Research Repository

Can soft information survive organizational distance? Evidence from SMB lending

Accepted for publication in European Accounting Review.

Can soft information survive organizational distance? Evidence from SMB lending

Paola Cerchiello[a], Stefano Filomeni[b], Alessandra Tanda[a]* and Anca Toma[a]

[a]*Department of Economics and Management, University of Pavia, Pavia, Italy;* [b] *Essex Business School, Finance Group, University of Essex, Colchester, UK*

*corresponding author, Department of Economics and Management, University of Pavia, Via San Felice al Monastero 5, 27100, Pavia, Italy. email: alessandra.tanda@unipv.it

Can soft information survive organizational distance? Evidence from SMB lending

This paper examines whether and how soft information in corporate loan applications influences the lending decision process, and whether organizational frictions moderate this influence. In this context, soft information refers to the text data collected, analyzed, and transmitted by loan officers to the bank's headquarters. Using a proprietary dataset from a large European bank, the study finds that soft information impacts both the outcomes and the speed of lending decisions. However, this effect is moderated by the organizational distance between the loan officer and the bank's headquarters. Additionally, we provide evidence that soft information embedded in comments serves as a strong predictor of company defaults, supporting its use in lending to small and medium-sized businesses (SMBs), particularly when branches are distant from the bank's headquarters. Our findings remain robust across additional analyses and tests, consistent with the presence of communication frictions within banking organizations that hinder the effective "hardening" of soft information as organizational distance increases..

Keywords: sentiment; corporate lending; organizational frictions; distance; credit scoring; text analysis

JEL codes: G21; G24; G14

1 Introduction

Communication frictions within complex banking organizations can affect both the transfer of soft information from loan officers to the bank's headquarters and its subsequent use (Filomeni, Udell, and Zazzaro 2020). Soft information helps reduce information asymmetries in lending decisions (Costello, Down, and Mehta 2020), but its transmission across such organizations can be costly, leading to potential information loss—particularly regarding uncodified soft information. While new technologies and models aim to evaluate borrower creditworthiness using more quantifiable and objective measures (Costello, Down, and Mehta 2020), soft information remains crucial in determining the final credit rating assigned to loan applications (Brown et al. 2015; Puri, Rocholl, and Steffen 2011). To preserve this valuable information, 'hardening' techniques are employed (Filomeni, Udell, and Zazzaro 2021).

This paper explores communication frictions within a large banking organization, focusing on whether and how soft information can be successfully 'hardened' and clearly transferred to the bank's headquarters. In our study, soft information is represented by the loan officer's sentiment embedded in comments on loan applications. Specifically, we aim to address our research questions by examining the challenges posed by hierarchical frictions. Our focus is on the effectiveness of communicating 'hardened' soft information to the bank's headquarters in a clear and unambiguous manner.

In the empirical analysis, we examine whether the soft information embedded by loan officers in comments on loan applications influences the outcome of the lending process, and whether this relationship is moderated by organizational frictions (e.g., hierarchical distance, delegation, related lending). We also control for characteristics of the firm, bank, and loan officer (e.g., gender, age, experience).

Consistent with Stein (2002), Liberti and Petersen (2019), and Filomeni, Udell, and Zazzaro (2020; 2021), we explore an additional factor that may explain the dominance of hard information over soft information in the underwriting and monitoring of business loans: the ‘hardening’ of soft information. Although the actual soft information is unobservable, we can analyze the loan officers’ textual comments attached to loan applications. If these comments are used to convey soft information to the decision-making entity, we can measure the hardening of this information through textual analysis and sentiment scores, which serve as proxies for the soft information. To assess the ability of hierarchical banks to harden soft information, we apply Stein’s theoretical framework by examining how the geographical distance between loan officers, who manage loan applications, and the bank’s decision-making centers affects their ability and willingness to harden soft information without losing its content in a commercial loan setting. This approach aligns with the findings of Fiordelisi, Monferrà, and Sampagnaro (2014), who argue that the geographical and cultural distances between a bank’s peripheral branches and its decision-making centers complicate the exchange of qualitative and soft information. This dynamic discourages local loan officers from hardening soft information and causes senior managers at the bank headquarters to be mistrustful of softened information from the periphery.

Our results indicate that sentiment, used as a proxy for soft information, plays a role in shaping the outcome of lending applications, with this relationship being moderated by organizational frictions (e.g., hierarchical distance, delegation, related lending). Further analysis reveals that these effects persist across both top and mid-level approval stages. Additionally, we find that managers located at greater organizational distances are less capable of processing subtle, qualitative soft information—particularly for borrowers in relationship-intensive lending situations. However, they are more inclined to rely on soft information when the loan is collateralized.

Our sample consists of 550 mid-corporate loan application documents obtained from the CIB division of a major European bank, which we refer to as ‘the Bank’ throughout the paper. The proprietary dataset includes all loan applications received during the analyzed time frame and was collected from the 24 CIB branches across Italy, one of the Bank's primary markets. The Bank's lending activities in Italy are representative of the overall banking industry, with a market share of approximately 15% in the country’s loan and deposit markets.

The dataset includes both structured and unstructured information. Structured information refers to the quantitative and objective data found in loan applications, as well as details about the outcome of the application (e.g., whether the loan is granted, the amount requested, and the amount granted). Unstructured information, on the other hand, is contained in the comments made by loan officers on the application, which is forwarded to the deliberating entity. This unstructured data can be viewed as a representation of soft information. In this paper, we use unstructured information to extract sentiment variables (as proxies for soft information) through text analysis techniques, categorizing sentiment into positive and negative classes. In these comments, loan officers can express their opinions and sentiment regarding the lending proposal, which can influence the decision of the deliberating entity and provide insights into the borrower. This is particularly important when considering loan applications from SMBs (Small and Medium-sized Businesses) being evaluated by complex organizations that may face communication frictions, especially when spatially distant (Filomeni, Udell, and Zazzaro 2020).

While the granularity and depth of our proprietary data make our empirical setting unique, we acknowledge that the analysis in this study is based on a limited time period and data from a single banking institution. However, our empirical framework aligns with those used in similar studies, such as Liberti and Mian (2009), Beck et al. (2013), Bose et al. (2024), and Filomeni et al. (2020, 2021, 2023), among others. Therefore, the results can be reasonably extended to other contexts where banks with a local presence and centralized decision-making processes assess SMB loan applications and aim to determine whether the inclusion of soft information is valuable and can be incorporated without losing effectiveness.

Additionally, due to limitations in data availability, this paper does not account for loan officers' incentives in evaluating loan applications. However, these incentives can influence the effort loan officers put into evaluating applications and providing more detailed comments to convey the correct soft information. Future research could focus on analyzing these specific factors and offer insights into loan officers' incentives and their impact on the transmission of soft information to the bank's headquarters.

This paper contributes to the financial intermediation literature on the determinants of credit origination. Specifically, it is part of a smaller group of studies that explore the role of loan officers in credit markets, highlighting that loan officers possess significant decision-making power and discretion in credit approvals. This discretion can be influenced by personal characteristics or private incentives (Agarwal and Ben-David 2018; Tzioumis and Gee 2013). Additionally, the literature finds that emotions and sentiment can impact financial decisions made by investors (Cortés, Duchin, and Sosyura 2016; Félix, Kräussl, and Stork 2020; Zhang and Li 2019; Alzahrani et al. 2023), and we suggest that this influence extends to the lending process and the loan officer.

Our novel contribution also lies in the literature on soft information in lending. We apply textual analysis to a micro-level dataset, focusing on comments made by the loan officer responsible for the specific firm-bank relationship associated with each loan application. To date, much of the literature on soft information has either sought suitable proxies for it or applied textual analysis in contexts that do not allow for the use of micro-level, granular information on corporate applications.

The results of this investigation are particularly relevant from both an academic and managerial perspective. The paper is organized as follows: Section 2 provides an overview of the lending environment and the credit scoring process at the Bank; Section 3 outlines the statistical methodology; Section 4 details the data characteristics and presents the main descriptive statistics; Section 5 discusses the results, includes additional cross-sectional analyses and robustness tests; and Section 6 concludes.

2 Soft information and organizational distance

Lending decisions are generally made by banks based on both hard and soft information available about the borrower. Organizational factors can influence the inclusion of soft information in the decision-making process, as complex organizations may face difficulties and costs in transferring soft information, particularly when dealing with uncodified soft information (Filomeni, Udell, and Zazzaro 2020). Including soft information in lending decisions enriches the information set and reduces information asymmetries (Costello, Down, and Mehta 2020), though it may not always improve lending decisions (Campbell, Loumioti, and Wittenberg-Moerman 2019; Banerjee, Cole, and Duflo 2009). Since Basel II, the use of sophisticated algorithms in lending has led to a shift toward more standardized and quantitative methods for evaluating creditworthiness and making lending decisions (Costello, Down, and Mehta 2020). At the same time, the distance between banks and small and mid-sized business (SMB) borrowers has steadily increased (Petersen and Rajan 2002), further favoring reliance on more objective and quantifiable evaluation methods. These methods may counterbalance the advantages that local banks have in assessing the creditworthiness of unlisted, small, and informationally opaque firms (DeYoung, Glennon, and Nigro 2008). In extreme cases, machine learning, AI applications, and the availability of new data sources may eventually reduce human involvement in lending decisions (Sutherland 2020; Liberti et al. 2022).

Despite this, both regulation and practitioners still allow discretion in the final lending decision (Brown et al. 2015; Puri, Rocholl, and Steffen 2011), and human skills and expertise remain highly valued in both the literature and the industry (Costello, Down, and Mehta 2020). The exclusion of soft information in new loan extensions can reduce the quality of the lending portfolio, leading to inefficient lending decisions (Agarwal and Ben-David, 2018). Soft information is particularly valuable for unlisted SMBs (Petersen and Rajan 1994, Cowan and Cowan 2006), especially when alternative information systems are unavailable (Cassar, Ittner, and Cavalluzzo 2015).

In complex banking organizations, the process of 'hardening' soft information may not fully eliminate communication frictions between branches and headquarters (Filomeni, Udell, and Zazzaro, 2021) and could lead to a loss of content. This loss does not occur immediately after the loan application is transferred to the next level; rather, it diminishes over distance, depending on factors such as the importance of the 'human touch' and socio-cultural proximity in communication (Qian, Strahan, and Yang 2015). For example, Qian, Strahan, and Yang (2015) show that internal ratings are more effective at predicting defaults when the loan officer and branch manager have worked together for a longer period, building mutual trust and familiarity.

Loan officers’ characteristics play a significant role in loan applications and the inclusion of soft information. Their discretion in the lending process directly influences the final credit rating assigned to a loan application, and their comments can convey important insights to the deliberating entity at the bank’s headquarters. Regarding discretion, the literature has found that loan officers' decisions are influenced by factors such as personal rotations (Hertzberg, Liberti, and Paravisini 2010), general regulatory concerns (Agarwal et al. 2012), cultural proximity to the borrower (Fisman, Paravisini, and Vig 2017), and private incentives (Agarwal and Ben-David 2018; Tzioumis and Gee 2013). Additionally, recent studies highlight the importance of emotions and sentiment in financial decisions (Cortés, Duchin, and Sosyura 2016; Félix, Kräussl, and Stork 2020; Zhang and Li 2019; Alzahrani et al. 2023), though little evidence has yet been provided on how these factors play a role in loan officers' comments on lending applications.

3 Bank lending environment and credit scoring process

The Bank's existing borrowers undergo an annual credit monitoring process to assess both their credit structure and credit quality. During this process, the Bank reviews its past credit decisions, either confirming or adjusting the previously granted credit—upward or downward. An upward credit renewal results in more credit being granted to a client, while a downward re-evaluation leads to credit rationing. In this case, the loan officer reassesses the borrower's overall position in terms of creditworthiness, current credit lines, actual credit usage, future liquidity needs, and loan pricing. If these parameters remain unchanged, the credit renewal process typically results in a confirmation of the current credit lines. However, if any of these parameters change, such as (i) a deterioration or improvement in creditworthiness, (ii) under- or over-usage of existing credit lines, (iii) pricing that is no longer aligned with updated credit risk, or (iv) new credit needs, the loan officer may propose modifying the credit structure by either increasing or rationing the borrower's credit lines.

The renewal process begins with the loan officer responsible for the firm-bank relationship. Before initiating the renewal, the loan officer completes the rating attribution procedure, resulting in the borrower’s updated final rating, which is then attached to the credit application. During the rating process, the loan officer collects hard information from the borrower’s financial statements, pro-forma quarterly financial statements, and future business plans, where available. Additionally, through interviews and plant visits, the loan officer gathers soft information. It is important to note that loan officers’ compensation includes both a fixed salary and a discretionary component based on the volume of loans they generate. Each year, loan officers are assigned a personal loan target to qualify for minimum extra bonuses, with larger bonuses associated with loan volumes above the target.

Furthermore, loan officers' activities are closely monitored at the bank’s headquarters, where their career prospects depend on evaluations by the bank’s risk management officers. These evaluations not only consider loan volume but also accuracy. As a result, loan officers are incentivized to gather and fully exploit soft information to improve the accuracy of their lending decisions, consistent with findings in Filomeni et al. (2023), which demonstrate the significant role of soft information in predicting corporate defaults. Credit decisions are not entirely centralized; rather, in line with common industry practices, the hierarchical level of approval is determined by a set of applicant and loan characteristics specified in the bank’s credit policy manuals. The lowest hierarchical level for approval remains the loan officer. These rules account for the total exposure the bank has to the applicant company (or, in the case of subsidiary corporations, to the economic group to which the applicant belongs), the amount of credit requested, the applicant’s credit score, and the strength of credit risk mitigation, such as collateral and personal guarantees.

The Bank uses a hybrid credit scoring process in which a borrower's final rating is based on both quantitative (hard) and qualitative (soft) information provided by the local loan officer responsible for the loan application. The credit scoring process begins with a loan officer at the local branch where the company applies. The outcome of the rating process is a final rating for the borrower, consisting of up to fifteen rating notches, divided into three main rating categories based on the borrower's creditworthiness. The first notch represents the highest risk and is equivalent to an S&P rating of CCC. The loan officer supports the application by providing non-codifiable soft information in the form of textual notes attached to the application. These notes are subject to inspection by the loan approval authority at the Bank, and any reputational consequences could affect the loan officer's career prospects. Once the loan officer generates the credit score and submits the loan application with the attached textual notes, they can no longer modify their appraisal. The loan application, along with the textual notes and credit score produced by the loan officer, is forwarded through the Bank's hierarchy to the manager(s) with the authority to make the final loan approval decision and determine the credit amount. The identity of the loan officer managing the firm-bank relationship is revealed to the loan approver. However, by controlling for the loan officer's age in all our model specifications, we mitigate potential concerns related to this. The Bank's hierarchical structure includes eleven approval levels, with the loan officer originating the credit score at level 1 (the lowest) and the executive board at level 11 (the highest), as shown in Figure 1.

<<Figure 1 around here>>

The geographical location of the loan decision-maker varies depending on the level of approval, with loan applications being reviewed at the branch, regional, or headquarters level based on the characteristics of the specific application. At the second hierarchical level, the loan officer who examines the loan application and submits the credit score works at the same branch as the final decision-maker, minimizing information transmission issues and facilitating the use of soft information. These loans are approved by a senior branch manager who works at the same branch as the originating loan officer. At the third and fourth levels of approval, the loan officer with decision-making authority may either be located at the same branch as the originating loan officer (i.e., the loan appraiser) or at one of the Bank's seven regional departments. From the fifth level of approval onward, the final loan decision-maker is based at the Bank’s headquarters. To account for other potential organizational frictions related to the distribution of loan approval authority, which may influence the discretion of loan officers in ‘hardening’ soft information related to sentiment in the loan application, we control for the hierarchical level of approval in all our model specifications. This includes loans approved at various hierarchical levels. This approach allows us to control for potential communication and organizational frictions with managers who use sentiment in the lending decision. Additionally, to the extent that loan approval delegation is linked to specific characteristics of loan applications and multiple hierarchical levels are located within the same branch or headquarters, this helps us isolate the impact of information transmission issues related to physical distances between the communicating parties.

The hierarchical level of approval is determined by a set of applicant and loan characteristics outlined in the Bank's credit policy manuals. The rules for approval delegation consider factors such as the total exposure of the banking organization to the applicant company (or, in the case of subsidiary corporations, to the economic group to which the applicant belongs), the amount of credit requested, the applicant's credit score, and the strength of credit risk mitigation, including collateral and personal guarantees.

4 Methodology

In this paper, we use two types of information: structured and unstructured. For the structured data, we employ several classical statistical models to explore the relationship between loan approval decisions or the granted amount and the written evaluation provided by the loan officer. Specifically, we apply different approaches, including linear regression, logistic regression, and survival models. The use of these approaches is driven by the need to investigate alternative research hypotheses, leading to the modeling of various target variables. These variables range from the approval of the loan (binary) to the granted amount (continuous) and the time to lending (continuous). Linear regression, logistic regression, and survival analysis are three standard and widely-used supervised methods that can be seen as alternative and/or complementary. As is well-known, linear regression is used when the target variable is continuous, logistic regression when it is binary, and survival analysis when time-to-event data are available. Regardless of the model chosen, the ultimate goal is to assess and quantify the impact and significance of the variable of interest—our sentiment score—while controlling for other relevant factors. In the following sections, we outline the three alternative models fitted according to the different nature of the target variables and the set of input and control variables.

Firstly, a logistic model in which we model the approved/not approved status of loan applications as follows:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1\, SentimentScore + \beta_2\, SentimentScore \times Branch\text{-}HQ + \beta_3\, Branch\text{-}HQ + \lambda\, X_{controls} + \eta\, Industry\ Effects + \gamma\, Area\ Effects + \epsilon \quad \text{(Eq. 1)}$$

where p_i is the probability of the event of interest, namely the approved status for loan i, $X_{controls} = (x_{i1}, \dots, x_{ij}, \dots, x_{iJ})$ is a vector of loans-specific control explanatory variables. As further control, we specify the parameter η for the Industry effect and the parameter γ for the Area effect. All regressions include four geographical area and industry dummies to control for unobserved characteristics of local credit market and credit demand that could be correlated with our distance measures.

Secondly, an ordinary OLS model has been estimated on the granted credit as follows:

$$GC_i = \alpha + \beta_1 Sentiment\ Score + \ \beta_2 Sentiment\ Score\ \times Branch\text{-}HQ + \beta_3 Branch\text{-}HQ + \lambda\, X_{controls} + \eta\ Industry\ Effects + \ \gamma\ Area\ Effects + \ \epsilon \quad \text{(Eq. 2)}$$

where GC$_i$ is the continuous target variable representing the granted credit for loan i, $X_{controls} = (x_{i1}, \dots, x_{ij}, \dots, x_{iJ})$ is a vector of loans-specific control explanatory variables. Once again, as further control, we specify the parameter η for the Industry effect and the parameter γ for the Area effect.

An additional logistic model is considered, where we binarize the time to approval based on the 75th percentile. In other words, we assess whether a longer time to approval is influenced by soft information, which we proxy using the sentiment expressed by the loan officer.

$$\ln\left(\frac{s_i}{1-s_i}\right) = \alpha + \beta_1\, SentimentScore + \beta_2\, SentimentScore\ \times Branch\text{-}HQ\ + \beta_3\ Branch\text{-}HQ + \ \lambda\, X_{controls} + \eta\ Industry\ Effects + \ \gamma\ Area\ Effects + \ \epsilon \quad \text{(Eq. 3)}$$

where s$_i$ is the probability of the event of interest, namely a time to approval greater than the third quartile (75% percentile) for loan i, $X_{controls} = \left(x_{i1}, \dots, x_{ij}, \dots, x_{iJ}\right)$is a vector of loan-specific control explanatory variables. As a further control, we specify the parameter η for the Industry effect and the parameter γ for the Area effect.

As a final step, we apply a Cox Proportional Hazards survival model to analyze the probability of corporate financial distress as follows:

$$h(t) = h_0(t) \times \exp\,(\beta_1 Sentiment\ Score + \beta_2 Sentiment\ Score \times Branch\ HQ + \beta_3\ Branch\ HQ + \lambda\, X_{controls} + \eta\ Industry\ Effects + \gamma\ Area\ Effects + \epsilon \quad \text{(Eq. 4)}$$

where h_0 is called the baseline hazard, $h(t)$ is the estimated hazard function determined by the set of considered covariates, $X_{controls} = (x_{i1}, \dots, x_{ij}, \dots, x_{iJ})$ is a vector of loans-specific control explanatory variables. As further control, we specify the parameter η for the industry fixed effects and the parameter γ for the area fixed effects.

4.1 Textual analysis

Textual analysis has become increasingly important in the financial applications literature (Loughran and McDonald 2020). One challenge in this field is accurately converting qualitative information into quantitative measures.

Sentiment analysis, also known as opinion mining, involves the use of natural language processing, text analysis, and computational linguistics to identify and extract subjective information from source materials (Liu and Zhang 2012). It is widely used in applications such as reviews and social media analysis, spanning areas from marketing to customer service. The goal of sentiment analysis is to evaluate the sentiments and opinions of a writer, either within a specific topic domain or across multiple topics. This is achieved by calculating the aggregated sentiment polarity, typically classified as positive or negative. Existing sentiment analysis approaches can be categorized into four main levels: word-level, sentence-level, document-level, and aspect/entity-level (Kumar and Abirami 2018).

The most critical part of the analysis relies on the sentiment classification. In general, two different approaches can be used:

- Score-dictionary based: The sentiment score is determined by counting the matches between predefined lists of positive and negative words and terms in each text source (such as a tweet, sentence, or paragraph);
- Score-classifier based: A statistical classifier is trained on a sufficiently large dataset of pre-labeled examples and is then used to predict the sentiment class of a new example.

However, the second option is rarely feasible because fitting a good classifier requires a large amount of pre-labeled data, which is particularly challenging when working with non-English documents, as is the case in our study. This is due to the limited availability[1] of textual analysis tools for the Italian language. Therefore, we chose to focus on a dictionary-based approach, adapting appropriate lists of positive and negative words using the well-known Loughran-McDonald (LM) dictionary, which contains words categorized

---

[1] Nicola (2018) addressed the issue of Italian sentiment analysis by developing SentITA, a system that participated in the ABSITA task proposed in Evalita 2018. The system is based on a Bidirectional Long Short-Term Memory (BiLSTM) network with attention, which utilizes word embeddings along with sentiment-specific polarity embeddings. It also incorporates grammatical information from POS tagging and NER tagging. SentITA participated in both the Aspect Category Detection (ACD) and Aspect Category Polarity (ACP) tasks, achieving 5th place in the ACD task and 2nd place in the ACP task.

by sentiment (Loughran and McDonald 2011).

These lexicons are based on unigrams, meaning they consist of single words. They include many English words, each labeled with sentiment scores for positive or negative sentiment, and possibly for emotions such as joy, anger, sadness, and more. For our purposes, we decided to focus solely on two emotion classes: positive and negative. The overall workflow for the textual analysis follows these steps: preprocessing, sentiment extraction, and visualization.

- Pre-processing phase: This involves removing unnecessary structure and extra text that is not relevant for the analysis. Each raw text may require different steps for data cleaning, often involving trial and error and exploration.
- Sentiment Analysis: Before calculating the sentiment score, two crucial aspects need to be addressed: the list of stopwords and the dictionary of positive and negative words. The positive and negative words from the LM dictionary were translated into Italian and checked individually due to mismatches in many constructions, such as words starting with "UN," "MIS," and "DIS" (e.g., unseen, mismatch, disabled).

Dictionaries of positive and negative words have been extended to enrich the algorithm capability to detect and score words into the positive and negative categories.

The dictionary was expanded using Kaggle resources for NLP in different languages (Chen and Skiena 2014)[2].

Negative words were increased by 1,375 terms, including specific words from the text, such as Italian comments in the loan application under the "proposal comment" section, like "revocation," "closure," "unresolved," "suffering," and "litigation." Regarding positive words, the original LM dictionary was expanded from 354 to 2,221 by adding context-specific words, such as "fairness," "improve," "repayment," "doubled," "renewed," "elevating," "expanding," and "capitalized." For stopwords, they were translated and adapted to the context, resulting in a list of 477 words to exclude. While this method has the advantage of simplicity, it does not account for polysemy, irony, or sarcasm. However, the nature of the documents and the formal language used by loan officers allow us to simplify the approach without compromising accuracy.

The sentiment calculation generates an index based on the algebraic sum of the scores associated with each token: +1 for positive words and -1 for negative words. In Figure 2, we display the sentiment score as a proxy for soft information, which shows an

[2] https://www.kaggle.com/rtatman/sentiment-lexicons-for-81-languages The sentiment lexicons in this dataset were generated through graph propagation based on a knowledge graph—a graphical representation of real-world entities and the connections between them. The general idea is that words closely linked on a knowledge graph are likely to have similar sentiment polarities. For this project, sentiments were derived from English sentiment lexicons.

asymmetric pattern, with a peak at the negative values and a long tail extending toward the positive values.

In Figure 3, we present the Tf-idf plot, which shows the term frequency-inverse document frequency for each translated token. This helps assess how relevant a word is to a document within the corpus. Additionally, we also include a word cloud in Figure 4.

<<insert Figure 2 around here>>

<<insert Figure 3 around here>>

<<insert Figure 4 around here>>

5 Data

5.1 Dataset

We examine the credit scoring and lending decisions of the lead bank (the ‘Bank’) in a large multinational European banking group, which has total assets of 646 billion euros, a market capitalization of approximately 50 billion euros, and subsidiaries in twelve central-eastern European and Mediterranean countries. In the country under analysis, the Bank operates over 1,900 traditional branches and holds a market share of around 15% in the loan and deposit markets. The group, as a whole, has 14 affiliated banks and about 4,500 branches.

The data used in this study were manually collected from the credit folders of 550 SMB corporate loan applications managed by the Corporate and Investment Banking (CIB)

Division of the Bank, covering the period from September 2011 to September 2012. The dataset pertains to one of the Bank's major markets—the Italian mid-corporate segment—which includes firms with annual turnover between 150 million and 1 billion euros. The CIB Division, responsible for the mid-corporate loan applications analyzed in this paper, is structured into 24 corporate branches spread across 13 regions in the country. These branches reflect a new organizational structure established by the Bank to separate relationships with medium and large enterprises from its traditional retail banking services for families and small businesses. Loan officers at CIB branches oversee loan applicants from the outset of the lending process, generating ratings and submitting loan proposals to the relevant authority within the Bank's hierarchy.

This segment of the loan market is typically less affected by information opacity compared to SMBs. As a result, lending to the mid-corporate segment should be less susceptible to issues related to the transmission of information across the bank's organizational structure. This may bias our findings against detecting an impact of organizational distance (i.e., against identifying frictions in information transmission) on credit scoring and lending decisions. Each credit folder contains highly granular information on the credit scoring process, including the final score and all intermediate scores. Additionally, each folder includes comprehensive data on the applicant and loan characteristics, the identity and location of the loan officer handling the application, and the hierarchical level at which the loan is ultimately approved (or denied).

5.2 Soft information and sentiment score formation

To leverage the unstructured data from the loan applications, we converted each document into text in order to extract unstructured data such as the comments within the loan applications.

The first challenge in working with PDF documents and extracting information from them is the inconsistency in the structure of the PDF files. As a result, the extraction process becomes a trial-and-error task, with the challenges and difficulties largely depending on the type of document.

Similar to scientific articles, loan application files are often stored in PDF format, which is primarily designed for printing and not optimized for searching or indexing. To extract the text, we used the R package *pdftools*, which allows for retrieving both text and metadata from PDF files. The key function used is pdf text, which returns a character vector with a length equal to the number of pages in the PDF. Each string in the vector contains the plain text version of the content on that page, which serves as our unit of interest.

Basic statistics on the minimum and maximum number of raw words are presented in Table 1.

<<Insert Table 1 around here>>

From the graphical visualization in Figure 5, which shows the distribution of the sentiment variables, we can observe the relationship between the number of words and the sentiment score for each approval class (0, 1). It is clear that approved loans (class 1) have higher sentiment scores. Additionally, the relationship between the sentiment score and the number of words for class 1 is notably positive and linear.

<<Insert Figure 5 around here>>

5.3 Main and control variables

The variables used in the analysis (see Table 2) are a subset of the 157 available variables, covering areas such as loan characteristics, borrower specifics, firm characteristics, rating indexes, macroeconomic indicators, sentiment variables to proxy soft information, and distance measures. The dependent variable is "approved," which is a dummy variable indicating 1 if the loan application is approved and 0 if it is rejected. The distribution is 13% for class 0 (rejected) and 87% for class 1 (approved).

We can classify input variables as follows:

- Soft information measured by the sentiment variable: The feature used in the model is the *sent per words* scaled accordingly. This variable represents the sentiment score extracted from the text, as introduced in Section 1. Each positive word is assigned a score of +1, and each negative word a score of -1. The sentiment score for each document was divided by the number of words in the relevant text section, and then the result was scaled.
- Rating measures: The variable that captures the loan rating is "final rating." This variable is part of the rating measures available for each loan and represents the integrated rating, along with any potential rating override by the loan officer (LO).

It is a step variable, taking values between 1 and 15, where 1 indicates the highest rating.

- Organizational distance: Our key organizational distance variable is Branch-HQ, which is measured as the logarithm of 1 plus the physical distance between the branch where the loan officer responsible for the credit score operates and the Bank's main headquarters (HQ). This variable reflects communication frictions resulting from spatial separation and the lack of personal contact, cultural affinity, common languages, and mutual trust between loan officers at local branches and senior managers at the Bank's headquarters. These factors may hinder the hardening and transmission of soft information within the banking organization. Specifically, Branch-HQ directly captures agency costs and communication frictions that arise due to the physical distance and the absence of personal interactions between bank officers, which can impede the flow of soft information. This organizational distance, measured as a continuous variable, captures the presence of communication frictions that prevent the soft information embedded in sentiment from traveling within the organization without loss of content. The choice of this variable aligns with Liberti and Mian (2009), who studied loans to large corporate customers made by a multinational bank in Argentina. They found that the sensitivity of the approved loan amount to subjective, soft information declined as the geographical distance between the loan officer and the higher-ranked officer responsible for the final decision increased, while sensitivity to objective, hard information was significantly higher for loans approved at a greater hierarchical distance. The use of this variable is also consistent with Filomeni et al. (2020), who demonstrated that the hardening of soft information is significantly sensitive to branch-to-headquarters distance. Additionally, Branch-HQ captures cultural factors that hinder information transmission and contribute to agency problems within the Bank's hierarchy. These include lower cultural affinity between officers at local and head branches, language differences, and reduced mutual trust. In our sample, the average Branch-HQ distance is 290.2 kilometers, with distances ranging from 2.5 to 1,482 kilometers[3].

From the credit folders, we extract various loan and firm characteristics that may influence the probability of loan approval. We include several variables that capture the accessibility and transmissibility of soft information, as well as the potential for agency problems. First, to capture the breadth of the bank-firm relationship, we control for scope of relationship, a binary variable that takes the value of 1 if the borrower purchases at least one additional product or service from the Bank beyond the loan, and 0 otherwise. Second, we consider the *logarithm of 1 plus the distance* between the branch where the loan officer

---

[3]We have descriptively examined borrower characteristics and sentiment scores, finding no systematic differences between centrally located and peripheral branches across the different percentiles of organizational distance. As a result, we have included three additional figures (A.1-A.4) in the Appendix.

works and the headquarters of the applicant company (referred to as branch-to-borrower distance).

Recent banking literature suggests that this distance reduces information asymmetries and monitoring costs (Petersen and Rajan 2002; Degryse and Ongena 2005; Agarwal and Hauswald 2010; Bellucci, Borisov, and Zazzaro 2013). Third, we control for the distribution of loan approval authority within the banking organization using *approval level*, which represents the hierarchical level at which loan approval occurs. This helps account for additional communication and organizational frictions with managers who use the credit score to make lending decisions. These frictions may influence the loan officers' discretionary decisions to "harden" soft information into the loan applications. Moreover, to the extent that loan approval authority is related to specific characteristics of loan applications and that multiple hierarchical levels are located in the same branch or headquarters, this allows us to isolate the impact of information transmission problems related to physical distances between the communicating parties.

We also control for a set of borrower and loan characteristics that reflect the firm’s financial health and information transparency. First, we include the final rating assigned to the borrowing firm. Second, we add the binary variable collateral, which takes the value of 1 if the credit line is collateralized and 0 otherwise. Additionally, we control for whether the firm is part of a group using the binary variable group belonging. Finally, we include several firm size measures, including the logarithm of total assets, long-term debt, and EBITDA.

Regarding loan officer characteristics, we control for the *gender* of the loan officer using the indicator variable female, which takes the value of 1 if the loan officer is female. Our sample includes data from 122 loan officers across 24 CIB branches, 78% of whom are male. Psychological, sociological, and economic studies suggest that women are generally more risk-averse and less overconfident than men (see Croson and Gneezy 2009) and tend to have lower job mobility due to societal roles and gender-based discrimination in both the workplace and labor market (see Loprest 1992; Fuller 2008; Del Bono and Vuri 2011). Women are often found to be less selfish and competitive (see Buser, Niederle, and Oosterbeek 2014) and more selfless and supportive than men (see Eckel and Grossman 1998, 2001; Dufwenberg and Muren 2006). These differences have been shown to influence loan officer behavior in areas such as loan origination, risk-taking, and lending relationships (see Bellucci, Borisov, and Zazzaro 2010; Agarwal and Ben-David 2018; Beck, Behr, and Guettler 2013; Bose, Filomeni, and Tabacco 2024). We also control for the *age* of the loan officer using the binary variable *aged*, which takes the value of 1 if the loan officer's age is equal to or greater than the median value of 50, and 0 otherwise. The average loan officer in our sample is 49 years old, with 21 years on the job, and the age range is from 20 to 60 years. Existing evidence on the impact of these variables on lending decisions is mixed. Agarwal and Wang (2009) and Agarwal and Ben-David (2018) find that older loan officers tend to approve more loans, but those loans are also more likely to default, suggesting that risk aversion and career concerns are more prominent early in a loan officer's career. In contrast, Beck, Behr, and Guettler (2013) find that loans underwritten and monitored by older officers are less likely to turn problematic. Therefore, the expected impact of age on lending decisions is ambiguous.

Finally, all regressions include four geographical area and industry dummies to control for unobserved characteristics of the local credit market and credit demand that may be correlated with our distance measures.

<<Insert Table 2 around here>>

6 Empirical Evidence

6.1 Soft information and lending decisions: the moderating effect of organizational distance

In Table 3, we present our results on the moderating effect of organizational distance on the relationship between the loan officer's soft information and the lending decision. Specifically, in column 1 of Table 3, we report the results for the loan approval rate, while in column 2, we show the results for the amount of granted credit. The main variable of interest is the coefficient of the interaction term between our proxy for soft information (Sentiment) and organizational distance: Sentiment score × Branch-HQ. This variable captures the moderating effect of organizational distance on the influence of the loan officer's sentiment on the lending decision.

As for the lower-order terms, we first find a significant and positive impact of the Sentiment score generated by the loan officers on both the loan approval rate and the amount of credit granted (coefficients of 1.317 and 0.062, respectively). This suggests that, on average, more positive sentiment is associated with better lending outcomes. The economic magnitude of these coefficients indicates that a one standard deviation increase in the Sentiment score improves the probability of loan approval by 3 times and increases the granted credit by 6.2%, respectively. Second, we find that Branch-HQ has a significant and positive effect on the loan approval rate, while having a significant and negative effect on the amount of credit granted (coefficients of 0.650 and -0.072, respectively). This suggests that loans originating from organizationally distant branches are, on average, of smaller amounts but more likely to be approved.

As for the interaction term, we find a significant and negative impact of Sentiment Score × Branch-HQ on both the loan approval rate and the amount of credit granted (coefficients of -0.081 and -0.021, respectively). This suggests that, on average, greater organizational distance diminishes the positive effect of good sentiment on lending outcomes, consistent with the presence of information frictions caused by communication at a distance (Filomeni, Udell, and Zazzaro 2021). Overall, our findings confirm that sentiment, which is intensive in soft information, has a fundamental spatial dimension. This makes internal communication more challenging over greater organizational distances between bank offices (Stein, 2002; Alessandrini et al., 2009), thereby confirming the persistence of spatially-based organizational frictions in transmitting loan officers' soft information, even in the context of modern credit-scoring-based lending technology.

Turning to other control variables, the *Female* variable captures the influence of gender bias in lending. We find that female loan officers have lower approval rates but grant larger amounts of credit. These results suggest that female loan officers may assess borrowers' soft information and creditworthiness more cautiously, granting more credit but with a lower approval rate compared to their male counterparts. This aligns with Beck et al. (2013), who highlight gender differences in risk aversion between women and men. Next, we find that older loan officers (Aged) tend to grant less credit and have lower approval rates. Additionally, Collateral is associated with higher approval rates but lower credit amounts. This result is consistent with Berger and Udell (1990), who found that collateral is often linked to riskier borrowers and riskier loans. We suggest that when risk is observable, high-risk borrowers are more likely to pledge collateral, which explains the negative relationship between Collateral and Granted credit. Higher hierarchical levels (Approval level) are associated with greater loan approval rates and larger amounts of credit. The depth of the bank-firm relationship (Scope of relationship) positively impacts both loan approval rates and granted credit, indicating the beneficial effect of soft information-intensive relationship banking on lending outcomes. Conversely, Group belonging is associated with higher approval rates and larger amounts of credit. Finally, Branch-borrower distance has a negative and significant effect on loan approval rates, as it exacerbates information asymmetries and increases monitoring costs (Petersen and Rajan 2002; Degryse and Ongena 2005; Agarwal and Hauswald 2010; Bellucci, Borisov, and Zazzaro 2013).

<<Insert Table 3 around here>>

6.2 Soft information and time to lending decision: the moderating effect of organizational distance

In Table 4, we present our results on the moderating effect of organizational distance on the relationship between the loan officer's soft information (measured by sentiment) and the time to the lending decision. The main variable of interest is the coefficient of the interaction term Sentiment score × Branch-HQ, which captures the moderating effect of organizational distance on the influence of the loan officer's sentiment on the time required to reach the lending decision, i.e., *Long* time to approval**.**

As for the lower-order terms, we first find a significant and negative impact of the Sentiment score generated by loan officers on the time to the lending decision (coefficient of -0.355). This suggests that, on average, more positive sentiment is associated with a shorter time to reach the loan approval decision. The economic magnitude of these coefficients indicates that a one standard deviation increase in the Sentiment score reduces the likelihood of a long approval time by 70%. Second, we find that Branch-HQ has a significant and negative effect on the likelihood of a long approval time (coefficient of -0.342), suggesting that loans from organizationally distant branches are, on average, more likely to be approved in a shorter time frame.

Regarding the interaction term, we find a significant and positive impact of Sentiment score × Branch-HQ on the time to the lending decision (coefficient of 0.105). This is consistent with the hypothesis that geographically distant loan officers have incentives to convey more favorable sentiment to increase the likelihood of loan approval.

As a result, senior managers at the Bank's headquarters, aware of this tendency to inflate sentiment, pay particularly close attention to loan applications with strong "hardened" soft information submitted by loan officers at distant branches. This further supports the existence of communication frictions within bank organizations (Stein, 2002; Alessandrini et al. 2009; Filomeni, Udell, and Zazzaro 2021).

Moving on to other control variables, we find that older loan officers (Aged) take more time to decide on a loan application, Collateral is associated with a shorter time to lending, and the higher the Approval level, the longer it takes to reach the lending decision.

<<Insert Table 4 around here>>

6.3 Soft information and corporate default probability: the moderating effect of organizational distance

In Table 5, we present our results on the moderating effect of organizational distance on the relationship between the loan officer's soft information (proxied by sentiment) and the probability of corporate default. The main variable of interest is the coefficient of the interaction term Sentiment score × Branch-HQ, which captures the moderating effect of organizational distance on the influence of the loan officer's sentiment on the probability of corporate default, i.e., Distress.

As for the lower-order terms, we first find a significant and negative impact of the Sentiment score generated by loan officers on the likelihood of firm default (coefficient of -0.316). This suggests that, on average, more positive sentiment is associated with lower corporate default probabilities. The economic magnitude of this coefficient indicates that a one standard deviation increase in the Sentiment score generated by loan officers reduces the likelihood of corporate default by 30%. Second, we find that Branch-HQ has a significant and negative effect on the likelihood of borrower bankruptcy (coefficient of -0.450), suggesting that loans from organizationally distant branches are, on average, of better credit quality.

Regarding the interaction term, we find a significant and negative impact of Sentiment score × Branch-HQ on corporate default probability (coefficient of -0.076). This is consistent with the hypothesis that geographically distant loan officers are incentivized to transmit more reliable sentiment to the higher levels of the Bank's hierarchy, being aware of the potential reputational effects associated with soft information transmission over greater organizational distances. This further supports the existence of communication frictions in the transmission of soft information across the banking organization (Stein, 2002; Alessandrini et al., 2009; Filomeni, Udell, and Zazzaro, 2021).

Moving to other control variables, we find that loans extended by female loan officers (Female) are associated with lower corporate default probabilities, consistent with their greater risk-aversion.

Next, we find that older loan officers (Aged) experience more corporate defaults, which is consistent with less career concern at stake for more senior officers. Additionally, we show that Scope of relationship and Branch-borrower distance are associated with higher corporate default probabilities. The former result suggests that deeper bank-firm credit relationships, which often involve more debt obligations, lead to an overall increase in corporate default probability. The latter result indicates that greater branch-borrower distance hinders the collection and accumulation of soft information, which is important in lending decisions (Petersen and Rajan 2002; Degryse and Ongena 2005; Agarwal and Hauswald 2010; Bellucci, Borisov, and Zazzaro 2013). The presence of Collateral and Group belonging are associated with lower corporate default probabilities, due to collateral pledges and the potential financial support from other companies within the same group. Specifically, when a firm is facing financial and/or operational distress and cannot access external financing, controlling shareholders may intervene to support the struggling unit (Friedman, Johnson, Mitton, 2003; Bae et al., 2012). In this regard, Lins, Volpin, and Wagner (2013) document that families, a common form of controlling shareholder in Europe, prioritize the survival of companies within their business groups during crises. This often involves reallocating resources from less affected group members to those in distress. Finally, higher Approval *levels* are associated with increased corporate default probabilities, as riskier loans are typically approved at higher levels in the Bank's hierarchy.

<<Insert Table 5 around here>>

6.4 Additional cross-sectional analyses

We now conduct additional cross-sectional analyses to explore the potential effects of Scope of relationship**,** Collateral, and Approval level on the relationship between soft information, organizational frictions, and lending outcomes. Specifically, we introduce a triple interaction term by interacting Scope of relationship, Collateral, or Approval level with our previous interaction term Sentiment score × Branch-HQ.

First, we test whether organizationally distant managers are better equipped to process soft information for relationship customers—i.e., those with an extended credit relationship, as captured by the Scope of relationship variable. Organizationally distant managers may have greater access to broader organizational resources and deeper insights into long-term customer behavior, which could enhance their ability to interpret qualitative information. This, in turn, may allow them to make more informed lending decisions based on soft, non-quantifiable data such as trust, reputation, and informal communication. Our findings show that the negative moderating effect of organizational distance on the sentiment-granted credit relationship is mitigated for relationship customers with a deeper firm-bank relationship. Empirically, this is reflected in the positive and significant coefficient of the triple interaction term Sentiment score × Branch-HQ × Scope of relationship, which indicates a positive effect on the amount of granted credit (column 1 in Table 6). However, this increase in granted credit for relationship customers is accompanied by a decrease in the loan approval rate, as shown by the negative and significant coefficient on the triple interaction term Sentiment score × Branch-HQ × Scope of relationship in column 2 of Table 6.

One possible explanation is that the increased reliance on "hardened" soft information for these customers, combined with organizational distance, leads to more cautious credit evaluation processes by managers at greater distances. This caution may stem from heightened scrutiny of qualitative factors, such as borrower trustworthiness or economic prospects, which are often subjective. Furthermore, managers' lending decisions for relationship customers at greater organizational distances exhibit a higher probability of corporate default. This may be due to the challenge of interpreting less salient qualitative soft information, especially since distant managers are not directly involved in collecting it. This is reflected in the positive and significant coefficient on the triple interaction term Sentiment score × Branch-HQ × Scope of relationship in column 4 of Table 6. Indeed, organizationally distant managers may struggle to accurately interpret qualitative insights when they are not directly engaged with relationship customers, potentially overlooking critical quantitative risk indicators. Overall, our findings, as reported in Table 6, highlight that while more credit is being granted, a seemingly more selective approval process may emerge, as reflected in lower approval rates for relationship customers. However, more credit is granted to high-risk borrowers, suggesting that organizationally distant managers often struggle to accurately interpret soft information collected by local loan officers. The lack of direct engagement can lead to misinterpretations of qualitative insights and a diminished understanding of the underlying context. This further supports the empirical evidence on the local nature of soft information and the limitations of "hardening" soft information.

<<Insert Table 6 around here>>

Second, we test whether organizationally distant managers are more likely to rely on soft information when a loan is collateralized, with the *Collateral* variable capturing this status. The reasoning behind this is that organizationally distant managers may feel that the presence of collateral mitigates the financial risk associated with the loan, making them more willing to consider qualitative factors (soft information) when assessing creditworthiness. Our findings show that the negative moderating effect of organizational distance on the sentiment-approval rate relationship is mitigated by the presence of collateral. Empirically, this is reflected in the positive and significant coefficient on the triple interaction term Sentiment score × Branch-HQ × Collateral, which highlights a positive effect on loan approval rates (column 2 in Table 7). Additionally, the presence of collateral reduces the time to reach the lending decision by organizationally distant managers (column 3 in Table 7). Overall, our findings, as reported in Table 7, show that organizationally distant managers are more likely to rely on soft information when a loan is collateralized.

<<Insert Table 7 around here>>

Lastly, we examine whether the negative moderating effect of organizational distance on the relationship between soft information and lending outcomes is limited to top approval levels or if it also affects mid-level approval processes within the banking organization. Understanding this distinction is crucial, as top-level approvals, typically made at the Bank's headquarters, involve senior managers whose ability to interpret soft information may be compromised by organizational distance from local branches and clients. This could amplify the negative impact on lending outcomes. In contrast, mid-level managers, who often have more direct interaction with relationship customers, may be better positioned to utilize soft information in their decision-making. To conduct this analysis, we create the binary variable Approval level HQ, which takes the value of 1 for loans approved at the Bank's headquarters and 0 otherwise. Our results show that organizationally distant HQ approval levels are associated with lower loan approval rates, as indicated by the negative and significant coefficient on the triple interaction term Sentiment score × Branch-HQ × Approval level HQ (column 2 in Table 8).

Interestingly, while organizationally distant HQ approval levels are linked to lower approval rates, organizationally distant low- and mid-level approval levels tend to reduce the amount of credit granted, possibly due to the difficulties HQ managers face in interpreting locally collected soft information. This is reflected in the negative and significant coefficient on the interaction term Sentiment score × Branch-HQ, as well as the non-significant coefficient on the triple interaction term Sentiment score × Branch-HQ × *Approval level* HQ for the amount of granted credit (column 1 in Table 8). Furthermore, the increased time required for lending decisions at organizationally distant HQ levels underscores the challenges in effectively interpreting transmitted soft information. This is indicated by the positive and significant coefficient on the triple interaction term Sentiment score × Branch-HQ × Approval level HQ, which shows a positive effect on *long* time to approval (column 3 in Table 8). Additionally, we find a negative and significant impact of Sentiment score × Branch-HQ × Approval level HQ on corporate default probability, consistent with the hypothesis that geographically distant loan officers are incentivized to transmit more reliable soft information to the HQ layers of the Bank's hierarchy, aware of the reputational effects of soft information transmission across greater organizational distances. This evidence further supports the existence of communication frictions in soft information transmission across the banking organization (Stein, 2002; Alessandrini et al. 2009; Filomeni, Udell, and Zazzaro 2021).

<<Insert Table 8 around here>>

6.5 Robustness checks

We conduct additional robustness checks to confirm our main empirical findings. First, we introduce two alternative soft information measures: a mood variable with three levels (-1 for negative, 0 for neutral, and +1 for positive), and a binary sentiment variable that distinguishes negative evaluations from neutral and positive ones. The mood variable helps assess the pure rational content expressed by loan officers, with the neutral level identifying comments that are balanced between positive and negative elements. This neutral class is determined by distinguishing comments with high positive or negative scores from those with near-zero scores, which indicate a balance between the positive and negative components. On the other hand, another classification might involve isolating only the negative comments to explore whether neutral and positive words, when considered together, can have a significant impact on predicting the approval status.

The results for the binary sentiment variable are reported in Table 9 and confirm our main findings across all models (columns 1-4, Table 9). The Sentiment score × Branch-HQ interaction term retains its sign and statistical significance, and the majority of relationships tested in the regressions remain consistent, including the first-order results for Sentiment score and Branch-HQ.

<< Insert Table 9 around here >>

The results for the three-level sentiment variable are reported in Table 10. Again, the coefficients associated with Sentiment score × Branch-HQ retain their sign and statistical significance in models (1), (2), and (4), while in model (3), the coefficient shows a positive sign but lacks statistical significance. Overall, we can conclude that the robustness checks further corroborate our findings from the baseline models.

<< Insert Table 10 around here >>

Secondly, to explore potential geographically-related linguistic differences across the sample (Cremer et al., 2007), we conduct an extensive exploratory data analysis to test for significant differences between the loan applications and their associated sentiment scores across the four geographical areas of the country. Not surprisingly, we found no evidence of a relationship between sentiment scores and geographic regions, suggesting that loan officers tend to use a common standard language, without regional variations in emphasis or tone.

Additionally, we conduct further robustness checks to test the consistency of our results. Untabulated results show that the main findings remain unchanged when we cluster standard errors at the branch level instead of the industry level. Furthermore, when we include an additional variable to capture the loan officer’s experience, alongside their age, the results remain consistent[4].

7 Conclusions

This paper investigates (i) the role of soft information in determining the outcome of loan applications by small and mid-sized businesses, in terms of approval rate, granted amount, time to decision, and the ability to predict company defaults; and (ii) whether organizational frictions impact the transmission of soft information throughout the decision-making process.

We analyze 550 corporate loan applications managed by a leading European bank. We extract soft information (measured as sentiment) from loan officers' comments on the applications and investigate how this affects lending outcomes, while also controlling for firm and loan officer characteristics. Regarding loan application outcomes, the sentiment score increases the likelihood of loan approval and the amount of credit granted. However, the ability of soft information to influence lending outcomes depends on organizational distance. As organizational distance increases, the positive effect of good sentiment on lending outcomes diminishes, consistent with the existence of information frictions caused by the distance between the loan officer and the deliberating entity. Our results also show

---

4 The high correlation between the age and experience of the loan officers prevents us from including both variables in the same model specification, due to potential multicollinearity concerns.

that a high sentiment score reduces the time to lending, but the interaction with distance shows a positive sign, further indicating frictions between the branch and the deliberating entity. This suggests that positive comments from distant loan officers may be scrutinized more carefully than those from officers at closer branches.

Finally, soft information serves as a good proxy for stronger companies, as positive sentiment is associated with firms that have a lower probability of default. Additionally, the interaction term shows a negative sign, suggesting that distant loan officers have incentives to communicate reliable positive sentiment due to reputational concerns. We also show that these effects hold for both top and mid-level approval stages, and that organizationally distant managers are less able to process subtle, qualitative soft information, particularly for soft-information-intensive relationship borrowers. Additionally, they are more inclined to rely on soft information when the loan is collateralized.

Overall, our results show that soft information, proxied by the loan officer's sentiment expressed in the comments attached to the application, plays a role in determining the outcome of lending applications. Additionally, organizational distance negatively impacts the ability of loan officers to effectively convey soft information to the deliberating entity.

Although our findings are limited to a single bank and a specific customer segment, they may be extended to credit institutions with similar organizational structures. However, different results could emerge in the context of leaner bank organizations or those that are geographically concentrated (e.g., savings and mutual banks). Further research could explore these alternative settings and assess whether sentiment also influences lending decisions in local banks.

References

Agarwal S, Ben-David I (2018) Loan prospecting and the loss of soft information. J Financ Econ 129(3):608-628.

Agarwal S, Benmelech E, Bergman N, Seru A (2012) Did the Community Reinvestment Act (Cra) Lead to Risky Lending? NBER Working Paper No. w18609.

Agarwal S, Hauswald R (2010) Distance and private information in lending. Rev Finan Stud 23(7):2757-2788.

Alessandrini P, Presbitero AF, Zazzaro A (2009) Banks, distances and firms' financing constraints. Rev Finan 13(2):261-307.

Alzahrani AI, Sarsam SM, Al-Samarraie H, Alblehai F (2023) Exploring the sentimental features of rumor messages and investors' intentions to invest. Int Rev Econ Finance 87:433-444.

Bae SC, Chang K, Kang E (2012) Culture, corporate governance, and dividend policy: international evidence. J Finan Res 35(2):289-316.

Banerjee A, Cole S, Duflo E (2009) Default and punishment: Incentives and lending behavior in Indian banks. Manuscript. Harvard Business School.

Beck T, Behr P, Guettler A (2013) Gender and banking: are women better loan officers? Rev Financ 17(4):1279-1321.

Bellucci A, Borisov A, Zazzaro A (2010) Does gender matter in bank–firm relationships? Evidence from small business lending. J Bank Financ 34(12):2968-2984. https://doi.org/10.1016/j.jbankfin.2010.07.008

Bellucci, A., Borisov, A., & Zazzaro, A. (2013). Do banks price discriminate spatially? Evidence from small business lending in local credit markets. J Bank Financ 37(11):4183-4197.

Berger AN, Udell GF (1990) Collateral, loan quality and bank risk. J Monet Econ 25(1):21-42.

Bose U, Filomeni S, Tabacco E (2024) Does Soft Information Mitigate Gender Bias in Corporate Lending? J Bus Ethics. https://doi.org/10.1007/s10551-024-05789-7

Brown M, Schaller M, Westerfeld S, Heusler M (2015) Internal control and strategic communication within firms–Evidence from bank lending. University of St. Gallen, School of Finance Research Paper 2015/04.

Buser T, Niederle M, Oosterbeek H (2014) Gender, competitiveness, and career choices. Q J Econ 129(3):1409-1447.

Campbell D, Loumioti M, Wittenberg-Moerman R (2019) Making sense of soft information: Interpretation bias and loan quality. J Account Econ 68(2-3):101240.

Cassar G, Ittner CD, Cavalluzzo KS (2015) Alternative information sources and information asymmetry reduction: Evidence from small business debt. J Account Econ 59(2- 3):242–263.

Chen Y, Skiena S (2014) Building sentiment lexicons for all major languages. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers):383-389.

Cortés K, Duchin R, Sosyura D (2016) Clouded judgment: The role of sentiment in credit origination. J Financ Econ 121(2):392-413. https://doi.org/10.1016/j.jfineco.2016.05.001.

Costello AM, Down AK, Mehta MN (2020) Machine + man: A field experiment on the role of discretion in augmenting AI-based lending models. J. Account. Econ. 70(2-3):101360. https://doi.org/10.1016/j.jacceco.2020.101360

Cowan CD, Cowan AM (2006) A survey-based approach of financial institution use of credit scoring for small business lending. Document de travail, Office of Advocacy, United States Small Business Administration, SBAH-04-Q-0021.

Cremer J, Garicano L, Prat A (2007) Language and the Theory of the Firm. The Q J Econ 122(1):373-407.

Croson R, Gneezy U (2009) Gender differences in preferences. J Econ Lit 47(2):448-474.

Degryse H, Ongena S (2005) Distance, lending relationships, and competition. J Finan 60(1):231-266.

Del Bono E, Vuri D (2011) Job mobility and the gender wage gap in Italy. Labour Econ 18(1):130-142.

DeYoung R, Glennon D, Nigro P (2008) Borrower–lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. J Financ Intermediation 17(1):113-143.

Dufwenberg M, Muren A (2006) Generosity, anonymity, gender. J Econ Behav Organ 61(1):42-49.

Eckel CC, Grossman PJ (1998) Are women less selfish than men? Evidence from dictator experiments. Econ J 108(448):726-735. https://doi.org/10.1111/1468-0297.00311

Eckel CC, Grossman PJ (2001) Chivalry and solidarity in ultimatum games. Econ Inq 39(2):171-188.

Félix L, Kräussl R, Stork P (2020) Implied volatility sentiment: a tale of two tails. Quant Finance 20(5):823-849. https://doi.org/10.1080/14697688.2019.1696018

Filomeni S, Udell GF, Zazzaro A (2020) Communication frictions in banking organizations: Evidence from credit score lending. Econom Lett 195:109412. https://doi.org/10.1016/j.econlet.2020.109412

Filomeni S, Udell GF, Zazzaro A (2021) Hardening soft information: does organizational distance matter? The Eur J Finance 27(9):1–31. https://doi.org/10.1080/1351847X.2020.1857812

Filomeni S, Bose U, Megaritis A, Triantafyllou A (2023) Can market information outperform hard and soft information in predicting corporate defaults? Int J Finance Econ: 1-26. https://doi.org/10.1002/ijfe.2840

Fiordelisi F, Monferrà S, Sampagnaro G (2014) Relationship lending and credit quality. J Financial Serv Res 46:295-315.

Fisman R, Paravisini D, Vig V (2017) Cultural proximity and loan outcomes. Am Econ Rev 107(2):457-492. DOI: 10.1257/aer.20120942

Friedman E, Johnson S, Mitton T (2003) Propping and tunneling. J Comp Econ 31(4):732-750.

Fuller S (2008) Job mobility and wage trajectories for men and women in the United States. Am Sociol Rev 73(1):158-183.

Hertzberg A, Liberti JM, Paravisini D (2010) Information and Incentives Inside the Firm: Evidence from Loan Officer Rotation. J Finance 65(3):795–828. https://doi.org/10.1111/j.1540-6261.2010.01553.x

Kumar AJ, Abirami S (2018) Aspect-based opinion ranking framework for product reviews using a Spearman's rank correlation coefficient method. Inf Sci 460:23-41. https://doi.org/10.1016/j.ins.2018.05.003

Liberti JM, Mian AR (2008) Estimating the effect of hierarchies on information use. Rev Finan Stud 22(10): 4057-4090. https://doi.org/10.1093/rfs/hhn118

Liberti JM, Petersen MA (2019) Information: Hard and soft. Rev Corp Fin Stud 8(1):1-41. https://doi.org/10.1093/rcfs/cfy009

Liberti JM, Sturgess J, Sutherland A (2022) How voluntary information sharing systems form: Evidence from a us commercial credit bureau. J Financ Econ 145(3):827-849. https://doi.org/10.1016/j.jfineco.2021.08.023

Lins KV, Volpin P, Wagner HF (2013) Does family control matter? International evidence from the 2008–2009 financial crisis. Review Finan Stud 26(10):2583-2619. https://doi.org/10.1093/rfs/hht044

Liu B, Zhang L (2012) A survey of opinion mining and sentiment analysis. In *Mining text data* (415-463). Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-3223-4_13

Loprest PJ (1992) Gender differences in wage growth and job mobility. Am Econ Rev 82(2):526-532.

Loughran T, McDonald B (2011) When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. J Finance 66(1):35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran T, McDonald B (2020) Textual analysis in finance. Annu Rev Financial Econ 12:357-375. https://doi.org/10.1146/annurev-financial-012820-032249

Nicola G (2018) Bidirectional attentional LSTM for aspect based sentiment analysis on Italian. EVALITA Evaluation of NLP and Speech Tools for Italian, 12(108).

Petersen MA, Rajan RG (1994) The benefits of lending relationships: Evidence from small business data. J Finance 49(1):3–37.

Petersen MA, Rajan RG (2002) Does Distance Still Matter? The Information Revolution in Small Business Lending. J Finance 57(6):2533–2570. https://www.jstor.org/stable/3094536

Puri M, Rocholl J, Steffen S (2011) Rules versus discretion in bank lending decisions. Available at SSRN 1786831.

Qian J, Strahan PE, Yang Z (2015) The impact of incentives and communication costs on information production and use: Evidence from bank lending. J Finance 70(4):1457-1493. https://www.jstor.org/stable/43612928

Sutherland AG (2020) Technology is changing lending: Implications for research. J Accoun Econ 70(2-3):101361. https://doi.org/10.1016/j.jacceco.2020.101361

Stein JC (2002) Information production and capital allocation: Decentralized versus hierarchical firms. J Finance 57(5):1891-1921. https://doi.org/10.1111/0022-1082.00483

Tzioumis K, Gee M (2013) Nonlinear incentives and mortgage officers' decisions. J Finan Econ 107(2):436-453. https://doi.org/10.1016/j.jfineco.2012.08.014

Zhang YJ, Li SH (2019) The impact of investor sentiment on crude oil market risks: Evidence from the wavelet approach. Quant Finance 19(8):1357-1371. https://doi.org/10.1080/14697688.2019.1581368

Tables

Table 1: Words and characters summary statistics.

Variable	Min	1^{st} Q	Median	Mean	3^{rd} Q	Max
Words raw	1.0	50.5	160.5	287.8	417.8	1088.0
Characters raw	5.0	853.2	2320.5	3696.3	5281.5	13440.0

Note: The number of words and characters in the 'proposal comment' section of the full PDF loan application. "Raw" refers to the extracted words before any cleaning procedures, such as stop-word exclusion, are applied.

Table 2: Descriptive statistics of the main used variables

Variable	Description	Obs	Mean	Std.dev	Min	Max
Dependent variables						
Approved	Binary variable equal to 1 if the loan is approved and 0 otherwise	550	0.86	0.34	0	1
Granted credit/Total assets (GC)	Ratio of borrower's granted credit over total assets	455	0.077	0.12	0	0.87
Long time to approval	Binary variable equal to 1 if the number of days taken to reach the lending decision is above the 75th percentile of the distribution (26 days) and 0 otherwise	550	0.23	0.42	0	1
Distress	Binary variable equal to 1 if the applicant firm is in financial distress and 0 otherwise	550	0.23	0.42	0	1
Sentiment variable						
Sentiment score	Sentiment score extracted from the given loan application	550	3.42	2.36	-10	10
Firm-Bank Relationship Variables						
Final rating	Final rating of a given borrowing firm	516	7.98	3.68	1	15
Collateral	Dummy equal to 1 if the credit line is col lateralized; 0 otherwise	550	0.39	0.49	0	1
Approval level	Step variable taking values between 1 and 11 according to the hierarchical level with the power of loan approval	550	4.14	2.76	1	11
Group belonging	Dummy equal to 1 if the borrower is part of an economic group; 0 if it is a stand- alone company	550	0.89	0.31	0	1
Scope of relationship	Dummy equal 1 if the borrower purchases at least one other banking product from the bank; 0 otherwise	550	0.52	0.50	0	1
Loan Officer Characteristics						
Female	Dummy equal to 1 if the loan officer responsible for the credit relationship is a female and 0 otherwise	539	0.22	0.41	0	1
Aged	Binary variable that takes the value of 1 if the age of the loan officer is equal or above the median value of 50, and 0 otherwise	520	0.53	0.50	0	1
Distance-related variables						
Branch-borrower distance	Log of 1 + distance between the branch in which the loan officer operates and the headquarters of the applicant company	550	4.39	2.09	0.18	9.59

	[km]		[795]	[2,523]	[0.2]	[14,753]
Branch-to-headquarters distance (Branch-HQ)	Log of 1 + distance in kilometers between the branch in which the loan officer operates and the Bank's headquarters	550	4.59	1.75	1.25	7.30
	[km]		[290]	[390]	[2.5]	[1,482]
Borrower's Characteristics						
Total assets	Logarithm of total assets	472	18.01	1.78	12.32	21.59
	[million euros]		[195]	[303]	[0.224]	[2,370]
Long-term debt	Ratio of the borrower's long-term debt over total assets	418	0.14	0.15	0	1.22
Ebitda	Ratio of the borrower's Ebitda over total assets	460	0.06	0.14	-1.37	0.60

Notes: This table presents the descriptive statistics for the variables used in the analysis, based on proprietary data collected from our data provider.

Table 3: Effects of Sentiment on Lending Decisions

VARIABLES	(1) Logit Model *Approved*	(2) OLS Model *Granted Credit*
Sentiment score	**1.317*** **	**0.062*** **
	(0.043)	**(0.005)**
Sentiment score × Branch-HQ	**-0.081*** **	**-0.021*** **
	(0.015)	**(0.002)**
Branch-HQ	0.650***	-0.072***
	(0.037)	(0.007)
Final rating	0.022	0.045**
	(0.032)	(0.011)
Female	-0.823***	0.298***
	(0.040)	(0.019)
Aged	-0.341***	-0.139***
	(0.030)	(0.016)
Collateral	1.964***	-0.084**
	(0.390)	(0.013)
Approval level	0.567***	0.121**
	(0.078)	(0.016)
Group belonging	2.452***	0.201***
	(0.524)	(0.011)
Scope of relationship	1.383***	0.279***
	(0.098)	(0.023)
Branch-borrower distance	-0.142***	-0.035
	(0.026)	(0.013)
Borrower's size	-0.512***	-0.441***
	(0.063)	(0.035)
Borrower's long-term debt	-0.156	0.622**
	(0.161)	(0.133)
Borrower's Ebitda	-0.105	-0.894**
	(-1.000)	(0.108)
Observations	326	326
Area FE	YES	YES
Industry FE	YES	YES
R^2 / Pseudo R^2	0.496	0.16

Robust standard errors in parentheses clustered at the industry level.
*** Significance at the 1% level
** Significance at the 5% level
* Significance at the 10% level.
Note: This table presents the impact of sentiment score and organizational distance on the likelihood of approval through a logit model (col 1) and on the amount granted by the bank by an OLS regression (col 2). Sentiment is extracted from loan officers' comments attached to the lending application folder sent to the deliberating entity for decision.

Table 4: Effects of Sentiment on Time to Lending

VARIABLES	Logit Model *Long Time to Approval*
Sentiment score	**-0.355****
	-0.199
Sentiment score × Branch-HQ	**0.105*****
	-0.036
Branch-HQ	-0.342***
	-0.138
Final rating	-0.001
	-0.019
Female	-0.274
	-0.459
Aged	0.993**
	-0.439
Collateral	-0.301**
	-0.163
Approval level	0.712***
	-0.178
Group belonging	0.879
	-0.775
Scope of relationship	-0.156
	-0.379
Branch-borrower distance	0.135
	-0.139
Borrower’s size	-0.246*
	-0.172
Borrower’s long-term debt	-0.878
	-1.313
Borrower’s ebitda	-0.182
	-1.404
Observations	326
Area FE	YES
Industry FE	YES
Pseudo R^2	0.391

Robust standard errors in parentheses clustered at the industry level
*** Significance at the 1% level
** Significance at the 5% level
* Significance at the 10% level.
Note: This table presents the impact of the sentiment score and organizational distance on the time to approval. Time to approval is measured by a dummy variable that takes the value of 1 if the decision time is particularly long. We define a 'long time to approval' as any time above the 75th percentile of the distribution. Sentiment is extracted from loan officers' comments attached to the lending application folder sent to the deliberating entity for decision.

Table 5: Effects of Sentiment on the Probability of Corporate Financial Distress

VARIABLES	Survival Analysis - Cox Proportional Hazards Model
	Distress
Sentiment score	**-0.316****
	(0.163)
Sentiment score × Branch-HQ	**-0.076****
	(0.036)
Branch-HQ	-0.450***
	(0.152)
Final rating	-0.437***
	(0.043)
Female	-0.725***
	(0.227)
Aged	0.624***
	(0.081)
Collateral	-0.247***
	(0.098)
Approval level	0.051***
	(0.017)
Group belonging	-0.162***
	(0.005)
Scope of relationship	0.960***
	(0.143)
Branch-borrower distance	0.013***
	(0.004)
Borrower's size	0.070
	(0.052)
Borrower's long-term debt	-0.020**
	(0.010)
Borrower's Ebitda	2.795***
	(0.258)
Observations	326
Area FE	YES
Industry FE	YES
Pseudo R^2	0.131

Robust standard errors in parentheses clustered at the industry level
*** Significance at the 1% level
** Significance at the 5% level
* Significance at the 10% level.
Note: This table presents the impact of the sentiment score and organizational distance on the probability of corporate financial distress. The distance to default is measured using a Cox Proportional Hazards Model.

Table 6. Additional cross-sectional analyses: Scope of Relationship

VARIABLES	(1) OLS Model *Granted Credit*	(2) Logit Model *Approved*	(3) Logit Model *Long Time to Approval*	(4) Survival Analysis Cox Proportional Hazards Model *Distress*
Sentiment score × Branch-HQ × Scope of relationship	0.050**	-0.250***	0.076	0.634***
	(0.010)	(0.053)	(0.099)	(0.119)
Sentiment score × Branch-HQ	-0.055**	-0.058	0.039	-0.652***
	(0.012)	(0.073)	(0.086)	(0.004)
Sentiment score × Scope of relationship	-0.342**	0.302	0.033	-4.017***
	(0.056)	(0.436)	(0.509)	(0.812)
Scope of relationship × Branch-HQ	-0.430**	1.786***	-0.150	-0.894
	(0.050)	(0.218)	(0.405)	(0.575)
Sentiment score	0.319**	1.348**	-0.211	3.282***
	(0.055)	(0.565)	(0.469)	(0.063)
Branch-HQ	0.228*	-0.272	-0.140	0.302*
	(0.058)	(0.418)	(0.218)	(0.181)
Scope of relationship	2.750***	-4.580***	-1.258	12.189***
	(0.254)	(1.756)	(1.805)	(3.471)
Final rating	0.039**	0.116***	0.033***	-2.552***
	(0.008)	(0.040)	(0.002)	(0.219)
Female	0.298***	-1.436***	-0.509***	-13.277***
	(0.024)	(0.046)	(0.092)	(0.391)
Aged	-0.150**	-0.585***	0.996***	4.734***
	(0.018)	(0.172)	(0.250)	(0.795)
Collateral	-0.120***	2.476***	-0.107	-54.175
	(0.012)	(0.473)	(0.265)	(0.000)
Group belonging	0.208***	-3.599***	0.758***	40.512
	(0.013)	(0.672)	(0.255)	(0.000)
Approval level	0.120**	0.627***	0.723***	-0.446***
	(0.018)	(0.094)	(0.047)	(0.058)
Branch-borrower distance	0.040	-0.136***	0.138**	0.172***
	(0.014)	(0.040)	(0.065)	(0.018)
Borrower's size	-0.439***	-0.672***	-0.307***	-0.193***
	(0.036)	(0.088)	(0.017)	(0.013)
Borrower's long-term debt	0.635**	-0.047***	-0.760	0.204**
	(0.136)	(0.011)	(1.099)	(0.080)
Borrower's Ebitda	-0.616***	-2.503**	-0.048	38.721***
	(0.052)	(1.064)	(0.201)	(2.671)
Observations	326	326	326	326
Area FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
R^2/Pseudo R^2	0.17	0.540	0.399	0.659

Table 7. Additional cross-sectional analyses: Collateral

VARIABLES	(1) OLS Model *Granted Credit*	(2) Logit Model *Approved*	(3) Logit Model *Long Time to Approval*	(4) Survival Analysis Cox Proportional Hazards Model *Distress*
Sentiment score × Branch-HQ × Collateral	0.064	0.434**	-0.116**	0.621
	(0.037)	(0.201)	(0.045)	(0.000)
Sentiment score × Branch-HQ	-0.041*	-0.140***	0.108***	-0.094***
	(0.014)	(0.002)	(0.022)	(0.005)
Sentiment score × Collateral	-0.383	-3.915***	0.531***	-4.016
	(0.158)	(1.334)	(0.171)	(0.000)
Collateral × Branch-HQ	-0.175	-3.022***	0.250	-3.720
	(0.153)	(0.855)	(0.168)	(0.000)
Sentiment score	0.167*	2.334***	-0.293**	-0.328***
	(0.051)	(0.108)	(0.149)	(0.057)
Branch-HQ	-0.002	1.245***	-0.255***	-1.746***
	(0.053)	(0.076)	(0.057)	(0.232)
Collateral	1.073	24.460***	-1.295	-32.789
	(0.664)	(5.799)	(0.862)	(0.000)
Final rating	0.046*	0.124***	0.014***	-2.444***
	(0.012)	(0.005)	(0.003)	(0.247)
Female	0.299***	-2.198***	-0.455***	-12.077***
	(0.021)	(0.130)	(0.079)	(0.699)
Aged	-0.151**	-1.123***	0.949***	5.947***
	(0.019)	(0.056)	(0.253)	(0.867)
Group belonging	0.211***	-5.733***	0.891***	46.693
	(0.015)	(0.522)	(0.197)	(0.000)
Approval level	0.129**	0.678***	0.695***	-0.624***
	(0.020)	(0.057)	(0.045)	(0.063)
Scope of relationship	0.262**	1.765***	-0.285	5.145***
	(0.035)	(0.123)	(0.271)	(0.462)
Branch-borrower distance	0.037	-0.125***	0.139**	0.457***
	(0.015)	(0.018)	(0.059)	(0.067)
Borrower's size	-0.453***	-1.051***	-0.259***	-0.086
	(0.040)	(0.081)	(0.030)	(0.082)
Borrower's long-term debt	0.647*	-1.683***	-0.940	0.149***
	(0.151)	(0.164)	(0.770)	(0.006)
Borrower's Ebitda	-0.974**	-5.171***	0.220	41.256***
	(0.106)	(0.021)	(0.287)	(3.608)
Observations	326	326	326	326
Area FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
R^2/Pseudo R^2	0.16	0.574	0.395	0.617

Table 8. Additional cross-sectional analyses: Approval level

VARIABLES	(1) OLS Model *Granted Credit*	(2) Logit Model *Approved*	(3) Logit Model *Long Time to Approval*	(4) Survival Analysis Cox Proportional Hazards Model *Distress*
Sentiment score × Branch-HQ × Approval level HQ	0.005	-0.082***	0.045***	-0.047***
	(0.002)	(0.014)	(0.005)	(0.001)
Sentiment score × Branch-HQ	-0.025**	0.027	-0.076***	-0.074***
	(0.004)	(0.051)	(0.013)	(0.010)
Sentiment score × Approval level HQ	-0.031**	0.485***	-0.363***	0.412***
	(0.006)	(0.039)	(0.056)	(0.058)
Approval level HQ × Branch-HQ	-0.044**	0.090*	-0.244***	0.330***
	(0.008)	(0.051)	(0.022)	(0.002)
Sentiment score	0.099***	0.766***	1.198***	-0.639***
	(0.010)	(0.109)	(0.024)	(0.194)
Branch-HQ	-0.007	0.430***	0.596***	-2.042***
	(0.017)	(0.089)	(0.060)	(0.324)
Approval level HQ	0.364**	0.037	2.565***	-3.081***
	(0.042)	(0.145)	(0.216)	(0.234)
Final rating	0.047*	0.069***	0.024***	-2.551***
	(0.011)	(0.018)	(0.001)	(0.291)
Female	0.304***	-0.936***	-0.427***	-12.753***
	(0.021)	(0.037)	(0.037)	(1.067)
Aged	-0.122**	-0.211***	0.967***	6.368***
	(0.019)	(0.055)	(0.274)	(1.080)
Collateral	-0.095**	1.893***	0.037	-52.450
	(0.014)	(0.202)	(0.204)	(0.000)
Group belonging	0.178***	-2.489***	0.896***	48.017
	(0.012)	(0.469)	(0.223)	(0.000)
Scope of relationship	0.271***	1.323***	-0.346	4.931***
	(0.025)	(0.021)	(0.308)	(0.467)
Branch-borrower distance	0.036	-0.158***	0.152***	0.565***
	(0.013)	(0.021)	(0.054)	(0.050)
Borrower's size	-0.431***	-0.515***	-0.292***	-0.203***
	(0.034)	(0.039)	(0.033)	(0.072)
Borrower's long-term debt	0.659**	-0.026	-1.272	0.148***
	(0.129)	(0.193)	(1.017)	(0.021)
Borrower's Ebitda	-0.984**	-1.120**	-0.935***	42.871***
	(0.101)	(0.558)	(0.166)	(4.415)
Observations	326	326	326	326
Area FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
R^2/Pseudo R^2	0.16	0.505	0.411	0.624

Table 9: Robustness tests. Alternative variable for Sentiment (Binary Variable)

VARIABLES	(1) OLS Model *Granted Credit*	(2) Logit Model *Approved*	(3) Logit Model *Long Time to Approval*	(4) Survival Analysis Cox Proportional Hazards Model *Distress*
Sentiment score	1.037***	5.386***	-1.933***	-16.862***
	(0.078)	(0.090)	(0.481)	(4.279)
Sentiment score × Branch-HQ	-0.257**	-0.506***	0.387***	-2.718***
	(0.026)	(0.106)	(0.107)	(0.653)
Branch-HQ	-0.060*	0.358**	-0.219**	-1.462**
	(0.019)	(0.173)	(0.088)	(0.576)
Final rating	0.045*	0.022	0.004	-1.638***
	(0.011)	(0.043)	(0.002)	(0.066)
Female	0.325***	-0.801***	-0.219***	-7.454***
	(0.019)	(0.133)	(0.035)	(0.372)
Aged	-0.147**	-0.422***	1.016***	3.593***
	(0.017)	(0.084)	(0.234)	(0.286)
Collateral	-0.057**	1.642***	-0.323	-48.066
	(0.012)	(0.081)	(0.211)	(0.000)
Approval level	0.118**	0.668***	0.732***	-0.637***
	(0.018)	(0.097)	(0.037)	(0.027)
Group belonging	0.210***	-1.820***	0.930***	41.328***
	(0.011)	(0.288)	(0.193)	(3.972)
Scope of relationship	0.275***	1.793***	-0.072	2.819***
	(0.023)	(0.128)	(0.282)	(0.060)
Branch-borrower distance	0.039	-0.024	0.116***	0.219***
	(0.015)	(0.066)	(0.030)	(0.013)
Borrower’s size	-0.444***	-0.605***	-0.288***	-0.454***
	(0.039)	(0.029)	(0.040)	(0.017)
Borrower’s long-term debt	0.599*	0.526	-0.980	0.191***
	(0.176)	(0.944)	(0.640)	(0.002)
Borrower’s ebitda	-1.072***	-2.505***	-0.124	21.336***
	(0.103)	(0.444)	(0.192)	(0.310)
Observations	326	326	326	326
Area FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
R^2/Pseudo R^2	0.16	0.438	0.387	0.609

Robust standard errors in parentheses clustered at the industry level

*** Significance at the 1% level
** Significance at the 5% level
* Significance at the 10% level.

Table 10: Robustness tests. Alternative variable for Sentiment (Three-Outcome Variable)

VARIABLES	(1) OLS Model *Granted Credit*	(2) Logit Model *Approved*	(3) Logit Model *Long Time to Approval*	(4) Survival Analysis Cox Proportional Hazards Model *Distress*
Sentiment score	0.536***	2.144***	0.188	-8.693***
	(0.043)	(0.037)	(0.279)	(1.356)
Sentiment score × Branch-HQ	-0.137**	-0.020**	0.026	-1.699***
	(0.016)	(0.010)	(0.053)	(0.265)
Branch-HQ	-0.069**	0.083	0.090	0.078
	(0.012)	(0.286)	(0.122)	(0.057)
Final rating	0.046**	-0.072	0.002	-1.990***
	(0.009)	(0.064)	(0.005)	(0.184)
Female	0.305***	-0.721***	-0.235***	-9.332***
	(0.025)	(0.120)	(0.068)	(0.255)
Aged	-0.147**	-0.229	1.014***	4.678***
	(0.019)	(0.172)	(0.227)	(0.770)
Collateral	-0.083*	1.514***	-0.313	-49.933
	(0.021)	(0.095)	(0.225)	(0.000)
Group belonging	0.214***	-0.804***	1.033***	39.255
	(0.008)	(0.146)	(0.205)	(0.000)
Approval level	0.120**	0.565***	0.693***	-0.675***
	(0.015)	(0.024)	(0.034)	(0.075)
Scope of relationship	0.274***	1.791***	-0.069	4.113***
	(0.020)	(0.045)	(0.288)	(0.361)
Branch-borrower distance	0.044**	-0.138	0.139***	0.190***
	(0.009)	(0.163)	(0.037)	(0.019)
Borrower’s size	-0.440***	-0.430***	-0.256***	-0.129**
	(0.033)	(0.051)	(0.043)	(0.063)
Borrower’s long-term debt	0.601*	-0.211	-0.978	0.105***
	(0.175)	(0.609)	(0.658)	(0.002)
Borrower’s ebitda	-0.924**	3.392***	-0.090	31.708***
	(0.103)	(0.019)	(0.112)	(2.300)
Observations	326	326	326	326
Area FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
R^2/Pseudo R^2	0.16	0.406	0.384	0.621

Robust standard errors in parentheses clustered at the industry level

*** Significance at the 1% level
** Significance at the 5% level
* Significance at the 10% level.

Figures

Figure 1: The Bank's organizational flowchart

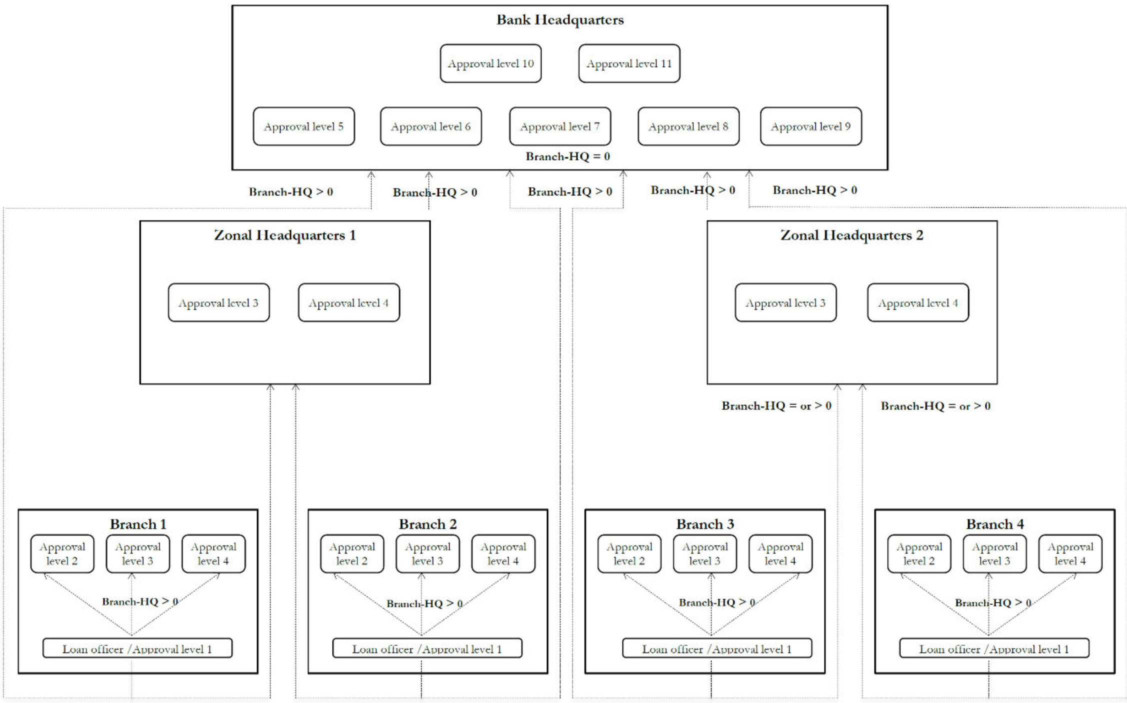


Figure 2: Histogram distribution of sentiment score

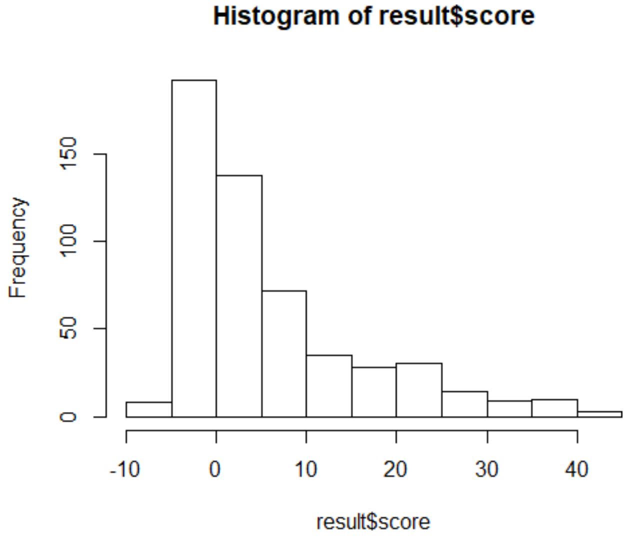


Figure 3: Tf–idf term frequency–inverse document frequency

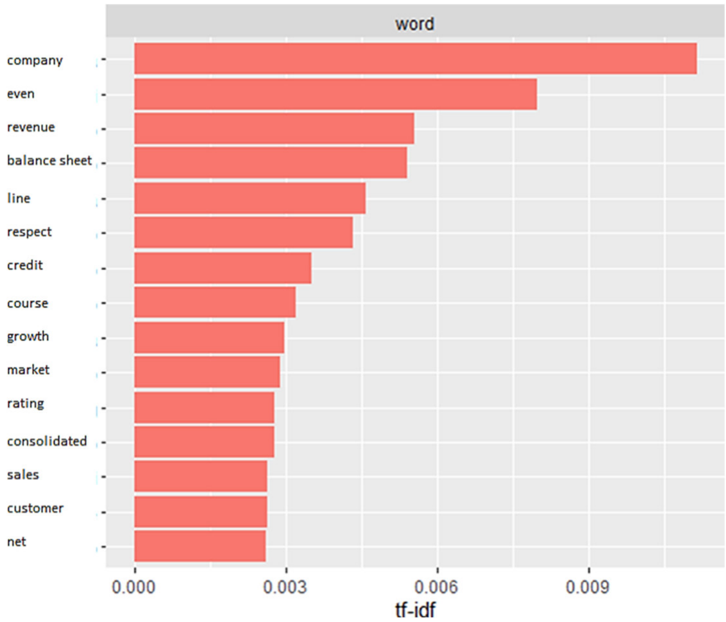
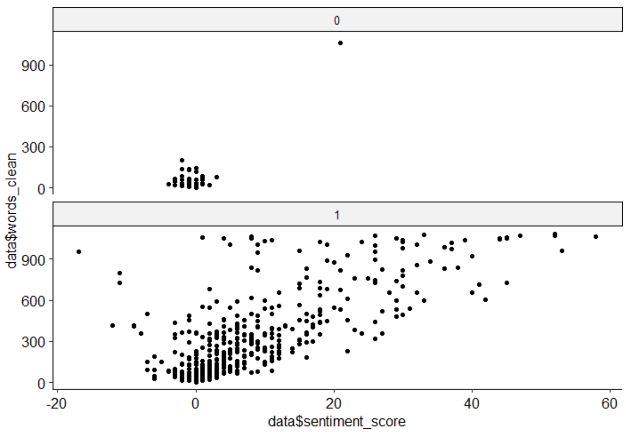


Figure 4: Word cloud of the considered data. Source: Authors' own elaboration.

Notes: dictionary of main words in the wordcloud: società=company; pari=equal; fatturato=revenue; bilancio=balance sheet; linea=line; rispetto=respect; credito=credit; corso=course/current; crescita=growth; mercato=market; rating=rating; consolidato=consolidated; ricavi=revenues/sales; cliente=customer/client; netto=net

Figure 5: Sentiment and word clean distribution per class of approved loans.



List of Figures

Figure 1: The Bank’s organizational flowchart

Figure 2. Histogram distribution of sentiment score

Figure 3. Tf–idf term frequency–inverse document frequency

Figure 4. Wordcloud of the considered data

Figure 5. Sentiment and word clean distribution per class of approved loans

Appendix

Figure A.1: Average sentiment score at the various percentiles of organizational distance.

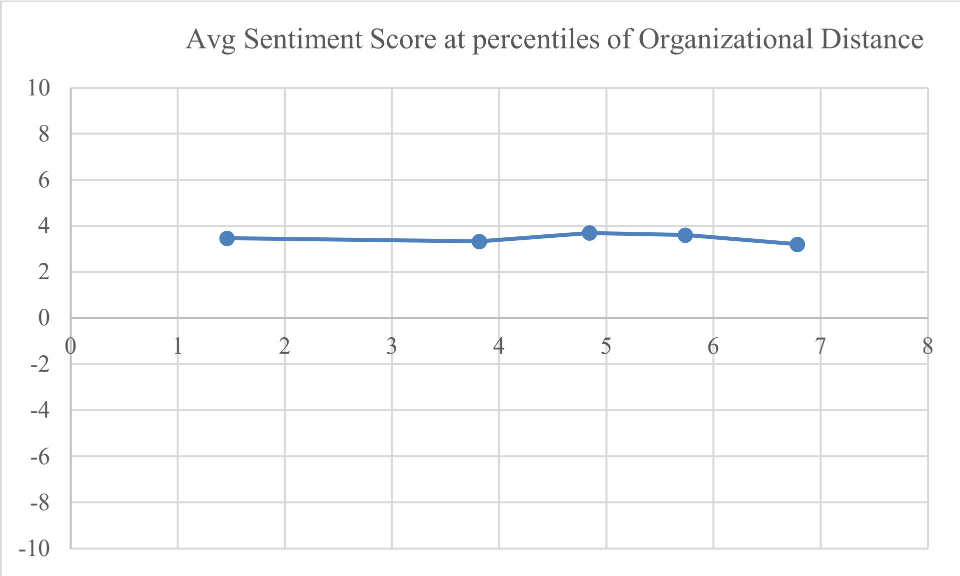


Figure A.2: Average total assets at the various percentiles of organizational distance.

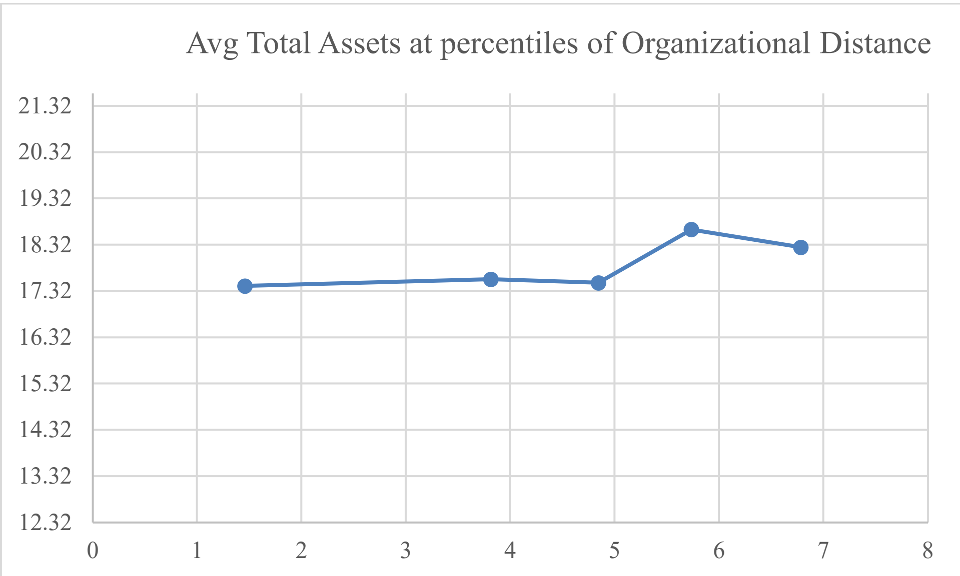


Figure A.3: Average leverage of borrowers at the various percentiles of organizational distance.

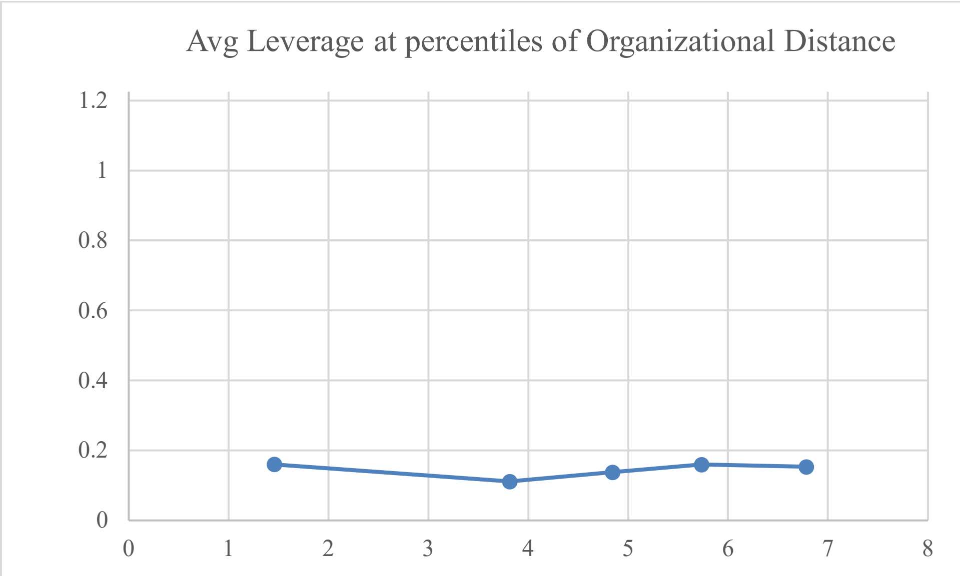


Figure A.4: Average Ebitda/TA at the various percentiles of organizational distance.