{t-2}^- + \\
& + \delta_{40}^+ N_t^+ + \delta_{41}^+ N_{t-1}^+ + \delta_{42}^+ N_{t-2}^+
\end{aligned} \tag{4}$$

It was discussed before that an increase in the news sentiment affects the returns positively. The same can be said about the number of positive news, that it affects the returns positively. However, an increase in the number of negative news would decrease the returns. These effects can be seen on the same day or with delay. Therefore, the restrictions on the news sentiment coefficients are: $\delta_{1i}^+, \delta_{2i}^-, \delta_{4i}^+ \geq 0$ and $\delta_{3i}^- \leq 0$, where

$i = 0, 1, 2$.

To proceed with the conditional volatility modelling, we first identify the model that fits the conditional returns best. For this purpose, we use the Akaike (AIC) and Bayesian (BIC) criteria. 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 models³.

When modelling the volatility, we specifically focus on the asymmetric effects. In the literature, asymmetric effects is a stylized fact observed in the volatilities of financial returns, which suggests that volatility increases much more when a negative return shock appears compared to a positive return shock of the same magnitude (Ghysels et al., 1996; Asai and McAleer, 2006). In the models we discuss below, we also explore the possible asymmetric effect of positive and negative news sentiments on conditional volatility.

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, which is modelled via GARCH (1, 1) and its extensions.

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

A commonly used model that takes into account the asymmetric effects is the Glosten, Jagannathan and Runkle - GJR - GARCH model (Glosten et al., 1993). We present the GJR-GARCH model with the news sentiment variables in the Equation 5.

$$\begin{aligned}
 h_t = & \gamma + (\alpha_1 + \theta I_{t-1}) \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} + \\
 & + \delta_{10}^- Reuters_t^- + \delta_{20}^- TheLocal_t^- + \delta_{30}^- Others_t^- + \delta_{40}^- Baltics_t^- + \\
 & + \delta_{11}^- Reuters_{t-1}^- + \delta_{21}^- TheLocal_{t-1}^- + \delta_{31}^- Others_{t-1}^- + \delta_{41}^- Baltics_{t-1}^- + \\
 & + \delta_{12}^- Reuters_{t-2}^- + \delta_{22}^- TheLocal_{t-2}^- + \delta_{32}^- Others_{t-2}^- + \delta_{42}^- Baltics_{t-2}^- + \\
 & + \delta_{10}^+ Reuters_t^+ + \delta_{20}^+ TheLocal_t^+ + \delta_{30}^+ Others_t^+ + \delta_{40}^+ Baltics_t^+ + \\
 & + \delta_{11}^+ Reuters_{t-1}^+ + \delta_{21}^+ TheLocal_{t-1}^+ + \delta_{31}^+ Others_{t-1}^+ + \delta_{41}^+ Baltics_{t-1}^+ + \\
 & + \delta_{12}^+ Reuters_{t-2}^+ + \delta_{22}^+ TheLocal_{t-2}^+ + \delta_{32}^+ Others_{t-2}^+ + \delta_{42}^+ Baltics_{t-2}^+
 \end{aligned} \tag{5}$$

In Equation 5, 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 $\alpha_1 + \theta + \alpha_2 < 1$. Asymmetric effects would occur if

³There 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.

$\theta > 0$, because negative errors increase volatility more than positive errors.

Model 2.2. NAGARCH(1, 1) model with asymmetric news, neutral news discarded

We also consider that the asymmetric effect of the return shocks could be in nonlinear form. For this reason, we employ the nonlinear asymmetric GARCH (NAGARCH) model (Robert and Victor, 1993; León et al., 2005). We present the NAGARCH model with news sentiment variables in Equation 6.

$$\begin{aligned}
h_t = & \gamma + \alpha_1 \left(\varepsilon_{t-1} + \delta h_{t-1}^{1/2} \right)^2 + \alpha_2 h_{t-1} \\
& + \delta_{10}^- Reuters_t^- + \delta_{20}^- TheLocal_t^- + \delta_{30}^- Others_t^- + \delta_{40}^- Baltics_t^- \\
& + \delta_{11}^- Reuters_{t-1}^- + \delta_{21}^- TheLocal_{t-1}^- + \delta_{31}^- Others_{t-1}^- + \delta_{41}^- Baltics_{t-1}^- \\
& + \delta_{12}^- Reuters_{t-2}^- + \delta_{22}^- TheLocal_{t-2}^- + \delta_{32}^- Others_{t-2}^- + \delta_{42}^- Baltics_{t-2}^- \\
& + \delta_{10}^+ Reuters_t^+ + \delta_{20}^+ TheLocal_t^+ + \delta_{30}^+ Others_t^+ + \delta_{40}^+ Baltics_t^+ \\
& + \delta_{11}^+ Reuters_{t-1}^+ + \delta_{21}^+ TheLocal_{t-1}^+ + \delta_{31}^+ Others_{t-1}^+ + \delta_{41}^+ Baltics_{t-1}^+ \\
& + \delta_{12}^+ Reuters_{t-2}^+ + \delta_{22}^+ TheLocal_{t-2}^+ + \delta_{32}^+ Others_{t-2}^+ + \delta_{42}^+ Baltics_{t-2}^+
\end{aligned} \tag{6}$$

In Equation 6, the volatilities h_t are positive if $\gamma > 0$, $\alpha_1, \alpha_2 \geq 0$, and stationary if $\alpha_1(1 + \delta^2) + \alpha_2 < 1$. Given that:

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

the non-linear asymmetry in Equation 6 is generated by $2\delta \varepsilon_{t-1} h_{t-1}^{1/2}$. As negative shocks are expected to increase volatility, asymmetric effect would occur if $\delta < 0$.

Both GJR-GARCH(1, 1) and NAGARCH(1, 1) discussed in Equations 5 and 6 nest the baseline GARCH(1, 1) model. This can be seen by setting θ and δ equal to zero in these equations, respectively.

In Equations 5 and 6 the effects of negative and positive news sentiments on the volatility are distinguished. Braun et al. (1995) and Malik (2011) state that volatility increases strongly in response to bad news and decreases in response to good news. Following this thought in Equations 5 and 6, an increase in the negative news sentiment would mean less bad news and this would decrease volatility, hence $\delta_{i,j}^- \leq 0$, for all $i = 1, 2, 3, 4$ and $j = 1, 2, 3$. On the other hand, if there is an increase in the positive news sentiment the volatility is expected to decrease, which implies that $\delta_{i,j}^+ \leq 0$, for all $i = 1, 2, 3, 4$ and $j = 1, 2, 3$. Incorporating this reasoning to the positivity restrictions, in Equations 5 and 6, the volatilities h_t are positive if $\gamma > 0$, $\alpha_1, \alpha_2 \geq 0$. Moreover, one could impose the restriction that $\gamma - \sum_i \sum_j \delta_{i,j}^- + \sum_i \sum_j \delta_{i,j}^+ > 0$. This condition comes from the fact that the news sentiment variables takes values between -1 and 1 and even in the extreme case where the squared residuals and volatility is very close to zero and the news sentiments take extreme values, the volatility stay positive. We didn't impose this restriction, but

we made sure that volatility stays positive in the estimations.

In all the ARMA models above, we assume that the error term is distributed normally. Typically, financial returns have thick tailed distribution. Ghysels et al. (1996) discusses the stylized facts about financial returns. Assuming that the returns are distributed normally has some statistical advantages, including that the estimation becomes easier. (Engle and Sheppard, 2001) In particular, when the errors are normally distributed, 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 Sheppard, 2001; Carnero and Eratalay, 2014). Since separating the estimation of mean and variance models would bring inefficiency due to the misspecification⁴, we use the Bollerslev-Wooldridge robust covariance matrix. As Bollerslev and Wooldridge (1992) point out, this is a consistent estimator of the White (1980) robust asymptotic covariance matrix, from which one can obtain the heteroscedasticity consistent standard errors.

4 Data

We utilize two data sets for this research. The first contains all news sentiments 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 Swedbank and SEB Bank, both operating in Sweden and the Baltic countries.

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.

To this end, 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.

⁴In this approach the misspecification is that when estimating the conditional mean parameters, the model is assumed to be homoscedastic.

To cover those periods when the main selected sources may 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 others⁵. It was highly desirable for further modelling 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 in general. Two horizontal lines ($y = 0.05$ and $y = -0.05$) were 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.

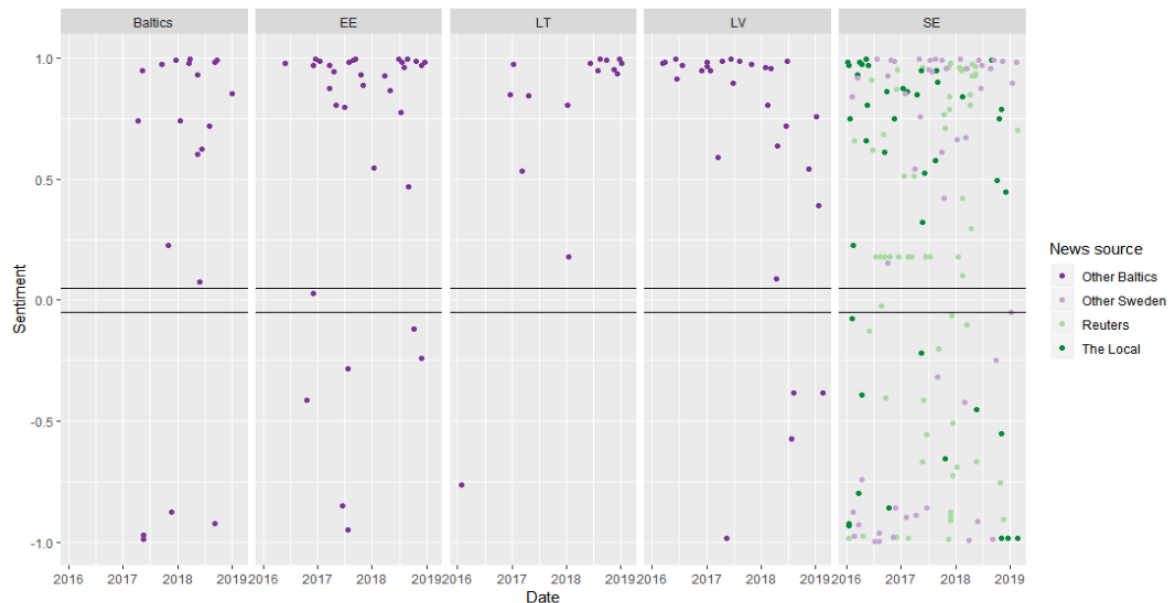
More positive Swedish news is provided by news portals other than "Reuters" or "The Local SE" but the difference is not considerable – all positive news are distributed almost equally. More negative Swedish news is provided by "Reuters", and the distribution is not as equal as for the positive news.

For Baltic news, we cannot divide the results by news portals due to the difficulty in finding such resources with sufficient news about the real estate market in these countries. Yet 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

⁵See Appendix A 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. The news between the horizontal lines are considered as neutral. Period: 04.01.2016 - 19.02.2019.

Source: Authors' calculations.

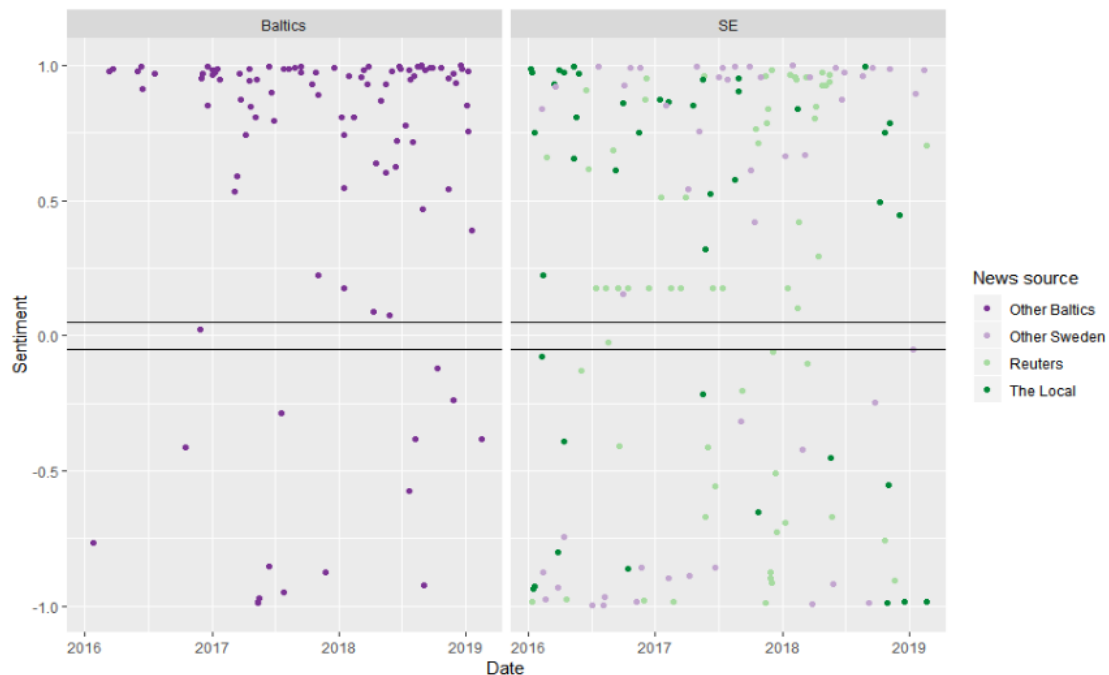
estate market. Furthermore, for all markets, the neutral amount of news is extremely low, and it was one reason why neutral sentiments were discarded for the estimation models, in which we considered negative and positive sentiments.

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: $(0.9, 1)$.

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 the fact that the standard deviation is big enough for all countries and represents that sentiment fluctuations are high. If we 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 a larger part of the sentiments presented in the data set for these countries are above the average value - mostly positive. Only for Swedish data can

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. The news between the horizontal lines are considered as neutral. Period: 04.01.2016 - 19.02.2019.

Source: Authors' calculations.

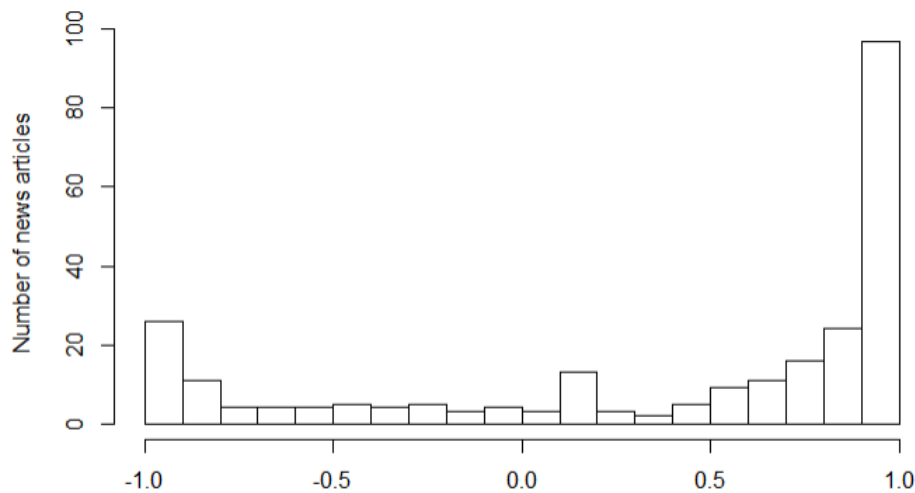
we state that the data is fairly symmetrical, as the skewness is closer to zero (-0.53). The second value to pay attention to is excess kurtosis - a way to notice the outliers in the distribution of data. Any value above 0 (Estonia, Latvia and in particular Lithuania) indicates leptokurtic or heavy-tailed news sentiment data, and there is a probability that we might have outliers. Regarding other excess 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.

The 10 most positive news sentiments are listed in the Table 2. For example, on 27.08.2018 an article gives a brief overview of the real estate market and what to expect in the future in Estonia. The author states that the Estonian real estate market is stable and low risks make it a good investment area⁶. Another example is that on 19.12.2018 an article writes about the attractive real estate market in Latvia, more specifically in Vilnius and Kaunas, and about potential interest of the investors in Klaipėda⁷. From

⁶Source: <https://blog.reinvest24.com/whats-going-on-with-the-housing-market-in-estonia>

⁷Source: <https://en.delfi.lt/archive/real-estate-experts-see-potential-of-klaipedacity.d?id=79906703>

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.

Table 1: Descriptive statistics of news sentiments data

Country	N	Mean	St.dev.	Med.	Min.	Max.	Skew.	E.kur.
Baltics	20	0.43	0.75	0.74	-0.99	1	-1.07	-0.58
Estonia	34	0.64	0.57	0.94	-0.95	1	-1.53	0.96
Lithuania	15	0.75	0.47	0.95	-0.76	1	-2.22	4.09
Latvia	29	0.64	0.55	0.95	-0.98	1	-1.63	1.43
Sweden	155	0.24	0.75	0.54	-1	1	-0.53	-1.35
All	253	0.38	0.72	0.75	-1	1	-0.85	-0.86

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, Med. - median, Min./Max. - minimum and maximum sentiment score, Skew. - skewness of the data, E.kur. - excess kurtosis of the data.

Source: Authors' calculations.

Table 2 it can be seen that half of these 10 most positive news were published during 2018. Out of these 10 news, 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 news sentiments are given in Table 3. These news sentiments are mostly for Sweden for different years. For example, on 04.07.2016 an article writes about

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.

the lack of housing in Sweden and its impact on immigrants and startups⁸. One other example is that on 03.08.2016 an article describes how Baltic states worry that Swedish banks, when in crisis, have the possibility of withdrawing money from Baltic subsidiaries to cover their losses in the domestic market, and this will negatively impact the economies of the Baltic states⁹. We also see that few Baltic news is also placed at the top of the negatives, but among them, there are only news related to the Latvian market or the Baltic market in general.

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.

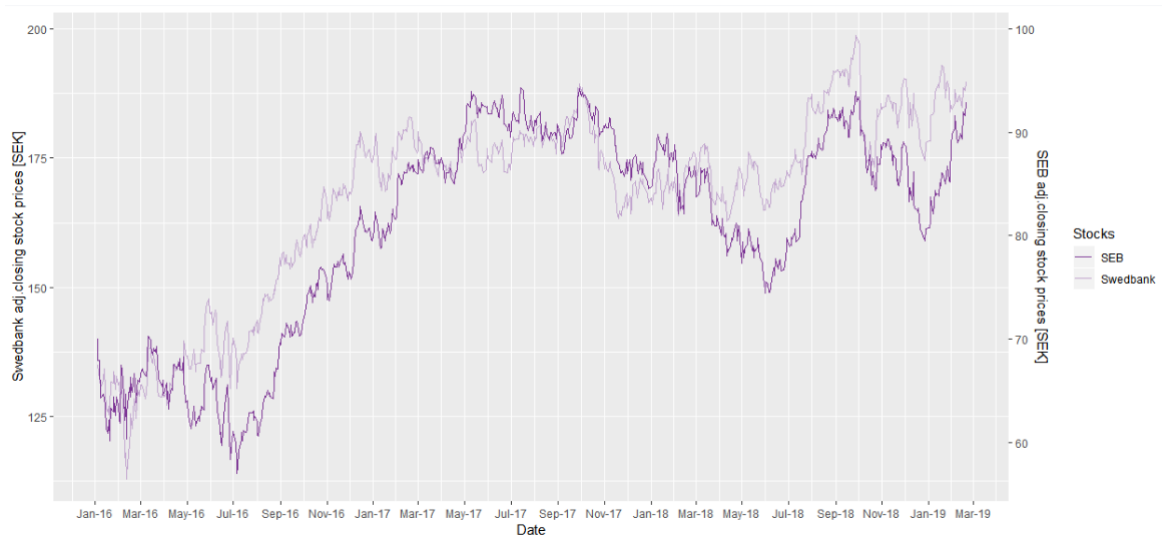
Source: Authors' calculations.

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 clouds shown in Figure 5. The most frequently used words have larger font size and

⁸Source: <https://www.ft.com/content/8961b57c-3872-11e6-a780-b48ed7b6126f>

⁹Source: <https://www.ft.com/content/33aa7868-595f-11e6-9f70-badea1b336d4>

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".

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. Past this date, 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. The stock prices of Swedbank and SEB Bank don't always move in the same direction. Around February of 2018, the SEB Bank stock price had a negative trend while Swedbank stock prices were rising. Also around March-April 2017, we can see that SEB Bank stock prices enter to a decline trend after Swedbank stock prices which may suggest late reaction. This is in line with some findings in Section 5.

We calculate the daily returns of the banks' stocks as follows:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \times 100 \quad (7)$$

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 the spread of these fluctuations change over time.

A statistical overview of the log returns for both stocks is provided in Table 4. From the skewness and excess kurtosis, we see that the distribution is not close to normal (skewness is -0.67 and -0.68 and excess kurtosis is 3.42 and 6.49 for Swedbank and SEB

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".

Bank, respectively). Returns data is negatively skewed: the left tail is longer than the right. This may suggest some extreme negative returns. The high excess kurtosis suggests that the distribution of the returns is fat-tailed, which is noticeable in particular with SEB Bank returns.

Table 4: Descriptive statistics of daily stock returns

Stocks	N	Mean	St.dev.	Med.	Min.	Max.	Skew.	E.kur.
Swedbank	792	0.04	1.24	0.01	-9.48	5.14	-0.67	3.42
SEB	792	0.04	1.37	0.00	-11.63	8.66	-0.68	6.49

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, Med. - median, St.dev. - standard deviation of sentiments, Min./Max. - minimum and maximum sentiment score, Skew. - skewness of the data, E.kur. - excess kurtosis of the data. All values are in percentages.

Source: Authors' calculations and "Yahoo! Finance".

Our analysis of the data starts with preliminary tests. We first check if the return series we obtained are stationary using the Augmented Dickey-Fuller (ADF) test. In this test the null hypothesis is that the return series is not stationary. From Table 5, we can see that this null hypothesis can be rejected for the returns of Swedbank and SEB Bank. Therefore we conclude that these return series are stationary.

We also look at the autocorrelation and partial autocorrelation structures of the returns, which allow us to determine the order of the MA(q) and AR(p) terms, respectively. Figure 8a shows that there are significant autocorrelations at lags 4 and 10 for Swedbank.

Table 5: Augmented Dickey-Fuller test for stationarity

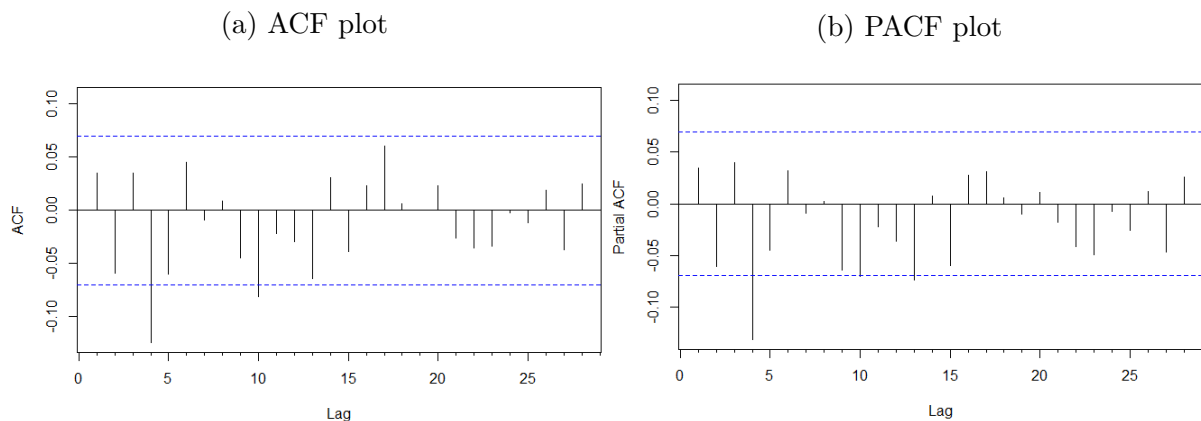
	Swedbank	SEB
Dickey-Fuller	-10.664	-9.8816
Lag order	9	9
P-value	0.01	0.01
Alternative hypothesis: stationary		

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.

The partial correlation plot in Figure 8b indicates 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). Figure 9b shows an increase at lags 1, 4, 10 and 16. These lags were taken into account when choosing the ARMA orders for Swedbank and SEB Bank returns.

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



Notes: This figure shows 8a autocorrelation function (ACF) and 8b 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.

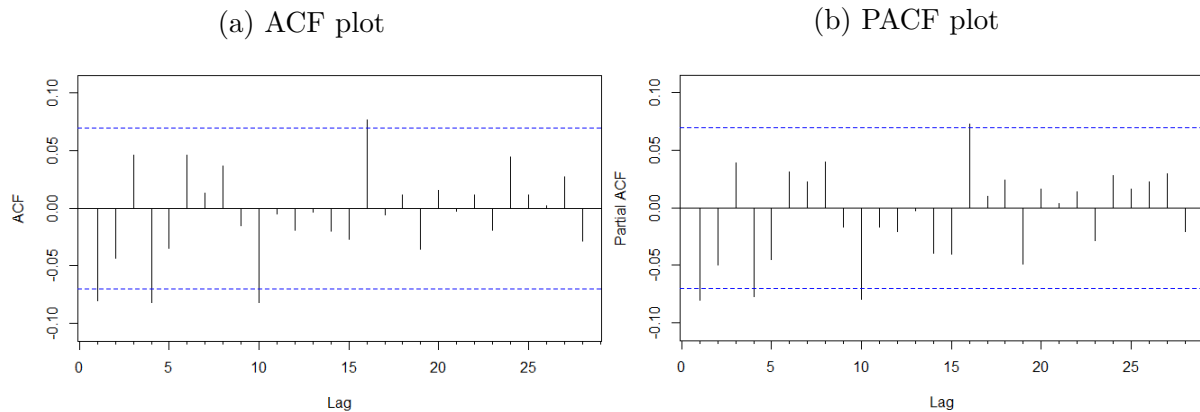
Source: Authors' calculations.

5 Results

The news vectors contain many zeros and relatively less non-zero values. Moreover, they are not autocorrelated¹², and therefore cannot explain or capture the autocorrelation

¹²Ljung-Box test was applied to the news vectors and there were no significant autocorrelations for any of the news vectors.

Figure 9: ACF and PACF plots of SEB Bank's daily stock returns data



Notes: This figure shows 9a autocorrelation function (ACF) and 9b 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.

structure in the return series. For this reason, when choosing the orders of the AR model, we don't take the news sentiment variables into account. Based on the AIC criteria from the models which yielded serially uncorrelated residuals¹³, we found that the AR(4) model is the most adequate for both Swedbank and SEB Bank returns. The 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 optimization procedure. In what follows, the estimation results that we want to focus on are presented in tables where the coefficients and their statistical significance are presented for each model.

5.1 Estimation Results

To save from space, we mention here that the AR and GARCH parameter estimates of all the models below satisfy the corresponding stationarity requirements. We discuss the restrictions on the coefficients of the news sentiment variables in detail in Section 3.3.

Model 1.0 AR(4) model (base-line model)

For both Swedbank and SEB Bank stock returns, the appropriate ARMA model was an AR(4). The estimated model is given in Table 10 in Appendix B. From the results, we can conclude that *AR* coefficients for both banks' stocks are significant at a significance level of 0.01 and stock returns are correlated with returns from the last 4 periods.

For the following models which extend the base model by including news sentiment

¹³The residuals of each model was tested for autocorrelation using the Ljung-Box test.

variables, we perform likelihood ratio tests comparing them with the nested smaller models. Although the likelihood ratio test results suggest that adding the news sentiment variables did not improve the fit significantly, some of the individual coefficients of interest are found to be significant¹⁴.

Model 1.1 AR(4) model with news component

The model AR(4) that contains news sentiment variables follows the Equation 2. The estimation results are given in Table 11 in Appendix B. For Swedbank and SEB Bank stock returns, we can see that the autoregressive parameter estimates changed slightly compared to the base model and the intercept is smaller.

The coefficient estimates suggest that the location (Baltic States or Sweden) of news matter. The Local and Others are news sources about the real estate market in Sweden and therefore it is not surprising to see that they lead to an increase in the Swedbank returns on the same day, while the effect of the former continues two days later. Reuters and Baltics news effect the Swedbank returns positively on the next day and the following day. For SEB Bank stock returns, Swedish news from The Local and Others make the returns increase on the same day and the positive effect of The Local continues two more days. On the other hand, Reuters news affects the SEB Bank returns positively in the next two days and Baltic news affect the SEB Bank returns positively in two days. The differences compared to the estimations with Swedbank returns are that (1) SEB Bank returns react slower to Baltic news: one day later, (2) the SEB Bank returns seem to depend comparatively more on the Others news sources, (3) in total, the SEB Bank returns are less exposed to the news from Reuters and The Local.

Model 1.2 AR(4) model with asymmetric effect of news, neutral news discarded

The AR(4) model with asymmetric effect of news follows from Equation 3 and the estimation results are given in Table 6.

Table 6 shows that when the news are distinguished for negative and positive sentiments, interesting results emerge. For both banks' stock returns, all the news sentiment coefficients are positive as the restrictions required, while some are not significant even at 0.1 significance level. We focus our attention on the significant coefficients.

For Swedbank stock returns, it is observed that less bad news¹⁵ from the Baltics mean higher returns on the same day but the main influence comes with a two-day delay. This is a result that Model 1.1 could not capture. Less bad news from Reuters affects returns positively with a delay of one and two days. The biggest effect of less bad news comes from The Local and Others on the same day and the impact continues two days more in

¹⁴In a model where the likelihood ratio tests or equivalently the F-test suggest that several coefficients in the larger model are jointly insignificant, if a few of those coefficients are individually significant, then they shouldn't be ignored by the researcher. See Wooldridge (2015) pp. 149-150 for a detailed discussion.

¹⁵Less bad news means an increase in the negative news sentiment variable. An alternative way to read the results would be *more bad news* which would mean a decrease in the same variable.

Table 6: Estimation results for the AR(4) model with asymmetric effect of news

Coefficients	Swedbank	SEB Bank
μ	0.0344 (0.0177) [*]	0.0385 (0.0101) [***]
β_1	0.0343 (0.0048) [***]	-0.0765 (0.0507)
β_2	-0.0699 (0.0175) [***]	-0.0659 (0.0194) [***]
β_3	0.0530 (0.0130) [***]	0.0333 (0.0084) [***]
β_4	-0.1221 (0.0595) [**]	-0.0799 (0.0135) [***]
<i>Reuters</i> s_t^-	0.0000 (0.0000)	0.1720 (0.0678) [**]
<i>TheLocal</i> s_t^-	0.8393 (0.2854) [***]	0.4885 (0.1912) [**]
<i>Others</i> s_t^-	0.5647 (0.1246) [***]	1.0128 (0.6627)
<i>Baltics</i> s_t^-	0.1399 (0.0609) [**]	0.0524 (0.0137) [***]
<i>Reuters</i> s_{t-1}^-	0.2619 (0.1026) [**]	0.0080 (0.0019) [***]
<i>TheLocal</i> s_{t-1}^-	0.4045 (0.1967) [**]	0.6765 (0.4587)
<i>Others</i> s_{t-1}^-	0.0000 (0.0000)	0.0000 (0.0000)
<i>Baltics</i> s_{t-1}^-	0.0000 (0.0000)	0.0000 (0.0000)
<i>Reuters</i> s_{t-2}^-	0.3357 (0.1218) [***]	0.1868 (0.0414) [***]
<i>TheLocal</i> s_{t-2}^-	0.0163 (0.0054) [***]	0.0414 (0.0200) [**]
<i>Others</i> s_{t-2}^-	0.0000 (0.0000)	0.1533 (0.0582) [***]
<i>Baltics</i> s_{t-2}^-	0.7157 (0.3739) [*]	0.8116 (0.5523)
<i>Reuters</i> s_t^+	0.0000 (0.0000)	0.0000 (0.0000)
<i>TheLocal</i> s_t^+	0.5539 (0.4448)	0.6210 (0.3680) [*]
<i>Others</i> s_t^+	0.0107 (0.0027) [***]	0.3555 (0.1959) [*]
<i>Baltics</i> s_t^+	0.0000 (0.0000)	0.0000 (0.0000)
<i>Reuters</i> s_{t-1}^+	0.4622 (0.3550)	0.3723 (0.4604)
<i>TheLocal</i> s_{t-1}^+	0.0000 (0.0000)	0.0000 (0.0000)
<i>Others</i> s_{t-1}^+	0.0000 (0.0000)	0.0699 (0.0303) [**]
<i>Baltics</i> s_{t-1}^+	0.1535 (0.1109)	0.1087 (0.0708)
<i>Reuters</i> s_{t-2}^+	0.2888 (0.1239) [**]	0.1101 (0.0699)
<i>TheLocal</i> s_{t-2}^+	0.2689 (0.2292)	0.0000 (0.0000)
<i>Others</i> s_{t-2}^+	0.0000 (0.0000)	0.0000 (0.0000)
<i>Baltics</i> s_{t-2}^+	0.0000 (0.0000)	0.1326 (0.1452)
LL	-1267.25	-1343.27
AIC	2592.51	2744.53
BIC	2728.07	2880.09
SSE	1144.12	1386.91
Pval. vs. AR(4)	0.6661	0.5681

Notes: The model is explained under Model 1.2. Standard errors are given in parenthesis. [*], [**], [***] show significance at 0.1, 0.05 and 0.01 significance levels, respectively.

Source: Authors' calculations.

the case of The Local. In the case of The Local, Others and Baltics news, the total effect of negative news is higher than that of the positive news. Finally, the positive impact of positive news from Reuters is relatively high, compared to the other news sources.

For SEB Bank stock returns, we can see that less bad news from Reuters, The Local and Baltics sources increase the returns significantly on the same day, while for the first two sources, the effect continues two days later. The effect of the bad news from Baltics is relatively small and comes on the first day. When looking at the positive news sentiments,

we see the positive impact of the news from The Local and Others news sources, where the effect from the latter continues in the next day.

As with the Swedbank returns, SEB Bank returns are largely, significantly and positively affected by the news from The Local and somewhat less by the news from the Baltics on the same day. Interestingly the SEB Bank returns are influenced positively by less bad news from Reuters on the same day, which is not the case of Swedbank returns. The news from the Others news sources affect the SEB Bank returns more compared to the Swedbank returns, but this coefficient is insignificant. Overall, there are many significant coefficients for the negative news sentiment variables, while not as much for the positive ones.

Model 1.3 AR(4) model with merged news data (positive and negative news distinguished)

The AR(4) model with merged news comes from Equation 4 and the estimation results are given in Table 7.

Table 7: Estimation results for the AR(4) model with merged news data (positive and negative news distinguished)

Coefficients	Swedbank	SEB Bank
μ	0.0500 (0.0053) [***]	0.0730 (0.0105) [***]
β_1	0.0386 (0.0049) [***]	-0.0804 (0.0087) [***]
β_2	-0.0657 (0.0072) [***]	-0.0631 (0.0107) [***]
β_3	0.0461 (0.0046) [***]	0.0270 (0.0081) [***]
β_4	-0.1273 (0.0295) [***]	-0.0809 (0.0161) [***]
$News_t^-$	0.0000 (0.0000)	0.0000 (0.0000)
$News_{t-1}^-$	0.0000 (0.0000)	0.0000 (0.0000)
$News_{t-2}^-$	0.1102 (0.0260) [***]	0.3125 (0.0699) [***]
$News_t^+$	0.0000 (0.0000)	0.0000 (0.0000)
$News_{t-1}^+$	0.0000 (0.0000)	0.0000 (0.0000)
$News_{t-2}^+$	0.0000 (0.0000)	0.0000 (0.0000)
N_t^-	-0.3865 (0.0914) [***]	-0.5116 (0.2101) [**]
N_{t-1}^-	-0.0372 (0.0087) [***]	-0.0000 (0.0000)
N_{t-2}^-	-0.0000 (0.0000)	-0.0000 (0.0000)
N_t^+	0.0000 (0.0000)	0.0083 (0.0029) [***]
N_{t-1}^+	0.0865 (0.0111) [***]	0.1278 (0.0386) [***]
N_{t-2}^+	0.0959 (0.0210) [***]	0.0574 (0.0295) [*]
LL	-1272.95	-1347.52
AIC	2579.89	2729.03
BIC	2659.36	2808.50
SSE	1160.73	1401.90
Pval. vs. AR(4)	0.6907	0.3209

Notes: The model is explained under Model 1.3. Standard errors are given in parenthesis. [*], [**], [***] show significance at 0.1, 0.05 and 0.01 significance levels, respectively.

Source: Authors' calculations.

For the Swedbank stock returns, the results interestingly suggest that the content of the news 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 rather than the content of news. The impact of one more piece of negative news affects the returns largely and negatively on the same day and to some extent on the next day. This is true even if the news of little importance. It seems that when there is large number of negative news from different sources, the returns are pulled down to a large extent on the same day. In contrast, if there is only one source providing bad news whose size might be large (which is perhaps speculative or fake news), the effect on the returns is relatively smaller and with a delay. Typically, speculative or fake news do not appear in many sources on the same day. Therefore, in this model the return process takes into account the probability that the news might be fake if it appears in one source¹⁶ Finally, the effect of the number of positive news affect the returns with delays of one and two days, but not on the same day.

For the stock returns of SEB Bank, the findings are similar to the case of Swedbank returns. The estimation results 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 negatively 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. Finally, we can say that SEB Bank returns are more exposed to the negative news and to the number of negative news compared to Swedbank returns.

In all these estimations for the conditional mean, the effect of news sentiments were restricted to be positive, for which the explanation was provided in Section 3.3. In fact, this restriction and the results which follow are in line with the findings of Arik (2011), where the author finds positive relationship between the news sentiments and S&P 500 excess returns.

We found out that the lowest AIC and BIC values among the conditional mean models we estimated belonged to the AR(4) model without news sentiment variables. Hence, we use the residuals taken from this model in the GARCH model estimations. As in the conditional mean models, we keep our attention on the extended models rather than the base GARCH model.

When we look at the AIC and BIC values of the Tables 8 and 9, we noticed that for Swedbank volatilities GJR-GARCH model with asymmetric news sentiments perform better, while for SEB volatilities NAGARCH model with asymmetric news sentiments is a better fit. For Swedbank volatilities GJR and NAGARCH results are similar for the coefficients of the news sentiment variables. However, for SEB volatilities, NAGARCH model with news sentiments variables have several more significant coefficients.

For Swedbank volatilities, the estimation results for Model 2.1 in Table 8 indicate that

¹⁶Another explanation could be that when some certain news appears in many different news sources, then it reaches to many investors who adjust their portfolios accordingly.

Table 8: Estimation results for the GJR-GARCH(1, 1) model with asymmetric news

Coefficients	Swedbank	SEB Bank
γ	0.1221 (0.0391) [***]	0.0691 (0.0111) [***]
α_1	0.0200 (0.0026) [***]	0.0440 (0.0152) [***]
α_2	0.8205 (0.0303) [***]	0.8921 (0.0189) [***]
θ	0.1595 (0.0118) [***]	0.0639 (0.0136) [***]
<i>Reuters</i> _t ⁻	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>TheLocal</i> _t ⁻	-0.0000 (0.0000)	-0.4677 (0.2241) [**]
<i>Others</i> _t ⁻	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Baltics</i> _t ⁻	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Reuters</i> _{t-1} ⁻	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>TheLocal</i> _{t-1} ⁻	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Others</i> _{t-1} ⁻	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Baltics</i> _{t-1} ⁻	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Reuters</i> _{t-2} ⁻	-0.2076 (0.0228) [***]	-0.0000 (0.0000)
<i>TheLocal</i> _{t-2} ⁻	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Others</i> _{t-2} ⁻	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Baltics</i> _{t-2} ⁻	-0.0308 (0.0048) [***]	-0.0000 (0.0000)
<i>Reuters</i> _t ⁺	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>TheLocal</i> _t ⁺	-0.0000 (0.0000)	-0.1932 (0.0318) [***]
<i>Others</i> _t ⁺	-0.2063 (0.0494) [***]	-0.0000 (0.0000)
<i>Baltics</i> _t ⁺	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Reuters</i> _{t-1} ⁺	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>TheLocal</i> _{t-1} ⁺	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Others</i> _{t-1} ⁺	-0.1111 (0.0158) [***]	-0.4089 (0.0724) [***]
<i>Baltics</i> _{t-1} ⁺	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Reuters</i> _{t-2} ⁺	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>TheLocal</i> _{t-2} ⁺	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Others</i> _{t-2} ⁺	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Baltics</i> _{t-2} ⁺	-0.0000 (0.0000)	-0.0000 (0.0000)
LL	-1214.47	-1275.20
AIC	2484.94	2606.39
BIC	2615.82	2737.28
Pval. vs. GARCH	0.3223	0.2987
Pval. vs. GJR-GARCH	0.9679	0.7615
Pval. vs. GARCH with asym. news	0.0027	0.0146

Notes: The model is explained under Model 2.1. Standard errors are given in parenthesis. [*], [**], [***] show significance at 0.1, 0.05 and 0.01 significance levels, respectively.

Source: Authors' calculations.

there is a significant asymmetric effect of the return shocks on the volatility, indicated by the value of θ . On the other hand there are only a few significant coefficients for the news sentiment variables. According to the results presented in Table 8, less bad news sentiments from Reuters and Baltics would decrease volatility two days later. On the other hand an increase in positive news sentiments in the Others news source would decrease volatility on the same day and next day. We should say that this result is not in line with the result of GARCH model with news sentiment variables (no asymmetry),

Table 9: Estimation results for the NA-GARCH(1, 1) model with asymmetric news

Coefficients	Swedbank	SEB Bank
γ	0.0998 (0.0211) [***]	0.0215 (0.0041) [***]
α_1	0.0945 (0.0141) [***]	0.0145 (0.0021) [***]
α_2	0.8191 (0.0221) [***]	0.8783 (0.0193) [***]
δ	-0.4934 (0.1056) [***]	-2.6273 (0.3116) [***]
<i>Reuters</i> s_t^-	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>TheLocal</i> s_t^-	-0.0000 (0.0000)	-0.4663 (0.1288) [***]
<i>Others</i> s_t^-	-0.0000 (0.0000)	-0.0126 (0.0011) [***]
<i>Baltics</i> s_t^-	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Reuters</i> s_{t-1}^-	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>TheLocal</i> s_{t-1}^-	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Others</i> s_{t-1}^-	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Baltics</i> s_{t-1}^-	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Reuters</i> s_{t-2}^-	-0.1988 (0.0296) [***]	-0.0829 (0.0168) [***]
<i>TheLocal</i> s_{t-2}^-	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Others</i> s_{t-2}^-	-0.0000 (0.0000)	-0.1430 (0.0370) [***]
<i>Baltics</i> s_{t-2}^-	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Reuters</i> s_t^+	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>TheLocal</i> s_t^+	-0.0000 (0.0000)	-0.1460 (0.0177) [***]
<i>Others</i> s_t^+	-0.2086 (0.0189) [***]	-0.0000 (0.0000)
<i>Baltics</i> s_t^+	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Reuters</i> s_{t-1}^+	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>TheLocal</i> s_{t-1}^+	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Others</i> s_{t-1}^+	-0.0750 (0.0079) [***]	-0.1902 (0.0352) [***]
<i>Baltics</i> s_{t-1}^+	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Reuters</i> s_{t-2}^+	-0.0000 (0.0000)	-0.1539 (0.0262) [***]
<i>TheLocal</i> s_{t-2}^+	-0.0000 (0.0000)	-0.0000 (0.0000)
<i>Others</i> s_{t-2}^+	-0.0000 (0.0000)	-0.0585 (0.0068) [***]
<i>Baltics</i> s_{t-2}^+	-0.0000 (0.0000)	-0.0000 (0.0000)
LL	-1215.91	-1270.69
AIC	2487.82	2597.38
BIC	2618.71	2728.27
Pval. vs. GARCH	0.5374	0.0297
Pval. vs. NAGARCH	0.9738	0.5455
Pval. vs. GARCH with asym. news	0.0135	0.0001

Notes: The model is explained under Model 2.2. Standard errors are given in parenthesis. [*], [**], [***] show significance at 0.1, 0.05 and 0.01 significance levels, respectively.

Source: Authors' calculations.

which revealed that an increase in the negative news sentiments in The Local news source would decrease the volatility on the same day. Since this model is clearly outperformed by the GJR-GARCH extension, we do not present its results in this paper. However they are available upon request from the authors. The probability value for the likelihood ratio test of the GJR-GARCH model with news sentiments variables versus the GARCH model with news sentiments variables is less than 1% significance level. Looking at the other p-values, we see that the news sentiment variables are jointly insignificant when compared

to a GJR-GARCH model.

For SEB volatilities we concentrate our attention to the results of Model 2.2 in Table 9. A very interesting result is that the nonlinear specification (NAGARCH) for the asymmetric effect of the return shocks works better for the SEB volatilities than the linear counterpart GJR-GARCH. In fact the asymmetry coefficient δ is quite large. The results indicate that less bad news from The Local and Others news sources would decrease the volatility significantly on the same day and the effect continues two days later for the Others news source. We also notice that less bad news from Reuters decreases volatility in two days. When looking at the positive news sentiments, we can see that the more good news from The Local would decrease volatility on the same day. Similar effect is seen for the Others news source on the next day and the third day. The effect from Reuters arrives in two days as it was for the negative news sentiments. We can notice that the NAGARCH coefficient and the news sentiments coefficients are jointly significant at 5% compared to a GARCH specification.

The similarity between the results for Swedbank and SEB Bank is that both volatility series maintain asymmetric effects from return shocks and the effect of Swedish news from Reuters source on the volatility of the returns arrive with two days lag.

The result that negative and positive news sentiments affect the volatility differently is in line with the works of Engle and Ng (1993) and Tetlock (2007). However, based on the likelihood ratio tests we couldn't find clear evidence that news sentiments jointly improve the fit over and above the asymmetric GARCH models: GJR and NAGARCH. Only a few coefficients were individually significant. On the other hand, the GJR and NAGARCH models clearly present a better performance compared to a GARCH(1, 1) model.

5.2 Further Discussion

We considered some other models in order to see whether different ways of introducing news sentiments would make a difference. In particular, we estimated smaller models which are nested in the ones we discuss in this paper. We were able to calculate the likelihood ratio tests from these estimations. We did not include the estimation results of these other models in this paper, however they are available upon request.

For the conditional returns, we extended the AR(4) model with news sentiments in Equation 2 by including dummy variables for extreme news. These dummy variables took value 1 when the news sentiment were 3 standard deviations away from their means. They were also distinguished for positive and negative news. We discarded this model since it presented high level of multicollinearity between the news sentiments variables and the dummy variables. In addition, most of the coefficient estimates for the dummy variables were equal to zero. We also considered a set up where we extended the AR(4) model with news sentiments in Equation 2 by adding the same news sentiment variables but multiplying them with an indicator that takes value 1 for negative news sentiments. This

yields a similar construction as in the GJR-GARCH model. We didn't include this model in our paper since the interpretation of the results we obtained for it were similar to that of Model 1.2 that considered negative and positive news sentiments separately.

For modelling the conditional volatility, we extended the GJR-GARCH and NA-GARCH models to include merged news as in Equation 4. Except for the number of negative news sentiments, all coefficients were restricted to be negative. Interestingly, almost all the coefficients for these news sentiments were estimated to be zeros for both banks' return volatilities. That's why we didn't include this model in the paper. The only exception was that in the GJR-GARCH results for the Swedbank volatilities, the coefficient of the number of negative news with two days lag was estimated to be 0.0751 and it was significant.

Another possible question could be about the distribution assumption on the return shocks. Financial returns usually have leptokurtic distributions. This is confirmed in the descriptive statistics of our returns data as well. In this paper we assumed that these shocks are normally distributed to be able to use the two step estimation procedure. We first estimated the conditional mean parameters, and after that we estimated the conditional volatility parameters. This idea was mentioned in Bauwens et al. (2006) and with that motivation numerically analyzed by Carnero and Eratalay (2014). De Almeida et al. (2018) also mention this estimation method that the returns or the residuals of a VARMA model can be used for the two step estimation procedure (volatility+correlation) of a DCC-GARCH model, which corresponds to the same estimation method. We followed this procedure to avoid possible convergence problems in numerical optimization during the implementation of the model, in particular when the models have too many parameters to estimate. If we had assumed a Student-t distribution for the return shocks, then we would have had to estimate the conditional mean and variance parameters altogether, which would have brought about this problem.

6 Conclusions

In this paper, we analysed the effect of the sentiment of news related to the real estate market in Sweden, Estonia, Latvia, and Lithuania, on the stock returns and volatilities of Swedbank and SEB Bank. In this analysis, we considered the period between 04.01.2016 and 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 considered 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.

We found that, for both banks the local Swedish news has more immediate and stronger

effect than the Baltic and Reuters news. Moreover, the positive and negative news sentiments affect the returns and volatilities of Swedbank and SEB Bank stocks differently. 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 of this is that the bank stock returns may not react as strongly to one big negative news but may react much more sharply to many small negative news. We also took into consideration the time lags, and we found that there was a difference between how soon the two banks react to the news. SEB Bank stock returns respond slower to Baltic news than Swedbank stock returns. On the other hand, SEB Bank stock returns are affected by negative Reuters news on the same day, while Swedbank stock returns are affected with a delay. Finally, we found out significant asymmetric effect of the return shocks on the volatilities of the two banks' returns. When this asymmetric effect was taken into account, the volatility of Swedbank stock returns was influenced by negative news from Reuters and Baltics with two days delay, and by positive news from local news sources on the same day and the next day. Similarly, the volatility of SEB returns depend on the news from Reuters with a delay of two lags. The volatility of SEB stock returns was influenced more by both negative and positive news from the local news sources on the same day and with two days lag. It is surprising to see that the volatility of SEB stock returns do not depend on Baltics news sentiments, and the volatility of Swedbank returns depend on it to a small extent.

This paper can be further extended 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 behaviour can be explained by this. Another extension could be to analyse the posts shared in social media such as Facebook, VK or 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.

References

- Acharya, V. V., Eisert, T., Eufinger, C., and Hirsch, C. (2018). Real Effects of the Sovereign Debt Crisis in Europe: Evidence from Syndicated Loans. *The Review of Financial Studies*, 31(8):2855–2896.
- Antweiler, W. and Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, 59(3):1259–1294.

- Arik, A. (2011). Modeling Market Sentiment and Conditional Distribution of Stock Index Returns Under GARCH Process. *Claremont Graduate University*.
- Asai, M. and McAleer, M. (2006). Asymmetric Multivariate Stochastic Volatility. *Econometric Reviews*, 25(2-3):453–473.
- Bailey, M., Dávila, E., Kuchler, T., and Stroebel, J. (2017). House Price Beliefs and Mortgage Leverage Choice.
- Balasubramaniam, K. (2018). How to Calculate a Stock's Adjusted Closing Price. <https://www.investopedia.com>.
- Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics*, 21(1):79–109.
- Beetsma, R., De Jong, F., Giuliadori, M., and Widiyanto, D. (2017). Realized (Co) Variances of Eurozone Sovereign Yields During the Crisis: The Impact of News and the Securities Markets Programme. *Journal of International Money and Finance*, 75:14–31.
- Beetsma, R., Giuliadori, M., De Jong, F., and Widiyanto, D. (2013). Spread the News: The Impact of News on the European Sovereign Bond Markets During the Crisis. *Journal of International Money and Finance*, 34:83–101.
- Bollen, J., Mao, H., and Zeng, X. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, 2(1):1–8.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Bollerslev, T. and Wooldridge, J. M. (1992). Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances. *Econometric Reviews*, 11(2):143–172.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. (1970). Time Series Analysis: Forecasting and Control Holden-Day San Francisco. *BoxTime Series Analysis: Forecasting and Control Holden Day1970*.
- Bradley, M. M. and Lang, P. J. (1999). Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. Technical report, Citeseer.
- Braun, P. A., Nelson, D. B., and Sunier, A. M. (1995). Good News, Bad News, Volatility, and Betas. *The Journal of Finance*, 50(5):1575–1603.
- Brook, C. and Burke, S. P. (2003). Information Criteria for GARCH Model Selection: An Application to High Frequency Data. *European Journal of Finance*, 9(6):557–580.

- Carmichael, B. and Coën, A. (2018). Real Estate as a Common Risk Factor in Bank Stock Returns. *Journal of Banking & Finance*, 94:118–130.
- Carnero, M. A. and Eratalay, M. H. (2014). Estimating VAR-MGARCH Models in Multiple Steps. *Studies in Nonlinear Dynamics & Econometrics*, 18(3):339–365.
- Cepoi, C.-O. (2020). Asymmetric Dependence Between Stock Market Returns and News During COVID19 Financial Turmoil. *Finance Research Letters*.
- De Almeida, D., Hotta, L. K., and Ruiz, E. (2018). MGARCH Models: Trade-off Between Feasibility and Flexibility. *International Journal of Forecasting*, 34(1):45–63.
- De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4):703–738.
- Deng, S., Huang, Z. J., Sinha, A. P., and Zhao, H. (2018). The Interaction Between Microblog Sentiment and Stock Return: An Empirical Examination. *MIS Quarterly*, 42(3):895–918.
- Ding, D., Huang, X., Jin, T., and Lam, W. R. (2017). *Assessing China's Residential Real Estate Market*. International Monetary Fund.
- Ederington, L. H. and Lee, J. H. (1993). How Markets Process Information: News Releases and Volatility. *The Journal of Finance*, 48(4):1161–1191.
- Engelberg, J. E. and Parsons, C. A. (2011). The Causal Impact of Media in Financial Markets. *The Journal of Finance*, 66(1):67–97.
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.
- Engle, R. F., Focardi, S. M., and Fabozzi, F. J. (2012). ARCH/GARCH Models in Applied Financial Econometrics. *Encyclopedia of Financial Models*.
- Engle, R. F. and Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of finance*, 48(5):1749–1778.
- Engle, R. F. and Sheppard, K. (2001). Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. Technical report, National Bureau of Economic Research.
- Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1):34–105.
- Finansinspektionen (2018). The Swedish Mortgage Market.

- Fisher, K. L. and Statman, M. (2000). Investor Sentiment and Stock Returns. *Financial Analysts Journal*, 56(2):16–23.
- Ghosh, C., Guttery, R. S., and Sirmans, C. F. (1997). The Effects of the Real Estate Crisis on Institutional Stock Prices. *Real Estate Economics*, 25(4):591–614.
- Ghysels, E., Harvey, A. C., and Renault, E. (1996). 5 Stochastic Volatility. *Handbook of Statistics*, 14:119–191.
- Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, 48(5):1779–1801.
- Groß-Klußmann, A. and Hautsch, N. (2011). When Machines Read the News: Using Automated Text Analytics to Quantify High Frequency News-Implied Market Reactions. *Journal of Empirical Finance*, 18(2):321–340.
- Gujarati, D. N. (2003). Basic Econometrics: Fourth Edition McGraw-Hill. *New York*.
- Gupta, K. and Banerjee, R. (2019). Does OPEC News Sentiment Influence Stock Returns of Energy Firms in the United States? *Energy Economics*, 77:34–45.
- Hansen, P. R. and Lunde, A. (2005). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH (1, 1)? *Journal of Applied Econometrics*, 20(7):873–889.
- Harris, Z. S. (1954). Distributional Structure. *Word*, 10(2-3):146–162.
- Hausler, J., Ruschinsky, J., and Lang, M. (2018). News-Based Sentiment Analysis in Real Estate: A Machine Learning Approach. *Journal of Property Research*, 35(4):344–371.
- He, L. T., Myer, F. C. N., and Webb, J. R. (1996). The Sensitivity of Bank Stock Returns to Real Estate. *The Journal of Real Estate Finance and Economics*, 12(2):203–220.
- Hu, M. and Liu, B. (2004). Mining Opinion Features in Customer Reviews. In *AAAI*, volume 4, pages 755–760.
- Hutto, C. J. and Gilbert, E. (2014). Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *Eighth International AAAI Conference on Weblogs and Social Media*.
- Hutto, C. J., Klein, E., Pantone, P., and Berry, G. (2019). Vader Lexicon. <https://github.com/cjhutto/vaderSentiment/tree/master/vaderSentiment>.
- Igan, D. and Pinheiro, M. (2010). Exposure to Real Estate in Bank Portfolios. *Journal of Real Estate Research*, 32(1):47–74.

- Kelly, S. (2016). *News, Sentiment and Financial Markets: A Computational System to Evaluate the Influence of Text Sentiment on Financial Assets*. PhD thesis, Trinity College Dublin.
- Kosapattarapim, C., Lin, Y.-X., and McCrae, M. (2012). Evaluating the Volatility Forecasting Performance of Best Fitting GARCH Models in Emerging Asian Stock Markets.
- Kothari, S. P. and Shanken, J. (1997). Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis. *Journal of Financial Economics*, 44(2):169–203.
- Kwan, S. (2019). Banks' Real Estate Exposure and Resilience. *FRBSF Economic Letter*, page 10.
- León, Á., Rubio, G., and Serna, G. (2005). Autoregressive Conditional Volatility, Skewness and Kurtosis. *The Quarterly Review of Economics and Finance*, 45(4-5):599–618.
- Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., and Chen, Y. (2014). The Effect of News and Public Mood on Stock Movements. *Information Sciences*, 278:826–840.
- Liu, W. and Morley, B. (2009). Volatility Forecasting in the Hang Seng Index Using the GARCH Approach. *Asia-Pacific Financial Markets*, 16(1):51–63.
- Maheu, J. M. and McCurdy, T. H. (2004). News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns. *The Journal of Finance*, 59(2):755–793.
- Malik, F. (2011). Estimating the Impact of Good News on Stock Market Volatility. *Applied Financial Economics*, 21(8):545–554.
- Martins, A. M., Serra, A. P., and Martins, F. V. (2016). Real Estate Market Risk in Bank Stock Returns: Evidence for 15 European Countries. *International Journal of Strategic Property Management*, 20(2):142–155.
- Mei, J. and Lee, A. (1994). Is There a Real Estate Factor Premium? *The Journal of Real Estate Finance and Economics*, 9(2):113–126.
- Mei, J. and Saunders, A. (1995). Bank Risk and Real Estate: An Asset Pricing Perspective. *The Journal of Real Estate Finance and Economics*, 10(3):199–224.
- Mitchell, M. L. and Mulherin, J. H. (1994). The Impact of Public Information on the Stock Market. *The Journal of Finance*, 49(3):923–950.
- Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica: Journal of the Econometric Society*, pages 347–370.
- NumFOCUS. Python Data Analysis Library. <https://pandas.pydata.org/>.

- Olowe, R. A. (2009). Modelling Naira/Dollar Exchange Rate Volatility: Application of GARCH and Assymetric Models. *International Review of Business Research Papers*, 5(3):377–398.
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. *PLOS ONE*, 10(9):e0138441.
- Richardson, L. Beautiful Soup 4.4.0 documentation. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.
- Robert, E. and Victor, N. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, 48(5):1749–1778.
- Sidorov, S. P., Date, P., and Balash, V. (2014a). GARCH Type Volatility Models Augmented With News Intensity Data. In *Chaos, Complexity and Leadership 2012*, pages 199–207. Springer.
- Sidorov, S. P., Revutskiy, A., Faizliev, A., Korobov, E., and Balash, V. (2014b). Stock Volatility Modelling With Augmented GARCH Model With Jumps. *IAENG International Journal of Applied Mathematics*, 44(4):212–220.
- Soo, C. (2015). Quantifying Animal Spirits: News Media and Sentiment in the Housing Market. *Ross School of Business Paper*, (1200).
- Soo, C. K. (2018). Quantifying Sentiment With News Media Across Local Housing Markets. *The Review of Financial Studies*, 31(10):3689–3719.
- Soroka, S. N. (2006). Good News and Bad News: Asymmetric Responses to Economic Information. *The Journal of Politics*, 68(2):372–385.
- Taylor, S. J. (1994). Modeling Stochastic Volatility: A Review and Comparative Study. *Mathematical Finance*, 4(2):183–204.
- Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62(3):1139–1168.
- Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms’ Fundamentals. *The Journal of Finance*, 63(3):1437–1467.
- Tumarkin, R. and Whitelaw, R. F. (2001). News or Noise? Internet Postings and Stock Prices. *Financial Analysts Journal*, 57(3):41–51.
- Verma, R. and Soydemir, G. (2009). The Impact of Individual and Institutional Investor Sentiment on the Market Price of Risk. *The Quarterly Review of Economics and Finance*, 49(3):1129–1145.

- Walker, C. B. (2016). The Direction of Media Influence: Real-Estate News and the Stock Market. *Journal of Behavioral and Experimental Finance*, 10:20–31.
- White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pages 817–838.
- Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*, volume 4. Almqvist & Wiksells.
- Yu, X. (2014). *Analysis of New Sentiment and Its Application to Finance*. PhD thesis.
- Zhang, X., Fuehres, H., and Gloor, P. A. (2011). Predicting Stock Market Indicators Through Twitter “I hope it is not as bad as I fear”. *Procedia-Social and Behavioral Sciences*, 26:55–62.

A 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>

B Estimation results presented in detail by equations

In what follows we present the estimation results of other models. In each table LL is the loglikelihood value, AIC and BIC are Akaike and Bayesian Information Criterion, respectively, and SSE is the sum of squared residuals of the estimation results.

Model 1.0. AR(4) model (base-line model)

Estimation results of AR(4) model, which is the base-line model, for Swedbank and SEB Bank stock returns.

Table 10: Estimation results for the AR(4) model (base-line model)

Coefficients	Swedbank	SEB Bank
μ	0.0489 (0.0040) [***]	0.0547 (0.0104) [***]
β_1	0.0422 (0.0066) [***]	-0.0811 (0.0167) [***]
β_2	-0.0692 (0.0097) [***]	-0.0637 (0.0083) [***]
β_3	0.0475 (0.0062) [***]	0.0295 (0.0025) [***]
β_4	-0.1299 (0.0321) [***]	-0.0793 (0.0086) [***]
LL	-1277.52	-1354.36
AIC	2565.04	2718.72
BIC	2588.41	2742.09
SSE	1174.24	1426.41

Notes: The model is explained under Model 1.1. Standard errors are given in parenthesis. [*], [**], [***] show significance at 0.1, 0.05 and 0.01 significance levels, respectively.

Source: Authors' calculations.

Model 1.1. AR(4) model model with news component

Estimation results of AR(4) model with news component for Swedbank and SEB Bank stock returns.

Table 11: Estimation results for the AR(4) model with news component

Coefficients	Swedbank	SEB Bank
μ	0.0087 (0.0017) [***]	0.0115 (0.0046) [**]
β_1	0.0350 (0.0074) [***]	-0.0811 (0.0131) [***]
β_2	-0.0677 (0.0109) [***]	-0.0623 (0.0237) [***]
β_3	0.0509 (0.0119) [***]	0.0297 (0.0030) [***]
β_4	-0.1250 (0.0345) [***]	-0.0798 (0.0258) [***]
<i>Reuters_t</i>	0.0000 (0.0000)	0.0000 (0.0000)
<i>TheLocal_t</i>	0.6350 (0.2026) [***]	0.5736 (0.1613) [***]
<i>Others_t</i>	0.2066 (0.0790) [***]	0.5816 (0.2274) [**]
<i>Baltics_t</i>	0.0000 (0.0000)	0.0000 (0.0000)
<i>Reuters_{t-1}</i>	0.3883 (0.1281) [***]	0.2663 (0.1944)
<i>TheLocal_{t-1}</i>	0.0000 (0.0000)	0.0716 (0.0317) [**]
<i>Others_{t-1}</i>	0.0000 (0.0000)	0.0000 (0.0000)
<i>Baltics_{t-1}</i>	0.0720 (0.0189) [***]	0.0000 (0.0000)
<i>Reuters_{t-2}</i>	0.3064 (0.1314) [**]	0.1345 (0.0362) [***]
<i>TheLocal_{t-2}</i>	0.2471 (0.0882) [***]	0.0543 (0.0153) [***]
<i>Others_{t-2}</i>	0.0000 (0.0000)	0.0015 (0.0009)
<i>Baltics_{t-2}</i>	0.1038 (0.0328) [***]	0.2323 (0.1281) [*]
LL	-1270.30	-1346.48
AIC	2574.59	2726.97
BIC	2654.06	2806.43
SSE	1152.97	1398.25
Pval. vs. AR(4)	0.2732	0.2028

Notes: The model is explained under Model 1.1. Standard errors are given in parenthesis. [*], [**], [***] show significance at 0.1, 0.05 and 0.01 significance levels, respectively.

Source: Authors' calculations.