[redacted]

Identification of Risks in the Banking Sector using Natural Language Processing

Master Thesis

Bela Tim Koch

Solothurn

[redacted]

Advisor:

[redacted]

Contents

1. Introduction

Supervisory agencies need to identify weaknesses of the institutions under supervision at an early stage to be able to react in a timely manner. The prevented collapse of the globally active and systematically important Swiss bank Credit Suisse by regulatory intervention in early 2023 have again demonstrated the necessity of banking supervision and the need for quick reactions to financial distressed banks to sustain financial and general economic stability. A few months prior to the 2023 banking crisis, openAI first launched ChatGPT, providing the general public access to their large language model (LLM) what drew huge public attention. As of August 2024, chatGPT has 200 million weekly users (Reuters, 2024), indicating the huge potential of the application of LLMs, which potentially could improve the banking supervision process. This process involves monitoring of a banks' risk management and controls, internal processes and procedures, governance and financial and operational soundness by reviewing banks' internal documents, discussion with banks' personnel and independent analysis (Hirtle and Kovner, 2022). In their work, supervisory agencies are often faced with resource constraints (Prenio, 2024). Hence, supervisory agencies such as the Swiss Financial Market Supervision Authority (FINMA) adopt a risk-oriented supervisory approach, concentrating on institutions that demand increased attention according to their supervisory category and a dynamic rating process (Swiss Financial Market Supervisory Authority, 2024).

To improve or extend this rating process, this study tries to leverage information published by news media providers. Given a set of news articles, a sentiment score is sought which captures the tone of the news media about the corresponding bank. Hence, the fast amount of text published by news media sources must be filtered for relevant information about individual banks, before the sentiment of this information can be classified as negative, neutral or positive. The goal is to construct an indicator which signals increased risk of the corresponding bank before traditional measures. This study hence contributes to the current initiatives by supervisory authorities in developing novel tools based on artificial intelligence and big data analytics to support their supervisory processes. Following Broeders and Prenio (2018), these novel methods are referred to as supervisory technology (suptech). As reported in Prenio (2024), a survery conducted by the Financial Stability Institute (FSI) and the Bank of International Settlements (BIS) Innovation Hub in 2023 shows that 47 of the 50 surveyed supervisory agencies have ongoing suptech initiatives, whereby 44 of them have already deployed a suptech tool.

Quantifying news sentiment and using the resulting sentiment scores for prediction is inspired by the financial literature, which has shown several promising results. Hence, besides focusing on sentiment analysis for improving the investment decision-making process, there is a branch of the literature which focuses on financial distress prediction. The literature can be categorised to corporation-expressed, media-expressed and internet-expressed sentiment analysis. Traditional literature in sentiment analysis for financial research questions uses traditional methods such as bag-of-words models to classify the sentiment of text and hence does not leverage the possibilities provided by novel methods. Additionally, Swiss banks which are not classified as global systemically important are not typically covered in research studies.

This study contributes to the media-expressed sentiment for financial distress literature by using a large language model for the classification of text and by testing the predictive performance of the retrieved sentiment indicator on data from 2022 until 2024 for a sample of Swiss and European banks. More specifically, we classify the sentiment using a Bidirectional Encoder Representations from Transformers (BERT) model developed by Google researchers in 2018 (Devlin et al., 2019), which was further fine-tuned on financial texts. Due to its design, the model is able to understand words in context of the remaining text and hence is expected to perform better when assessing the sentiment of a text compared to traditional methods. We assess the predictive power of this sentiment indicator for common risk proxies to determine whether the sentiment indicator is able to forecast the riskiness of the corresponding bank. The results show, dependent on the sample and risk proxy, that the news sentiment indicator has predictive power. Some further modifications in the construction of the indicator and in the application for forecasting are necessary to implement it in the context of an early warning system.

This paper is structured as follows. First, we present the proposed procedure to obtain the sentiment scores in Section 2. Next, in Section 3, we provide an overview of the literature on the foundations of sentiment analysis in finance, the current research of sentiment analysis relevant in a supervisory context, and the methods used to measure the financial distress of banks. In Section 4, we describe the methods used to construct the news sentiment indicator and discuss the benefits and limitations. We present the data used in this analysis, the experimental design as well as its results in Section 5. Finally, we conclude in Section 6.

2. Sentiment Analysis for Supervision

First published in 1997 by the Basel Committee on Banking Supervision (BCBS), the 29 Core Principles for effective banking supervision, or the Basel Core Principles (BCPs), are the minimum global standard for the sound prudential regulation and supervision of banks and banking systems (Bank of International Settlements, 2024). These BCPs can be categorised to BCPs 1-13 which define the powers, responsibilities and functions of supervisors and BCPs 14-29 about prudential regulations and requirements for banks. Relevant for this study are BCPs 8-13, which outline various tools and techniques aimed at facilitating robust on-site and off-site supervision of banks and banking groups. Particularly BCP 8 which defines the supervisory approach states, that the supervisor develops and maintains a forward-looking assessment of the risk profile of individual banks.

The turmoils in the banking sector in early 2023 highlighted the importance of this forward-looking risk assessment. As Prenio (2024) states, a lack of adequate resources and effective tools affects the ability of supervisory authorities for timely interventions. Besides investments in human resources, supervisory authorities are investing in innovative tools using novel technology (Prenio, 2024). Broeders and Prenio (2018) defined the use of innovative technology by supervision agencies to support their processes as supervision technology, or suptech. According to Prenio (2024), there is much hope that suptech helps to enhance supervisory ability, whereby the developments in generative artificial intelligence might provide the potential of suptech to be a transformative force in financial supervision.

Facing this need for effective tools in banking supervision and opportunities offered by artificial intelligence, we introduce an indicator which measures the sentiment of the news media on banks under supervision. Given a corpus of news articles, we want to determine a single score which quantifies the sentiment of media coverage for the corresponding bank for a given time period. This score should range from -1 for the most negative to 1 for the most positive, with 0 indicating neutrality. To construct this indicator, the following problems must be addressed. First, the corpus must be filtered for articles which contain relevant information about the bank in question. Then, within these articles, the specific passages of the articles containing relevant content must be identified. All these relevant passages must then be classified as positive, neutral or negative. These classifications must then be combined and processed to generate a single sentiment score of the corresponding news article. Subsequently, the articles must be further aggregated to generate a single sentiment score of the corresponding bank for a given time period. The resulting news sentiment indicator reduces the vast amount of published information to a single score and prevents the necessity of manually processing the news about multiple banks. Forward-looking risk assessment could benefit from this news sentiment indicator by proxying new information which is not or not yet reflected in other metrics.

3. Literature Review

3.1 Foundations of Sentiment Analysis in Finance

While the application of sentiment analysis in banking supervision is a relatively new field, De Bondt and Thaler (1985) and Cutler et al. (1988) already showed that financial markets react to news. De Bondt and Thaler (1985) show that they do not only react but tend to overreact to unexpected and dramatic news events. Tetlock (2007) was one of the first which constructed a quantitative indicator based on news article sentiment to analyse the interactions between news articles and the stock market by using a bag-of-words algorithm. The results show that high media pessimism predicts falling stock prices. Du et al. (2024) show that most literature focuses on applications in predictive analytics for financial metrics such as asset prices or the corresponding volatility. Applications mainly include investment management, with coverage of the stock market, foreign exchange market, and cryptocurrency market.

3.2 Sentiment Analysis for Banking Supervision

While research in financial sentiment analysis methods often focuses on the investment decision-making process, they may also have potential applicability in the context of banking supervision. The body of literature directly related to sentiment analysis in banking supervision remains limited. Hence, we also cover studies of sentiment analysis for general financial distress prediction, which could easily be applied to the banking sector specifically. Following Kearney and Liu (2014), we categorise the sentiment analysis for banking supervision literature according to their corresponding data source to corporation-expressed sentiment, media-expressed sentiment and internet-expressed sentiment.

Corporation-expressed sentiment analysis tries to capture the tone of the corporate executives regarding the performance, management and the strategy of a firm based on various corporate disclosures. Hajek and Munk (2023) and Huang et al. (2023) show that the predictive performance of general financial distress prediction methods for firms can be improved by extending them with sentiment measures which are based on firms' annual reports. By analysing the U.S. banking sector, Gandhi et al. (2019) suggests that the early warning systems of banking supervision authorities which are based on financial statement data could be extended by an indicator derived from the frequency of negative words in the banks' annual reports. While these studies focus on classifying the sentiment of text, Hajek and Munk (2023) show that speech emotion recognition can be applied on transcripts of earning conference calls and show that managerial emotions recognition can improve financial distress prediction.

Focusing on content published by news agencies, media-expressed sentiment tries to capture the tone in a news article. Smales (2016) analyses the media-expressed sentiment on banks using news sentiment data from Thomson Reuters News Analytics (TRNA). The algorithm used by TRNA identifies all sentences where the corresponding bank is mentioned and analyses the related content. The algorithm considers the order of words, adjectives, and common financial expressions. It is trained using several thousand randomly selected news stories which were manually labeled by former market participants (Smales, 2016). TRNA classifies each story as negative, neutral or positive. Smales (2016) computed the sentiment indicator by summing up the classified sentiment of all novel, highly-relevant news stories for the bank considered, whereby negative stories have a score of -1, neutral stories of 0 and positive stories of 1. Smales (2016) uses a sample of major international banks for the observation period from the first of November 2004 until the 31st of December 2010. The results of Smales (2016) suggest that while Credit Default Swap (CDS) spreads as a marked determined measure react to news sentiment, the spread between the London Interbank Offered Rate (LIBOR) and Overnight Indexed Swaps (OIS) as a bank determined measure does not. Furthermore, by analysing the relationship between the sentiment of certain news topics and CDS spreads, the studies of Roeder et al. (2020) suggest that further categorising news sentiment into topics might improve the predictive power of the indicator. They use sentiment scores and topic classifications by RavenPack News Analytics for a sample of systemically important banks which are tracked in the KBW Nasdaq Global Bank Index. The observation period spans from the first of January 2009 until the 31st of May 2018. Both Smales (2016) and Roeder et al. (2020) show that negative sentiment in news articles could indicate higher risk of the corresponding bank. However, as Agoraki et al. (2022) note, positive sentiment can also lead to increased financial instability by increased risk appetite by bank investors. Borovkova et al. (2017) extend the sample to systemically important financial institutions and show that an indicator for systemic risk constructed from news sentiment leads other systemic risk measures by as long as 12 weeks.

Internet-expressed sentiment tries to capture the tone of the discussion in social media posts which can be written by the bank itself, public or private institutions or natural persons. Fernandez et al. (2021) derive a sentiment indicator for the Mexican financial sector as a whole based on tweets, which captures sources of financial stress which are not reflected in quantitative risk measures and show that this indicator correlates with measures for financial risk. Illia et al. (2021) analyse the relationship between the sentiment of tweets about a single bank and its daily business performance and find that tweet sentiment might affect bank performance when embedded in a larger conversation. The recent banking crisis in 2023 opened the discussion whether social media is a new factor which accelerates bank run behavior. Indeed, by analysing the bank run on Silicon Valley Bank in 2023, first results in Cookson et al. (2023) indicate that social media sentiment could amplify bank run risk.

3.3 Measures for Financial Distress of Banks

Since the riskiness of a bank cannot be measured directly, the empirical literature about financial distress of banks relies on indirect measures. Hence, different measures have been developed in the literature which should proxy the health of a bank. Some measures rely on accounting data of the corresponding banks. For example, Sinkey (1978) addresses the question of identifying problematic banks and observes that nearly all failed banks exhibited low net capital ratios. However, most banks with low net capital ratios did not fail. Other studies such as Chiaramonte and Casu (2016) analyse the predictive performance of the Altman Z-score in the banking sector. The main limitation of such measures is the relative low frequency due to the dependency on accounting data. Market measures in contrast are available in higher frequency and hence enable nowcasting the risk of banks. Market measures are derived from financial markets such as the stock or options market and try to get an implied proxy of the perceived risk of market participants. Sarin and Summers (2016) give an overview of market measures such as stock price volatility, option implied volatility, credit default swaps, the price to earnings ratio and preferred stock yields. They state that all indicators are only an imperfect proxy but looking at multiple different indicators enables the assessment of market beliefs.

3.4 Contribution to the Literature

This study builds on existing studies about explaining and predicting risk proxies of banks. First, we test whether news sentiment is an additional determinant of CDS spreads. Annaert et al. (2013) analysed the determinants of CDS spreads for 32 banks in the Euro area for the observation period spanning from December 2003 to September 2010. We extend this model by adding a news sentiment indicator and evaluate the model for European Globally Systematically Important Banks (G-SIBs) for the observation period from September 2023 to October 2024 as well as for UBS and Credit Suisse from January 2022 to June 2023. Next, we modify the model to predict CDS spreads for sovereign debt by Cathcart et al. (2020) to be suitable for banks and evaluate the results on the same sample as before. Additionally, we take up the research question of Roeder et al. (2020) and test whether we can confirm the significant effect of non-financial news sentiment on CDS spreads using our sample. Compared to the studies by Smales (2016) or Roeder et al. (2020), we do not use sentiment scores provided by an external sources but rather leverage LLMs to construct the sentiment scores by ourselves. Additionally to the analysis of CDS spreads, we evaluate the predictive power of news sentiment for stock price decreases as well as stock price volatility. To the best of our knowledge, this is the first study to examine the construction of a sentiment indicator based on Swiss news articles using LLMs to be applied for measuring idiosyncratic risks of Swiss banks.

4. Proposed News Sentiment Indicator

4.1 Overview

As shown in the literature, the sentiment in news articles can be a leading indicator for movements in financial markets. The goal of this analysis is to construct a sentiment indicator from news media articles using a large language model (LLM) to classify the sentiment from a given text. We use Bidirectional Encoder Representations from Transformers (BERT) models which were fine-tuned on financial texts for sentiment classification. Due to its ability to understand the context of words within the remaining text provided, it is expected that a BERT model performs better in correctly classifying the sentiment of text compared to traditional methods such as bag-of-word algorithms. Using news articles from different sources, we construct sentiment indicators in daily and weekly frequency. These indicators should be used as a leading indicator for several risk proxies of a corresponding bank and hence deliver early warning signals for financially distressed banks. In this Chapter, we present the process of estimating the sentiment score of a bank for a given time period. In a first step, described in Section 4.2, we prepare the news article to be further processed. In Section 4.3, we introduce the method of classifying the sentiment. Using this classification, we construct a single sentiment score as described in Section 4.4. The process starting from the news article to the corresponding sentiment score is visualised in Figure 4.1. Additionally, in Section 4.5 we present the possibility to further classify the articles by topic. The Chapter concludes with Section 4.6, where we discuss the benefits and limitations of the constructed indicators.

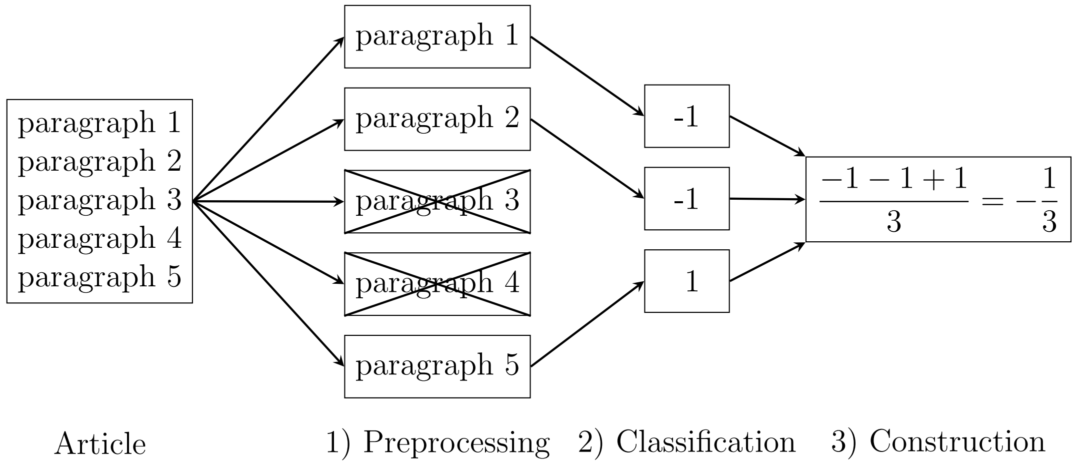

Figure 4.1: The Process from Article to Sentiment Score

4.2 Step 1: Preprocessing of News Articles

Depending on the underlying data source used for this study, it is not guaranteed that an article contains relevant information in the context of the research question. Since the articles are queried from the data source by the banks name, its ticker or a common abbreviation, there will be articles in the sample which contain no relevant content about the financial health of the corresponding bank. There might be articles in the sample that, for example, describe a sports event where the corresponding bank is mentioned as a sponsor. Additionally, it is possible that an article covers different topics or different companies and hence only a part of the article might contain relevant information about the corresponding bank.

Hence, to avoid introducing noise through irrelevant content, a preprocessing procedure is applied to identify articles which contain information about the corresponding bank. First, the article is split into its paragraphs. Using pattern matching and predefined identifying patterns for the corresponding bank, we check for each paragraph whether the bank is mentioned in it and remove paragraphs with no occurrence of the banks identifying pattern. Then, again using pattern matching, it is checked whether the content of these paragraphs is relevant from a financial perspective by checking for the occurence of words from a predefined list of economic and financial keywords. We then classify the sentiment as described in Section 4.3 only on these remaining paragraphs from the article.

4.3 Step 2: Sentiment Classification of Text

Once the preprocessing of the news articles is finished, we have a list of paragraphs extracted from a specific article which we assume contain relevant information about the corresponding bank b at time t. Next, for each paragraph, we want to assess whether the sentiment is positive, neutral or negative. For this classification, we use BERT models based on Devlin et al. (2019). Devlin et al. (2019) use a masked language model (MLM) and next sentence prediction (NSP) as two unsupervised tasks to train the BERT model. The MLM randomly masks 15% of tokens in a sentence and predicts the masked tokens using the remaining tokens in the sentence to obtain a bidirectional pre-trained model, and hence gains the ability to understand the relationship between words in a sentence. To further understand sentence relationships, Devlin et al. (2019) pre-train the BERT model with a next sentence prediction task. Hence, the main advantage of using a BERT model in this analysis compared to traditional methods such as lexicon-based models is the ability of the BERT model to understand context. Additionally, fine-tuning a pre-trained BERT model to perform a specific task does not require substantial task-specific modifications of the architecture and is hence relatively easy (Devlin et al., 2019). To classify both German and English text, this analysis utilises a BERT model which was fine-tuned for financial sentiment classification on the corresponding language. Both models are available on `https://huggingface.co`. The English model is described in Araci (2019) and available on `https://huggingface.co/ProsusAI/finbert`. The German model is described in Scherrmann (2023) and available on `https://huggingface.co/scherrmann/GermanFinBert_SC_Sentiment`. Both Araci (2019) and Scherrmann (2023) evaluated the performance of their BERT model using the Financial PhraseBank dataset by Malo et al. (2014). Araci (2019) report an accuracy of 0.84 and an F1-score of 0.84 when evaluating his model over the whole dataset and an accuracy of 0.97 and an F1-score of 0.95 when evaluating his model on a subset of the data with 100% annotator agreement. Scherrmann (2023) evaluates his model on a machine translated version of the Financial PhraseBank dataset by Malo et al. (2014) and reports an accuracy of 0.96 and a macro F1-score of 0.93. Both the model of Araci (2019) and the model of Scherrmann (2023) outperformed the benchmark models shown in their studies. Table 4.1 demonstrates the ability of the German BERT model to correctly classify the sentiment of the text given the context. Note that these texts are only slightly different, yet they express different sentiments, what the model correctly identifies.

Table 4.1 Examples of Sentiment Classification

Input	Classification
Diese Neuigkeiten über die schlechte Entwicklung der allgemeinen Wirtschaftslage hat einen beträchtlichen Einfluss auf die Credit Suisse.	Negative
Diese Neuigkeiten über die schlechte Entwicklung der allgemeinen Wirtschaftslage hat keinen beträchtlichen Einfluss auf die Credit Suisse.	Neutral
Diese Neuigkeiten über die schlechte Entwicklung der allgemeinen Wirtschaftslage hat einen beträchtlich vorteilhaften Einfluss auf die Credit Suisse.	Positive

The sentiment s towards bank b at time t of each paragraph p and article a in the list resulting of the preprocessing procedure as described in Section 4.2 is classified by the fine-tuned BERT model as follows:

$$s_{b,t,a,p} = \begin{cases} -1 & \text{if the paragraph } p \text{ is classified as negative} \\ 0 & \text{if the paragraph } p \text{ is classified as neutral} \\ 1 & \text{if the paragraph } p \text{ is classified as positive} \end{cases} \tag{4.1}$$

Hence, the sentiment scores of the paragraphs p are discrete with $s_{b,t,a,p} \in \{-1, 0, 1\}$. Both the fine-tuned German and English BERT models were executed locally on a standard consumer laptop, enabling secure processing of sensitive data and practical implementation without requiring special hardware. Hence, these methodologies could be applied within a supervisory process without significant investments in IT infrastructure.

4.4 Step 3: Construction of the Sentiment Indicator

After the sentiment of all paragraphs P in article a about bank b at time t is classified as described in Section 4.3, the overall sentiment of article a ($s_{b,t,a}$) is calculated as the arithmetic mean of all corresponding paragraphs $p \in P$. To define the overall sentiment towards bank b at time t, the sentiment is further aggregated by the arithmetic mean of the sentiment $s_{b,t,a}$ for all articles $a \in A$ about bank b at time t:

$$s_{b,t} = \frac{1}{A} \sum_{a=1}^{A} s_{b,t,a} \quad \text{whereby} \quad s_{b,t,a} = \frac{1}{P} \sum_{p=1}^{P} s_{b,t,a,p} \tag{4.2}$$

As the literature shows, sentiment scores tend to be noisy, which is why smoothing techniques or aggregation on lower frequency are applied to reduce the noise observed. In this analysis, when working in weekly frequency, we assume that the noise is adequately reduced by aggregating all articles from the corresponding week using the arithmetic mean. We assume that information contained in articles before t remain relevant at time t, whereby the relevancy is decreasing over time. When working in weekly frequency, we incorporate this assumption into our models by including lagged values of the sentiment indicator $s_{b,t}$ into our models. However, when working with daily frequency, we construct the sentiment indicator $\tilde{s}_{b,t}$ as a N-period backward-looking weighted moving average:

$$\tilde{s}_{b,t} = \frac{1}{\sum_{n=1}^{N} n} \sum_{n=0}^{N-1} (N-n) s_{b,t-n} \tag{4.3}$$

As a result, we get a weekly and a daily indicator which proxies the sentiment in news articles, wereby $s_{b,t} \in [-1, 1]$ and $\tilde{s}_{b,t} \in [-1, 1]$. Values in $[-1, 0)$ indicate negative sentiment, 0 neutral sentiment and $(0, 1]$ positive sentiment. Furthermore, we define an adjusted N-period backward-looking weighted moving average:

$$\tilde{s}_{adj,b,t} = \begin{cases} 0 & \text{if } \tilde{s}_{b,t} > -0.25 \\ \tilde{s}_{b,t} & \text{otherwise} \end{cases} \tag{4.4}$$

This indicator $\tilde{s}_{adj,b,t}$ is a signal for negative sentiment of a sufficient degree only. Sentiment scores which are above the threshold of -0.25 are hence classified as neutral. When deciding on the threshold in Equation (4.4), we are faced with following trade-off. When setting the threshold too low, we might loose too much information. When setting the threshold not low enough, we might fail to effectively distinguish between marginally negative and significantly negative news. The threshold of -0.25 appears to be a reasonable choice. However, a more sophisticated approach could be considered to refine this threshold further.

4.5 Additional Step: Topic Classification

To construct the sentiment indicator, all content which was available after preprocessing as described in Section 4.2 was used to construct one single indicator. News sentiment can be further categorised by the topic of the content. The results of Roeder et al. (2020) suggest, that news sentiment of different topics have different relations to CDS spreads, whereby negative news topics are associated with significant changes in CDS spreads. They also find that not only financial topics but also non-financial topics such as sanctions and legal issues are associated with CDS spread changes.

In this additional step, the articles were classified by topic using another BERT model, which was trained for financial topic classification. Again, this model is publicly available on `https://huggingface.co/nickmuchi/finbert-tone-finetuned-finance-topic-classification`. This classifier was trained on a dataset with tweets about financial content which have a topic label from a list of 20 different topics. The author reports an accuracy of 0.91 and an F1-score of 0.91 on his evaluation set. For this study, we focus on bank specific non-financial topics, since this would be a bigger contribution to measures which already proxy the financial market. We classify the topic of article a as follows:

$$\text{topic}_a = \begin{cases} \text{Company, Product News} \\ \text{Legal, Regulation} \\ \text{Personnel Change} \\ \text{other} \end{cases} \tag{4.5}$$

This classification was performed once using the title and once using the lead of the article. The lead, usually at the beginning of the article, summarises the key content of the article and can be identified using HTML-tags included in the dataset. Since the BERT model used for this classification was trained on english text, both the title and the lead are translated to english before this classification using a general LLM.

4.6 Discussion

One of the main challenges of constructing a sentiment indicator from unstructured text is to determine whether the content within the text is relevant for the specific task. The simple pattern matching decision rule used in this study as described in Section 4.2 could be further extended to filter out more noise from the indicator. The decision rule as used in this study could be designed more strictly by requiring the surpassing of a predefined quote of matched patterns to total words in the article. However, there is a trade-off between being too strict and hence filtering out too much information or being not strict enough and hence introducing too much noise in the indicator. Another possibility to reduce the noise in the sentiment are filtering techniques as used in Equation (4.3). There might be other smoothing techniques such as the CUMSUM-filter or the exponential moving average which could perform better in this application. Additionally, the sentiment indicator could also be constructed to consider changes rather than levels in news sentiment by applying crossing moving averages or crossing exponential moving averages.

The sentiment classification using a BERT model appears to be promising. Traditional methods for sentiment classification require extensive preprocessing of the text such as word stemming and lemmatising as well as the maintenance of a word list with the associated sentiment. The BERT model used in this study can classify the text successfully without requiring further preprocessing. The main advantage however is the ability of the BERT model to identify the context of words and classifying the sentiment accordingly. As displayed in Table 4.1, the model correctly changes the classification of the texts which differ only slightly but contain a different sentiment.

The sentiment indicator could be further adjusted by incorporating additional meta information about the article or the publisher of the corresponding article. If available, articles could be weighted by readership figures of the publisher, giving more weight to publications with a larger readership. The sentiment indicator could also be adjusted by incorporating the total number of articles published on a given day. An unusually high volume of articles published on a specific day might indicate increased relevance of the information contained in the articles. There might also be differences in the writing styles of publishers that could be leveraged. Some publishers might have a more exaggerated reporting approach, while others might be more neutral. Consequently, if the readers are aware of the corresponding publishers way of reporting, their reactions on articles from different publishers might be different even if the sentiment scores are the same.

5. Experimental Analysis

5.1 Overview

Turning now to the implementation of the sentiment indicator as discussed in Chapter 4, in Section 5.2 we describe the dataset used for constructing the news sentiment indicator as well as the sample of banks considered in our analysis. We present the hypotheses to be analysed in Section 5.3. In Section 5.4 we evaluate the results of testing these hypotheses. We conclude by discussing the benefits and limitations of this study in Section 5.5. The implementation and instructions for replication are available at `https://github.com/bt-koch/newsews/`. In accordance with the terms of use, the data however cannot be shared.

5.2 Data

Depending on the queried bank, news media articles are retrieved from either the Swissdox@LiRI database or from the Eikon Data application programming interface (API). Swissdox@LiRI is a database which includes about 23 million media articles from Swiss media sources, mainly from the German and French speaking parts of Switzerland. It provides an interface specifically designed for big data applications and hence allows to query the whole database with the option to filter for language, time interval, keywords, sources and document types and returns all matching news articles in a machine-readable format. The whole content of the article is available. For this analysis, a query for each publicly traded Swiss bank is submitted to the database, whereby the database is filtered using the banks name, stock ticker or a common abbreviation of the bank name as keywords for a time interval starting in January 2022 until June 2023. Only German articles are used for the analysis however.

The sample of publicly traded Swiss banks is extended to all European Globally Systematically Important Banks (G-SIBs) from the 2023 list published by the Financial Stability Board (FSB), which are identified by the FSB in consultation with the Basel Committe on Banking Supervision as well as with national authorities. To retrieve news articles about the European G-SIBs, the Eikon Data API as a second provider for news media articles is used, which provides the option to request news articles from their news feed. This provider also offers the whole content of news articles in a machine-readable format. However, since the interface is not specifically designed for big data applications, the procurement of each item requires a separate API request. Therefore, corresponding API limits are quickly reached, which is why only english articles labeled as significant news by an algorithm by Eikon are requested. Furthermore, if there are more than five articles available for a specific bank on a given day, only five randomly chosen articles are queried. This rule was necessary for 13.3% of all requested days. Furthermore, using the Eikon Data API, only news articles from the past 15 months can be retrieved. Thus, the observation period starts in September 2023 and spans until October 2024.

5.3 Experimental Design and Hypotheses

First, the analysis focuses on Credit Default Swap (CDS) spreads as a proxy of the risk associated with a bank. We start by analysing whether the sentiment indicator is a determinant of the changes in CDS spreads. This analysis is based on the work of Annaert et al. (2013), extended by the inclusion of the sentiment indicator as a potential determinant. Annaert et al. (2013) try to explain and not to predict weekly changes in CDS spreads using contemporaneous explanatory variables. The explanatory variables include the risk free rate, leverage and the equity volatility as credit risk variables, the bid-ask spread as a liquidity variable and the term structure slope, swap spread, corporate bond spreads, the market return and market volatility as business cycle and market wide variables. Annaert et al. (2013) run their analysis for 32 listed banks from the Euro area in the period from 2004 to 2010. Their results show that the proposed determinants can explain CDS spreads. However, the effect and significance of these determinants is not stable over the observation period but rather display highly dynamic characteristics.

We now introduce the sentiment indicator as described in Equation (4.2) as a new explanatory variable in the model of Annaert et al. (2013). Due to limited data availability, the explanatory variable bid-ask spread from Annaert et al. (2013) is not considered in our analysis. The multivariate panel regression model is then given as follows:

$$cdsspread_{b,t} = \alpha_b + \sum_{k=1}^{K} \beta_k x_{b,k,t} + \theta s_{b,t} + \sum_{g=1}^{G} \gamma_g z_{g,t} + e_{b,t} \tag{5.1}$$

whereby b identifies the bank and t the time period. Note that we use a bank specific effect measured by the intercept α_b, K time-varying bank-specific explanatory variables ($x_{b,k,t}$) with corresponding coefficients β_k and G time-varying common explanatory variables ($z_{g,t}$) with corresponding coefficients γ_g as in Annaert et al. (2013). Additionally, we extend the model by the time-varying bank-specific sentiment indicator $s_{b,t}$ with corresponding coefficient θ. The error is given by $e_{b,t}$. Using this regression model, we test following hypothesis:

Hypothesis 1 *The sentiment indicator is not a determinant of CDS spreads.*

Hence, we test whether the coefficient for sentiment θ is equal to zero. We expect $\theta < 0$, meaning that negative sentiment $s_{b,t} < 0$ is associated with increased CDS spreads and hence higher risk of the corresponding bank b.

The main interest of this analysis is however, whether the indicator constructed from news articles could serve as an early warning signal for distressed banks. Hence, rather than evaluating whether the sentiment indicator determines the current CDS spread, we are interested whether the indicator can predict future changes in risk proxies such as the CDS spread, the maximum decrease in stock price in a forward looking observation window or the equity volatility of the banks in our samples.

First, we assess the predictive performance of the news sentiment indicator on CDS spreads. We follow Cathcart et al. (2020), which find that their news sentiment indicator has predictive power for CDS spreads on sovereign debt. They construct a global news sentiment indicator using the Thomson Reuters News Analytics (TRNA) database and focus on 25 developed and less developed countries from various geographies from 2003 to 2014. Besides finding that their sentiment indicator is a determinant of CDS spreads using panel regression techniques, Cathcart et al. (2020) find that their news sentiment indicator can predict CDS spreads using panel vector autoregression methods.

Based on this analysis, we adjust the panel vector autoregressive model proposed by Cathcart et al. (2020) to be suitable for estimating the relationship between the news sentiment of the banks in our sample on their corresponding CDS spreads. Note that our analysis has a bank specific independent variable $s_{b,t}$ for the news sentiment. Additionally, since high frequency bank-specific controls based on fundamentals or other data sources despite market data are difficult to include, we have no further bank specific controls in the regression model. However, the model includes fixed effects, which is why time invariant bank specific characteristics are controlled for. Hence, the model is given as follows:

$$cdsspread_{b,t} = \alpha_b + \sum_{k=1}^{K} \beta_k x_{k,t-1} + \sum_{\tau=1}^{5} \delta_\tau cdsspread_{b,t-\tau} + \sum_{\tau=1}^{5} \theta_\tau s_{b,t-\tau} + e_{b,t} \tag{5.2}$$

whereby we include a bank specific effect α_b. $x_{k,t-1}$ are the K explanatory variables stock market, volatility premium, term premium, treasury market, investment grade and high yield following Cathcart et al. (2020). Their effect is captured with the corresponding coefficients β_k. The other explanatory variables from Cathcart et al. (2020) are not included since they are only suitable when analysing sovereign debt and are not replaced as discussed. Following Cathcart et al. (2020), we include the CDS spread and the sentiment indicator lagged up to five periods in the model. δ_τ is the coefficient for the CDS spread of lag order τ, θ_τ is the coefficient for the sentiment indicator of lag order τ. The error is captured by $e_{b,t}$. Using this specification of the panel vector autoregressive model we test following hypothesis:

Hypothesis 2 *The sentiment indicator does not predict CDS returns.*

Hence, we test whether the coefficients for the sentiment indicator of different lag orders are zero. We expect negative coefficients for the sentiment indicator which would suggest that negative sentiment increases the risk of a bank. However, as already established in the literature, there are both over- and underreactions to news, which is why we might observe different magnitudes and sign changes in the coefficients for the sentiment indicator of different lag orders.

In the model as described in Equation (5.2), the sentiment indicator is treated as an exogenous variable. However, there is high potential for reverse causality. News content might affect the behavior of market participants which might change CDS spreads. In the same way, changes in CDS spreads might affect the news reporting about the corresponding bank. Because of this potential bidirectional causality we fit the model as described in Equation (5.2) with the sentiment indicator as a second endogenous variable while the remaining controls remain exogenous. We therefore further analyse the following hypotheses:

Hypothesis 3 *There is no bidirectional relationship between CDS spreads and the news sentiment indicator.*

Hypothesis 4 *The sentiment indicator does not support the theory of over- and underreaction.*

We test for the bidirectional hypothesis by conducting a Granger Causality test as well as looking at the impulse response functions (IRFs). To test the over- and underreaction hypothesis, we look at the effect of news sentiment on CDS spreads using the corresponding IRF.

Since the models so far require that there are CDS traded on the banks, a lot of banks cannot be included in the sample. Hence, to include the remaining banks of our Swiss sample in the analysis, we need to use other proxies for the risk perceived in the financial markets about the corresponding banks. Since all banks are publicly traded, risk proxies constructed from the corresponding stock price allow all banks to be included in the analysis without high demands on data availability. Hence, the remaining models focus on risk proxies derived from the stock price and are fitted on the whole sample of Swiss banks as well as the European G-SIBs.

First, we assume that the sentiment of the news media is reflected in the stock price of the corresponding bank in the foreseeable future. As the literature has shown, the effect of news sentiment is most pronounced for media pessimism, which is able to predict falling stock prices (for example as in Tetlock 2007). Additionally, from the perspective of banking supervision, we care most about increased risk. Hence, we use the maximum decrease md at time t of the stock price S over the next 14 days of the corresponding bank as a risk proxy:

$$md_t = \min\left\{0, \min_{i\in\{1,\ldots,14\}} \frac{S_{t+i} - S_t}{S_t}\right\} \tag{5.3}$$

This model is implemented in daily frequency. Hence the weighted moving average of the news sentiment as described in Equation (4.3) is used. Additionally, since md is a measure for falling stock prices only, we also fit the regression model with the adjusted news sentiment indicator as defined in Equation (4.4). We therefore fit the following panel autoregressive models:

$$md_{b,t} = \alpha_b + \sum_{\tau=1}^{5} \delta_\tau md_{b,t-\tau} + \theta\tilde{s}_{b,t} + e_{b,t} \tag{5.4}$$

$$md_{b,t} = \alpha_b + \sum_{\tau=1}^{5} \delta_\tau md_{b,t-\tau} + \theta\tilde{s}_{adj,b,t} + e_{b,t} \tag{5.5}$$

whereby $md_{b,t}$ is the $t+14$ maximum stock price decrease of the stock price of bank b at time t. Similarly to the model described in Equation (5.2), md as the independent variable is included as lagged control variables for up to five lags in the model with corresponding coefficients δ_τ. We fit the model once with the daily sentiment indicator as described in Equation (4.3) and once with the adjusted daily sentiment indicator as described in Equation (4.4) and capture their associated effect by θ. To keep the model relatively simple and due to the high frequency, we do not include further controls. However, again fixed effects control for time invariant bank specific characteristics. A bank specific effect is captured by the intercept α_b while the error is captured by $e_{b,t}$. Using this model, we check the following hypothesis:

Hypothesis 5 *The sentiment indicator does not predict the maximum decrease of a banks stock price over the following 14 days.*

We test whether the coefficients for the sentiment score of different lag orders are zero. We expect the coefficients to be non-zero or more specifically $\theta > 0$, indicating that negative sentiment increases the maximum stock decrease within the following 14 days.

Another risk proxy which we can directly derive from the stock price is the corresponding volatility. We assume that negative news sentiment increases risk and uncertainty about the corresponding bank. Hence, we expect that negative news sentiment is associated with higher volatility in the following periods. To analyse this relationship, we fit both a generalised autoregressive conditional heteroscedasticity (GARCH) model with covariates, GARCH-X, as well as a heterogeneous autoregressive (HAR) model with covariates. Following the implementation of the GARCH-X model as in Sucarrat (2021) and the HAR model as in Boudt et al. (2022), only one additional covariate can be included in the model. Hence, we have the following representation of the GARCH(1,1)-X model:

$$\sigma_{b,t}^2 = \alpha_b + \beta_{b,1} e_{t-1}^2 + \beta_{b,2}\sigma_{t-1}^2 + \theta_b \tilde{s}_{b,t-1} \tag{5.6}$$

whereby $\sigma_{b,t}^2$ is the realised volatility. The intercept α_b, the squared white noise shock from period $t-1$, e_{t-1}^2, as well as the lagged realised volatility σ_{t-1}^2 follows the GARCH(1,1) model first introduced by Bollerslev (1986). The weighted moving average of the news sentiment indicator as described in Equation (4.3), $\tilde{s}_{b,t-1}$, is the additional explanatory variable for the implementation of the GARCH(1,1)-X model described in Sucarrat (2021).

Additionally, we have the following representation of the HAR model:

$$\sigma_{b,t}^2 = \alpha_b + \beta_{b,1}\sigma_{b,t-2:t}^2 + \beta_{b,2}\sigma_{b,t-6:t}^2 + \beta_{b,3}\sigma_{b,t-21:t}^2 + \theta_b \tilde{s}_{b,t-1} + e_{b,t} \tag{5.7}$$

whereby $\sigma_{b,t}^2$ is the realised volatility at t. We follow the HAR model by Corsi (2009) who identifies the short-term, mid-term and long-term agents in financial markets as the three primary volatility components. Hence, we include the volatility of the stock price of the past 3, 7 and 22 days ($\sigma_{b,t-2:t}^2$, $\sigma_{b,t-6:t}^2$ and $\sigma_{b,t-21:t}^2$) to represent the short-term, mid-term and long-term agents in our regression. Again, a bank specific effect is given by the intercept α_b. Additionally, we include the weighted moving average of the sentiment indicator $\tilde{s}_{b,t-1}$ according to Equation (4.3). The error is again captured in $e_{b,t}$. For both models, we test the following hypothesis:

Hypothesis 6 *The sentiment indicator does not predict stock price volatility.*

We test whether the coefficient for the sentiment indicator is zero. We expect $\theta_b < 0$ which would indicate that negative sentiment increases stock price volatility. Note that both the GARCH-X model and the HAR model can only be fitted for one bank at once. Hence, we do not estimate the whole panel, rather we focus on Credit Suisse and UBS as the two biggest banks in the Swiss sample, as we expect highest media coverage and market activity for these banks.

As discussed in Section 4.5, the news sentiment indicator might be of different significance when we further differentiate by the topic of the article. Following Roeder et al. (2020), we fit the regression as in Equation (5.2) using a news sentiment indicator which was constructed separately for each topic. We compare whether the coefficients differ significantly between the different sentiment indicators. To fit the regression, we use the same specification as in Equation (5.2). We test the following hypothesis:

Hypothesis 7 *The sentiment of news has no different impact on the risk of a bank depending on the corresponding topic.*

5.4 Evaluation of Hypotheses

First, according to Hypothesis 1, we check whether the sentiment indicator is a determinant of CDS spreads. In Table 5.1 we show the results of the panel regression of CDS spreads following Annaert et al. (2013), whereby the main interest is in the estimated coefficient for the sentiment indicator. As we can see, the estimated coefficient is not statistically different from zero. Still, as expected, the coefficient is negative which would mean that negative sentiment is associated with increased risk of the corresponding bank. Noticeable is the difference in magnitude of the coefficient and its standard error between the two samples. Given the results of the panel regression following Annaert et al. (2013), we cannot reject Hypothesis 1 that the current news sentiment of a bank is no determinant of the corresponding CDS spread.

Table 5.1 Determinants of CDS spread

	Model 1	Model 2
sentiment	-1.353 (1.24)	-12.415 (22.12)
corporate bond spread	-53.269*** (11.21)	-74.285** (33.74)
equity volatility	0.239 (0.21)	1.578*** (0.49)
leverage	-0.353*** (0.07)	0.706*** (0.22)
market return	-16.228 (21.88)	-42.564 (114.22)
market volatility	0.141*** (0.02)	0.627*** (0.17)
risk free rate	5.701** (2.41)	-27.972*** (10.06)
term structure	36.069*** (8.48)	5.06 (35.45)
R^2	0.46	0.47
adj. R^2	0.44	0.44
Sample	European Banks	Swiss Banks
Groups	10	2
Number of Obs.	514	144
Obs. Period	Sep 2023-Oct 2024	Jan 2022-Jun 2023

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors reported in parantheses.

Next, by evaluating the output of the panel vector autoregressive model of CDS spreads following Cathcart et al. (2020), we check whether the sentiment indicator has predictive power on CDS spreads. Table 5.2 shows the model according to Equation (5.2). For the

European sample, we do not observe any coefficients of the different lagged values of the sentiment indicator which are statistically different from zero. For the Swiss sample, the coefficient for the sentiment indicator of lag order one is statistically different from zero and of economically significant magnitude. The sentiment indicators of higher lag order are still not statistically significantly different from zero. Still, the sign and magnitude might indicate a possible correction of the strong reaction of the lag order of zero. Hence, we can only partially reject Hypothesis 2 that the sentiment indicator has no predictive power for CDS spreads. To improve inference, we may also want to test whether the coefficients combined are statistically different from zero.

Table 5.2 Panel VAR of CDS on Sentiment

	Model 1	Model 2
sentiment $(t-1)$	-0.905 (1.8)	-68.882** (34.19)
sentiment $(t-2)$	2.364 (1.83)	-17.451 (33.52)
sentiment $(t-3)$	0.646 (1.84)	26.074 (31.63)
sentiment $(t-4)$	-1.438 (1.87)	13.795 (32.31)
sentiment $(t-5)$	2.767 (1.9)	16.423 (30.94)
cds $(t-1)$	-0.011 (0.07)	-0.316*** (0.11)
cds $(t-2)$	-0.047 (0.05)	-0.21** (0.1)
cds $(t-3)$	-0.034 (0.05)	0.061 (0.1)
cds $(t-4)$	-0.099* (0.05)	0.074 (0.11)
cds $(t-5)$	-0.046 (0.05)	0.026 (0.1)
high yield spread $(t-1)$	8.869 (6.28)	76.555** (38.88)
investment grade spread $(t-1)$	-32.306* (17.23)	-95.09* (50.45)
stock market $(t-1)$	0.433* (0.25)	-1.447 (1.01)
term premium $(t-1)$	4.942*** (1.78)	6.271 (4.52)
treasury market $(t-1)$	-0.164** (0.08)	0.003 (0.01)
volatility premium $(t-1)$	-0.022*** (0.01)	0.024 (0.02)
Sample	European Banks	Swiss Banks
Groups	10	2
Number of Obs.	509	144
Obs. Period	Sep 2023-Oct 2024	Jan 2022-Jun 2023

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors reported in parantheses.

To further analyse this hypothesis, we conduct a Granger causality test to assess whether the sentiment indicator helps to predict the CDS spreads and vice versa. Note that this test is conducted for one bank at a time without additional control variables. We run the Granger causality test for both Credit Suisse and UBS and report the results in Table 5.3. The results of this test suggests that the news sentiment indicator Granger-causes the CDS spreads for Credit Suisse, for UBS only the specification with 2 lags suggests this Granger-causality. Additionally, there seems to be no Granger-causality of CDS spreads on the news sentiment. Hence, this results do support Hypothesis 3 about no bidirectional relationship between CDS spreads and the news sentiment indicator and suggest that there is Granger-causality of news sentiment on CDS spreads only.

Table 5.3 Granger Causality Test

	Credit Suisse		UBS	
	s causes $cdsspread$	$cdsspread$ causes s	s causes $cdsspread$	$cdsspread$ causes s
1	0.05**	0.52	0.99	0.79
2	0.05**	0.69	0.10*	0.85
3	0.06*	0.72	0.20	0.62
4	0.10	0.47	0.31	0.59
5	0.16	0.51	0.31	0.53

Notes: *** p<0.01, ** p<0.05, * p<0.1.

Figure 5.1 further supports this conclusion. As we can see, the generalised impulse response functions (IRFs) suggest that an unit shock in the news sentiment indicator is followed by a reaction in the CDS spreads while an unit shock in the CDS spreads does not seem to be followed by a reaction in the news sentiment indicator. In the following two periods after an unit shock in the sentiment indicator, we observe a negative reaction of CDS spreads, indicating that positive sentiment shocks reduce the risk proxied by CDS spreads and vice versa. This negative relationship is followed by a positive reaction before a complete reversal. Hence, we can observe similar dynamics as in the results of Cathcart et al. (2020) and hence reference their behavioral story of self-attribution bias. Participants in financial markets are likely to overestimate the accuracy of their own projections of the implication the news event will have on the fundamentals of the corresponding bank. Hence, they overreact to the news event. This overreaction is followed by a underreaction, triggered by the now available information about the materialised impact of the news event. This behavioral bias of market participants is theoretically discussed by Daniel et al. (1998). Additionally, and in contrast to Cathcart et al. (2020), the effect of a shock in sentiment on CDS spreads according to the IRF suggests a complete reversal. This would indicate that the news sentiment does not contain pure information but rather only noise. In summary and in contrast to Hypothesis 4, this results suggest that the sentiment indicator does support the theory of over- and underraction of market participants to news events.

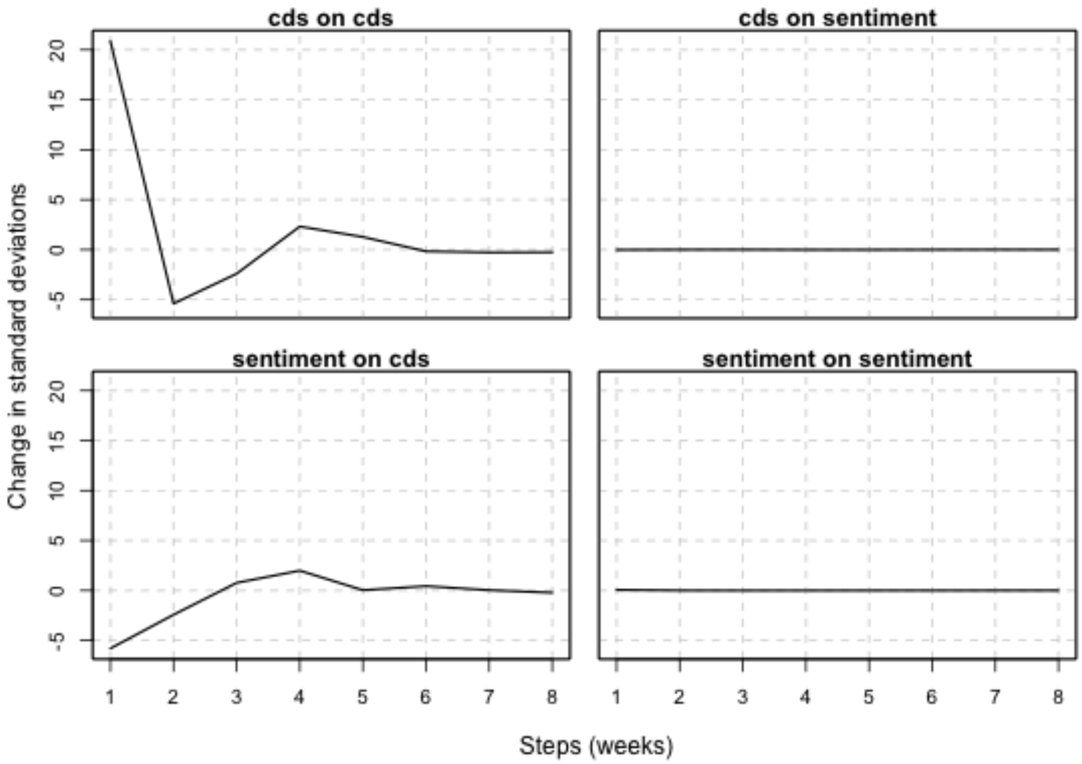

Figure 5.1: Generalised impulse response function

Having analysed the relationship between the sentiment indicator as proposed in Equation (4.2) and CDS spreads, we now turn to the hypothesis whether the sentiment indicator as proposed in Equation (4.3) has predictive performance for the maximum stock price decrease over the next 14 days of the corresponding bank. For the European G-SIBs, the estimated coefficients for the sentiment indicator are not statistically different from zero, hence the results of the model according to Equation (5.5) are not reported. Table 5.4 reports the results of this regression for the Swiss sample. While the coefficient of the unadjusted sentiment indicator is not statistically significantly different from zero, the coefficient of the adjusted sentiment indicator is. The results indicate that negative sentiment is associated with a higher drop in the stock price. Hence, on the Swiss sample, we can reject Hypothesis 5 that the news sentiment indicator has no predictive power of the maximum drawdown.

Table 5.4 Panel VAR of MD on Sentiment

	Model 1	Model 2
sentiment wma	0.001 (0)	
sentiment wma adj		0.017** (0.01)
md $(t-1)$	0.911*** (0.02)	0.911*** (0.02)
md $(t-2)$	0.007 (0.02)	0.006 (0.02)
md $(t-3)$	0.063*** (0.02)	0.063*** (0.02)
md $(t-4)$	-0.047** (0.02)	-0.047** (0.02)
md $(t-5)$	-0.07*** (0.02)	-0.072*** (0.02)
Sample	Swiss Banks	Swiss Banks
Groups	20	20
Number of Obs.	3449	3449
Obs. Period	Jan 2022-Jun 2023	Jan 2022-Jun 2023

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors reported in parantheses.

When evaluating both models for predicting the volatility of the banks stock price, neither the GARCH-X model as reported in Table 5.5 nor the HAR model as reported in Table 5.6 resulted in estimated coefficients for the sentiment indicator which are statistically different from zero. Hence, with this analysis we cannot reject Hypothesis 6 that the sentiment indicator has predictive performance in forecasting future stock price volatility of corresponding banks.

Table 5.5 GARCH-X using Sentiment

	Credit Suisse	UBS
σ^2_{t-1}	0.478*** (0.11)	0.762*** (0.07)
$\tilde{s}_{b,t-1}$	0 (0)	0 (0)
e^2_{t-1}	0.344** (0.18)	0.134* (0.08)
Number of Obs.	360	374
Obs. Period	Jan 2022-Jun 2023	Jan 2022-Jun 2023

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors reported in parantheses.

Table 5.6 HAR using Sentiment

	Credit Suisse	UBS
α_b	-0.005** (0)	0.001 (0)
$\sigma^2_{b,t-2:t}$	-0.102 (0.1)	-0.113 (0.08)
$\sigma^2_{b,t-6:t}$	0.242 (0.16)	0.211 (0.23)
$\sigma^2_{b,t-21:t}$	-0.257** (0.13)	-1.104** (0.43)
$\tilde{s}_{b,t-1}$	0.022 (0.04)	0.017 (0.02)
Number of Obs.	360	374
Obs. Period	Jan 2022-Jun 2023	Jan 2022-Jun 2023
Sample	Credit Suisse	UBS

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors reported in parantheses.

Table 5.7 shows the regression output from the same regression as in Table 5.2 on the Swiss sample, but now the regressions were fitted using the sentiment indicators which only considered articles with the associated topic. Compared to the general sentiment indicator, the first lag of these sentiment scores are not significantly different from zero. Additionally, the lag of order three for the topic "Product News" is significant. These results suggest that the sentiment of news has a different impact on the risk of a bank depending on the corresponding topic. However, in contrast to Roeder et al. (2020), using these results we cannot confirm that non-financial topic such as legal news do significantly influence CDS spreads. Again, to improve inference, we may also want to test whether the coefficients combined are statistically different from zero.

Table 5.7 Panel VAR of CDS on Topic Sentiment

	Legal	Product News	Personnel Change
sentiment ($t-1$)	-0.465 (13.25)	-8.891 (15.12)	18.878 (17.22)
sentiment ($t-2$)	-13.27 (13.41)	-24.948 (16.6)	8.662 (17.31)
sentiment ($t-3$)	8.627 (12.91)	29.365* (16.27)	8.041 (18.01)
sentiment ($t-4$)	8.539 (13.16)	15.848 (16.19)	17.088 (17.82)
sentiment ($t-5$)	-2.56 (12.87)	-8.379 (15.52)	14.856 (17.41)
cds ($t-1$)	-0.266** (0.11)	-0.3*** (0.11)	-0.271** (0.11)
cds ($t-2$)	-0.174* (0.1)	-0.191* (0.1)	-0.194** (0.1)
cds ($t-3$)	0.04 (0.1)	0.054 (0.1)	0.021 (0.1)
cds ($t-4$)	0.055 (0.1)	0.06 (0.1)	0.052 (0.1)
cds ($t-5$)	0.046 (0.09)	0.034 (0.09)	0.043 (0.09)
high yield spread ($t-1$)	59.418 (39.79)	81.531** (39.8)	74.391* (39.19)
investment grade spread ($t-1$)	-53.222 (47.96)	-71.025 (46.83)	-62.549 (47.43)
stock market ($t-1$)	-2.008** (1.01)	-1.674* (0.96)	-1.482 (0.97)
term premium ($t-1$)	5.339 (4.61)	4.815 (4.56)	7.441 (4.73)
treasury market ($t-1$)	0.002 (0.01)	-0.001 (0.01)	0.004 (0.01)
volatility premium ($t-1$)	0.021 (0.02)	0.02 (0.02)	0.034* (0.02)
Sample	Swiss Banks	Swiss Banks	Swiss Banks
Groups	2	2	2
Number of Obs.	144	144	144
Obs. Period	Jan 2022-Jun 2023	Jan 2022-Jun 2023	Jan 2022-Jun 2023

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors reported in parantheses.

5.5 Discussion

All hypotheses, the result of their evaluation as well as the method used to test the hypotheses are summarised in Table 5.8:

Table 5.8 Summary of Evaluation of Hypotheses

	Hypotheses (H_0)	H_0 rejected	Method used
1	The sentiment indicator is not a determinant of CDS spreads.	no	Panel VAR
2	The sentiment indicator does not predict CDS returns.	rejected on Swiss sample	Panel VAR
3	There is no bidirectional relationship between CDS spreads and the news sentiment indicator.	rejected on Swiss sample	Granger-causality tests, IRFs of Panel VAR
4	The sentiment indicator does not support the theory of over- and underreaction.	rejected on Swiss sample	IRFs of Panel VAR
5	The sentiment indicator does not predict the maximum decrease of a banks stock price over the following 14 days.	rejected on Swiss sample	Panel VAR
6	The sentiment indicator does not predict stock price volatility.	no	GARCH-X, HAR
7	The sentiment of news has no different impact on the risk of a bank depending on the corresponding topic.	no	Panel VAR

Particularly striking is the difference between the results of the European G-SIB sample and the Swiss Banks sample. One difference between the samples is the observation period used for the analysis. As the results of Annaert et al. (2013) have shown, estimated coefficients and their significance are not stable over time. In particular, their results suggest that the variables included in their analysis were more influential in time of crisis. The Swiss Bank sample does include the 2023 banking crisis while the European sample does not, which might be a reason for the difference in results.

Another key difference is the news article database used for calculating the sentiment indicator for the two subsamples. The news sentiment indicator constructed for the Swiss banks sample is built on much more articles which, through aggregation, might reduce the noise by giving less weight to potential articles which do not contain much impactful information about the state of the corresponding bank. Since the general public which consumes mainstream media as contained in the database used for the Swiss sample does not care as much about recent developments as the reader of news provided by specialised financial news vendors such as Refinitv, only highly relevant information might be published in the mainstream media. Hence, news articles published from mainstream media as in the Swiss banks sample might contain information of greater importance compared to financial news on average. Note however that as the news articles from the Refinitv Eikon API are labeled with the Refinitiv Identification Code (RIC) of the corresponding bank as well as of the nature of the database itself, we assumed for all articles that the content is relevant for the corresponding bank. Building a more sophisticated data preprocessing pipeline for the articles retrieved from the Refinitiv Eikon API might further reduce noise and hence increase the predictive performance.

The results in our analysis can be interpreted as explorative studies in correlation and simple causality testing. It tries to answer whether the content of financial news articles can be quantified and used as a predictive indicator in applications relevant for banking supervision or risk management. We can see that for some periods, samples and risk proxies, there is predictive performance in the sentiment of content from media articles. To build a reliable indicator, more detailed analysis regarding sample size, observation period and robustness as well as out of sample analysis is required. As we have seen in the analysis of Hypothesis 5, there might also be the need of further adjusting the indicator by adapting the classification rule to distinguish more drastically between positive, neutral and negative signals.

Finally, to build an early warning indicator, we would need to implement a decision rule on when the corresponding bank should be classified as risky enough to be further monitored by the supervisor. If adverse events of regular severity are occurring and the financial markets are reacting accordingly, there should not be a signal for the supervisor. For example, if quarterly results are not as good as expected but still not as worse to threaten the financial stability of the bank, a sound bank should not become financially distressed. However, this could still be reflected in the news coverage as well as in the risk proxies. Thus, a threshold of the corresponding risk proxy could be defined. Crossing this threshold could then be the decision rule to initiate a more in-depth monitoring process. There might also be a threshold which could be implemented which decides whether a predicted increase in the corresponding risk proxy could be a normal market reaction or a sign of increased financial distress.

The sentiment indicator itself could also be constructed in a different way such that the indicator itself could be used as a decision rule. The sentiment indicator as of now does not distinguish on news on higher or lower relevance in the supervisory context.

This could be done by training an LLM by providing a metric for each textual input which indicates the level of relevance. Another possibility would be to define classes of topics which are highly relevant in the supervisory context and construct the sentiment indicator only on news about these predefined categories. Additionally, the construction of the sentiment indicator could further be extended such that the textual inputs could be classified not only as negative but also in slightly negative, negative or highly negative (for example by assigning values -1, -2 and -3 accordingly). For this, it would be required to train a novel LLM which could classify the relevance and severity of negative news in a supervisory perspective. Given the BERT model used in this analysis, one could construct the sentiment indicator by multiplying the assigned value of -1 for negative and 1 for positive news with the probability of the predicted class which is provided in the output of the model. This should give less weight to less clearly positive or negative articles and vice versa.

Furthermore, the analysis could be extended by predicting spreads between the risk proxies of the different banks. For example, one could define the bank with the lowest risk proxy or the average risk proxy as a benchmark. Then, it could be analysed whether the spread of the banks to the benchmark can be predicted using a sentiment indicator. One could also analyse whether the spread of the sentiment score of a single bank to the general sentiment towards the banking sector as a whole has predictive power. Rather than just predicting the risk proxy, this might reveal if the risk of a specific bank deteriorates abnormally in comparison to its peers.

Note that a lot of measures used for nowcasting the current riskiness of a bank are derived from market data, since other measures such as accounting based metrics are only available in a relatively low frequency. However, banks under supervision are not always publicly traded, making market derived risk proxies unavailable. By leveraging the news media data source which enables the construction of the high frequency news sentiment indicator for both listed and unlisted banks under supervision, this gap could potentially be filled. Additionally, even for publicly traded banks, metrics such as Credit Default Swaps or metrics derived from the option market might still not be available since the corresponding financial product is not traded at all or does not have a sufficiently high trading volume. Furthermore, the finance literature has shown that news sentiment can be a leading indicator in the financial markets and therefore might signal increased risk earlier than market derived measures. Although the sentiment indicator does not require that the bank is publicly traded, there should be sufficient media coverage of the corresponding bank. Else, the construction of this indicator in a high frequency would not be possible or could be driven by a few articles with extreme sentiment.

6. Conclusion and Outlook

This study tries to answer the question whether the sentiment of content published in news articles about individual banks can be quantified and whether the resulting metric has predictive power for the risk of the corresponding bank. The underlying issue is whether an indicator derived from news media articles can improve the decision-making process for banking supervisors by helping to determine whether a particular bank requires closer monitoring. The study shows that current developments in the research area of large language models can be applied to classify the sentiment of financial news articles. To assess the predictive power of the sentiment indicator, various models for different risk proxies such as Credit Default Swap spreads, the maximum drop of stock prices in a leading observation window or the stock price volatility are fitted and evaluated. The results suggest that there is predictive power depending on the news media data source, the sample and observation period as well as the risk proxy considered. One main advantage of the proposed indicator is that it enables to nowcast the level of financial distress of a bank in a high frequency. For this, it does not require that the corresponding bank is traded publicly and hence expands the group of banks for which the risk can be nowcasted already using market derived proxies. Furthermore, as seen in the literature as well as in some results in this study, sentiment indicators can be leading to market derived metrics, hence signaling increased risk earlier. To utilise this sentiment indicator as an early warning signal from a supervisory perspective, additional study and further extensions of the methods presented are necessary. A decision rule needs to be implemented which defines when additional supervisory activity is required due to the signals from the sentiment indicator. There should only be a signal if the negative sentiment does not represent a regular adverse event, but rather a threat to the financial health of a bank. Additionally, if textual data is expected to play a significant role in further optimising the supervisory process, it may be beneficial to train a large language model specifically for supervisory purposes, rather than relying on a model trained for general financial applications. The methods presented could further be extended to include additional textual data sources such as publications of the banks itself or social media content.

Bibliography

[1] Agoraki, M.-E. K., Aslanidis, N.and Kouretas, G. P. (2022), U.S. banks' lending, financial stability, and text-based sentiment analysis. *Journal of Economic Behavior and Organization 197*, 73–90

[2] Annaert, J., De Ceuster, M., Van Roy, P.and Vespro, C. (2013), What determines Euro area bank CDS spreads? *Journal of International Money and Finance 32*, 444–461

[3] Araci, D. (2019), FinBERT: Financial sentiment analysis with pre-trained language models
URL: *https://arxiv.org/abs/1908.10063*

[4] Bank of International Settlements (2024), Core principles for effective banking supervision - executive summary. Last accessed: 2024-12-22
URL: *https://www.bis.org/fsi/fsisummaries/cps_for_banking.pdf*

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

[6] Borovkova, S., Garmaev, E., Lammers, P.and Rustige, J. (2017), SenSR: A sentiment-based systemic risk indicator.
URL: *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2951036*

[7] Boudt, K., Kleen, O.and Sjørup, E. (2022), Analyzing intraday financial data in R: The highfrequency package. *Journal of Statistical Software 104*(8), 1–36

[8] Broeders, D.and Prenio, J. (2018), Innovative technology in financial supervision (suptech) - the experience of early users. *FSI Insights on Policy Implementation*

[9] Cathcart, L., Gotthelf, N. M., Uhl, M.and Shi, Y. (2020), News sentiment and sovereign credit risk. *European Financial Management 26*(2), 261–287

[10] Chiaramonte, L.and Casu, B. (2016), The determinants of bank CDS spreads: Evidence from the financial crisis. *Contemporary Issues in Financial Institutions and Markets* pp. 51–77

[11] Cookson, J. A., Fox, C., Gil-Bazo, J., Imbet, J. F.and Schiller, C. (2023), Social media as a bank run catalyst. *Available at SSRN 4422754*

[12] Corsi, F. (2009), A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196

[13] Cutler, D. M., Poterba, J. M.and Summers, L. H. (1988), What moves stock prices? *National Bureau of Economic Research Cambridge, Massachusetts 487*

[14] Daniel, K., Hirshleifer, D.and Subrahmanyam, A. (1998), Investor psychology and security market under- and overreactions. *The Journal of Finance 53*(6), 1839–1885

[15] De Bondt, W. F.and Thaler, R. (1985), Does the stock market overreact? *The Journal of Finance 40*(3), 793–805

[16] Devlin, J., Chang, M.-W., Lee, K.and Toutanova, K. (2019), BERT: Pre-training of deep bidirectional transformers for language understanding
URL: *https://arxiv.org/abs/1810.04805*

[17] Du, K., Xing, F., Mao, R.and Cambria, E. (2024), Financial sentiment analysis: Techniques and applications. *ACM Computing Surveys 56*(9), 1–42

[18] Fernandez, R., Palma Guizar, B.and Rho, C. (2021), A sentiment-based risk indicator for the mexican financial sector. *Latin American Journal of Central Banking 2*(3)

[19] Gandhi, P., Loughran, T.and McDonald, B. (2019), Using annual report sentiment as a proxy for financial distress in US banks. *Journal of Behavioral Finance 20*(4), 424–436

[20] Hajek, P.and Munk, M. (2023), Speech emotion recognition and text sentiment analysis for financial distress prediction. *Neural Computing and Applications 35*(29), 21463–21477

[21] Hirtle, B.and Kovner, A. (2022), Bank supervision. *Annual Review of Financial Economics 14*(1), 39–56

[22] Huang, B., Yao, X., Luo, Y.and Li, J. (2023), Improving financial distress prediction using textual sentiment of annual reports. *Annals of Operations Research 330*(1), 457–484

[23] Illia, L., Colleoni, E.and Meggiorin, K. (2021), How infomediaries on twitter influence business outcomes of a bank. *International Journal of Bank Marketing 39*(5), 709–724

[24] Kearney, C.and Liu, S. (2014), Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis 33*, 171–185

[25] Malo, P., Sinha, A., Korhonen, P., Wallenius, J.and Takala, P. (2014), Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology 65*(4), 782–796

[26] Prenio, J. (2024), Peering through the hype - assessing suptech tools' transition from experimentation to supervision. *FSI Insights on policy implementation* (58)

[27] Reuters (2024), OpenAI says ChatGPT's weekly users have grown to 200 million. Last accessed: 2024-12-11
URL: *https://www.reuters.com/technology/artificial-intelligence/openai-says-chatgpts-weekly-users-have-grown-200-million-2024-08-29/*

[28] Roeder, J., Palmer, M.and Muntermann, J. (2020), Utilizing news topics for credit risk management: The explanation of bank CDS spreads. *Journal of Decision Systems 29*, 32–44

[29] Sarin, N.and Summers, L. H. (2016), Understanding bank risk through market measures. *Brookings Papers on Economic Activity 2016*(2), 57–127

[30] Scherrmann, M. (2023), German FinBERT: A German pre-trained language model. *arXiv preprint arXiv:2311.08793*

[31] Sinkey, J. F. (1978), Identifying "problem" banks: How do the banking authorities measure a bank's risk exposure? *Journal of Money, Credit and Banking 10*(2), 184–193

[32] Smales, L. A. (2016), News sentiment and bank credit risk. *Journal of Empirical Finance 38*, 37–61

[33] Sucarrat, G. (2021), Garchx: Flexible and robust GARCH-X modeling. *The R Journal 13/1*, 276–291

[34] Swiss Financial Market Supervisory Authority (2024), Banking supervision. Last accessed: 2024-12-16
URL: *https://www.finma.ch/en/supervision/banks-and-securities-firms/*

[35] 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

Statement of authorship

I hereby declare that I have written this thesis without any help from others and without the use of documents and aids other than those stated above. I have mentioned all used sources and cited them correctly according to established academic citation rules. I am aware that otherwise the Senat is entitled to revoke the degree awarded on the basis of this thesis, according to article 36 paragraph 1 letter o of the University Act from 5 September 1996.

Bela Tim Koch

Solothurn, January 2, 2025