LUISS

Master's Degree in Corporate Finance

Double Degree

Course of **Risk Management**

Sentiment Analysis-based Approaches to Downgrade Risk

Large Language Models application study

Prof. Daniele PENZA – *Supervisor*

Prof. Giacomo MORELLI – *Co-Supervisor*

Francesco MANCUSO #760831 - *Candidate*

Academic Year 2023/2024

Table of Contents

Abstract

On August 7, 2023, Moody's downgraded a set of American regional banks. The downgrades came amid widespread turmoil in the international banking sector, rocked by the Credit Suisse crisis. The United States' regional banking sector was already going through troubled times, experiencing the aftermath of Silicon Valley Bank and First Republic Bank failures. The downgrades were motivated by the banks' exposure towards the commercial real estate industry; however, the confidence around the regional banking sector's stability was already on a steady decline since the SVB and FRB failures.

This paper attempts to leverage existing knowledge about the effectiveness of sentiment analysis to propose a dynamic risk measure targeting downgrade risk, building on the current regulation framework targeting credit risk management. This framework will be developed and tested on the case study of five of the regional banks downgraded. Given a set of assumptions, we simulate the deployment of the proposed dynamic downgrade risk add-on by a financial institution exposed to downgrade risk on a portfolio of American regional banks.

The first section covers the current BCBS approach to cover downgrade risk under Basel IV. We point out how there is currently no requirement capturing downgrade risk directly.

In the second section we illustrate the effectiveness of sentiment analysis and the informational value of sentiment. We exhibit literature to corroborate how sentiment contains downgrade-relevant information.

The empirical study is described in the third section. We first lay out the data used for the analysis and discuss the rationale behind the use of different sentiment sources. Then, we describe in detail the methodology for the construction and calculation of the sentiment index, and we discuss the rationale behind the threshold determination method we propose. We finally show the results with graphs and comment the findings.

Discussion of the results obtained closes the paper.

Downgrade risk requirements, regulations, and remarks

This section is dedicated to a general presentation of the current treatment of downgrade risk under Basel IV, in the context of the regulation regarding credit risk. We explain the method for calculating the Default Risk Charge under MAR22, and then go on to briefly present the rationale behind CreditMetrics and its possible applications across the banking book and the trading book to address default risk. We also briefly provide context on the Monte Carlo and bootstrapping methods for simulating distributions.

From Basel 2.5 to the DRC

Downgrade risk is the risk of mark-to-market losses associated with the credit downgrade of an issuer which has not defaulted (Basel Committee on Banking Supervision, 2009). In other words, downgrade risk is the downside portion of migration risk, which more widely refers to the risk of changes in credit ratings on a security or obligor.

By nature, downgrade risk exists across the banking book and trading book of a bank, impacting credit risk and market risk. Downgrade risk affects the banking book, as when an obligor's credit profile deteriorates, the bank has a higher chance of losses on its credit exposure. This leads to higher provisions for non-performing loans and increased credit risk capital requirements. In the trading book, downgrade risk affects the bank's exposure to individual companies through credit sensitive instruments, like bonds and credit derivatives (Hull, 2015).

The original approach towards building a capital requirement for downgrade risk was defined in Basel 2.5, which defined an Incremental Risk Charge (IRC), intended to capture all residual risks incurred on securities held in the trading book, not covered by the existing Value at Risk framework (Basel Committee on Banking Supervision, 2009). In this instance, downgrade risk was intended to be captured together with default risk.

The rationale of this choice was that empirical evidence suggested that default and credit rating migration events tend to be clustered around specific periods of increased market stress (Basel Committee on Banking Supervision, 2009) (Kladakis & Skouralis, 2024).

Upon implementation of the Fundamental Review of the Trading Book, the IRC was replaced by the Default Risk Charge (DRC) model, which focuses only on default risk. This model is still able to capture downgrade risk thanks to the longer liquidity horizons considered in the measurement, but it does not include a specific provision of downgrade risk (Basel Committee on Banking Supervision, 2019).

In the context of the new revised standardized approach, the credit positions held by a financial institution are allocated to a set of default, pre-set risk buckets, based on credit ratings. The DRC is then calculated by aggregating the individual estimated losses on each position in the event of default. The aggregation is done through a weighted sum of the individual positions' estimated losses, where the weights are assigned depending on the risk bucket to which each position is allocated (Basel Committee on Banking Supervision, 2019).

The revised standardized approach also includes a Residual Risk Add-on, which partly compensates for the lack of consideration of downgrade risk in the DRC calculation. The residual risk add-on is intended to capture any other risk factors not addressed by the DRC requirements.

The residual risk add-on is the simple sum of gross notional amounts of instruments with "residual risks", like instruments having an exotic underlying (Basel Committee on Banking Supervision, 2019), multiplied by default pre-set risk weights. The residual risk add-on does not have a specific provision to account for downgrade risk. In the new standardized approach, downgrade risk is therefore captured only through the increased liquidity horizons considered in the measure.

Effective from January 1st, 2022, the Basel Committee on Banking Supervision has released an update to the DRC (MAR22) directed at clarifying the calculation of Risk weighted assets for positions exposed to default risk on the trading book.

The DRC is calculated by weighted sum of the individual net Jump to Default (JTD) risk positions. This regulation recognizes hedging and short positions, thus makes a distinction between Net and Gross JTD, with the net JTD being used to finally calculate the capital requirement.

For non-securitization exposures, gross JTD is a function of the face value of the position, or notional amount, i.e. "the amount of the instrument against which the loss of principal is determined" (Basel Committee for Banking Supervision, 2022), the Loss Given Default (LGD), i.e. the loss registered on the position if the issuer defaulted, and the cumulative mark-to-market profit and loss realized on the exposure. Net JTD is calculated by offsetting long positions with hedging and short exposures for each individual obligor and instrument. The net JTD is allowed to be zero, in the presence of derivative positions which would protect from default risk completely (Basel Committee for Banking Supervision, 2022).

In the MAR22, the LGD figures for each position are pre-assigned, depending on the seniority of the instrument considered. More specifically, equity and non-senior debt instruments have a LGD of 100%, senior bonds have LGD of 75%, covered bonds have LGD 25%.

The risk buckets for the weighted net JTD calculations are assigned through credit quality categories, which are based on external credit ratings. The default weighting scheme is presented in Annex 2.2.1.

Annex 2.2.1. Default weighting scheme for the calculation of net weighted JTD exposures for non-securitization portfolios. Source: Basel Committee on Banking Supervision.

Credit rating	Default risk weight
AAA	0,5%
AA	2%
A	3%
BBB	6%
BB	15%
B	30%
CCC	50%
Unrated	15%
Defaulted	100%

The sum of all weighted net JTD for each bucket is calculated by taking into account the net short JTD positions adjusted by the Hedge Benefit Ratio, i.e. the ratio between the total long JTD exposures and the total long and short JTD exposure for each obligor.

The current DRC framework touches on downgrade risk, but does not ultimately have an explicit provision covering the risk of mark-to-market losses due to credit rating changes.

Quantitative methods for credit migration risk – CreditMetrics and random sampling techniques

Although there isn't a risk charge capturing downgrade risk specifically, the estimation of credit rating transition probabilities is central in modern risk management. By far the most popular visualization of credit migration are credit transition matrices. Transition matrices show the probability of an issuer migrating from one credit rating to another over a period of time. They are at the center of many topics in risk management, and are notably a basic input in a host of methods used to calculate Value at Risk in the banking book for credit risk (Hull, 2015).

Credit transition matrices work on the assumption that credit rating changes are independent from each other across periods. This has been observed not to be the case, as credit rating momentum is a well-documented phenomenon (Hull, 2015). Moreover, the estimation of migration probability is based on historical data only.

One of the prominent methodologies that makes use of credit transition matrices as an input is J.P. Morgan's CreditMetrics, a method to estimate the probability distribution of losses arising from credit events, including through credit downgrades. CreditMetrics uses probabilities calculated in transition matrices to carry out a Monte Carlo simulation of ratings transitions on a portfolio of loans over a one-year period. Calculating the credit spread losses arising from the downgrades simulated at the end of the year, banks are able to infer the most likely level of credit losses.

CreditMetrics is applicable to market risk on the trading book too. The credit rating transition matrix is scaled down to a period of 10 days, and the same Monte Carlo simulation is carried out to determine the distribution of possible losses and Value at Risk.

In this context, it is worth it to briefly discuss the merits of two of the most popular methods for simulating distributions: the Monte Carlo simulation, used in CreditMetrics, and the similar Bootstrapping method. Although both techniques ultimately rely on random sampling to estimate distributions, they operate on different assumptions.

As mentioned, the Monte Carlo simulation is a computational technique that allows to evaluate possible scenarios by generating a large number of random samples, given a probability distribution. A common approach is to estimate the probability distribution of a variable from historical data and then applying that distribution to analyze possible outcomes of that same variable over a period of time.

The Monte Carlo simulation does not draw from historical values; rather it assumes their probability distribution to remain constant over the period of the simulation. Producing a reliable estimation of the probability distribution requires a reasonably large amount of historical data.

The bootstrappnig method, on the other hand, rests on different assumptions. While it is also employed to simulate variables' possible developments, it uses resampling from an historical dataset of observed values, i.e. it draws random subsamples, with replacement, from the historical distribution and reshuffles them to simulate alternative trajectories.

Differently from the Monte Carlo simulation, bootstrapping does not require a known historical probability distribution to operate, as it draws directly from the empirically observed distribution of the variable to generate random samples.

Sentiment Analysis applications to banking and risk modeling: literature review and hypothesis formulation

In this section, we first introduce Sentiment Analysis and briefly focus on the advantages and disadvantages of each approach to this technique. We proceed to give an overview of existing literature studying the informative properties of sentiment signals and their applications to risk management. We focus on works analyzing the relevance of sentiment data in the context of credit ratings. We then formulate this paper's research question and discuss its rationale.

Introduction

Sentiment Analysis refers to the computational study of opinions, attitudes and emotions towards a topic or entity (Walaa, Ahmed, & Hoda, 2014). Other, more general definitions, as reported by Kearney and Liu (2014), describe the wider concept of "content analysis" as *"any technique that enables inference by objectively and systematically identifying specified characteristics within text"*.

The financial sphere is one of the largest contributors to Sentiment Analysis (SA) research. Regulators, institutional investors, and individual market participants are always on the lookout for hints about future business conditions and trends. Behavioral scientists have studied how sentiment impacts on decision-makers. A wide host of methodologies have been applied to forecast and monitor macroeconomic variables and portfolio performances; in the context of risk management, SA has been used for systemic risk factors estimation and forecasting and anticipating down phases in the business cycle, among a wide range of other useful applications, including investigating the informative value of credit action reports and credit ratings.

The two main traditional approaches to textual sentiment detection are the dictionary-based approach and the machine learning approach.

The first approach makes use of very large words lists called “dictionaries”. Dictionaries contain scores of words and phrases that are pre-classified in defined dictionary categories. An algorithm mechanically matches the words it reads in a document to the dictionary and classifies the tone of the text by word counting.

This approach has the advantage of easier implementation with respect to machine learning. There exists a wealth of well-established programs and dictionaries to perform SA through this method, including highly effective sector-specific dictionaries that allow for tailored SA for specific research needs.

However, the dictionary-based approach has several potential fault lines. First, its effectiveness is highly dependent on the appropriate choice of dictionary. The dictionary’s degree of relevance to the document under examination, its dimension and degree of detail are crucial factors for the outcome of the analysis.

Second, the dictionary-based approach ignores any linear ordering of terms. This is known as the “bag-of-words” issue. The model is only counting the occurrences of some pre-classified words in its dictionary, ignoring the document’s actual meaning and scope. A document’s informative value can hardly be reduced to its word count. This problem can be tackled by implementing weighting schemes to word counting algorithms, to some success (Kearney & Liu, 2014).

On the other hand, the machine learning approach relies on statistical inference to classify words in a document (Li, 2010) (Kearney & Liu, 2014). This approach too uses datasets similar to dictionaries, with pre-set sentiment categories. These databases are used as "training sets" where algorithms register rules for sentiment classification. The programs are then provided with the "unsupervised" corpus of datasets, where they apply the rules learned in the training set.

Machine learning methods for sentiment analysis have been showed to be more accurate and versatile than dictionary-based approaches (Kearney & Liu, 2014). However, they too have significant drawbacks.

The main obstacle of traditional machine-learning based methods for SA is their cost. Text in the training set must be manually classified, and large training sets are required to ensure better model performance. Machine learning algorithms cannot be standardized and retrieved easily like mechanical word-classification programs, as every algorithm needs to be trained on the relevant training set from scratch. That makes them hard to recalibrate. Moreover, the considerable complexity of machine learning models makes them more costly to implement for most organizations.

We explore the effectiveness of several SA methods applied to banking and finance with a special focus on SA applications to credit risk monitoring and forecasting, and the link between credit ratings and investor and media sentiment to showcase the solid theoretical background of such practices.

Literature review

There is a wealth of extant literature covering the informative value of news sentiment for macroeconomic variables and systemic risk.

A study conducted by the Federal Reserve of San Francisco on financial newspaper articles ranging from 1980 to 2015 shows that positive sentiment shocks are correlated with controlled inflation and increased consumption (Shapiro, Sudhof, & Wilson, 2020).

They find that their news sentiment model anticipates survey-based consumer sentiment measures and outperforms other models which do not make use of sentiment data.

The study notes the remarkable amount of macro-relevant information contained in news sentiment, both in scheduled and non-scheduled releases, respectively financial stability reports and financial newspaper articles. Correa et al. (2020) relate the macroeconomic variables forecasting to crises in the banking sector.

Similarly to other studies, they build a sentiment index to forecast macroeconomic downturns. They analyze the correlation between sentiment extracted from financial stability reports released by central banks across thirty countries over twelve years. They too note how sentiment captured by their index significantly explains movements in the business cycle. They then apply the same framework more specifically to the banking sector and find that negative sentiment in financial stability reports is a relevant predictor of banking crises (Correa, Garud, Londono, & Mislang, 2020).

The topic of the effectiveness of sentiment contained in news articles in conveying information about risk in the financial system is expanded by Borovkova et al. (2017). They conduct an extensive study on news articles sentiment about systemically important financial institutions and a set of major asset managers worldwide over a span on seventeen years to create a comprehensive systemic risk indicator based on sentiment data. The systemic risk sentiment indicator they propose, called SenSR, is showed to lead other systemic risk indicators using macroeconomic fundamentals in signaling periods of increased systemic risk by as much as twelve weeks (Borovkova, Garmaev, Lammers, & Rustige, 2017).

Their approach notably makes use of Natural Language Processing (NPL) algorithms made available by the Thomson Reuters News database to assign sentiment values to each article. They calculate the systemic risk indicator adjusting for the novelty and relevance of the news, hence suggesting that recent articles should hold more weight in the computation of a news sentiment index to account for the public's shifting attention and information incorporation dynamics, with more recent news being more informative and older news gradually losing weight in opinion formation.

Fernandez et al. (2021) elaborate on this approach to build a systemic risk sentiment index focused on the Mexican banking sector. Differently from previous studies, their approach uses Twitter data. Leveraging big data analysis tools, they gather text mentioning a set of individual Mexican banks and on the wider Mexican financial sector from the entire timeline of Spanish language twitter.

Their results show that the index captures information about financial stress not otherwise detectable from quantitative risk measures based on past financial data, notably money laundering and customer service satisfaction. Similarly to other studies, they conclude that inclusion of sentiment data improves forecasting and monitoring of systemic financial stress measures (Fernandez, Guizar Palma, & Rho, 2021).

Elaborating further on the process for assigning sentiment scores, they implement three different SA methodologies to assign polarity at a document level, including a neural network, and then use a majority voting mechanism to correct for possible inaccuracies to determine the final score.

The literature shows that sentiment data's informativeness goes beyond the macro variables referring to the whole economy or the entire banking sector. Many studies focus on using SA to extract information about risks for individual banks.

Nopp and Hanbury (2015) study a measure of individual financial institutions' attitude towards risk by analyzing CEO letters and forward-looking guidance sections of banks' annual reports. They find that a significant correlation between uncertain and negative language in annual reports and wider sector performance, and a weaker correlation to the individual institution's performance (Nopp & Hanbury, 2015).

Their study shows that scheduled corporate releases contain meaningful information about the sector's outlook. The weaker effectiveness of their model with respect to individual banks' performance is probably reconcilable with their choice of data. Scheduled company releases are remarkably low frequency and often need to be complemented with other sentiment sources to be effective in time series modeling (Kearney & Liu, 2014).

Going back to big data analysis, Accornero and Moscatelli (2018) propose a sentiment-based model to measure depositors' trust in a financial institution, to improve accuracy in forecasting retail deposit flows by using Italian language twitter data. Their paper shows that adding sentiment indicators significantly improves the effectiveness of models for investor discipline, especially when applied to smaller banks, perceived to be weaker by the consumers. They note that one of the largest advantages of sentiment data extracted from tweets is its continuous availability in real-time, allowing for “nowcasting” of an indicator, observing the actual value of an indicator in real time without lags. (Accornero & Moscatelli, 2018).

Their study highlights the relevance of introducing an “interconnection indicator” in the model for improving precision of sentiment data. Sentiment around financial institutions is likely to have spillover effects on other banks perceived to be interconnected. Interconnection in the banking system is a well-documented phenomenon and should be accounted for when measuring sentiment around financial institutions; their study however introduces an interesting focus on the perception of interconnection among financial institutions by consumers.

This measure of interconnection might not be immediately apparent in models based only on financial data. Industry-wide informative contagion turns out to be a highly significant input in their model, helping to gauge depositors’ trust in an institution and consequently retail deposit flows (Accornero & Moscatelli, 2018).

Another important field of experimentation with sentiment analysis considers the informational value of sentiment contained in analyst reports around credit ratings and in credit ratings themselves. Ratings released by specialized agencies like Moody’s, S&P and Fitch help reduce asymmetric information in the market, thus allowing institutions to effectively manage their credit risk. A lot of regulations regarding credit risk capital requirements and fund mandates indeed orbit around credit ratings, making their impact on banks’ and asset managers’ balance sheet considerable (Hull, 2015).

In particular, the literature shows that sentiment contained in analyst reports does contain credit-sensitive information, going beyond the informational value of the credit rating alone.

Research released by S&P Global (Oyeniyi, 2022) shows that the tone and sentiment of analyst reports accompanying negative rating actions are significant performance indicators when applied to an equity portfolio. In a basket of recently downgraded companies, equity returns tend to worsen proportionally to the magnitude of the negative sentiment measured in the analyst reports. The article shows the existence of a strong link between negative sentiment and rating actions.

On this note, Loffler et al. (2021) investigate the impact of negative news on tone in Moody's rating reports. They find that negative sentiment news significantly amplifies the impact of a credit downgrade on stock price, and the overreaction is corrected more slowly as investor attention increases. Their research that tone in rating reports is a significant predictor of future credit downgrades (Loffler, Norden, & Rieber, 2021).

Argawal et al. (2016) go further in exploring this relationship. They study the informational value of credit action reports going beyond the meaning conveyed just by the credit rating action itself. The paper argues that tone and sentiment information contained in analyst reports is free from the widely discussed problem of conflicts of interest in credit ratings, thus allowing to filter through the credit rating inflation problem. They argue that such reports thus contain additional information that can be used to assess an issuer's default risk beyond the credit rating itself (Agarwal, Chen, & Zhang, 2016).

Their paper quantifies net tone of analyst reports as the arithmetical sum of negative and positive tone on a phrase level, establishing a relevant positive relationship between analyst report sentiment and credit risk. It is argued that the net tone measure is effective at forecasting credit downgrades in the one year and two-year horizons, thus confirming that analyst reports' sentiment contains credit and default risk-relevant information while not being biased by conflicts of interest (Agarwal, Chen, & Zhang, 2016).

Smales (2014) focuses on the relationship between financial news sentiment and market-determined measures of credit risk for financial institutions, proxied by historical Credit Default Swap (CDS) spreads. His paper shows the markets' tendency to proportionally overreact to negative news with respect to positive news, with a highly significant correlation between news sentiment and an institution's credit risk (Smales, 2015).

His paper goes on to compare the effectiveness of sentiment information on bank-determined measures of credit risk, proxied by the LIBOR rate. In this case, the correlation he finds is not as strong as in market-determined credit risk measures (Smales, 2015).

The literature shows that sentiment data is highly informative towards downturns in the business cycle, financial cycle, macroeconomic variables, firm-specific risk events, and credit downgrades.

Hypothesis formulation

Under the current framework, downgrade risk is largely addressed as a byproduct of credit and market risk; that is to say, downgrade risk is expected to be covered by the regulations currently in place covering market and credit risk and lacks a rule specifically and directly addressing it. This is not a large issue under the given assumptions; it can be shown that the credit losses distribution estimated by using a method accounting for downgrades and a method not accounting for downgrades should in theory be the same, everything else being equal and with the same set of starting assumptions (Hull J. , 2012). It is the timing of losses arising from the need of revaluations and rebalancing that make downgrade risk relevant and pervasive across the banking and trading book.

As mentioned above, current approaches hinge very heavily on credit transition matrices, and in turn on the assumptions behind them. Transition matrices' failure to address credit rating drift, paired with the complete reliance on historical data on rating changes, makes the current approach remarkably backward-looking and likely to leave a sizeable amount of useful information on the table.

On a higher level, these methods have the credit ratings themselves as their only source of information, however they may be computed. A "hidden assumption" of the current approach towards credit risk management, and by reflection towards downgrade risk, is that credit ratings incorporate all available information regarding an institution's future probabilities of experiencing a credit event.

The literature consensually shows that sentiment contains default and downgrade-relevant information. Moreover, different sources of sentiment data (analyst reports, company press releases, and simple news articles) have been shown to each add additional information beyond that conveyed by credit ratings alone.

Studies show that adding sentiment data improves models' predictive and explanatory power on a wide array of different fields of application; those studies conducted SA through the traditional methods, whose limitations were discussed above. Dictionary-based methods are relatively cheaper and simpler to implement but lack precision and suffer from the "bag-of-words" problem; neural networks and machine learning approaches have shown way better results but are cumbersome to implement (Kearney & Liu, 2014). Overall, sentiment analysis as a method lacked a straightforward framework of implementation due to the overall complexity of the tools required to carry it out. This likely decreased the cost-effectiveness of implementing sentiment-based methods in risk management.

However, recent technological advancements in the Artificial Intelligence and Natural Language Processing (NLP) fields solve many of the problems previously facing widespread implementation of SA based methods. Large Language Models (LLMs) are "*a category of foundation models trained on immense amounts of data making them capable of understanding and generating natural language*" (IBM, 2024).

LLMs are significantly faster and cheaper to deploy than traditional machine-learning methods for SA, with very advanced generalist models available even online for a small fee. They don't need the extensive calibration required for constructing training sets for machine-learning, and reach remarkable levels of precision in detecting nuances in text meaning; the author's own estimate of these engines' accuracy in classifying tone from text varies, based on practical experience, but is generally close to the 95% mark with respect to human assessment, compared to the reported 75% accuracy compared to human assessment registered in traditional machine learning methods (Kearney & Liu, 2014).

The extensive existing framework could be adapted to employ sentiment-based approaches, now easier to implement thanks to artificial intelligence, and to explicitly tackle downgrade risk as a standalone risk category to better reflect its widespread implications across the trading and banking book activities of financial institutions, introducing a risk charge explicitly capturing downgrade risk.

This would have the advantage of adding a forward-looking dimension to risk charges capturing downgrade risk. Leveraging new technologies to enrich current risk management frameworks is in line with the Basel Committee on Banking Supervision goals for future development of prudential mechanisms, namely introducing more risk-sensitive capital requirements and reinforce banks' ability to withstand potential shocks (Basel Committee on Banking Supervision, 2019).

As mentioned, studies show that employing sentiment data to construct a forward-looking, dynamic charge to capture downgrade risk has the potential to make the current framework include residual information not captured by historical data. Such an approach would also give a larger edge to capital requirements in the incorporation of such information. Sentiment-based risk measures have been shown to lead other indicators only based on macroeconomic or historical data in forecasting financial stress and credit events, notably downgrades.

In addition to the current regulatory prudential framework, sentiment analysis fits well with widely accepted calculation methods in the field of credit risk. The Monte Carlo simulation approach employed by CreditMetrics is very versatile and effective across risk classes and is already incorporating downgrade risk in its default uses.

Monte Carlo and bootstrapping-based simulations are already a widely accepted practice in the risk management field and would be a good fit to help define a downgrade risk provision.

Another advantage of sentiment data is their availability in real-time. Thanks to recent developments in LLM technology, news articles and analyst reports' tone can be extracted as soon as they come out, potentially allowing for "nowcasting" of sentiment around a financial institution, improving risk-sensitiveness of existing credit risk measures.

Taking all these possible advantages into account, we propose a basic framework for a sentiment-based, dynamic downgrade risk measure. The risk measure will build on existing approaches; to add dynamicity and respect the current framework, it will exist as a dynamic add-on provision on top of the already existing Default Risk Charge (DRC) as defined under MAR22.

The proposed risk measure will also build on the pre-defined default weights assigned to rating buckets under MAR22; however, it will act as a dynamic buffer between one rating bucket and another, attempting to allow financial institutions to reach credit events with sufficient capital requirements already in place.

Empirical study

In this section, we first describe the data on which the study is based, their collection process and sources. We then discuss the methodologies adopted to analyze the data and go on to present the rationale behind the calculation of the sentiment indices and dynamic thresholds. We present the results obtained by applying the framework on the data at our disposal and discuss the findings with the help of graphs. We finally draw our conclusions.

The data

This study considers scheduled and non-scheduled news releases over seven months, from January 1st to July 31st, 2023, covering five financial institutions downgraded by Moody's in August 2023: M&T Bank (Ticker: MTB), Fulton Financial (Ticker: FULT), Prosperity Bank (Ticker: PB), BOK Financial (Ticker: BOKF), Webster Financial (Ticker: WBS). Furthermore, the analysis also takes into consideration research, news, opinions and interviews on the general state and outlook of the U.S. regional banking sector, and on a portfolio of comparable institutions which were not downgraded in the same instance: Zions Bancorp (Ticker: ZION), Fifth Third Bancorp (Ticker: FITB) and Huntington Bancshares (Ticker: HBAN). Differently from other studies, we use Large Language Models (LLMs) to extract the information about sentiment from the sources gathered.

The news, sourced from Bloomberg, include pricing actions from analysts, analyst reports, research releases, opinions, interviews, including opinions and interviews from company management, insider transaction reports, and news articles directly citing the institutions. Bloomberg employs artificial intelligence to gather all relevant news material in a specified timeframe and language about a specified ticker through the "CN" function. The material is stored in plain text format and arranged in time order according to the release date.

Overall, the analysis took into consideration 669 individual sources. Upon review, the database was cleaned from duplicate news, "news about news", i.e. an outlet releasing an article directly citing another outlet, and noise; however, opinions discussing the contents of articles in the databases were kept in consideration and not considered duplicates. Likewise, same-sign pricing actions published by different analysts were not considered duplicates. News about businesses, clients, industries to which the examined banks had significant exposure during the period in consideration were included when they directly mentioned the banks in question. Sources include Bloomberg intelligence, Bloomberg news, the Wall Street Journal, Financial Times, The Banker, SeekingAlpha (for forward-looking opinion) and local newspapers, among others.

Bloomberg allows to access thousands of news releases with impressive granularity and ease of access, also having the advantage of filtering *ab origine* only reliable and trusted news sources, without the need to implement complex web scraping and data cleaning algorithms which could have been outside the scope of this work. Furthermore, using a wide array of sources when analyzing news sentiment decreases the probability of extracting biased sentiment from one source which does not reflect the general attitude of the market.

For this study, we chose to combine three of the most popular sentiment sources, i.e. corporate disclosures, financial news, and analyst reports. This combination allows for several advantages, as detailed by Kearney and Liu (2014). Corporate disclosures are likely to convey sentiment from company management and insiders. As unlikely as it may be that they might want to convey negative sentiment about their company, insiders are the most knowledgeable people about the firms they serve, hence their disclosures are relevant in studying firm performance, and sentiment conveyed by insiders' forward-looking guidance in annual reports has been shown to be a good predictor of sector performance in the financial industry (Nopp & Hanbury, 2015). However, corporate releases are remarkably low frequency, being released quarterly or annually.

News stories are a flexible information source that sits on the other end of the spectrum with respect to insider releases. They reflect the points of view by pure outsiders and are generally higher frequency than company reports, and their release pattern is assumed to be as random as the events they portray. This allows them to be used in modeling daily sentiment levels. News stories cover all types of events, which give a much more complete picture of the markets' perception of a company. Furthermore, they are easier to extract and analyze for LLMs, which have limited context windows, i.e. support only a limited number of characters to form their responses; hence, a lengthy earnings call transcript is not ideal to submit for analysis to an LLM. However, news stories only cover backwards-looking opinions, i.e. they comment on past events only.

Analyst reports and opinions bridge the gap between insider releases and news articles. Analysts have a pivotal role in the forming of opinions by the markets, as they provide forward-looking information from an outsider's point of view, combining advantages from news articles and insider releases (Kearney & Liu, 2014).

We chose to not include online message boards high-frequency data, namely Twitter data. This type of sentiment source is the one best representing small investor sentiment; however, being online platforms often chaotic and unregulated, the signals coming from that source would have added considerable noise to the database. Furthermore, messages on Twitter tend to be less clear and formal then news releases, using numerous abbreviations, slang, links, et cetera. These characteristics make Twitter data not the ideal data source to examine through an LLM. Processing a database of Twitter big data requires bearing a considerable cost in time and effort and would have required data analysis techniques which would have been outside the scope of this study.

One of the main points of this work is demonstrating that SA methodologies' implementation costs are now close to zero, thanks to new AI tools. Engaging in massive database maintenance would have defeated the purpose of this study. Making use of the first three sentiment sources allows for an optimal combination of their respective strengths while keeping the execution costs at a reasonable level.

Textual analysis and sentiment extraction were performed by Large Language Models (LLMs). These AI chatbots are especially well-suited for performing sentiment analysis, as they are extensively trained to detect tone in user messages and then replicate it in their responses, when appropriate. With respect to previous neural network or machine learning approaches to sentiment analysis, these models don't need a specific training environment for the task at hand and they can be easily calibrated by submitting new prompts.

The AI models used for this analysis were chosen based on performance history and availability. One of the main points of this project being proposing a more cost-effective, accessible method for leveraging sentiment data for risk management, we used the most widely available, state-of-the-art generalist models currently developed: OpenAI's ChatGPT 4o, Anthropic's Claude 3 Opus and Sonnet, and Mistral's Large model.

To streamline the analysis across multiple LLM models, we developed a set of standard prompts detailing the task to be performed by the engine, including a pre-defined response template which improved comparability and standardization of output data.

In this study we use a high-level subtype of SA, document-level SA. This methodology assigns a single value to a whole document, as opposed to sentence or aspect level SA, which take a more granular approach to evaluating sentiment, respectively assigning sentiment values to single sentences and single words. We deemed document-level SA to be more fitting to the characteristics of this project, as it is the most widely used method among works on the same note examined in the literature, and it considerably lowers computational power and model complexity requirements.

The bots are instructed to examine the articles provided for each date ranging from 01/01/2023 to 31/07/2023, and then assign a score of 1, 0, or -1, corresponding to positive, neutral/uncertain, and negative sentiment to each individual article published for each date; the signals extracted are then aggregated through arithmetic sum to obtain a net sentiment value (NSV) for the date.

Thus, for each database of articles we obtain three separate vectors of daily NSVs, corresponding to the three models used. These are in the end aggregated through a majority voting mechanism, i.e. we compose the final NSV vector by comparing, for every day, the scores assigned by the three engines, and take the most represented value. If the bots give three different verdicts, the final daily NSV is the simple average between the three values proposed. The NSVs are then time-ordered and used as the basic input for the analysis.

Methodology

The methodology in this study revolves around the calculation of a simple "sentiment index" for each bank under consideration, derived from sentiment signals as extracted by the AI models from provided sources.

It's worth noting the difference between investor sentiment and textual sentiment. Investor sentiment refers to beliefs in a company's future performances, cash flows, and investment risks which are not directly supported by observable facts at hand; textual sentiment is more narrowly defined as the degree of positivity or negativity conveyed by text data.

The former is a subset of the latter; textual sentiment might well convey semi-rational investor expectations, but it also includes more objective information about firms and markets (Kearney & Liu, 2014). The index proposed by this methodology attempts to track textual sentiment, not only investor sentiment.

The index is then examined against a set of thresholds. The thresholds will be calculated via an approach inspired by the CreditMetrics methodology, i.e. through a bootstrapping simulation. We use the bootstrapping method to generate a large number of trajectories for each sentiment index and estimate the lower percentiles of their distribution.

The methodology attempts to build on the existing Default Risk Capital (DRC) requirement for non-securitization portfolios as outlined in the MAR22 document effective in January 2022 (Basel Committee for Banking Supervision, 2022) to include downgrade-relevant information conveyed by sentiment in the measure, to increase its effectiveness by adding in some predictive power. This framework will be adapted to the method described and applied to a portfolio of generic vanilla bonds, issued by the banks under scrutiny, with original rating equal to that of the respective issuer.

Sentiment Index Calculation and Rationale

Most of the relevant literature dealing with financial newspaper article sentiment identifies three main data types extracted from textual tone analysis: sentiment, relevance, and novelty. Several relevant studies note that accounting for such news characteristics improves model performance (Borovkova, Garmaev, Lammers, & Rustige, 2017) (Smales, 2015).

This study conforms to the mainstream trend in the literature of defining sentiment as a discrete variable taking one of three values (-1, 0, 1), indicating whether the text's attitude towards the financial institution in question is negative, neutral, or positive. This method has the advantage of assigning a straightforward label to each source, considering only the text tone and no other information like intensity of sentiment.

Novelty refers to the degree of outdatedness and originality of a source. Our index relies only on novel sources in the sense that duplicate articles, analyses, and other sources that reported the same exact information as a source already in the database were excluded. Our model also accounts for the degree of outdatedness of an analysis or news release. Other studies (Fernandez, Guizar Palma, & Rho, 2021) (Nopp & Hanbury, 2015) focus only on real-time gathering of data, without accounting for source outdatedness; others (Correa, Garud, Londono, & Mislang, 2020) (Borovkova, Garmaev, Lammers, & Rustige, 2017) mechanically exclude news from the calculation of their index as soon as new data comes in. Our index attempts to simulate the public's perception of the importance of a news release, and how that perception is used in forming expectations about the future, by giving more weight to the latest news developments while gradually phasing out older information.

Relevance refers to how prominently an institution is mentioned in a source. In his study Smales (2015) builds on an existing approach by Groß-Klußmann et al. (2011), quantifying relevance as a continuous variable [0, 1] and using it as a threshold to filter out irrelevant news; however, in this regard we recognize more closely the approach by Borovkova et al. (2017).

Bloomberg's function "CN" allows to use a built-in NLP engine which selects the most relevant news. The user has the option to adjust the algorithm's tolerance, with the default option set on "balanced". We opted to maintain this level, which optimizes the ratio between the amount of data sources gathered and their average relevance.

Hence, our model does not explicitly account for news relevance, because we don't need to filter out irrelevant news at the computational level, having relevance been already incorporated at the data gathering stage.

The sentiment index employed in this study aims to transform raw sentiment data into a time-series capable of representing the trajectory of news sentiment around a financial institution. For each time-ordered series of simple sentiment values, we calculate the exponentially weighted moving average. We use a balanced approach, setting a range of 6 days, resulting in an alpha of 28.57%. The EWMA accounts for information assimilation by the public, giving more importance to more recent news while retaining progressively less information as the news ages. The index at date t_x is then calculated as the cumulative sum of the EWMAs from dates t_0 to t_{x-1}.

The index aims to capture effects of sentiment around the banks themselves and the effects of perceived interconnection between institutions and the wider sector. To achieve this, the data for the individual banks was crossed with a separate vector of news not referring to any bank directly but discussing the state and outlook of the whole regional banking sector.

The sentiment values for the whole sector were extracted using the same methodology as the institution-specific signals and given the same weight in the computation of the NSVs for each bank. Adding a common set of values to all individual bank vectors simulates the institutions' real and perceived shared exposure to sector-wide trends, which has been shown to improve predictive accuracy when applied to the highly interconnected banking industry (Accornero & Moscatelli, 2018).

Perhaps the most crucial step in the functioning of the index is the approach taken towards "null" days, i.e. days when no relevant news was released. Given the random nature of news events and the relatively moderate coverage awarded to the institutions in question, the datapoints' pattern is highly volatile. This results in periods of missing data followed by clusters of releases around salient points in the narrative.

The several ways of addressing the problem all work on the assumption that over time the sentiment score will converge to a value; the "no news is good news" approach assumes that the score would converge to a small but positive value; vice versa, the "no news is bad news" approach assumes that it should converge to a small and negative value; finally, the "no news is no news" approach proposes that the score should converge to zero over time.

We choose a variation of the last. This approach works on the assumption that if there are no new inputs, then the sentiment will stay constant, so the public doesn't remove past information completely, but instead incorporates it in their beliefs and uses it to form assumptions about the future. Instead of the null days having a value of zero, hence the EWMA values registering the day as a neutral day and the index gradually resetting to zero, we take the null days' marginal contribution to the index to be zero.

That means that we take the EWMA of the last non-null date as a baseline value for the following null days, until a new non-null day updates the sentiment value; so that when expanding the sample over the full period to include null days, the index's rate of change gradually becomes zero when no news come out. This approach is proportionally punitive towards institutions which observe streaks of negative sentiment values, as it does not gradually phase out past effects of negative sentiment. The resulting full vector of sentiment values and the time series of the index is therefore the full date range from 01/01/2023 to 31/07/2023, with null days taking the last non-null day's value.

For a non-null date t, the Net Sentiment Value is therefore expressed as the arithmetic sum of bank-specific and sector-wide news articles sentiment values $i_x \in [-1; 0; 1]$, such that $NSV_t = \sum_{x=1}^{n} i_x$; and the index will be defined by $SI_t = \sum_{x=0}^{t} EWMA_x$, where $EWMA_t = \alpha NSV_t + (1-\alpha)EWMA_{t-1}$; $\alpha = \frac{2}{N+1}$.

The sample is expanded such that for a null day $EWMAt_{n=null} = 0$ thus $SI_{n=null} = SI_{n-v}$ where v is the number of null days until the last non-null day on record.

Thresholds, Capital Requirements Definition and Rationale

The other central element to the analysis is the method we use to determine what value of the index corresponds to an increase of downgrade risk. This task requires establishing a set of thresholds that mark different bands, or regions, each with an associated increment of capital requirements.

The capital requirements calculation is based on the Jump to Default (JTD) charges in the context of the Default Risk Charge (DRC) framework, as defined in the MAR22 document effective 01/01/2022. This approach is particularly fitting for the purposes of this study, as it defines a provision intended to capture risk arising from credit spread shocks which may not be represented in other provisions. It also puts in place a weighting scheme based on credit rating buckets (BIS, 2022), which is instrumental in setting a solid theoretical groundwork for a dynamic weighting scheme to capture downgrade risk on a position.

More specifically, assuming no hedging and no short positions for simplicity: the JTD charge is computed for every single exposure, considering each position's Loss Given Default (LGD), set at a default value of 100%, 75%, or 25% according to the debt's rank; the nominal value of the exposure, i.e. for bonds the notional/face value (N); and the market P&L on the position. Hence, the JTD charge for a long position on a generic vanilla bond with maturity of more than one year will be $\max\,[LGD * N + P\&L, 0]$. To calculate DRC, we take the weighted sum of the individual net JTD positions, where the weights (W) are default values set according to the credit rating bucket assigned to the bond, such that:

$DRC = \max\,[\sum W * net\ JTD, 0]$.

We will borrow from this approach, changing the risk weighting mechanism from the default one to a dynamic one, that will respond to the level of the sentiment index previously established with respect to a set of thresholds. This will be developed into an add-on risk charge which will attempt to cover the position for downgrade risk, setting an appropriate buffer before the credit event occurs.

For the purpose of this study, we will consider a portfolio composed of five long positions on generic plain vanilla, non-preferred bonds. We assume that the portfolio is held to maturity and that there are no short or hedging positions. The resulting add-on provision after these assumptions will effectively be expressed simply as the weighted sum of the bonds' face values, where the weights will be adjusted depending on the relevant index' performance with respect to the thresholds.

The thresholds' goal is to define a set of bands, or regions that give us a sense of the state of the sentiment with respect to a model scenario. We aim to make the thresholds responsive to real-time data, to leverage the "nowcasting" capabilities of sentiment indicators, while not relying on extensive historical data. One of the main features of the model we propose is significant malleability vis à vis possible constant updating with real-time sentiment detection via web-scraping algorithms and similar data-mining methodologies.

To construct the dynamic threshold, we make use of the sentiment signals gathered on the downgraded banks and sector-wide news, plus a set of non-downgraded banks, for which the sentiment extraction and data-crossing process is analogous to the one used for the downgraded banks. We then use block bootstrapping to generate a large set of random sentiment vectors based on the actual recorded values for each bank, downgraded and not downgraded, with the addition of a small but strictly positive "control vector". For each bootstrapped vector, we then apply the sentiment index construction process to generate a large set of simulated index trajectories for each bank. Finally, we calculate the dynamic thresholds by taking the 0.1, 0.2, 0.3, 0.4, and 0.5 percentiles of the cross-sample daily distribution of simulated trajectories. The simulation uses daily resampling to reflect the indices updating daily.

We deem the bootstrapping method to be more appropriate for our experiment. One reason is that the add-on downgrade risk charge we proposed is not fully based on historical data from a fixed period but is intended to constantly update the underlying sample to allow for "nowcasting" of percentile thresholds. Bootstrapping is fitting to this approach; for a set block size, the algorithm will leave the module of the block size and total sample size out of the simulation. By updating the database in real-time, whenever enough new observations are submitted, the algorithm will create a new block and enrich the simulation in real time.

Moreover, there is no evidence in the literature that sentiment values follow any common distribution. Potentially each institution's sentiment values might follow a different distribution, hence making the Monte Carlo simulation of all trajectories less accurate. With bootstrapping we can simulate all index trajectories together and derive the cross-sample percentiles without worrying about the individual underlying distributions.

The addition of the control vector to the simulation database is crucial to reduce the buffer's procyclicality. Through trial and error, we note that without the control vector, the dynamic thresholds are too sensitive to sharp downturns and lose their effectiveness. The presence of a constant, strictly positive subsample helps support the indicator and makes it more robust in times of increased volatility such as the months following the regional bank crisis in the USA.

This approach awards a twofold advantage. First, it lets us compare the sentiment indices' performance with respect to random scenarios, i.e. alterative sentiment trajectories based nonetheless on empirically observed values; second, it's possible to regulate the degree of simulated samples' spread with respect to the realized trajectories, and therefore the model's tolerance for underperformance, by changing the bootstrapped subsamples' block size.

A larger block size preserves more of the news releases' autocorrelation and leads to more uniform (less volatile) simulation outcomes, with the thresholds representing realized trajectories more closely; vice versa, choosing a smaller block size increases randomness and leads to a more spread-out set of simulated trajectories and less "forgiving" thresholds. By taking the lower percentiles, we ensure that the capital buffer will be deployed only when the sentiment indices' performance is significantly negative, worse than random, and not when the index is positive or only slightly negative.

The weight determination mechanism is based on the level of the sentiment index with respect to the regions delineated by the thresholds corresponding to the 0.1, 0.2, 0.3, 0.4, and 0.5 percentiles of the cross-sample daily distribution of simulated trajectories. We relate the indices' crossing between two bands to a change in the risk weight assigned in the DRC calculation.

Building on the default risk buckets assigned under MAR22, we take the difference between the issuer's current risk bucket weight and the immediate lower one as the range in which our downgrade risk charge weight is allowed to fluctuate. Each passing of a band corresponds to a symmetrical change in the risk weight, equal to 1/5 of the difference range, such that if an issuer's sentiment index crosses below the 0.1 percentile threshold, the combination between the downgrade risk charge and current DRC assigned with the standard weighting scheme will equal the standard weighting scheme DRC for the lower bucket. For instance, an A rated bond falls in the 3% risk bucket; if it gets downgraded to BBB, it falls in the 6% risk bucket (+3%). The proposed downgrade risk charge adds $\frac{1}{5} * 3\% = 0.6\%$ to the risk weight for each band passed, until the weight reaches the 6% mark, equal to the weight assigned to BBB rated bonds under the default framework, thus meeting the credit event with the risk charge already in place (Annex 1.2.2.1).

Annex 1.2.2.1. - the combination between the downgrade risk charge and current DRC assigned with the standard weighting scheme will equal the standard weighting scheme DRC for the lower bucket. Source: Author's elaboration.

MAR22 Risk Bucket	Threshold region	Increment	DRC+Downgrade Risk Charge	Cumulative increment
A (3%)	>0.5	0%	3,0%	0%
	0.5-0.4	0,6%	3,6%	0,6%
	0.4-0.3	0,6%	4,2%	1,2%
	0.3-0.2	0,6%	4,8%	1,8%
	0.2-0.1	0,6%	5,4%	2,4%
BBB (6%)	<0.1	0,6%	6,0%	3,0%

Results

We apply the described methodology to the news data gathered about M&T Bank (Ticker: MTB), Fulton Financial (Ticker: FULT), Prosperity Bank (Ticker: PB), BOK Financial (Ticker: BOKF), Webster Financial (Ticker: WBS) over the 01/01/2023-31/07/2023 timeframe, supported by the Sectorwide news, Zions Bancorp (Ticker: ZION), Fifth Third Bancorp (Ticker: FITB) and Huntington Bancshares (Ticker: HBAN) news. The risk charge calculation will refer to a portfolio composed only of long positions on generic plain vanilla bonds by the five downgraded issuers, which we assume is held for more than one year.

The banks in question were downgraded on 07/08/2023 by Moody's, from an original rating of A3/Baa1 to Baa1/Baa2. Outlooks varied from Negative to Stable. Under MAR22, the DRC for such a position would be max $[\sum 3\% N, 0]$ before downgrade and max $[\sum 6\% N, 0]$ after. Through the downgrade risk add-on provision proposed, we gradually build up DRC before the credit event happens.

Applying the methodology, we manage to fully achieve this result for four out of five banks (FULT, BOKF, PB, WBS) while MTB, the largest institution in the sample, reaches the 1.2% increment the week before the downgrade. In March 2023 a flurry of negative sentiment news impacted the whole regional banking sector due to a new wave of concerns about their loan book quality. MTB compensated the strongly negative sector-wide sentiment by posting better than expected earnings at the end of the quarter and reassuring the public about the solid nature of their well-diversified balance sheet.

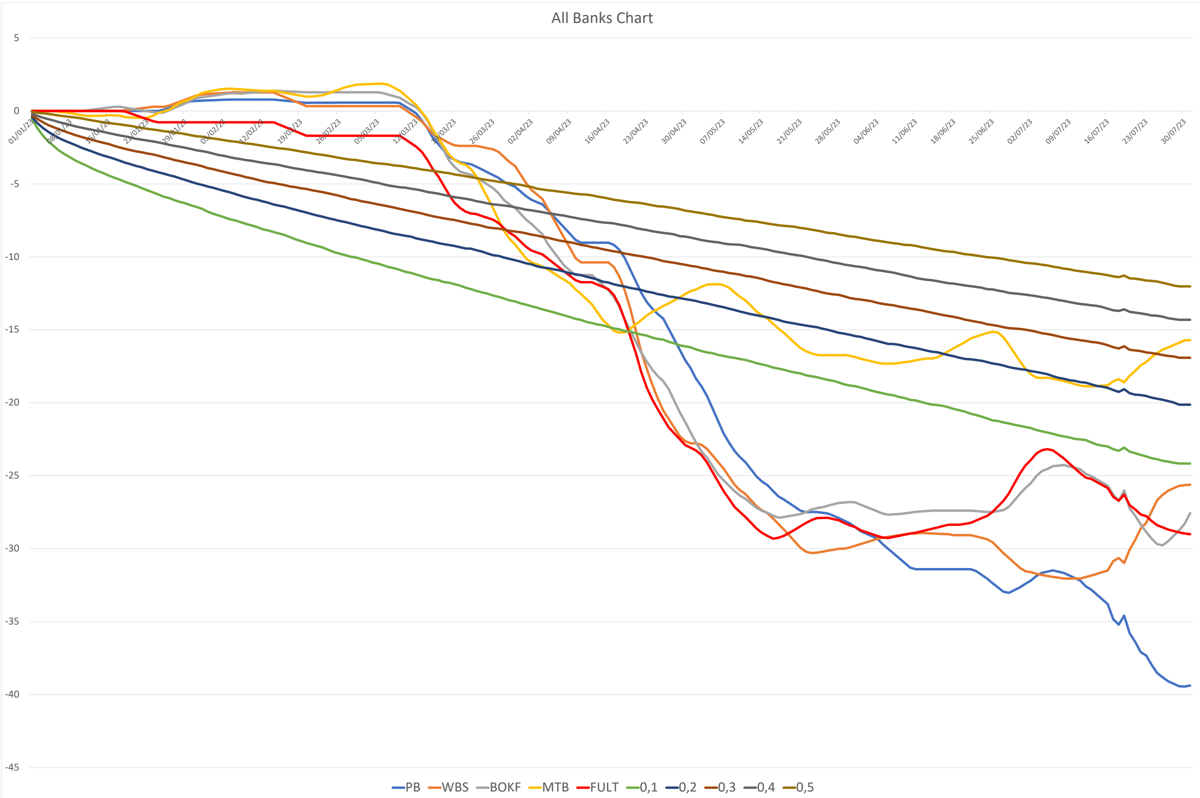
Nonetheless, for all institutions we succeed in reaching the credit event with a higher capital buffer than mandated by default rules. This is consistent with literature findings of negative sentiment having stronger effects than positive sentiment, and sentiment indicators being proportionally more effective on small banks that are perceived to be weaker (Accornero & Moscatelli, 2018) (Smales, 2015).

Our risk measure is also consistent with findings by Argawal et al., of sentiment risk indicators being predictive of downgrades in the one-year horizon (Agarwal, Chen, & Zhang, 2016). As the US regional banking crisis was closely followed by worldwide media, the correction of overreaction by the markets was possibly slowed even more by investor attention (Loffler, Norden, & Rieber, 2021). Moreover, as shown in previous literature, credit ratings respond to a general increase in systemic risk (Kladakis & Skouralis, 2024). These results show that perceived systemic risk might accelerate this process.

The methodology developed confirms the literature’s consensus on the informational power of sentiment data, including default risk and downgrade risk relevant information. Our study in particular shows that there is potential to further develop sentiment analysis-based methodologies specifically applied to risk management, and that the latest technologies offer the possibility to greatly simplify and streamline the infrastructure and competences required to carry out the necessary analyses.

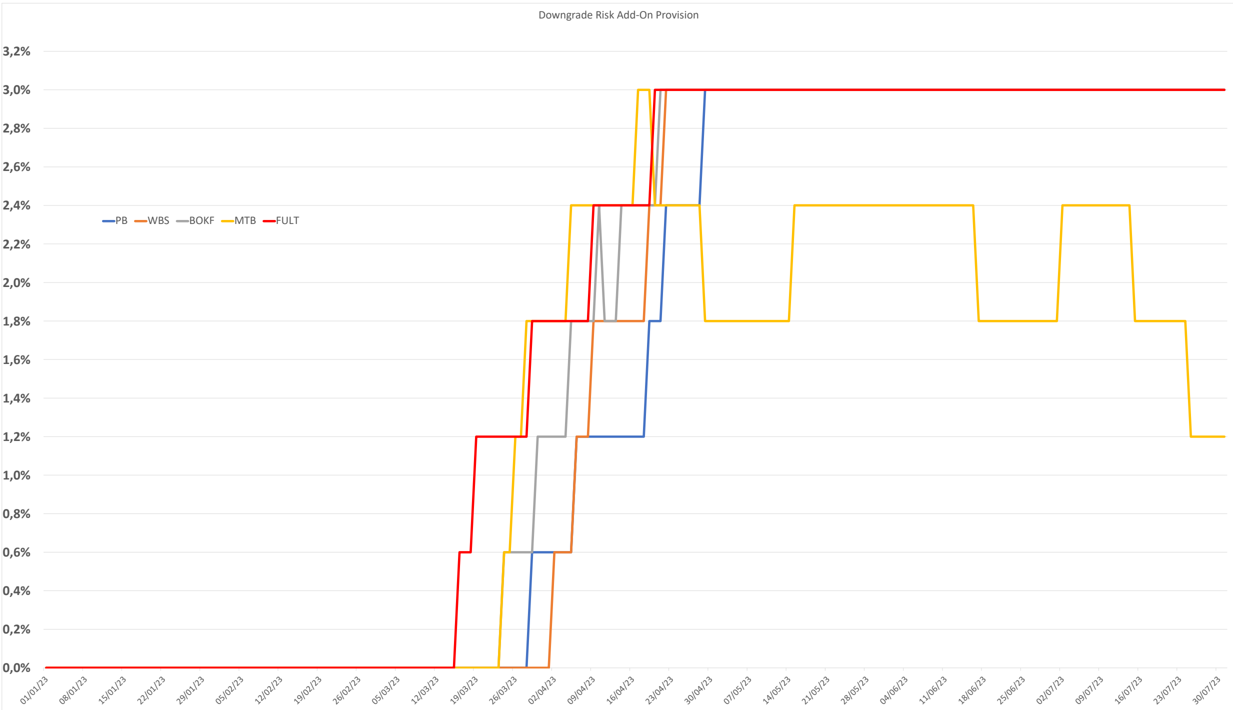
1.1.1. Graphs and discussion of results

Picture 1. - Applying the methodology, we manage to fully achieve this result for four out of five banks (FULT, BOKF, PB, WBS) while MTB, the largest institution in the sample, reaches the 1.2% increment the week before the downgrade… Source: Author's elaboration.



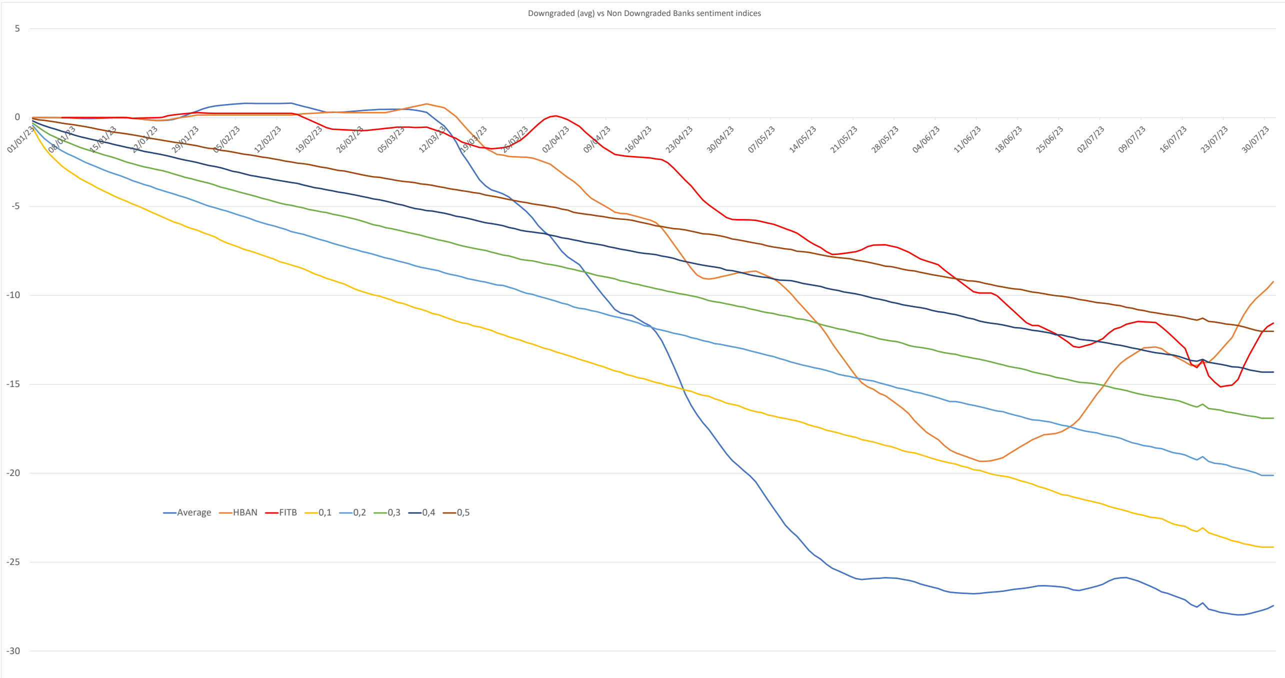
Picture 1 shows the evolution of sentiment indices trajectories for examined banks, over the period ranging from 01/01/2023 to 01/07/2023. Note the general downturn beginning in March, marking the beginning of a generally very negative Q2 for all banks examined. Sentiment around MTB is supported in April by the better-than-expected Q1 earnings posting, while the four smaller institutions continue the fall. The sentiment indices stabilize towards the beginning of June and reach the end of the observation period well below the bottom 10% threshold in four out of five cases. Applying the downgrade risk add-on provision, this would have meant meeting the downgrade in the first week of August with an already elevated Downgrade risk add-on charge across the portfolio according to our model.

Picture 2. …nonetheless, for all institutions we succeed in reaching the credit event with a higher capital buffer than mandated by default rule. Source: Author's elaboration.



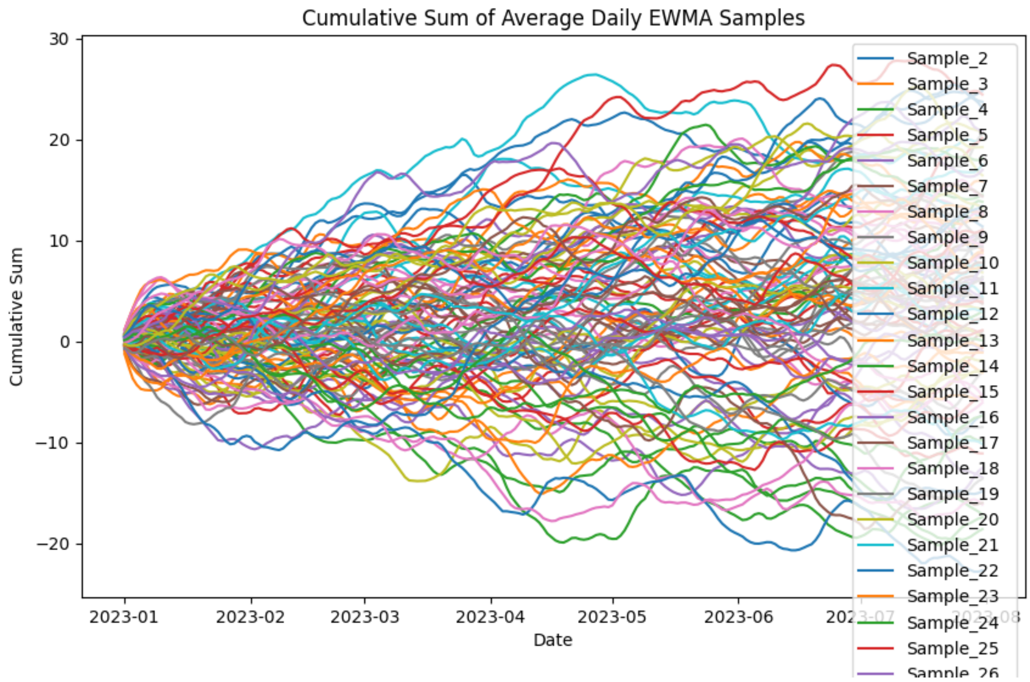
Picture 2 shows the progression of the proposed Downgrade Risk Add-on Provision. Each change corresponds to a band crossing by the relevant sentiment index. Here we can visualize the risk charge building up as time progresses. Note how downgraded banks already reach the maximum level DRAP, thus building a DRC equal to that of the lower MAR22 bucket, around the last week of April at the latest, hence anticipating the default rule by 2-3 months. MTB's sentiment index then experiences a series of crossings back and forth between the 2.4% and 1.8% thresholds, finally arriving at the downgrade in August with an augmented DRC of $4.2\% * N$ under our set of assumptions.

Picture 3. Comparison between downgraded and non-downgraded banks' sentiment index progression. Source: Author's elaboration.



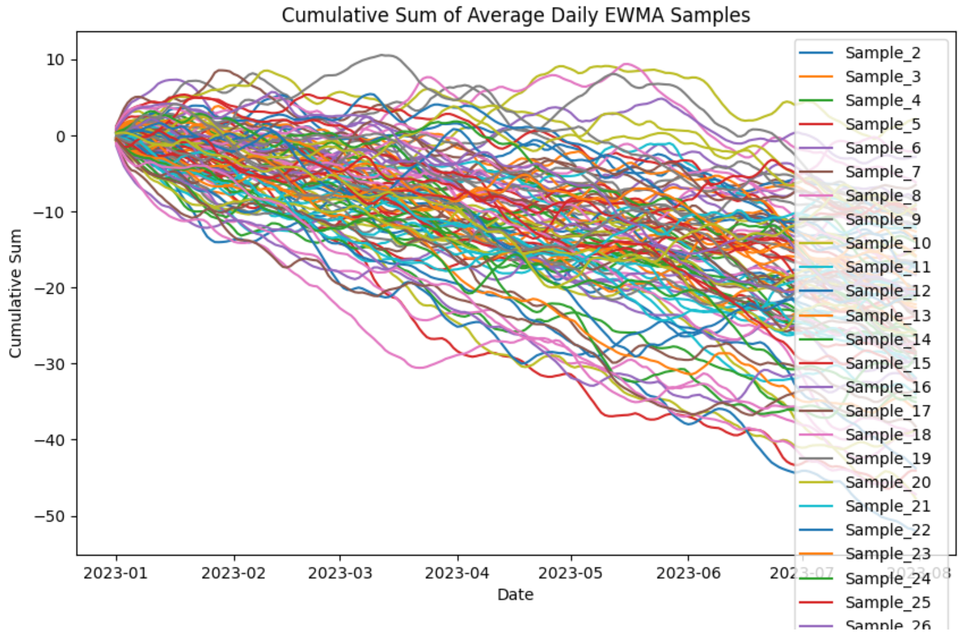
Picture 3 shows the downgrade-relevant information contained in news sentiment. In blue, the average sentiment index calculated across banks in the downgraded sample can be seen experiencing a downturn between March and April and remaining in downgrade territory well before August. On the other hand, the two other sentiment indices show the sentiment around two non-downgraded institutions of comparable size, HBAN and FITB. Note how both non downgraded banks' sentiment indices show the effects of negative sentiment contagion, becoming negative around March and April. However, FITB and HBAN sentiment was held up by these banks' limited exposure to commercial real estate and limited connection to recently problematic institutions like SVB and First Republic with respect to the downgraded banks. Both HBAN and FITB do conform to the general downward trend from March 2023, but at the end of July their DRAP charge is at 0, above the highest threshold. This outcome confirms the remarkable informative value of non-scheduled releases, which would not have been captured by other methodologies.

Picture 4. Bootstrapping – simulation of sample trajectories for threshold determination. Source: Author's elaboration.



Picture 4 shows the first 100 of 90000 sample indices simulated to calculate the dynamic thresholds. The simulation was carried out in Python extracting source data from Excel through a block bootstrapping mechanism as described above. For the final results, we chose a block size of 3, i.e. the sample NSV sentiment vectors were simulated by randomly rearranging the observed NSV sentiment vectors in groups of three consecutive values. Through trial and error, it was determined that such a block size was best balancing the randomness of simulated trajectories with respecting temporal dependencies between sentiment data extracted from news articles. Higher block sizes showed a less volatile set of simulations and much lower thresholds, due to negative sentiment streaks less likely to be broken up by the block bootstrapping process resulting in a higher tolerance for bad performance by the sentiment indices, as shown in picture 5.

Picture 5. Larger block sizes lead to more lenient dynamic thresholds. Source: Author's elaboration.



Picture 5 shows the simulated trajectories resulting from setting the block size to 5. Predictably, the indices tend to reflect the actual observations much more closely, as observed successions of NSV values are less likely to be broken up by the bootstrapping. With the original dataset being not so vast, consisting of a few hundred individual news articles, even a small increment in the block size leads to less volatile (and useful) simulations. Smaller block sizes are also more effective at incorporating new signals quickly if the method is applied to real-time monitoring of the threshold levels.

Conclusions

The paper shows that new Artificial intelligence tools could potentially be employed to improve the risk-sensitiveness and informational efficiency of the existing risk management framework. Large Language Models (LLMs) are very effective at capturing sentiment and tone in text documents, especially of medium length, like news articles. LLMs could well improve SA approaches' precision; their relative ease of use and inexpensiveness make them a very attractive alternative to traditional machine learning-based SA approaches. Indeed, banks and regulators could deploy significantly more nuanced and precise engines and methodologies to capture even more informational value than showed in this paper utilizing a rather simple framework.

Better, faster, cheaper sentiment analysis would allow current risk management frameworks to access and reflect more of the available information, improving banks' resilience and preparedness to shocks. We know that periods of increased systemic risk and credit events such as downgrades are significantly correlated, and sentiment data is very effective in forecasting increases in system-wide financial stress. This paper, in line with the literature, further shows that sentiment-based measures are effective in delivering more precise information at the institutional level, hence not only helping to foresee the increase in systemic risk, but also identifying which institutions are relatively more likely to be affected.

The predictive efficacy of the measure proposed in this paper is consistent with results obtained by previous studies on the matter. The sentiment indicators show a decline approximately 12 to 14 weeks before the credit event. This range could probably be refined by further increasing the number of sources employed in the analysis.

The proposed measure also would make the current framework more attentive to downgrade risk. Instead of assuming downgrade risk to be covered by credit and market risk measures just based on the assumption that credit ratings and historical data about credit migrations include all available information about the future creditworthiness of a financial institution, we could refine the approach by including forward-looking information, capturing residual risks not addressed precisely by the current framework.

In conclusion, including sentiment data is now easier and cheaper than ever. It’s worth to leverage its proven informational value to make existing risk measures more forward looking and based on more realistic assumptions.

Bibliography

Accornero, M., & Moscatelli, M. (2018, February). Listening to the buzz; social media sentiment and retail depositors' trust. *Temi di discussione (working papers)*. Banca d'Italia.

Agarwal, S., Chen, V., & Zhang, W. (2016). The information value of credit rating action reports: A textual analysis. *Management Science*.

Basel Committee for Banking Supervision. (2022, January). Standardised approach: default risk capital requirement. *MAR Calculation of RWA for market risk MAR22* . Bank for International Settlements.

Basel Committee on Banking Supervision. (2009). Guidelines for computing capital for incremental risk in the trading book. Bank for International Settlements.

Basel Committee on Banking Supervision. (2019). Explanatory note on the minimum capital requirements for market risk. Bank for International Settlements.

Borovkova, S., Garmaev, E., Lammers, P., & Rustige, J. (2017). SenSR: A sentiment-based systemic risk indicator. *De Nederlandsche Bank Working Paper*.

Correa, R., Garud, K., Londono, J., & Mislang, N. (2020). Sentiment in Central Banks' Financial Stability Reports. *Oxford Review of Finance*.

Fernandez, R., Guizar Palma, B., & Rho, C. (2021). A sentiment-based risk indicator for the Mexican financial sector. *Latin American Journal of Central Banking*.

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

Hull, J. (2012). *Risk Management and Financial Institutions (third edition).* Wiley.

Hull, J. C. (2015). *Risk Management and Financial Institutions.* Wiley.

IBM. (2024). *What are LLMs?* . From ibm.com.

Kearney, C., & Liu, S. (2014). textual sentiment in Finance: A survey of methods and models. *International Review of Financial Analysis*.

Kladakis, G., & Skouralis, A. (2024). Credit rating downgrades and systemic risk. *Journal of International financial markets, Institutions and Money*.

Lando, D., & Skodeberg, T. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance*.

Li, F. (2010). The information content of forward- looking statements in corporate filings- A naïve bayesian machine learning approach. *Journal of Accounting Research*.

Loffler, G., Norden, L., & Rieber, A. (2021). Negative news and the stock market impact of tone in rating reports. *Journal of banking and finance*.

Nopp, C., & Hanbury, A. (2015). Detecting Risks in the Banking System by Sentiment Analysis. *Conference on Empirical Methods in Natural Language Processing.* Lisbon: Association for Computational Linguistics.

Oyeniyi, T. (2022, October). Hanging on every negative word: Natural Langage Processing Analysis of Credit Action Reports. *S&P Global Markets Intelligence*.

Shapiro, A. H., Sudhof, M., & Wilson, D. (2020). *Measuring News Sentiment.* Federal Reserve Bank of San Francisco.

Smales, L. A. (2015). News sentiment and measures of bank credit risk. School of Economics and Finance, Curtin University.

Walaa, M., Ahmed, H., & Hoda, K. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*.