# Essays on
# Financial Markets and Information Technology

Dissertation

zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
Moritz Sefried, M.Sc.
aus Tübingen

Tübingen
2025

| | |
|---|---|
| Tag der mündlichen Prüfung: | 14.10.2025 |
| Dekanin: | Prof. Dr. Taiga Brahm |
| Dekan: | Prof. Dr. Dominik Papies |
| 1. Gutachter: | Prof. Dr. Werner Neus |
| 2. Gutachter: | Prof. Dr. Christian Koziol |

# Contents

# List of Figures

# List of Tables

# List of Derivations

# Chapter 1
# Introduction

55 years ago, Eugene Fama published his influential review on the efficient market hypothesis, in which he defined an efficient market as one in which prices fully reflect all available information (Fama 1970). Besides being an elegant framework, Fama (1970) himself states that this efficient market hypothesis is an extreme null hypothesis and further refinements are needed. Later, Grossman and Stiglitz (1980), for example, argue that perfectly efficient markets are logically impossible - if prices would instantly reflect all available information, there would be no incentive for market participants to collect (costly) information in the first place. Therefore, Grossman and Stiglitz (1980) established a more nuanced view on information and market efficiency. Markets are still powerful in aggregating information, but the process is noisy and often distorted. Heading in the same direction, Merton (1987) later emphasizes that investors rarely possess complete knowledge. Instead, they act on only a subset of available options, leading to inefficiencies.

However, since the first foundational insights mentioned above, the way information in financial markets is accessed, distributed, and utilized has evolved dramatically. The volume and variety of available data have grown at an enormous rate in recent years (Fortune Business Insights 2025). At the same time, the challenge has shifted from not simply accessing information to managing and making sense of overwhelming quantities of data. Here, information technology (IT) comes into play. With increasing (and still rising) computing power and constantly evolving analytical methods such as machine learning, new tools for data or information management have been developed. The ways in which IT can be applied in financial markets are diverse and cover a wide range of applications. For example, IT can be used to collect and store vast amounts of data from various sources, including market data and transaction records, and allows for fast, detailed processing (Hauswald and Marquez 2003, Grasshoff et al. 2019). In general, the more information is stored, the more can be analyzed. In addition, technological interfaces such as Application Programming Interfaces (APIs) allow real-time access to gather external or distribute internal data with market participants. Thereby, APIs enable a faster and more seamless way to access data. Moreover, IT can be used to identify patterns in data at the analytical level. Thus, IT could be used to predict, for example, trends in customers' behavior. Each of these possible IT applications offers opportunities for market participants to improve their efficiency and insights, but also unique challenges in terms of, e.g., data quality or model risk.

Compared to the pre-Internet era, information itself is now abundant and more accessible than ever before. Instead, the use and interpretation of information or data have become more and more important and evolved to a key driver of decision-making processes in an economic setting (Varian 2014). However, besides obvious opportunities coming with advancements in the technological progress, new vulnerabilities have also been introduced. In the latest annual report of the German Council of Economic Experts (GCEE 2024), this becomes clear. On the one hand, the report highlights that enhancing technological capabilities is not only essential for maintaining competitiveness but also for fostering innovation in financial markets. For example, the ability to process information quickly and accurately has become a critical determinant of a market participant's success in today's financial system (BIS 2024). On the other hand, the annual report also states that an unreflective adoption of technological tools may threaten financial stability by introducing systemic risks that are only partly understood at this stage. As a result, it becomes clear that enhancing technological capabilities for information analysis is important, but not a straightforward process. The proliferation of data increases the temptation but also the danger of overfitting, mistaking noise for signal, and thus drawing false conclusions from accidental patterns. Spurious correlations and algorithmic biases are therefore no longer niche concerns but central issues that financial practitioners must be aware of in the future (EBA 2022). In addition, these challenges have not gone unnoticed by regulatory bodies, which increasingly view digital innovation as both a source of progress and a potential threat to financial stability. The Financial Stability Board recently outlined the stability implications of machine learning and artificial intelligence and warned that the widespread use of the same or similar machine learning models and data sources could lead to increased correlations in the market, increasing the risk of, for example, flash crashes (FSB 2024). Similarly, the European Banking Authority issued guidelines to strengthen information and communication technology (ICT) risks in response to growing digital dependencies in the financial sector (EBA 2025). These regulatory responses highlight that the use of advanced IT is not only a matter of innovation and efficiency, but also one of regulatory oversight and systemic resilience. As IT systems become more embedded in financial decision-making, understanding the structure of the underlying information becomes increasingly important.

It is crucial to recognize that information itself is not a homogeneous construct. The risks associated with misinterpreting large-scale data also depend on the type and nature of the information being used. In this context, distinguishing between hard and soft information becomes essential (Liberti and Petersen 2019). "Hard" information subsumes objective and verifiable financial data. It can be easily stored and processed. Examples of hard information include price data, audited financial statements, or credit scores. In contrast, soft information is more qualitative or even subjective and therefore harder to formally capture. Examples of soft information are borrowers' perceived trustworthiness or relational knowledge, for instance, when deciding whether to grant a loan (Petersen and Rajan 1994; Berger and Udell 2002). Even hard to measure management skills or character of a borrower can be subsumed under soft information and be relevant for decision-making (Grunert and Norden 2012). Soft information plays a vital role when formal, hard metrics prove insufficient. Especially in the context of banking and lending, relationship lending becomes key to reducing information asymmetries (Petersen and Rajan

1994). Moreover, Liberti and Petersen (2019) emphasize that the distinction between hard and soft information has significant implications for the structure of financial organizations and their technological systems. While hard information can easily be standardized and automated, soft information often requires human judgment or sophisticated tools. Therefore, understanding different types of information and their limitations is fundamental when determining which kind of IT is most suitable for the market participants' specific situation and needs.

These considerations set the stage for the empirical analyses of this thesis. The first part of the following analyses (Chapter 2, *"The Benefits of Banks' IT Investments in Times of Trouble: Evidence from Loan Loss Accruals During the COVID-19 Pandemic"*) concentrates on banks and how they process information about their borrowers. We investigate whether banks with higher IT expenditures are able to generate favorable insights out of their available data. By analyzing call report data on U.S. commercial banks during an exogenous shock on the economy (COVID-19), we reveal the role of IT on banks' loan risk assessments. Thereby, we add to the literature by providing empirical evidence on a topic where results, up to now, strongly diverge (Beccalli 2007; DeYoung et al. 2007; Koetter and Noth 2013; Buchak et al. 2018; Pierri and Timmer 2022).

While the first part focuses on banks by using hard information from call reports, the second part (Chapter 3, *"The Role of Geographical Distance for Investment Decisions in Crowdinvesting"*) shifts to a different yet related setting: we analyze a crowdinvesting setting in which investment relevant information is shared via an online platform. We use a proprietary dataset on a Mexican crowdinvesting platform to investigate the role of geographical distance between investors and startups on investment decisions in crowdinvesting, especially the investment probability and investment amount. Since the digital platform is expected to reduce the role of geographical distance, we examine how distance interacts with the investor-startup-relationship and with observable, yet partly subjective, signals on the platform (namely the number of Likes of a crowdinvesting project), which have been shown to correlate with investors' decision making (Angerer et al. 2017; Courtney et al. 2017). By exploiting detailed data on investor and startup characteristics, we add empirical evidence on the relation between these factors.

The last part of the analyses then returns to banks' loan business (Chapter 4, *"The Role of Societal Sentiment on Commercial Banks' Loan Portfolio"*). This time, we use information from U.S. commercial banks and combine it with a sentiment measure which we derive from the music listening behavior of the U.S. population from an online streaming platform. We look at the interplay between hard information from banks' call reports and soft information from music sentiment. Musical sentiment has been shown to influence consumers' behavior and perceptions (e.g., Benetos et al. 2022; Hunter et al. 2011; Saarikallio and Erkkilä 2007; North and Hargreaves 1996). Thus, we use a sentiment measure derived from Spotify and analyze its correlation to banks' loan business in its entirety – ranging from loan volume changes to loan charge-offs.

In doing so, we fill a gap in the literature by examining the association between banks' loan business and (societal) sentiment on a larger scale.

Summing up, the relevance of IT is becoming ever more critical in a digital and information-intensive financial market. This thesis highlights the need for a nuanced understanding of how IT and the handling of information might interact with financial market outcomes. Therefore, this thesis provides new evidence and might be a source of inspiration for future research and financial market supervision despite the already well-recognized importance of IT. However, as the German Council of Economic Experts already stated, the increasing use of digital technologies in the financial system can give rise to new systemic risks, such as the danger of abrupt liquidity outflows or excessive risk-taking by banks (GCEE 2024). In other words, it is essential not only to embrace the benefits of technological progress, but also to manage the risks that come with it.

# Chapter 2
# The Benefits of Banks' IT Investments in Times of Trouble: Evidence from Loan Loss Accruals During the COVID-19 Pandemic.

**Abstract** Motivated by diverging results from the literature, we investigate whether investments in information technology (IT) help banks to assess their loan portfolio. More specifically, we focus on the consequences of accumulated expenses for data processing on banks' ability to estimate their loan loss accruals. We further test for differences when the banks' borrowers get hit by the economic trouble from the COVID-19 pandemic. Using a sample of U.S. commercial banks before and during the COVID-19 pandemic, we find more precise estimates of loan loss accruals during these troublesome times in banks that accumulated higher data processing expenses. Surprisingly, we do not find significant differences in the precision of loan loss accruals by banks' IT investments during normal times. Our findings contribute to consolidate previously diverging results by showing that IT investments help banks following a structural break, such as the COVID-19 pandemic.

> *"It is therefore no surprise to me that European banks were technologically ready to handle the coronavirus crisis. [. . . ] Their IT systems were up to the job of keeping the show on the road."*
>
> – Pentti Hakkarainen, Member of the Supervisory Board of the ECB, September 16th, 2020.

**Acknowledgments** This chapter is adapted from my published paper "The benefits of banks' IT investments in times of trouble: evidence from loan loss accruals during the COVID-19 pandemic" in the Journal of Business Economics 2023. It was joint work with Prof. Jan Riepe from the University of Paderborn.

## 2.1 Introduction

Information technology (IT) has become increasingly important to banks. Banks have spent heavily on IT, for example, by developing a clear digital strategy, redesigning core processes, and making significant investments in IT infrastructure. Appropriate IT systems do not only help banks save time and effort in the faster extraction of more fine-grained information, but they are also a vital resource in assessing risks within the banks' core business activities (Grasshoff et al. 2019; Hauswald and Marquez 2003).

The COVID-19 pandemic that spread throughout the world in 2020 has made efficient information processing even more pressing because expectations for and outlooks on the clients' economic wellbeing have become highly volatile, and banks have had to question previously identified trends each and every day. This holds true for firms in general and banking firms specifically. The OECD highlights the importance of adapting to the new circumstances coming with the pandemic. If firms do not sufficiently invest in IT and do not implement IT appropriately, they would risk falling behind in terms of productivity (OECD 2020). In a similar vein, Andrea Enria, Chair of the Supervisory Board of the ECB, argues that digitalization does not only help banks enhance their revenues but also improve banks' cost-efficiency (Enria 2021). According to the IDC survey, bank managers were also highly aware of the importance of IT systems during the COVID-19 pandemic and have prioritized process automation in the face of spiking workloads as a result of the COVID-19 crisis (Reuters 2020). Furthermore, the internal business processes within banks face new challenges, including lockdowns and remote work from home, which is easier to cope with for banks with higher IT expertise.[1] The build-up of IT expertise thereby did not happen overnight but required long-term investments. Jerry Silva, global banking research director at IDC, refers to a digital divide in the banking industry during the COVID-19 pandemic: "Sometimes I call it the predatory gap, because those banks are going to be able to steal market share from those that were not prepared prior to 2020." (Reuters 2020).

However, our understanding of the bank-level consequences of better IT is limited. The research frequently fails to isolate the direct implications of higher IT investments for performance, and the empirical evidence on these implications strongly diverges (Beccalli 2007; DeYoung et al. 2007; Koetter and Noth 2013; Buchak et al. 2018; Pierri and Timmer 2020). Plausible reasons for the inconclusive evidence are a lack of proper data as well as a well-suited empirical identification strategy (Beccalli 2007; DeYoung et al. 2007; Buchak et al. 2018).

In this study, we investigate the link between IT investments and the internal business processes in banks during the COVID-19 pandemic. Our study investigates how IT capabilities are helpful for banks during the COVID-19 pandemic in coping with and managing loan risk after a loan was already granted to the borrowers. Therefore, our investigation uses the absolute abnormal loan

[1] Mike Dargan, global head of group technology at UBS: "We had four main focus areas, remote working to enable the employees at UBS, system stability, as we saw a lot of volatility, cyber security, and operations continuity" (Reuters 2020).

loss provisions ( |*ALLP*| )[2] as main proxy (Dal Maso et al. 2018). We investigate banks because they have heavily invested in information technologies and have become technology-intensive (Berger 2003; Beccalli 2007). Consequently, the bank-level consequences of IT investments are highly meaningful and relevant. We explore the COVID-19 pandemic because it provides a unique empirical identification. The COVID-19 pandemic represents a plausibly exogenous shock to banks' information quality when screening and monitoring borrowers. Because different industries got hit by the pandemic in remarkably different ways and indirect effects from the highly volatile stock market affected borrowers' collateral and liquidity needs, the economic situations of borrowers during the pandemic changed more frequently and in less predictable ways. Duan et al. (2021) provide empirical support for the effect of the COVID-19 pandemic on loan risk. Moreover, Pierri and Timmer (2022) study consumer spendings during the COVID-19 pandemic and find that IT can play an important role as a mitigating (short-term) factor when a pandemic hits the economy. By using a database with information from telephone research interviews, Kwan et al. (2020) study the effect of IT capabilities on banks' ability to serve customers. They claim that banks with higher IT capabilities are, on the one hand, better positioned for the future, where probably fewer bank branches and more digital banking are present, and also for extreme shocks like the COVID-19 pandemic itself.

Overall, we expect IT investments to positively correlate with banks' quality of loan risk assessments. Especially in a field where data have such pivotal importance, banks' reliance on well-implemented processes is key to their decision-making. Since the amount of data continues to grow, having IT systems that can handle this data is critical. In times of relatively stable economic conditions, banks can use their experience to estimate future loan losses and, thus, build up accruals based on that experience. If this is the case, IT investments could still be helpful in decision-making but not crucial since banks' managers can orient themselves on the values from previous periods. However, and following the literature on the business level consequences of better IT capabilities in banks, we expect that banks with better IT capabilities more precisely estimate their loan loss provisions in general. Moreover, we use the COVID-19 pandemic as a shock to banks' information environment and argue that banks with better IT capabilities could better cope with that shock. Thereby, our identification relates to the approach by Pierri and Timmer (2020), who also use a shock to identify the bank-level consequences of banks' IT capabilities. In contrast to Pierri and Timmer (2020), we do not use a shock to the financial system and the regulatory capital from the 2008 financial crisis, but we rely on the COVID-19 pandemic that primarily influences the banks' clients but not directly the financial sector. Nevertheless, and in line with the results from Pierri and Timmer (2020) and due to banks' capability to process data quicker and in more detail (Hauswald and Marquez 2003; Grasshoff et al. 2019), we expect banks with higher levels of IT investments do better assess their (loan) risk in highly uncertain times compared to banks with lower IT investments. Since

[2] We regress loan loss provisions on their determinants which we gather from the respective financial statement. |*ALLP*| are the absolute values of the residuals of this regression model. Higher |*ALLP*| indicate a lower quality risk assessment.

the pandemic forces banks to adapt quickly to a new business environment, we expect that data processing is even more advantageous in such unstable times. Additionally, when this pandemic occurred, banks with lower IT investments could not use their experience on loan losses during stable times before but needed to make new assumptions. During those times, IT investments are, consequently, most valuable.

We find that banks with higher IT capabilities can better assess their loan risk in times of the pandemic or, more generally, in times of high uncertainty after a structural break. In our setting, this is true for the first two quarters of the year 2020. Afterwards, the influence reduces and it comes to an alignment of banks with high and low IT capabilities. Surprisingly, however, IT capabilities do not seem to play a significant role in stable times, which we hark to the Bayesian learning theory.

Our study relates and contributes to the literature in different ways. Our study closely relates to Beccalli (2007). She investigates whether banks' IT investments improve their performance using a sample of 737 commercial banks from Europe in the pre-dotcom era. She only finds a very weak and partially negative link between banks' IT spending and profitability. Based on her findings, she articulates a "profitability paradox". In contrast to Beccalli (2007), who relies on expert estimates of IT investments and some voluntary disclosures on IT investment, we use mandatory data from the FDIC on all commercial banks in the United States (U.S.). We further relate to DeYoung et al. (2007). They study the consequences of banks' internet adoption on different balance sheet and income statement items based on a sample of U.S. commercial banks around the turn of the millennium. They find a positive link between early internet adoption and current profitability. Koetter and Noth (2013) use banks' productivity to measure performance rather than the net income. Koetter and Noth (2013) measure IT expenditures as the sum of costs for software, hardware, third-party services, shared service centers, and information transmissions. They use a dataset that comprises over 400 German savings banks between 1996 and 2006. Estimating this relationship with five different alternative output definitions, they find a significant and positive contribution of IT investments to banks' output and conclude that IT can help improve the screening and monitoring of banks' borrowers. Since data on IT investments is hard to gather and frequently neither publicly available nor structured and detailed, Kriebel and Debener (2019) try to measure banks' IT investments by examining their annual reports with a textual analysis. They find a positive link between more IT-related words in annual reports and items in the income statement. In other related work, Buchak et al. (2018) investigate the consequences of IT on FinTechs' growth, Fuster et al. (2019) on FinTechs loan processing abilities, and Di Maggio and Yao (2020) on FinTech's loan screening ability. Fu and Mishra (2022) study finance-related mobile app market in times of the COVID-19 pandemic.

The work of Pierri and Timmer (2020) was one of the first studies on the risk-consequences of better IT capabilities and, consequently, very closely relates to our study. Pierri and Timmer (2020) investigate banks' loan quality around the 2008 financial crisis for a sample of U.S.

commercial banks. Pierri and Timmer (2020) find evidence that banks with more IT capabilities (as measured by the share of personal computers in that bank) were able to select more solvent borrowers. However, our study complements the work by Pierri and Timmer (2020) by focusing more on monitoring loans rather than screening customers, which means that loans have already been granted to the customers in our setting.

We contribute to the literature by providing new evidence on the performance level consequences of IT capabilities on banks' business using standardized data by the FDIC. We contribute to the literature by providing new empirical evidence on the relation between IT investments and the quality of risk assessments in banks with an advantageous setting and dataset.

Thereby, we also relate to the work by Berg et al. (2020) that investigates one underlying channel that helps to explain a link between banks' IT capabilities and their loan defaults. They analyze whether banks can use their customers' digital footprint to better forecast their likelihood of default on loan obligations. In addition to a credit bureau score, Berg et al. (2020) find that the digital footprint can indeed support the company in its lending decisions.

## 2.2 Methodology

### 2.2.1 Measurement of Information Technology Investments

The empirical research on the consequences of banks' IT investments suffers from a lack of IT expense information in the popular databases. Furthermore, strong endogeneity concerns regarding the direction of the relationship (Koetter and Noth 2013) call for comprehensive data and a suitable empirical setting. Compared to other countries, a huge advantage of the U.S. banking sector is that IT expense data is available and disentangled on bank-level and not on group-level. Additionally, the level of service and the customer approaches of smaller commercial banks in the U.S. are very similar to those in other parts of the world. Therefore, we are convinced that our results can be transferred to group-level and consequently other countries.

Banks have improved their capabilities in terms of information technologies and can use them to their economic advantage. Bharadwaj (2000) calls this capability a firm's *IT capability* and defines it as the "ability to mobilize and deploy IT-based resources in combination or copresent with other resources and capabilities". Bhatt and Grover (2005) split capability into three different types: value, competitive, and dynamic. Value capability refers to firms' IT investments; competitive capability refers to firms' IT business experience and the relationship between IT and business managers; and dynamic capability refers to the firm's knowledge about and adaptation to technological changes and new opportunities. In this study, we mainly refer to banks' value capability for IT in the sense of Bhatt and Grover (2005) as our empirical proxy measures the accumulated investments in banks' data processing. Hereafter, we will use the term *IT capability* (Bharadwaj 2000) when discussing the underlying concept of how IT influences banks' business activities and the term *IT investments* when referring to our empirical measure for data processing related IT capability.

Nevertheless, the measurement of banks' IT capabilities still is a crucial challenge for empirical research. This study relies on banks' investment in IT and specifically the data processing expenses from their quarterly call reports to measure IT capabilities. The link between bank IT capability and banks' IT investments has been frequently argued (Bhatt and Grover 2005) and used in previous empirical literature (Beccalli 2007; Koetter and Noth 2013; Xin and Choudhary 2019). The argument follows the idea that banks can acquire IT capabilities by investing in commercially available IT, which means IT adoption is not exclusive (Xin and Choudhary 2019). In our main specification, we directly refer to those IT capabilities that are closely related to data processing, as this is closer to the actual acquisition of IT capabilities. Data processing expenses is a mandatory separate line item in banks' quarterly income statements in the other non-interest expense section.

The use of quarterly information thereby allows us to capture changes in banks' IT capabilities over time. Using such a time-variant measure for IT capabilities is one innovation in this study. It enables us to fully use our panel data that is impossible when using other frequently used proxies like a snapshot of a ranking or a cross-sectional survey (Bharadwaj 2000). In this way, we can measure whether changes in banks' IT investments have business consequences. We can thereby rule out biases from the time-invariant factors that simultaneously influence banks' IT capabilities, such as geographic or institutional factors.

Nevertheless, we openly acknowledge that past IT investments do not perfectly predict current IT capabilities because IT projects might fail (Xin and Choudhary 2019). Second, specific IT hardware and knowledge might take some time to be adequately implemented (Campbell 2012). But at the same time, IT hardware loses its value over time because of technological change. Failing IT projects will thereby create noise to our measure and bias our empirical results against finding anything. Therefore, our empirical evidence for our measure is a conservative estimate of the actual underlying relationship. Furthermore, we explicitly use IT investments related to data processing to mitigate the effects of large investments in banks' IT infrastructure, whose failure directly results in the recognition of the expenses on their balance sheets. Technological change and the implementation time call for the use of lagged information on banks' IT investments but require accumulating those investments for only a few quarters. In this study, we decided to use the average IT investments from the first quarter of the year 2015 to the last quarter of 2019, just before the COVID-19 Pandemic started, as our main explanatory variable. Nevertheless, our empirical inferences remain qualitatively unchanged if we use the average IT investments from 2017, or a rolling IT measure as seen in the robustness checks in Section 2.5.

### 2.2.2 Testing the Relationship between IT Investments and Quality of Banks' Loan Risk Assessments

LLPs are banks' most important, loan-related accrual (Liu and Ryan 2006; Kanagaretnam et al. 2010*a*; Beatty and Liao 2014). On average, IT investments are banks' third-largest non-interest expense and are just as large as their marketing expenses, legal fees, accounting, and consulting expenses taken together. Banks' loan loss provisions (LLP) are economically important because they are their largest accrual and are tied to a broad range of other outcomes (Beatty and Liao 2014). Studies have frequently used them to measure banks' transparency or earnings quality (Kanagaretnam et al. 2010*b*; Beatty and Liao 2014; Jin et al. 2018) since bank managers can and potentially want to engage in steering LLP. This steering might reduce the overall quality of earnings (Jin et al. 2021). We control for this discretionary behavior by adding control variables to our LLP specification. Afterward, we can determine the absolute misestimation of banks' LLP, which we refer to as the quality of risk assessments. Thus, the more accurately banks assess their LLP, the higher the quality of risk assessment we perceive for these banks.

We closely follow the two-step approach from Jin et al. (2021) and Dal Maso et al. (2018) and concentrate on the magnitude of abnormal LLP to represent the quality of banks' loan risk assessments. First, we estimate the LLP for each quarter using an LLP model, which closely follows Beatty and Liao (2014). We extend their model by adding *EBLLP*, *RegCap*, and *LLA* to account for manager discretion. Overall, our first stage regression is as follows:

$$\begin{aligned} LLP_{i,t} = {} & \beta_0 + \beta_1 dNPL_{i,t} + \beta_2 dNPL_{i,t-1} + \beta_3 RegCap_{i,t-1} + \beta_4 CO_{i,t} + \beta_5 CO_{i,t-1} \\ & + \beta_6 EBLLP_{i,t} + \beta_7 dLoans_{i,t} + \beta_8 LLA_{i,t-1} + \beta_9 Size_{i,t-1} + \alpha_j \end{aligned} \tag{2.1}$$

where $LLP_{i,t}$ stands for the loan loss provisions scaled by total loans of the bank, $dNPL_{i,t}$ is the change in nonperforming loans from the previous to the current year at the bank level that is divided by total loans, and $RegCap_{i,t-1}$ is the previous year's share of regulatory capital that is scaled by risk-weighted assets. $CO_{i,t}$ represents the ratio of charge-offs in the current year to loan loss allowances, and $EBLLP_{i,t}$ represents the earnings before loan loss provisions that are scaled by total loans. $dLoans_{i,t}$ is the change in loans from the previous to the current year that is scaled by total assets, $LLA_{i,t-1}$ stands for the amount of loan loss allowances in the previous year that is divided by total loans, and $Size_{i,t-1}$ represents the natural logarithm of total assets from the beginning of the period. The values of the profit and loss statement variables reflect the actual amount added in the respective quarters. Further, $\alpha_j$ represents the fixed effect on the state-level. Technically, we conduct a regression for each quarter itself. In this way, we have a regression model that is equivalent to a state-by-time fixed effects model since the respective coefficients can vary for each quarter and, therefore, capture regional-specific influences per period.

In the next step, we calculate |*ALLP*|, the absolute residuals of the first-stage regressions. We use the absolute value because we are not interested in whether banks under- or overestimate their LLP but whether they have any deviations. The |*ALLP*| is our final measure of the quality of banks' assessments of loan risk. Since we are interested in the relation between banks' loan risk assessments and IT investments, our main explanatory variables are *IT investments* from before the pandemic and the interaction term for *IT investments* during the COVID-19 pandemic. Therefore, after calculating |*ALLP*|, we regress our *IT investments* variable together with call report items on |*ALLP*| to test their relationship. As already mentioned, we include an interaction term for *IT investments* and the year of the pandemic to gather the correlation of *IT investments* during a structural break. Additionally and to account for serial correlation in our model, we cluster the standard errors on bank-level. Consequently, our regression model looks as follows:

$$\begin{aligned}|ALLP_{i,t}| = \beta_0 + &\beta_1 ITinvestments_{i,t} \times COVIDCrisis_t + \beta_2 ITinvestments_{i,t} \\ &+ \beta_3 COVIDCrisis_t + \beta_4 EBLLP_{i,t} + \beta_5 RegCap_{i,t} \\ &+ \beta_6 AssetGrowth_{i,t} + \beta_7 LoanstoAssets_{i,t-1} \\ &+ \beta_8 DepositstoAssets_{i,t} + \beta_9 DepositstoAssets_{i,t-1} \\ &+ \beta_{10} dNPL_{i,t} + \beta_{11} dNPL_{i,t-1} + \beta_{12} RealEstateLoans_{i,t} \\ &+ \beta_{13} CommercialLoans_{i,t} + \beta_{14} RetailLoans_{i,t} + \alpha_j + \tau_t\end{aligned} \tag{2.2}$$

where $ITinvestments_{i,t}$ stands for our eight quarter IT measure; the binary variable $COVIDCrisis_t$ represents the year 2020 when the outbreak of COVID-19 occurred. It equals one for each quarter in the year 2020 and zero otherwise.

The literature uses a large set of control variables in a very heterogeneous way. Some variables appear in many studies, while others occur in one or two empirical models (see Beatty and Liao (2014) for a discussion of the differences in the early LLP models). We closely follow Dal Maso et al. (2018) in our selection of control variables. We differentiate by refraining from including constant state variables, but our state-fixed effects account for this exclusion. In a nutshell, we control for bank characteristics with different loan-related variables because they are closely related to the level of LLP and thus allow us to more precisely capture the influence of our explanatory variable *IT investments*. Namely, we include banks' *EBLLP* because, for example, Kilic et al. (2021) argue that income smoothing is especially important in bank accounting. We also add the ratio of *Regulatory Capital*, lagged *Loans to Assets* ratio, *Deposits to Asset* ratio of the current and prior year, and changes in *NPL* of the past two years as well as *Asset Growth*. We also add the *Real Estate*, *Commercial*, and *Retail Loans* of the respective bank since the manager's discretion over LLP differs across loan types (Liu and Ryan 1995; Bhat et al. 2020). Further, $\alpha_j$ is the fixed effect on the state level, and $\tau_t$ is the fixed effect per quarter. The variables mentioned above allow us to control for banks' specific business focus and reduce the chance of an omitting variable bias in our estimation of the IT investments correlation.

The coefficients of interest are $\beta_1$ and $\beta_2$ in *Equation (2.2)*. We expect IT investments to reduce |*ALLP*|. Thus, we expect a negative coefficient for $\beta_2$. Thereby, it captures the overall link between IT investments and |*ALLP*|. $\beta_1$ identifies the differences in the link between IT investments and |*ALLP*| during normal and crisis periods. We expect that past IT investments are even more beneficial in reducing |*ALLP*| during crisis times that should lead to a negative $\beta_1$. As articulated by Dal Maso et al. (2018), the error terms from *Equation (2.1)* are used in *Equation (2.2)* and thus can influence the coefficient estimates (especially by extreme values). To address this concern, we use a robust regression estimation.

## 2.3 Data

We use quarterly call reports from the Federal Financial Institutions Examination Council (FFIEC) for all U.S. commercial banks in the period from 2015 to 2020. We drop bank observations with missing or negative values on core balance sheet items, such as equity, total assets, and IT investments.[3] Additionally, we limit our sample to commercial banks because LLP only provide meaningful information on the business-level consequences of IT investments for these firms. We especially eliminate financial service firms and investment banks with a loan to asset ratio or deposits to total assets ratio below 50%. We drop small banks with total assets below 100,000 USD because small banks operate in a different regulatory environment. Furthermore, we delete observations characterized by extensive merger activities and financial distress to limit any confounding effects from corporate restructuring on the bank-level outcomes. After cleaning up our sample, we have 8,522 observations from 665 banks with data on all variables left for our regressions. The remaining banks in our sample are typical commercial banks. To further account for outliers, we winsorize all variables at the 1% level.

## 2.4 Empirical Results

### 2.4.1 Univariate

The descriptive statistics of all variables are provided in Table 2.1. Banks in our sample have, on average, 84.7% of deposits and 74.6% of loans on their balance sheets. Our primary variable of interest is *IT investments* which is a line-item of other non-interest expenses. We scale IT investments by banks' total other non-interest expenses to alleviate competing influences from

[3] Banks are only required to report IT investments that exceed $100,000 or 7% of other non-interest expenses. Because we cannot observe the exact value of IT investments for these banks, these observations would dilute the results.

banks' loan portfolio, that is, LLP and charge-offs. There are considerable differences in our measure of IT investments for each bank. While banks from the lower quantile have on average a ratio below 13‰, banks belonging to the highest quantile have on average a ratio of more than 25‰.

Table 2.1 also displays the pairwise correlations of IT investments and other bank characteristics. The correlations do not indicate any issues related to multi-collinearity and are all in line with our expectations. We find a negative correlation between *IT investments* and *Size* that indicates that IT investments are easily scalable. *EBLLP* shows no statistically significant correlation with *IT investments*, while *Regulatory Capital* and *IT investments* are positively correlated. The correlations are not in line with IT investments depending on banks' free cash flow but indicate a more planned and conscious decision by banks on IT investment.

To strengthen our findings from the correlation analysis, we further look at variance inflation factors (VIF) to assure that our variables are not affected by multicollinearity. The highest VIF value in our regression model is 3.87 for the *Deposits to Asset* variable of the current period. Unfortunately, there is in general no critical threshold for the VIF that indicates multicollinearity. As a rule of thumb, a VIF between 5 and 10 or greater than 10 is likely a sign for multicollinearity. Since this is not the case for our regression, the VIF analysis is in line with the results from the correlation analysis.

**Table 2.1:** *Descriptive Statistics and Correlation Matrix.*[4]

| | Mean | Std. Dev | p25 | p50 | p75 | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) | (17) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st stage | | | | | | | | | | | | | | | | | | | | | | |
| (1) LLP | 0.001 | 0.001 | 0.000 | 0.000 | 0.001 | | | | | | | | | | | | | | | | | |
| (2) dNPL | 0.000 | 0.005 | -0.001 | -0.000 | 0.001 | 0.08* | | | | | | | | | | | | | | | | |
| (3) RegCap | 0.138 | 0.035 | 0.116 | 0.129 | 0.147 | -0.02* | -0.05* | | | | | | | | | | | | | | | |
| (4) CO | 0.039 | 0.073 | 0.003 | 0.015 | 0.042 | 0.51* | 0.01 | -0.05* | | | | | | | | | | | | | | |
| (5) EBLLP | 0.016 | 0.013 | 0.012 | 0.017 | 0.023 | -0.66* | -0.06* | 0.07* | -0.36* | | | | | | | | | | | | | |
| (6) dLoans | 0.015 | 0.028 | -0.001 | 0.011 | 0.024 | 0.05* | 0.02 | -0.13* | -0.04* | -0.00 | | | | | | | | | | | | |
| (7) LLA | 0.012 | 0.005 | 0.009 | 0.011 | 0.014 | 0.45* | -0.02 | 0.12* | 0.10* | -0.17* | -0.14* | | | | | | | | | | | |
| (8) Size | 13.954 | 1.350 | 12.923 | 13.892 | 14.773 | 0.15* | 0.06* | -0.24* | 0.12* | -0.01 | 0.08* | -0.13* | | | | | | | | | | |
| 2nd stage | | | | | | | | | | | | | | | | | | | | | | |
| (9) ALLP | -0.035 | 6.158 | -3.434 | -0.156 | 2.968 | 0.45* | -0.01 | 0.00 | -0.03* | 0.03* | 0.01 | 0.05* | 0.01 | | | | | | | | | |
| (10) \|ALLP\| | 4.460 | 4.671 | 1.428 | 3.208 | 5.889 | 0.42* | 0.06* | 0.08* | 0.28* | -0.26* | 0.03* | 0.22* | 0.05* | 0.14* | | | | | | | | |
| (11) IT Investments | 0.189 | 0.086 | 0.127 | 0.188 | 0.245 | -0.04* | 0.03* | 0.04* | -0.05* | 0.01 | 0.00 | -0.03* | -0.18* | 0.01 | -0.05* | | | | | | | |
| (12) COVID Crisis | 0.227 | 0.419 | 0.000 | 0.000 | 0.000 | 0.21* | 0.09* | 0.01 | -0.04* | -0.16* | 0.05* | 0.10* | 0.11* | 0.00 | 0.23* | -0.04* | | | | | | |
| (13) Asset Growth | 0.024 | 0.046 | -0.000 | 0.016 | 0.036 | 0.08* | 0.04* | -0.10* | -0.01 | -0.02 | 0.67* | -0.08* | 0.08* | 0.02 | 0.05* | 0.01 | 0.22* | | | | | |
| (14) Loans to Assets | 0.746 | 0.088 | 0.685 | 0.749 | 0.810 | 0.06* | 0.06* | -0.26* | 0.02 | -0.14* | 0.40* | -0.16* | 0.08* | -0.02 | -0.02 | 0.03* | -0.04* | 0.31* | | | | |
| (15) Deposits to Assets | 0.847 | 0.067 | 0.808 | 0.852 | 0.889 | 0.01 | 0.02 | -0.18* | -0.04* | 0.05* | 0.29* | 0.08* | -0.22* | 0.03* | -0.02 | 0.06* | 0.15* | 0.52* | 0.03* | | | |
| (16) Real Estate Loans | 0.521 | 0.166 | 0.418 | 0.533 | 0.634 | -0.23* | -0.04* | 0.06* | -0.17* | -0.02 | 0.16* | -0.36* | 0.08* | -0.06* | -0.10* | -0.07* | -0.07* | 0.10* | 0.39* | -0.14* | | |
| (17) Commercial Loans | 0.115 | 0.080 | 0.058 | 0.097 | 0.151 | 0.08* | 0.07* | -0.36* | 0.06* | -0.02* | 0.19* | -0.04* | 0.16* | 0.03* | 0.04* | 0.02* | 0.21* | 0.20* | 0.19* | 0.20* | -0.31* | |
| (18) Retail Loans | 0.033 | 0.054 | 0.004 | 0.013 | 0.033 | 0.34* | 0.01 | -0.04* | 0.29* | -0.16* | 0.00 | 0.20* | 0.09* | 0.07* | 0.10* | 0.03* | -0.03* | 0.00 | 0.03* | 0.03* | -0.38* | 0.01 |

*Table 1 provides the summary statistics on the left-hand side of the table - columns (1) – (17) state the Pearson correlation coefficients between the variables. Continuous variables are winsorized at the 1% level. In Appendix A, all variables are defined. *p indicates significance at the 5% level.*

[4] For representational reasons, *ALLP* and |*ALLP*| in Table 2.1 to Table 2.5 are multiplied by $10^4$.

***Table 2.2:*** *Regression of Absolute Abnormal Loan Loss Provisions on IT Investments.*

| Dependent Variable:<br>\|ALLP\| | (1)<br>Pre COVID-19 pandemic | (2)<br>COVID-19 pandemic | (3)<br>Full sample |
|---|---|---|---|
| IT investments × COVID Crisis | | | -5.30838** |
| | | | (2.11218) |
| COVID Crisis | | | 3.31228*** |
| | | | (0.49868) |
| IT Investments | -1.37245 | -5.87718*** | -1.50134 |
| | (1.11008) | (1.93654) | (1.12459) |
| State & Time FE | Yes | Yes | Yes |
| Observations | 6,591 | 1,931 | 8,522 |
| $R^2$ | 0.1391 | 0.1517 | 0.1538 |

*Table 2.2 contains three OLS regression specifications with standard errors that are robust and clustered on bank level. The dependent variable is |ALLP|. Column (1) only comprises the Pre COVID-19 crisis timeframe, whereas column (2) only comprises the COVID-19 crisis year, namely, the year 2020. However, column (3) includes the full sample. We regress IT investments solely on |ALLP| to test their relationship for the first two specifications. We incorporate an interaction term in column (3) since the Pre COVID-19 crisis and the crisis period itself are included. Additionally, we conduct the regressions with state- and time-fixed effects. Continuous variables are winsorized at the 1% level. Robust standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.*

Table 2.2 complements the correlation analysis by providing empirical evidence from a regression of IT investments on |*ALLP*| for different time periods. We observe a significant and negative correlation between the IT measure and |*ALLP*| only during the COVID-19 pandemic. For the years before the pandemic, the correlation is negative but far smaller and not statistically significant. The results are not in line with our prediction for stable times but provide initial support for the value of IT in times of high uncertainty.

### 2.4.2 Multivariate

Table 2.3 shows the results of the OLS estimations of |*ALLP*| on *IT Investments* with control variables. In column (1), we only consider the years before the outbreak of the COVID-19 pandemic. Column (2) is limited to the observations of the pandemic year only, and column (3), our main specification, has all observations since the year 2016 and an interaction term for the COVID-19 pandemic and *IT investments*.

Column (1) shows *IT investments* have a negative but statistically non-significant coefficient in normal times. The coefficient changes in column (2) and supports our hypothesis that IT investments are more helpful in times of a crisis. Since we excluded the years before the COVID-19 pandemic in column (2), the pandemic simultaneously affects the coefficients for *IT investments* and the control variables. To isolate the incremental benefits of IT investments conditional on the pandemic period, we use the interaction term between *IT investments* and the *COVID Pandemic* dummy. In column (3) of Table 2.3, we observe a negative and statistically highly significant coefficient for the interaction term and a negative but statistically non-significant main effect for *IT investments*. The results from Table 2.3 show the benefits from IT investments for banks' risk assessments during a structural break, such as the COVID-19 pandemic. Without a structural

***Table 2.3:*** *Regression of Absolute Abnormal Loan Loss Provisions on IT Investments and Control Variables.*

| Dependent Variable: $\|ALLP\|$ | (1) Pre COVID-19 pandemic | (2) COVID-19 pandemic | (3) Full sample |
|---|---|---|---|
| IT investments × COVID Crisis | | | -4.97753** |
| | | | (2.08462) |
| COVID Crisis | | | 3.19506*** |
| | | | (0.50835) |
| IT Investments | -1.78099 | -5.19143*** | -1.76565 |
| | (1.12302) | (1.87027) | (1.14386) |
| *Controls:* | | | |
| dNPL | 12.31328 | 78.09683** | 26.93091* |
| | (13.26871) | (33.83121) | (14.05298) |
| Lagged dNPL | 9.47364 | 32.3366 | 10.79672 |
| | (13.14489) | (35.19397) | (12.94717) |
| RegCap | 9.76714*** | 9.31151* | 9.01669*** |
| | (2.51508) | (5.64969) | (2.72012) |
| EBLLP | -59.90706*** | -113.44291*** | -74.32631*** |
| | (13.94126) | (22.83986) | (15.21513) |
| Asset Growth | 5.98414*** | 7.42889* | 5.2041*** |
| | (2.04721) | (4.39928) | (2.01126) |
| Lagged Loans to Assets | -0.68257 | -1.21702 | -1.49504 |
| | (1.20706) | (2.86646) | (1.36502) |
| Deposits to Assets | -2.35615 | -12.13982** | -4.78267** |
| | (1.8491) | (5.38401) | (2.16516) |
| Lagged Deposits to Assets | -0.67845 | 2.79325 | 0.46479 |
| | (1.43999) | (4.18671) | (1.73982) |
| Real Estate Loans | -1.05162 | -1.69304 | -0.80649 |
| | (0.94128) | (1.85896) | (1.00803) |
| Commercial Loans | -0.50356 | -0.78833 | 0.41306 |
| | (1.46567) | (2.55385) | (1.58325) |
| Retail Loans | 0.90592 | -1.80632 | -0.56595 |
| | (2.81759) | (4.61755) | (2.97823) |
| State & Time FE | Yes | Yes | Yes |
| Observations | 6,591 | 1,931 | 8,522 |
| $R^2$ | 0.18091 | 0.23082 | 0.19752 |

*The table contains three OLS regression specifications with standard errors that are robust and clustered on bank level. The dependent variable is |ALLP|. Similar to Table 2.2, column (1) only comprises the Pre COVID-19 crisis timeframe, whereas column (2) only comprises the COVID-19 crisis year, namely, the year 2020. However, column (3) includes the full sample. Here, we regress IT investments and control variables on |ALLP| to test whether our findings from Table 2.2 still hold. We incorporate an interaction term in column (3) since the Pre COVID-19 crisis and the crisis period itself are included. Again, we conduct the regressions with state- and time-fixed effects. Robust standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.*

break, we do not find a statistically significant benefit from IT investments on banks' forecasting accuracy of loan losses. This finding is somehow surprising and contradicts our first expectation. Overall, these results show that IT investments are not significantly beneficial in normal times but allow banks better assess their loan risk in situations of greater uncertainty.

In non-pandemic times, one standard deviation increase in IT investments reduces the dispersion of |*ALLP*| by approximately 3%.[5] Since the coefficient is not statistically significant, we cannot rule out the possibility that this result occurred by chance. However, one standard deviation increase in IT investments reduces the dispersion of |*ALLP*| by slightly more than 12% in the pandemic period. Consequently, banks with *IT Investments* that are one standard deviation

[5] We calculate this dispersion by multiplying the standard deviation of *IT Investments* with the sum of the coefficients for the IT main- and interaction-effect and divide the effect by the standard deviation of |*ALLP*|. For the evaluation in the pre-pandemic time, we only consider the main-effect coefficient.

higher can estimate 1 out of 10 loan impairments more accurately in economically unstable times.

### 2.4.3 Pre-Treatment Balance between Treatment and Control Group

***Figure 2.1:*** *Course of the Interaction Coefficient of IT Investments and the Current Quarter over Time.*

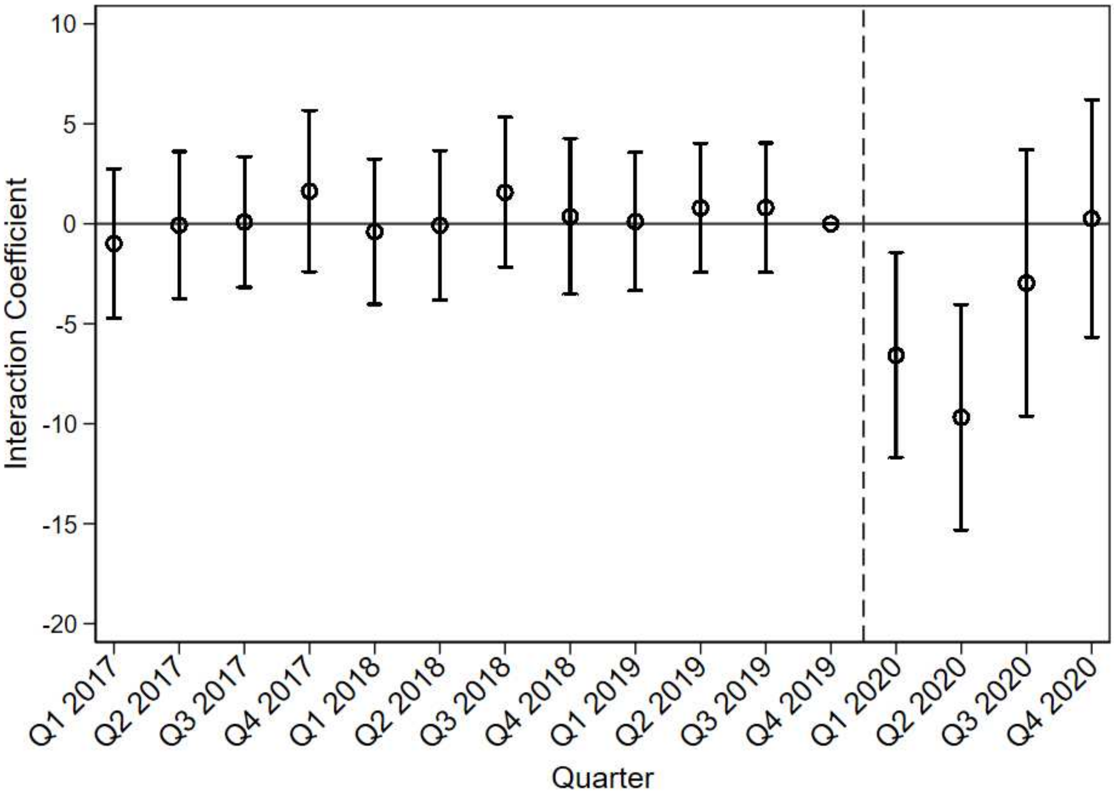



*Figure 2.1 shows the course of the interaction coefficient of IT investments and the respective quarters over time from the beginning of the year 2017 to the end of 2020. The circles describe the average coefficients and the corresponding lines represent the standard errors. We use the fourth quarter of 2019, the last quarter in non-pandemic times, as our reference group.*

Diverging trends in the pre-pandemic period are a major concern for our empirical identification strategy. Consequently, if banks with high and low IT investments show large differences in the trends for |*ALLP*| already long before the crisis, our model would not capture the time trends appropriately. Similarly, if the trend divergence does not happen around the outbreak of the COVID-19 pandemic but long after that, whether it would be responsible for the observed changes in |*ALLP*| would be unclear. Figure 2.1 shows the interaction coefficient of IT investments and the respective quarter. It clearly shows that the coefficient is not significantly different from zero before the pandemic hit the economy. The first time a negative coefficient is visible, is in the first two quarters of the year 2020. Furthermore, the pre-pandemic standard errors are definitely smaller than in the year of the pandemic. Even though higher standard errors within a pandemic are not surprising, the coefficients before the pandemic has occurred are quite precisely estimated and not statistically significant from zero. Furthermore, we see that the benefits of banks with higher IT investments diminish in the following quarters. This diminishment harkens to the Bayesian learning theory – explaining why we do not find support for IT benefits in stable times. We argue that banks with higher IT investments generally estimate their LLP more precisely and adapt more quickly than banks with lower IT investments. Nevertheless, the latter can adapt to the new situation and learn from the past quarters after the structural break, too, just more slowly than banks with higher IT investments. This newly gained experience then helps banks more precisely estimate the LLP in the upcoming (and eventually more similar) periods, and the difference between banks with higher and lower IT investments diminishes (again) over time.

***Figure 2.2:*** *Absolute Abnormal Loan Loss Provisions for Banks with Respect to Their IT Investments (Median Split).*

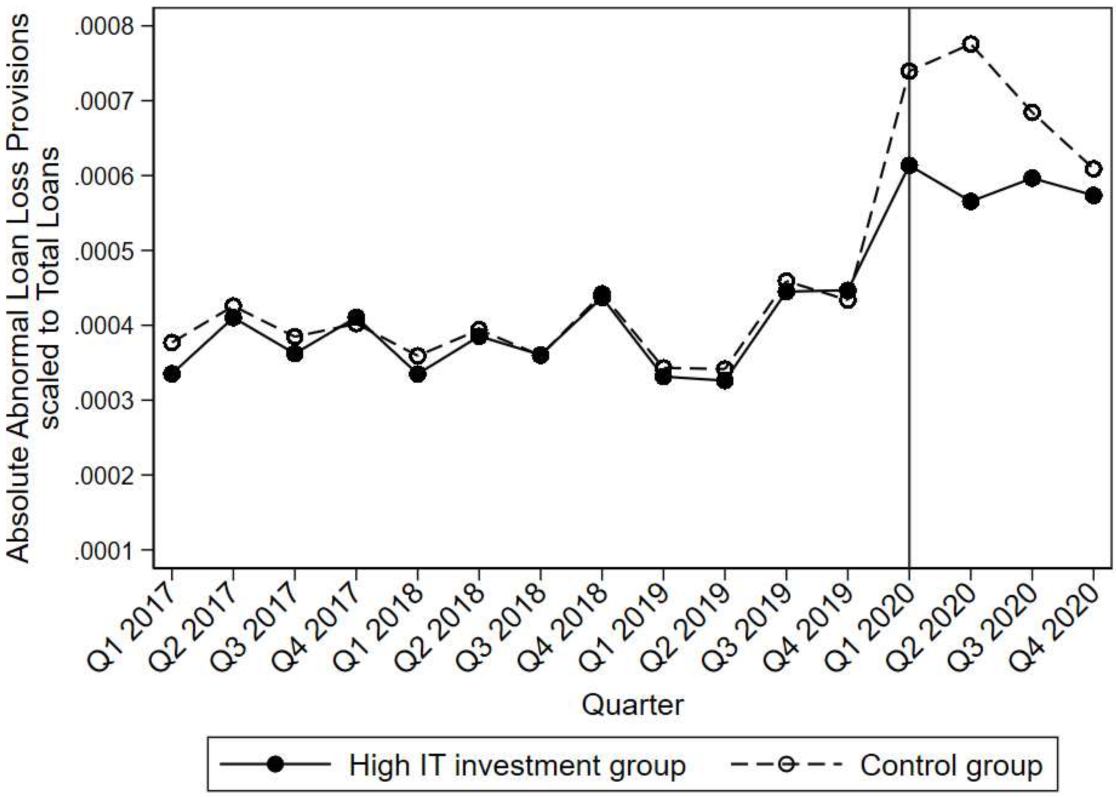



*Figure 2.2 shows the course of |ALLP| over time from the beginning of the year 2017 to the end of 2020. We split the sample into two groups. One group contains banks with above-median, and the other group contains banks with below-median IT investments. The solid black line describes the average |ALLP| for banks with above-median IT investments, and the dashed line represents the average |ALLP| for banks with below-median IT investments. The horizontal solid black line marks the first quarter of 2020, which we define as the first quarter in the COVID-19 pandemic in the U.S.*

An alternative representation of the parallel trend can be found in Figure 2.2. It displays the result of a median split in *IT investments* where we observe both groups' mean |*ALLP*| and not only the interaction coefficient. Before the outbreak of the COVID-19 virus in the U.S., $|ALLP|$ is very close to each other, independent of the respective IT group. After the COVID-19 outbreak, we see an increase in $|ALLP|$ for both groups, which is not surprising due to the increasing uncertainty in economic outlooks and borrowers' situations. However, this increase in $|ALLP|$ is much lower for banks belonging to the high IT investments group. Therefore, banks with more IT investments can better assess their loan risks and better react to changing environments or structural breaks. Figure 2.2 supports the results of our multivariate regressions that there is indeed a significant difference between banks that invest more in IT and banks that do not. In line with the multivariate regressions, higher IT investments have a decreasing influence on the dispersion of the LLP; in other words, they have a positive correlation with higher quality assessments of loan risk. Again, we see that the benefits of banks with higher IT investments diminish in the following quarters.

## 2.5 Robustness Tests

Placebo tests are a possible instrument to test the validity of a Difference-in-Differences approach (Cunningham 2021). Our results could be influenced by other confounding events that happened around the outbreak of the COVID-19 pandemic or are attributable to general time trends. In Figure 2.1, we can already see these placebo tests. Since the interaction coefficient between our IT measure and the respective quarters is not significantly different for any period before the outbreak of the pandemic, we conclude that our finding is only valid for the first two quarters of the pandemic timeframe. Overall, the results from the placebo tests show that our main empirical results in Table 2.3 cannot be attributed to other events but most likely come from the COVID-19 pandemic.

Additionally, our empirical identification strategy is exposed to various concerns, for example, potential endogeneity between the main variables of interest. We tackle the different concerns in different ways. The first concern comes from possible anticipation effects. Suppose firms anticipated the COVID-19 pandemic and thus invested more and earlier in IT investments but simultaneously granted different types of loans with lower $|ALLP|$, then endogeneity could bias our results. To address this concern, Column (1) of Table 2.4 shows an alternative specification if we use IT investments from the year 2017 as our instrument for IT capabilities. We argue that banks' IT investments in 2017 are unlikely to be affected by any anticipations of the COVID-19 pandemic. Our results are qualitatively unchanged compared to the results of our main specification, and IT investments still have a negative and significant coefficient for the period before the COVID-19 pandemic. Moreover, we conduct the regression also with a rolling IT investments measure over the previous eight quarters to account for changes in the IT strategy in banks over time. Since this measure is not static, we gather results from a "fuzzy Difference-in-

Differences" approach which has some important implications for the interpretation of the results as explained by de Chaisemartin and D'Haultfœuille (2018). In case there is no stable treatment over time for the control group, they show that the regression results rest on the assumption that the treatment effects are stable and homogenous over time. Since our treatment variable is, even though variable and a rolling measure over time, rather stable, we are still convinced that the results are correct, at least direction-wise.

***Table 2.4:*** *Robustness Tests: Static IT Measure from 2017 and Rolling IT Measure.*

| Dependent Variable:<br>\|ALLP\| | (1)<br>IT investments 2017 | (2)<br>IT investments 8 quarter |
|---|---|---|
| IT investments x COVID Crisis | -4.19339** | -4.71836** |
| | (1.84588) | (1.85487) |
| COVID Crisis | 3.05301*** | 3.34643*** |
| | (0.46704) | (0.5186) |
| IT Investments | -1.43388 | -1.56087 |
| | (1.05939) | (1.06815) |
| *Controls:* | | |
| dNPL | 26.97183* | 26.74905* |
| | (14.07151) | (14.06055) |
| Lagged dNPL | 10.35812 | 10.74867 |
| | (12.95374) | (12.91138) |
| RegCap | 8.83426*** | 9.05268*** |
| | (2.70599) | (2.72461) |
| EBLLP | -74.39943*** | -74.22252*** |
| | (15.23483) | (15.181) |
| Asset Growth | 5.21572*** | 5.28317*** |
| | (2.01301) | (2.00568) |
| Lagged Loans to Assets | -1.5841 | -1.61358 |
| | (1.36324) | (1.36) |
| Deposits to Assets | -4.8261** | -4.76708** |
| | (2.16755) | (2.15238) |
| Lagged Deposits to Assets | 0.49417 | 0.38998 |
| | (1.74744) | (1.73191) |
| Real Estate Loans | -0.78641 | -0.85831 |
| | (1.01145) | (1.00577) |
| Commercial Loans | 0.3893 | 0.33273 |
| | (1.58624) | (1.58207) |
| Retail Loans | -0.48658 | -0.69663 |
| | (2.99085) | (2.98153) |
| State & Time FE | Yes | Yes |
| Observations | 8,522 | 8,522 |
| $R^2$ | 0.197 | 0.198 |

*The table contains the results of the robustness tests. There are two OLS regression specifications with standard errors that are robust and clustered on bank-level. The dependent variable is |ALLP|. Column (1) shows the results from the robustness test with the static IT Investments from the year 2017. The regression specification itself is similar to the regressions whose results we present in Table 2.3. Column (2) contains the results from the regression with the rolling IT measure. Again, the specification we use here is the main specification shown in Table 2.3. Each column contains data for the entire sample. Robust standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.*

Furthermore, functional form misspecification might bias our results. If banks with high IT investments structurally differ from those with low IT investments, and if the control variables cannot adequately capture these differences, Shipman et al. (2017) show that the empirical results might be biased. To alleviate the concerns of functional form misspecification, we conduct a propensity score matching (PSM) analysis. The PSM can be used to estimate causal treatment

***Table 2.5:*** *Propensity Score Matching - First Stage and Average Treatment Effect on the Treated.*

| Above-median IT investments | Coef. | Std.Err | t-value | p-value | [95% Conf | Interval] |
|---|---|---|---|---|---|---|
| LLP | 1.186 | 13.08 | 0.09 | 0.928 | -24.452 | 26.823 |
| RegCap | 1.268*** | 0.446 | 2.84 | 0.004 | 0.394 | 2.142 |
| Strictness | 0.591** | 0.29 | 2.04 | 0.042 | 0.022 | 1.161 |
| NPL | -3.618* | 2.032 | -1.78 | 0.075 | -7.601 | 0.365 |
| Loans to assets | 0.124 | 0.164 | 0.76 | 0.448 | -0.197 | 0.445 |
| Deposits to assets | 1.155*** | 0.22 | 5.24 | 0 | 0.723 | 1.587 |
| Size | -0.181*** | 0.016 | -11.64 | 0 | -0.211 | -0.15 |
| EBLLP | 1.024 | 1.14 | 0.90 | 0.369 | -1.21 | 3.258 |
| Loans | 0*** | 0 | 3.47 | 0.001 | 0 | 0 |
| Fed Chartered | 0.141*** | 0.036 | 3.94 | 0 | 0.071 | 0.212 |

| | | | |
|---|---|---|---|
| Mean dependent var | 0.497 | SD dependent var | 0.500 |
| Pseudo r-squared | 0.027 | Number of obs | 8398 |
| Chi-square | 311.633 | Prob>chi2 | 0.000 |
| Akaike crit. (AIC) | 11,352.094 | Bayesian crit. (BIC) | 11,429.487 |

| ALLP | Treated | Controls | Difference | S.E | T-Stat |
|---|---|---|---|---|---|
| Unmatched | 4.3422 | 4.664 | -0.3219 | 0.1022 | -3.15 |
| ATT | 4.3387 | 4.693 | -0.3542 | 0.1487 | -2.38 |

| | Off Support | On Support | Total |
|---|---|---|---|
| Untreated | 1 | 4226 | 4227 |
| Treated | 29 | 4142 | 4171 |
| Total | 30 | 8368 | 8398 |

*The upper part of Table 2.5 shows a probit estimation result. The dependent variable is a binary variable indicating whether a bank belongs to the group of banks that invests more than the median bank in IT - this is indicated by a value of 1. With this probit estimation, we conduct the Propensity Score Matching. The result of the Propensity Score Matching is shown in the middle part of Table 2.5. Both groups, namely high IT investment banks (Treated) and low IT investment banks (Control), are then compared. The ATT Difference, together with the respective T-statistic, is the relevant item. ATT stands for average treatment effect on the treated. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.*

effects (Caliendo and Kopeinig 2008) and is also a powerful tool if there is misspecification of the functional form in the regression model (Shipman et al. 2017). In our case, the PSM matches banks with comparable, observable variables to generate a control group as similar as possible to the treatment group. The treatment in our framework is whether banks belong to the 50% of banks that spend more than the median bank on IT investments. The first step is to conduct a probit regression of our treatment variable. The result of this regression is provided in Table 2.5. We include bank-level as well as geographic variables as control variables in our probit regression specification: the amount of *LLP* is scaled to total loans, the *Regulatory Capital* ratio, the amount of *NPL* is scaled to total loans, the *Loans to Assets* and *Deposits to Assets* ratio, the *Size* of the bank, and the *EBLLP*. We also add the *Strictness* Index developed by Agarwal et al. (2014). This index captures the differences between states and their regulatory strictness. Lastly, we add quarter-fixed effects.

Our matching procedure allows for the replacement of the control group observations. According to Shipman et al. (2017), this procedure leads to better quality matching. For the matching, we use a caliper width of 0.2, as suggested by Wang et al. (2013). The balancing property between the two groups is satisfied that indicates they significantly differ in the treatment variable but not in the variables for which we carried out the matching.

Again and in line with our previous findings, we calculate a highly significant and negative coefficient for the ATT, stated in Table 2.5. ATT describes the so-called average treatment effect on the treated – or in our case, the effect of having above-median IT investments. To summarize, a negative ATT is in line with our previous results and strengthens them.

In addition, tax incentives might provide a confounding effect for our sample of banks. Since banks can use LLP to manage earnings and their regulatory capital, we test whether our results hold when we consider tax incentives for banks. For this purpose, we first exclude firms that had losses in a single period from our regression model and observe that our results still hold and remain qualitatively unchanged. The same applies when considering different firm types – namely, whether a bank is a C- or S-corporation. C- and S-corporations differ in their tax regulations because C-corporations are taxed under subchapter C of the IRS and S-corporations are taxed under subchapter S. Thus, we add a triple interaction term to our regression that gives us the coefficient for the firm types during the pandemic that is interacted with *IT Investments*. The unreported results from the triple interactions show that our results are robust for the different tax statuses of banks. While the interaction term between *IT investments* and the pandemic indicator is qualitatively unchanged, the coefficient for the (triple) interaction term is insignificant. Also, the single coefficient for S-corporation banks is, though negative, not significant. Thus, we do not observe a significant coefficient regarding tax motives for our IT investment results and, accordingly, do not expect tax motives to be a driver of banks' IT investments.

## 2.6 Limitations

By using standardized data on banks' IT provided by the FDIC, we gather results supporting the positive value of IT for their loan risk assessments. However, as explained previously, we are only able to retrieve the so-called value capability (Bhatt and Grover 2005). Since IT capability is not only determined by value capability, we cannot make statements about banks' competitive or dynamic capabilities and their relation to banks' loan risk assessments. Nevertheless, we expect our measure to provide a conservative estimate of the actual effect of IT capabilities on banks whenever the failure probability does not systematically differ from our dependent variable. Only if the factors that drive the success of IT projects also enable banks to better forecast their loan portfolios in crisis situations could they affect our empirical results. However, it would definitely be insightful to empirically disentangle the different types of IT capabilities and reveal their relation to banks' loan risk assessments.

## 2.7 Discussion and Conclusion

In this study, we raise the question of the benefits from IT investments for banks. While other studies find diverging results, we provide new empirical evidence and show that IT investments indeed make sense for banks, at least from a loan risk perspective, which is crucial for commercial banks. We investigate whether banks with more IT investments can more precisely estimate their loan risks and thus, assess their loan loss provisions as exactly as possible. While assessments of loan risk in normal times can be anticipated quite well due to low uncertainty in the economic environment, we do expect a benefit from additional IT during normal times, but not a huge one. This is why we include a period of economic distress in our analysis, namely the COVID-19 pandemic. The structural disruption caused by the pandemic creates new dynamics in the economic and business environment that have made it necessary for banks to estimate their loan risks without being able to draw on experience from previous periods.

The final sample of this study comprises 8,522 bank-quarter observations from 665 banks. Our study shows that IT investments are indeed helpful for banks' loan risk assessments in times of economic distress. Banks with higher IT investments estimated loan risks more precisely in terms of LLP than banks with a lower level of IT investments which means that IT investments are unquestionably relevant for banks from a loan risk perspective. Contradicting our first expectation, IT investments are not statistically or significantly beneficial in normal times when speaking in terms of loan risk assessments. This is surprising at first, but keeping in mind that banks can use their experience from previous quarters when no structural break occurs, it appears plausible and relatively straightforward. When no structural disruption occurs, banks can update their expectations and estimations with experience from a similar business environment from past quarters regardless of their IT investments which means they could use last year's LLP as an anchor and adjust according to the managers' gut feelings.

The results from this study have important consequences for banks and their digital transformation. First, banks should invest more in IT, even though it might not be profitable immediately. IT investments or capabilities make the bank more resilient to external shocks and structural breaks. Therefore, higher IT investments, or digital transformation in general, are crucial for the stability of the overall banking system. As data becomes more extensive and detailed, there is no other option than to implement or build a business structure that allows more and more data to be used and processed for many bank-specific tasks.

# Chapter 3
# The Role of Geographical Distance for Investment Decisions in Crowdinvesting

**Abstract** Startups rely on alternative forms of financing, such as crowdinvesting, as they usually miss reliable credit ratings and bank loans. We explore the role of geographical distance between investors and startups, and analyze the role of signals and the investor-startup-relationship in investment decisions. Using a unique proprietary dataset from the largest crowdinvesting platform in Mexico, we show that distance between investors and startups negatively affect the investment probability and investment amount. An investor-startup-relationship only diminishes this effect in specific instances while objective signals do in any case. Evidence is in line with the idea that investors with a close relationship to a startup should be used as fallback options instead of exclusive investors in startups' early development stage. Furthermore, it underscores that signals are of high importance in modern forms of financing.

*The new source of power is not money in the hands of a few, but information in the hands of many.*
– John Nasbitt, U.S. author

**Acknowledgments** This chapter is adapted from my published paper "The role of geographical distance for investment decisions in crowdinvesting" in the Journal of Small Business Management 2024. It was joint work with Dr. Kristina Uhl from the Eberhard Karls University of Tübingen.

## 3.1 Introduction

The rising use of the Internet puts online platforms more and more into the center of attention not only for online sales. The importance of online platforms in the financial sector also constantly increases (OECD 2019). In crowdinvesting, also called equity crowdfunding, platforms enable investors to easily choose from one of multiple equity investments which are offered in a single place. Crowdinvesting platforms lead to a change in investment decisions of private investors as they do not only save time for investors but also increment their access to alternative investment opportunities (Konovalova et al. 2020). They ease the access to information in general as well as through features and details, for example, comments or answers on frequently asked questions, which can more easily be made available and disseminated online.

Many studies focus on investment decisions in established investment types, usually conducted without online platforms involved. This research strand consistently shows that the physical distance between investors and investees is negatively correlated with the investment probability and investment amount (Bae et al. 2008; Degryse and Ongena 2005; Dejean 2020). In crowdinvesting, most steps in the investment process are arranged online. This fact might suggest that the comparison and evaluation of investment possibilities have become easier in recent years, leading to a lower relevance of the physical distance between the investment parties. However, studies show that also in this comparably new form of financing, investors experience larger information asymmetries with greater physical distance. This informational disadvantage plays a major role in investors' decision-making in crowdinvesting (Kleinert and Volkmann 2019). A close relationship between investor and investee might decrease information asymmetries, and foster investors' confidence in startup success so that the investment probability and investment amount increase (Mollick 2014; Angerer et al. 2017). Besides information asymmetries reflected through physical distance as a driver of investors' decision making, social distance might also be proxied by distance. For example, Agrawal et al. (2008) conclude that social aspects are able to substitute the physical distance between investors and investees. In line with the so-called mere-exposure effect, investors might be biased toward physically closer startups since they are exposed more often to the startups or the area the startups are located in and therefore feel, for example, more connected to them. Moreover, objectively observable signals by other investors may decrease uncertainty in decision-making. In turn, these signals may positively influence investment decisions (Courtney et al. 2017).[6]

In this study, we investigate the link between geographical distance between investors and startups and the investment probability and investment amount in crowdinvesting. Using detailed data on investor and startup characteristics, we conduct logit regressions with 407,497 observations (investment and no investment observations) as well as OLS regressions with fixed effects

[6] Sah and Stiglitz (1986, p. 716) subsume such phenomena under "the architecture of an economic system [which] affects the errors made by individuals within the system, as well as how these errors are aggregated". Ultimately, these interdependencies describe Bayesian arguments.

with 11,054 observations (investment observations only) from the largest Mexican crowdinvesting platform. Thereby, we investigate if distance still plays a role in investment decisions. More specifically, we concentrate on the investment probability and investment amount. We also analyze the moderating effect of the investor-startup-relationship in this context. If any kind of relationship exists, on the one hand, informational disadvantages of investors should decrease as more information is directly transferred from the better informed party, that is, the startup, to the less informed one, that is, the investor. On the other hand, an existing relationship might lead to more trust in the startup through previous encounters and, thus, reduced uncertainty and a feeling of connectedness. Hence, we expect that relationships mitigate the negative effect of distance on the investment size. In the final part of this study, we evaluate the effect of observable signals on investment decisions of future investors in crowdinvesting. We assume that a high number of likes proves to be a signal of confidence by other investors who have already had good experience with the investee or are convinced of a startup's success in the near future. In turn, we expect potential investors to be motivated to invest more, leading to a higher investment amount.

Our study helps deepening the understanding of existing links between physical distance, personal relationships, and signals in the field of entrepreneurial research. With the given dataset containing information on investors and startups, we provide insights into how rationally observable characteristics influence investment decisions in crowdinvesting, that is, the investment probability and investment amount.

First, we find that a greater distance is associated with a decrease in the investment probability and investment amount. A 10% increase in distance corresponds to a decrease in investment probability by 1.42 pp and the investment amount by 0.39%.

Second, we provide evidence that the investor-startup-relationship correlates with the investment probability as well as the investment amount. A given relationship shows a positive correlation with investment decisions. However, a relationship does not fully counteract the negative role of being farther away from the startup in any case. In addition, this study confirms the finding by Angerer et al. (2017), that family and friends are frequent investors in startups' early development stage.

Last, we explore the role of likes, that is, an observable signal, on investment decisions. The investment probability rises with the number of likes received. However, we find a significantly negative correlation of likes on the investment amount, indicating that private investors prefer to hop on projects with an expected positive outcome (indicated by a high number of likes from other investors) with small amounts of money. As the outcome is still uncertain in crowdinvesting, this is in line with the idea of risk diversification. This result shows that investors consider other investors' opinions in their decision-making process.

Our study relates and contributes to the literature in different ways. First, this study relates to the strand of literature on the effect of geographical distance between investors and startups in investment decisions (Chen et al. 2010; Coval and Moskowitz 1999). Furthermore, we add to the literature on the effects of relationships and third-party signals on the link between distance and investment decisions. Second, our study relates to the general literature on crowdinvesting which is typically focused on the U.S. and Europe. Our study expands these results by providing new empirical evidence and insights into the development in Mexico. Mexico, with an expected yearly growth rate of 4.75% in the upcoming years (Statista 2022*b*), constitutes one of the fastest growing markets for crowdinvesting—to compare, Germany expects a growth rate of 4.49% (Statista 2022*a*) and the U.S. of 4.07% (Statista 2022*c*). Third, we use a unique dataset including detailed information on crowdinvestment activities, investor, and startup characteristics. Thus, we provide new insights into investment decisions within this increasingly demanded modern way of financing. We also shed light on the role of startups' public evaluation (that is, likes) in investment decisions. This topic has become more and more important as the digitalization proceeds with big steps. In turn, online activities in all areas, especially in financial transactions, become of higher focus and a more and more relevant option to consider in investors' decision-making.

The paper proceeds as follows. We review the literature and develop our hypotheses, and describe the sample, methodology, descriptive statistics, and the research design. Then, we present and discuss our empirical results followed by robustness tests and a description of the limitations of this study. A summary of the main findings concludes.

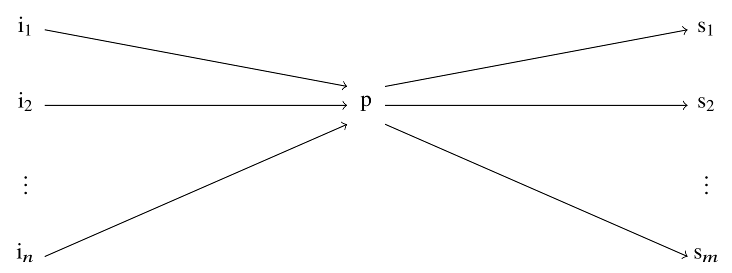
## 3.2 Institutional Background

Crowdinvesting describes a subcategory of crowdfunding (e.g., Leboeuf and Schwienbacher 2018).[7] It constitutes a form of equity financing and is also known as equity crowdfunding. In crowdinvesting, usually a large number of private investors support early-stage firms, and especially startups, with a relatively small amount of money per person (e.g., Moritz and Block 2014). Similar to other equity investments, investors purchase equity shares in the firm. In return, they receive dividends in later years depending on the success of the investee (e.g., Ahlers et al. 2015).

Crowdinvestment deals are usually arranged via online platforms on the Internet. *Figure 3.1* displays the operating principle of a crowdinvesting platform $p$ which potentially connects $n$ investors ($i$) with $m$ startups ($s$). On the platform, startups can register and create a profile with information on themselves and their business idea. Investors looking for an investment opportunity can browse these startups, and decide (not) to support a particular startup financially.

[7] "Crowdfunding" describes a modern form of financing which initiates an open call with or without an intermediary. Crowdfunding consists of three subgroups: Crowdinvesting (= equity capital), crowdlending (= debt capital), and reward-/donation-based crowdfunding.

***Figure 3.1:*** *Operating Principle of Crowdinvesting*




Since high information asymmetries exist between (potential) investors and startups (e.g., Kleinert and Volkmann 2019) and the success of early-stage firms is per se uncertain, crowdinvesting is considered relatively risky (e.g., Angerer et al. 2018). Crowdinvesting platforms bundle and ease the access to information. They provide as much information as possible publicly available, and create, e.g., rankings in terms of risk for the startups listed on their website. However, investors looking for investment opportunities still remain with some uncertainty about the quality of their potential investments as they cannot assess the correctness of the information given on the website about the startups (e.g., Goethner et al. 2021). Still, crowdinvesting constitutes a convenient way for investors to participate in a potentially profitable project with a small amount of money (e.g., Moritz and Block 2014).

*Figure 3.2* depicts the starting page of a crowdinvesting platform.[8] All investment possibilities, i.e., startups, are listed next to each other. The starting page is accessible without registering on the platform. On the top, the current amount of likes which a startup has received from registered users of the platform until that moment is displayed. Subsequently, the startup's name and a short own description provide information on the startup's business model. This information is followed by details created by the crowdinvesting platform. Based on the startup's financial information, the platform internally assesses the estimated yearly return (in percent) and the startup's risk (on a 15-point-scale from low, stable, moderate, fluctuating, to high). Additionally, the amount of money already collected (in Mexican Pesos) and the number of engaged investors is displayed. The former information is also depicted in a bar chart as the share of the financing goal, i.e., as the share of the minimum amount of financial resources which a startup strives to raise within this particular investment period.

For startups entering an investment period in the short run, the same information is displayed at the bottom of the crowdinvesting platform's starting page. Upcoming investment possibilities are announced two weeks before they get started. Instead of the bar chart displaying the already collected funds, the startup's minimum financing goal is depicted.

---

[8] *Figure 3.2* and *Figure 3.3* depict the website layout of our data-providing platform. All information in the text is also platform-specific. Other crowdinvesting platforms are structured similarly.

***Figure 3.2:*** *Starting Page on the Crowdinvesting Platform*

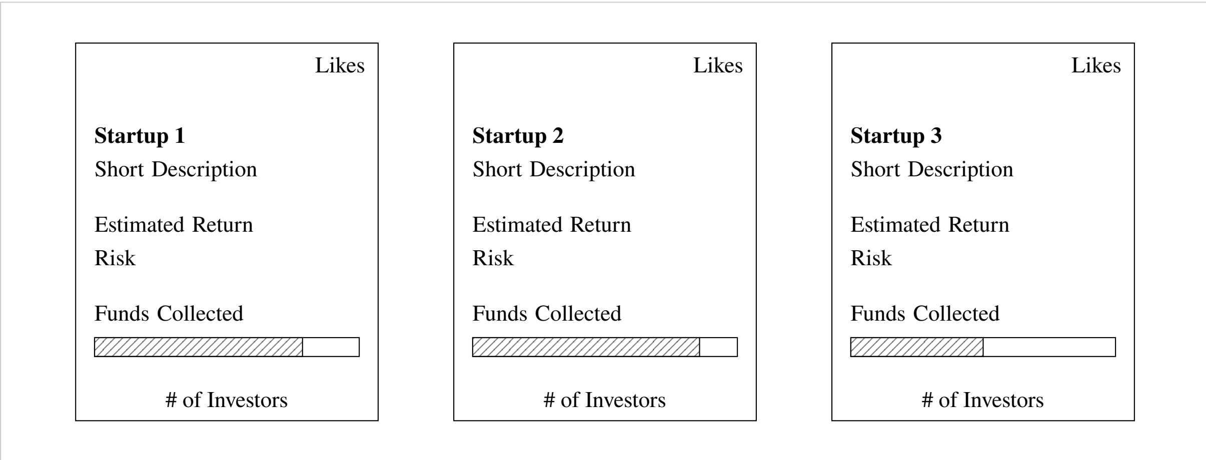
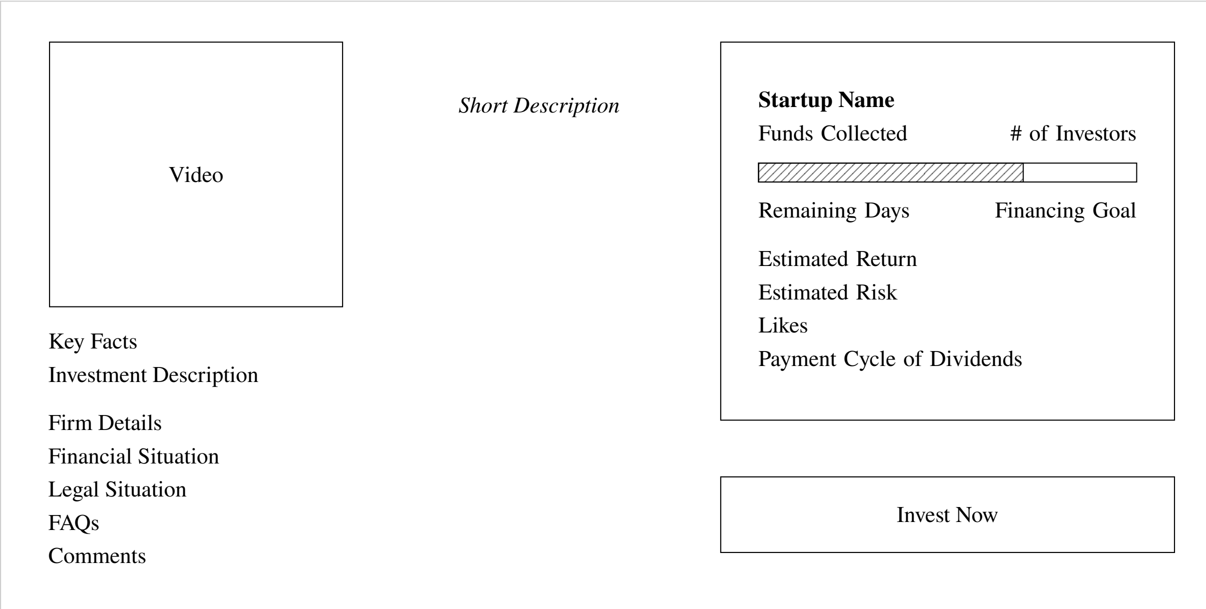



*Figure 3.3* displays the structure of startups' individual websites which can be accessed via the starting page. These individual pages are also accessible without registering on the platform. On the top, the startup presents itself in a short video of approximately two minutes. Additionally, the short description from the starting page is again depicted. Below, further details are shown including key facts on the startup and a description of the investment opportunity. Exemplarily, this might be details on the firm's leading employees, or the startup's business model. Subsequently, further information follow on the startup, its financial and legal situation, some answers on frequently asked questions, and comments about the startup from former users of the platform.

More information on the specific investment possibility is displayed on the right of the individual startup websites. First, the amount of money already collected and the number of engaged investors are shown. The bar chart again displays the collected funds as proportion of startup's overall financing goal. The minimum financing goal is depicted just below the bar chart as well as the remaining days till the investment period expires. The duration of an individual investment period ranges between approximately one and eight months, with a mean of four months. Further below, details on the estimated yearly return and the risk of investment are depicted. It is followed by the number of the startup's likes and the payment cycle of dividends (mostly per trimester but may also be defined ex ante as per semester or yearly by the startup).

***Figure 3.3:*** *Individual Startup Page on the Crowdinvesting Platform*




For investing in a startup, investors need to register on the platform. In the first step, it is sufficient to enter their first and last name, a personal email address, and to choose a password to enter the internal area of the crowdinvesting platform. When investing for the first time, more information is demanded, e.g., investors' date of birth, their highest educational degree, and the state of residence. The crowdinvesting platform also tries to draw investors' attention to the risk taken through crowdinvesting activities by asking them some questions on their general risk preference. Thereafter, investors can freely choose the amount of money which they want to invest.[9] The amount invested stays between three and six years with the startup, and gets repaid to the investors with a one-time payment at the end of the predefined time period. The crowdinvesting platform charges 5% of the investment amount as fee for their services.

If the minimum financing goal of the startup is reached within the investment period, investors are requested to transfer their investment amount to the crowdinvesting platform after the termination of the investment period. The platform then transfers the money to the startup. Thereby, investors purchase a contractual share of future revenues in the startup. This share is defined as the proportion of the individual investment amount divided by the total funds collected.[10] In subsequent years, investors receive dividends which are defined as the respective share of the startup's earnings from sales or other business activities. Dividends are paid in a regular rhythm, starting immediately after the startup has received the funds collected. The exact size of the dividends paid out may vary between payment periods, depending on the success of the startup. To decrease uncertainty for investors, startups define a minimum and maximum dividend (in

[9] Some startups require a minimum amount of money from an individual investment. The investment amount can be increased or withdrawn by the investor at any time during the predefined investment period.

[10] The total amount collected is by definition equal or larger than the startup's minimum financing goal.

percent of earnings). During the payment cycle of dividends, investors receive information on the startup's progress, e.g., in sales, and can communicate with the startup via the crowdinvesting platform.

If the startup's minimum financing goal is not reached within the investment period, investors are not requested to transfer their individual investment amount to the startup (via the crowdinvesting platform).

## 3.3 Literature Review and Hypotheses Development

Unsurprisingly, lacking financial resources restrict startup growth (Beck et al. 2006; Levine 2005). In addition, an absent credit and firm history often impedes early-stage startups from access to traditional ways of financing (Carpenter and Petersen 2002). In crowdinvesting, large information asymmetries exist between investors and investees (Kleinert and Volkmann 2019). These information asymmetries between startups as the better and investors as the less informed party play a major role in investors' decision-making. They stem from market imperfections and result in market inefficiencies (Akerlof 1970; Rosser 2003). Information asymmetries are especially relevant for borrowers in crowdfunding compared to the more traditional ways of financing (Piva and Rossi-Lamastra 2018). Blum and Goldfarb (2006) show that even in the case of purely digital goods sold and used via the Internet, that is, in the absence of trading and other opportunity costs, informational differences seem to play a role in investment decisions. The physical distance between investors' and startups' place of domicile may therefore increase information asymmetries between the two parties, for which the information hypothesis, assuming local investors to receive more and timelier information regarding local firms, could lead to more trust about the quality of a startup and lead to higher or more frequent investments in startups located nearby.

And indeed, several researchers find a preference for local investments compared to physically more distant ones (Bae et al. 2008; Malloy 2005). Dejean (2020, p. 337) underscores this finding with the explanation of "a regional attachment or a distant relationship with a particular community." A broad literature confirms this idea in equity and startup investments. First, Coval and Moskowitz (1999) find a preference for local equity investments in the U.S. When investigating the portfolio composition of stocks, investment managers show a strong bias toward local firms. Chen et al. (2010) display the preference toward geographically closer investments in the venture capital market in the U.S. This effect persists even though nonlocal investments outperform local venture capital investments in terms of return. Lutz et al. (2013) confirm this finding for the German venture capital environment. Besides currency and legal frameworks, Niemand et al. (2018) also observe the preference for physically close investments in the equity crowdfunding context. However, they call for further research on this financing method on a national level. Hornuf et al. (2022) find a local bias for individual equity crowdfunding investments and investment portfolios. This effect appears to be largest for family, friends, and angel-like investors.

However, since the invested amounts are typically relatively small and therefore not much is at stake financially for investors, information asymmetries might not be solely the driver of this phenomenon in crowdfunding. Moreover, in crowdfunding, investors' participation in the startup is low compared to, for example, venture capitalists. Thus, besides information asymmetries, the so-called mere-exposure effect (Zajonc 1968) could be a further explanation for investors providing money close to their location. The mere-exposure effect is a form of cognitive bias stating that the repeated exposure to a stimulus (in this case, for example, the startup itself or the location the investors live in) changes the way people think and feel about the stimulus itself, even though the stimulus does not change (Bornstein and Craver-Lemley 2022). Through these (multiple) exposures, investors might develop a feeling of liking for that stimulus (Stafford and Grimes 2012) and familiarity with others (Kwan et al. 2015). Hence, geographical distance can also be a proxy for cultural or social preferences and, for example, connectedness, and can result in some kind of familiarity between investors, the area, and also the firms located in this geographical area, leading to a higher likelihood or a higher ambition to invest in startups nearby (Lindblom et al. 2018). Consequently, we expect:

***Hypothesis 1a:*** *Higher geographical distance between investors and startups negatively correlates with investors' decision to invest.*

***Hypothesis 1b:*** *Higher geographical distance between investors and startups negatively correlates with the investment amount.*

In line with the mere exposure effect, Agrawal et al. (2008, p. 259) state that "social proximity [...] substitutes for [...] spatial proximity", indicating that social proximity, next to geographical distance, does also play a role in investment decision making. According to the mere-exposure effect, repeated exposures to a stimulus strengthens the attitude regarding this stimulus. Through their repeated interaction, the attitude toward that startup can thus be enhanced (Bornstein and Craver-Lemley 2022) and lead to a positive investment decision compared to a case where the startup is unknown to them. Hence, investors might feel drawn to startups to whom they were exposed in the past, or in other words, those they have already engaged with in the past. Accordingly, based on the mere exposure effect, any gap — be it social or informational — between investors and startups due to geographical distance could be alleviated by previous interactions with the startups.

Empirically, several researchers support the idea that previous interactions or close relationships are associated with a higher investment amount (for example, Angerer et al. 2017).[11] Banerji and Reimer (2019) show that the more connections a founder has, the higher the amount of money raised. They state "the social connectedness of founders [to be] the best predictor of funds raised" (Banerji and Reimer 2019, p. 46). Mollick (2014) observes similar effects in his exploratory study. Measuring the effect of entrepreneurs' social network by the number of friends online, he shows that larger personal networks are associated with a higher success probability of crowdfunding projects. Leyden et al. (2014) and Vismara (2016) underscore this idea. They explain the effect by arguing that social ties decrease uncertainty. As several studies have already revealed the positive effect of relationships on the investment probability, we exploit the advantage of this dataset of having detailed information on investors, investments, and startups, and focus on these links and respective interactions to the investment amount.[12] Thus, we hypothesize:

***Hypothesis 2:*** *An existing investor-startup-relationship positively mitigates the role of geographical distance between investors and startups on the investment amount.*

Due to, on average, small individual investment amounts in crowdinvesting, the costs of evaluating a startup for crowdinvestors are disproportionately high compared to the amount invested. Following the underlying idea of Bikhchandani et al. (1992) arguing that people use information from others, when this newly gained information outweighs one's own information, the absence of easily available information in crowdinvesting could result in a fertile field for some kind of herd behavior. Herd behavior can be divided into two aspects. First, persons discount their own information and rely less on it compared to the observed behavior of others. This may indicate the assumption that predecessors are better informed or conducted high cognitive efforts to analyze, for example, the startup. Second, Bandura (1977) already describes that people tend to imitate others meaning that investors engage in herd behavior simply because previous investors did.

For herd behavior to occur, the investors' previous behavior must be observable for the potential investors. Since we are in an online scenario where no face-to-face communication or direct observation of others is possible, likes are a conceivable way to transfer information between investors. Li and Wu (2013) state that the volume of likes can represent the aggregated approval of a product and can reduce quality uncertainty. Thus, likes act as a positive costless signal for investors' peers. If new investors think that other investors, in aggregation, have more or even superior information or experiences about the startup represented by the number of likes received, likes as a signal could lead to herd behavior and result in more frequent or even higher

[11] In studies analyzing relationships in investment decisions, a "relationship" may also describe the situation of having already invested in a startup (e.g., Sorenson and Stuart 2001). We define the term "relationship" for this study in our section on Variables.

[12] The same argument holds for Hypothesis 3 when investigating the role of likes.

investments in said startup. For example, Cornelis et al. (2022) state that the cognitive perception influences potential investors' funding intention. Moreover, it is not even necessary for a like to be accompanied by an investment, as Bikhchandani et al. (1992) describe that also a similar behavior, in our case investing in the startup, can be triggered by the signal.

Empirically, this notion is confirmed. Signals are able to reduce uncertainty (Ahlers et al. 2015). In their empirical studies, Davila et al. (2003) and Courtney et al. (2017) confirm that signals from startups as well as from third parties increase startups' possibilities for success. Nitani et al. (2019) confirm signals on, for example, firm and owner attributes, to influence the success probability of crowdinvesting campaigns. Block et al. (2018) confirm that signals which are easier to interpret show higher effects on the crowd. Several researchers observe that likes, describing observable "one-click cues" by former users, are a popular signal to use today (Ahlers et al. 2015). Likes are common in social media (Sumner et al. 2020) and indicate that individuals positively evaluate the content of a post, comment, or picture (Eranti and Lonkila 2015). Thus, they communicate individuals' "overall positive affect, or sentiment, toward the organization's message" (Saxton and Waters 2014, p. 287).

While one could argue that likes are only cheap talk and thus not relevant for investment decisions, in recent years, they have received increasing attention by the crowd and have shown higher importance in investors' decision-making process with the rising use of the Internet (Di Pietro et al. 2020; Kleinert and Volkmann 2019). Nitani et al. (2019) conclude that crowdfunding investors act rational and give strong weight to observable signals when taking investment decisions. This link is stronger for simple signals and nonexpert private investors (Monti et al. 2014). Carr et al. (2018) reveal the importance of the quantity of likes on social media platforms, and Mochon et al. (2017) display that online likes positively influence customer behavior. Thus, since likes are independent from the physical distance, they could mitigate the negative role of physical distance in the decision-making process and we hypothesize:

***Hypothesis 3:*** *The received likes positively mitigate the role of geographical distance between investors and startups on the investment amount.*

## 3.4 Data and Research Design

### 3.4.1 Data

The dataset for this study comprises data from the largest crowdinvesting platform in Mexico between 2015 and 2019. It includes information on startups and (potential) investors, as well as investors' individual decisions (not) to invest in a particular startup. In contrast to the databases of many former studies, our dataset contains information on the exact investment amount per investor per startup as well as private details, for example, investors' and startups' state of domicile. This information enables us to analyze factors influencing the concrete financial support provided while simultaneously looking at, for example, the relationship between startups and investors and other investment-specific characteristics.

Mexico as our country of data provenance represents an ideal environment for crowdinvesting research for several reasons. First, Mexico appears to be a country with very heterogeneous states in terms of the economic surrounding (see OECD 2022). Second, similar to blockchain technologies which might reduce problems of inefficient governance and corruption in Mexico (Zbinden and Kondova 2019), crowdinvesting allows money flow between lenders (= investors) and borrowers (= startups). Thus, the financial support provided depends entirely on the online presentation of the startup as there is no previous bilateral communication between investors and startups.

Third, Mexico corresponds to other countries in terms of the development of crowdinvesting in the past couple of years. Since 2018, the transaction volume in Mexico has almost doubled, reaching USD 8.39 million in 2022. Experts of the business data platform "Statista" project the transaction volume in the Mexican crowdinvesting market to further grow by 4.75% per year in the short run (Statista 2022*b*). This fact underscores the importance of our analyses.

### 3.4.2 Sample

The dataset includes observations from the platform's internal management system for which investees are characterized as "startups". We do not consider observations for which investees are registered as "SMEs", "revenue-based funding" or "real estate funding" as these categories did not exist throughout the whole sample period. Furthermore, we focus on startups' successful investment rounds, that is, when the minimum financing goal was reached within the investment period, as it ensures the explanatory power of our results with regard to the investment success. In addition, we do not have investor information from nonsuccessful investment rounds, and thus refrain from including these observations to allow a detailed analysis. We also limit the database to investors and startups located inside Mexico to avoid biased results due to a different macroeconomic environment which might drive our results. We focus on the first investment within each particular bilateral investor-startup combination. This restriction rests upon the idea that investors learn from former investments in a respective startup (Hornuf and Neuenkirch 2017). Furthermore, keeping these, in our dataset, 26 subsequent investments would not allow additional meaningful analyses due to the low number of observations and would, on top, bias our coefficient of the investor-startup relationship as relationships and knowledge flows in the entrepreneurial market evolve over time (Wadhwa and Kotha 2006).

After these selection decisions, we create all possible investor-startup combinations. This procedure results in our initial database with 618,120 investor-startup combinations (= observations) from 6,120 investors and 101 startups from the years 2015 to 2019. We subsequently drop observations with illogical values in any of our variables of interest. Exemplarily, this involves observations with objectively wrong information on investors' date of birth (for example, from investors characterized as born in the year 1887). This step decreases our sample by 5,454 observations (that is, 54 investors). Furthermore, we do not include observations for which information on dates does not logically fit our analyses. Exemplarily, this restriction relates to observations for which a startup was funded before an investor profile was created virtually, which logically prevents a possible investment. This excludes 205,169 investor-startup combinations from the sample. Overall, our final dataset for Hypothesis 1a, in which we investigate the factors influencing the investment probability, contains 407,497 observations. For the subsequent analyses of the investment amount (Hypothesis 1b, Hypothesis 2, and Hypothesis 3), we only include those observations where an investment has taken place. This results in a final sample of 11,054 individual investor-startup combinations.

Figure 3.4 displays the distribution of investors in Mexico. Investors appear widely spread throughout the country. Still, Mexico City (CMX), the capital city of Mexico, encompasses by far the highest number with approximately one third of all investors in the dataset. It is followed by the State of Mexico (MEX) with approximately half as many investors residing in this state (990).[13]

Figure 3.5 displays the distribution of startups throughout the Mexican states. We find a huge proportion of startups (61 of 101 startups) to be located in Mexico City. The other 50 startups in our dataset are distributed throughout the other states. All values are reported in Table A.3.1 of the Appendix.

[13] The State of Mexico describes the surrounding area of Mexico City. The capital cities of Jalisco and Nuevo Léon are Guadalajara and Monterrey, which are the second and third largest city in Mexico, respectively. Thus, this finding is not surprising.

***Figure 3.4:*** *Local Distribution of Investors in Mexico*

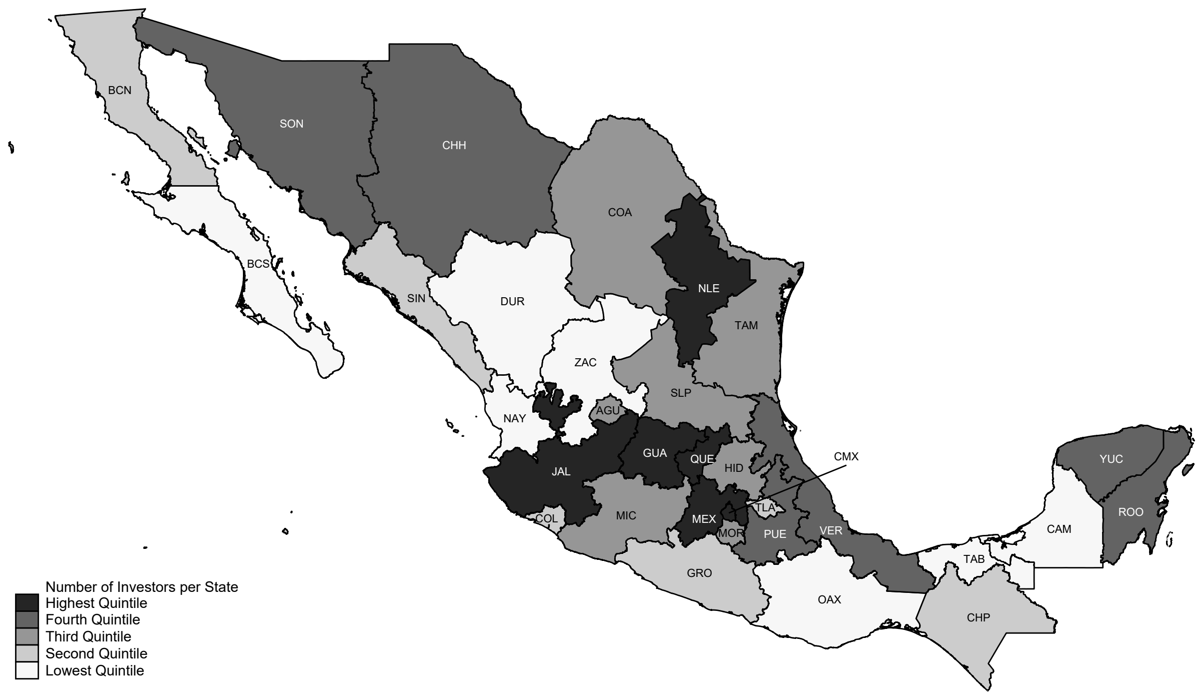
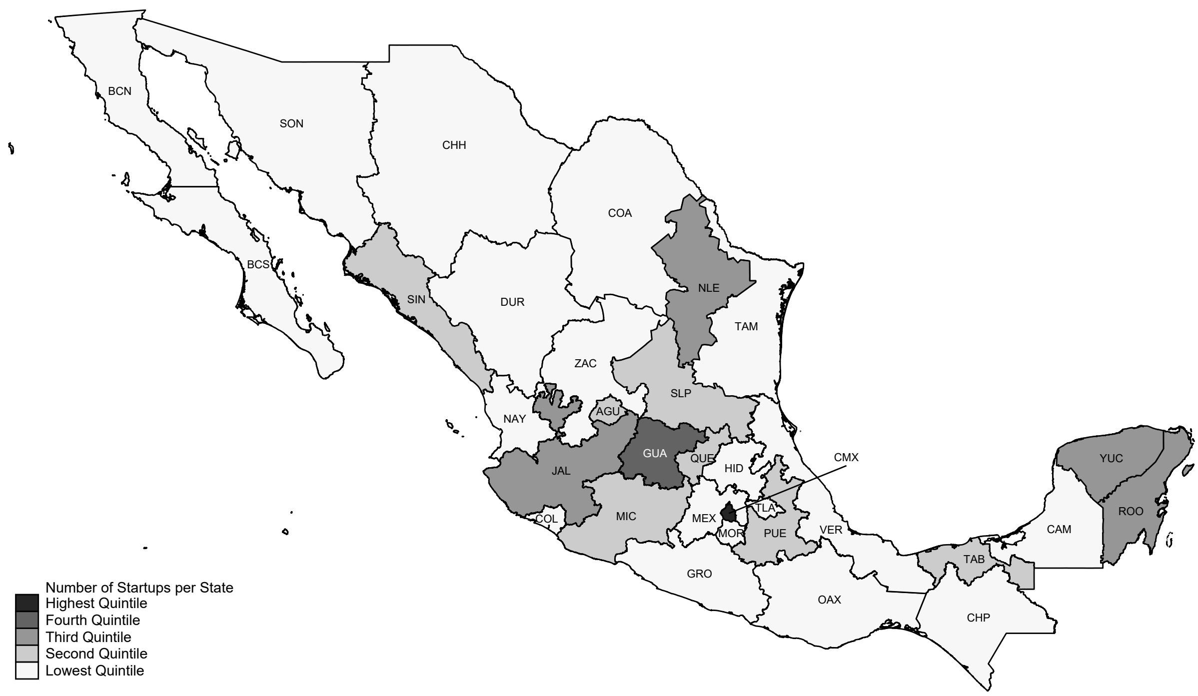



***Figure 3.5:*** *Local Distribution of Startups in Mexico*



### 3.4.3 Variables

#### 3.4.3.1 Investment Dummy

In Hypothesis 1a, we investigate the role of distance between investors and startups on the investment probability. To measure the investment probability in crowdinvesting, we employ a dummy variable as the dependent variable for whether an investment has taken place (= 1) or not (= 0). An "investment" is characterized through the investor's decision to financially support a startup via the crowdinvesting platform, that is, clicking "invest now", deciding on the investment amount, and entering all personal information on the platform.

#### 3.4.3.2 Investment Amount

In Hypothesis 1b, Hypothesis 2, and Hypothesis 3, we investigate factors influencing the investment amount per investor per startup. The investment amount is measured in Mexican pesos, enabling us to immediately observe the real effects of the explanatory variables. We use the natural logarithm to deal with the heterogeneous distribution of the amount invested and to manage outliers, ensuring the normality and homoscedasticity of the distribution.

#### 3.4.3.3 Distance

The geographical distance may capture a myriad of factors that might be associated with the investment decision. On the one hand, for investors located nearby, access to information is easier as they can, for example, talk to the investee's employees, customers, or competitors. Furthermore, investors can get information on the firm's current development through local media or might even have personal connections to the investee's management team (Bae et al. 2008; Coval and Moskowitz 1999). On the other hand, geographical distance may reflect social distance. Dejean (2020, p. 337) speaks of "a regional attachment or a distant relationship with a particular community", showing that distance is not solely a proxy for information asymmetries, but also (social) connectedness and attitude toward, for example, a region are reflected in the geographical distance. Empirically, we use the variable Distance in logarithmic terms.[14]

We measure the physical distance between one specific startup and one specific investor using the midpoints of the startup's and the investor's state, rounded to full kilometers. This procedure is based on the nonavailability of data regarding the exact place of residence of the investors. However, due to the large number of Mexican states (32) and the large size of Mexico (1,973,000 km$^2$), this proxy shows widely varying values.

[14] We add one to the investor-startup-distance of all observations. Thereby, we keep investor-startup-combinations from the same state. We proceed similar for all other variables which are measured using the natural logarithm (investment amount, financing goal, investor experience, and startup likes).

#### 3.4.3.4 Investor-Startup-Relationship

We also include the relationship between investors and startups in our regression analyses. Studies show that closer relationships, on average, lead to higher investment amounts (Angerer et al. 2017). Within the registration process on the crowdinvesting platform, investors must characterize their relationship to the startup into one of five groups: Customer, event, family/friends, maker network, or stranger. For the purpose of this study, we create a dummy variable which takes the value of one if the investor and startup have any kind of relationship, being it as family/friends, as customers, or knowing each other from an event or maker network. In contrast, this variable takes the value of zero if investors denote themselves as strangers to the startup.

#### 3.4.3.5 Likes

Within the analysis of Hypothesis 3, we take a closer look at the moderating role of the startups' likes received on the platform and the link between distance and investment decisions. On the platform, registered users can "like" a startup if they support a startup's idea. The current number of likes per startup is displayed on the platform as one characteristic feature on the list of potential investees presented. For each startup, we know the number of likes received at the time the data was extracted from the database. Likes received represent an observable signal to the less informed party, i.e., (potential) investors (Block et al. 2018). They might be seen as "a sense of community [...] who share a similar interest [...] that may result in increased positive attitudes towards the brand" (Coursaris et al. 2016, p. 3547). We include this variable into the analyses using the natural logarithm to account for smaller effects of numerically equal changes for a larger compared to a smaller number of likes.

#### 3.4.3.6 Control Variables

As control variable, we first include the financing goal within a predefined crowdinvesting period set by startups into our analysis. The financing goal describes the minimum amount of financial resources, measured in Mexican pesos, which a startup strives to get within the respective investment period. Including this variable, we follow several researchers (Ahlers et al. 2015; Kleinert et al. 2020; Vismara 2016). The presentation of a startup's financing target might correlate with investors' behavior with regard to providing financial support or not, and regarding size. We include this variable using the natural logarithm.[15]

In addition, we account for three investor-specific characteristics within our analyses. We include a dummy for investors' gender. It takes a value of one for male and a value of zero for

[15] One could also imagine taking the already collected proportion of the financing goal as control variable into the analysis. However, we miss information on the overall investment amount by all investors in each particular point in time. In addition, we cannot sum up individual investments to get this amount as we only include first investments per bilateral investor-startup-combination into our analysis.

female investors. Including this variable accounts for the lower risk aversion of men compared to women (Rieger et al. 2015). In turn, it is of high relevance in the comparatively risky investment form of crowdinvesting. We also include investors' age grouped by decades into the regression. It results in seven dummy variables for investors under or equal to 30, between 31 and 40, between 41 and 50, between 51 and 60, between 61 and 70, between 71 and 80, and above 80 years of age. Thereby, we account for different stages of life that come with different financial needs and, thus, different savings or investment attitudes. The age group of investors under or equal to 30 serves as our reference group within the analyses and the interpretation of the results. Moreover, we employ the natural logarithm of investors' time on the crowdinvesting platform in days to account for the possibility to invest in other startups listed on the crowdinvesting platform.[16] We measure this experience as the time between the day of investors' registration on the crowdinvesting platform until the day when the investment by the investor is registered in the platform's system (= investment date). We assume that the change in investors' experience becomes lower with a longer personal history on the platform

### 3.4.4 Descriptive Statistics

Table 3.1 displays the summary statistics on observation level for all variables. The average distance between investor and startup amounts to 451.66 km (and 355.26 km for the smaller sample in Hypothesis 1b, Hypothesis 2, and Hypothesis 3).[17] Considering the comparatively high standard deviation, our proxy for information asymmetries between investors and startups is very heterogeneously distributed. After creating all investor-startup combinations, we find an investment in approximately 3% of our observations. The average investment amount which startups receive amounts to 16,126.53 MXN (equals approximately 800 USD) with individual investments ranging from 160 to 1,490,000 MXN. The number of likes received by one startup ranges from 21 to 4,194 likes with an average amount of 1,081 likes. In 72% of all observations, investors and startups show a relationship if an investment takes place.

[16] This study concentrates on the first bilateral investor-startup interaction. Thus, investors cannot collect experience from former investments into the same startup.

[17] For understandability purposes, we describe absolute numbers in the text instead of the natural logarithms which are depicted in *Table 3.1*.

***Table 3.1:*** *Descriptive Statistics.*

| | N | Mean | StdDev | Min | p25 | p50 | p75 | Max |
|---|---|---|---|---|---|---|---|---|
| **Main Variables (H1a)** | | | | | | | | |
| ln(Distance) | 407,497 | 4.49 | 2.72 | 0 | 3.89 | 5.56 | 6.54 | 8.00 |
| InvestmentDummy | 407,497 | 0.03 | 0.16 | 0 | 0 | 0 | 0 | 1 |
| **Main Variables (H1b, H2, H3)** | | | | | | | | |
| ln(Distance) | 11,054 | 3.97 | 2.81 | 0 | 0 | 4.90 | 6.50 | 8.00 |
| ln(InvestmentAmount) | 11,054 | 8.94 | 1.14 | 5.08 | 8.41 | 8.92 | 9.62 | 14.21 |
| ln(Likes) | 11,054 | 6.30 | 1.23 | 3.09 | 5.52 | 6.25 | 7.17 | 8.34 |
| Relationship | 11,054 | 0.72 | 0.45 | 0 | 0 | 1 | 1 | 1 |
| **Control Variables (H1b, H2, H3)** | | | | | | | | |
| ln(Experience) | 11,054 | 3.57 | 2.37 | 0.00 | 1.10 | 4.16 | 5.58 | 7.55 |
| Gender | 11,054 | 0.84 | 0.37 | 0 | 1 | 1 | 1 | 1 |
| ln(Goal) | 11,054 | 14.7 | 1.22 | 11.61 | 13.60 | 14.90 | 15.80 | 16.31 |
| InvestorAge | 11,054 | 34.25 | 8.72 | 16 | 28 | 33 | 39 | 89 |

*This table displays the descriptive statistics including all variables of the empirical analysis. N describes the number of observations which are included in the analyses. The variable $InvestorAge$ constitutes the base for creating the dummy variables used in the following analyses. Table A.3.7 in the Appendix displays the definitions of all variables.*

Looking at the control variables, 75% of the investors in our sample are equal or under 39 years of age. Furthermore, 84% are male. This is in line with previous studies on the typical characteristics of crowdfunding investors (Hervé et al. 2019). The average investor in our sample has been registered on the platform for 195 days (that is, approximately 0.5 years) before investing in a particular startup. The financing goal by startups ranges from 109,800 to 12,100,000 MXN (equals approximately 5,500 to 605,000 USD). Table A.3.2 of the Appendix displays detailed descriptive statistics for all variables.

Demir (2009) and Berre and Le Pendeven (2021) state that the general macroeconomic situation matters for investment decisions in crowdinvesting. Unreported results show that we find a significant effect of most macroeconomic variables within the startups' states of domicile whereas the macroeconomic situation in the investors' states does not seem to matter as much in investment decisions. However, these findings lead us to include fixed effects for both investor and startup states into our subsequent analyses to capture all cross-sectional and time-invariant characteristics on both sides.

Our observation period covers the years 2015 to 2019. The duration of startups' individual investment period ranges between 41 and 243 days with a mean of 126 days. Figure 3.6 exemplarily displays the distribution of five investment periods of different startups throughout the year 2015. We see that investment periods are differently long and can terminate within a calendar year or continue beyond. We split all individual investment periods of our sample (from the years 2015 to 2019) in quintiles, that is, five equivalently long time frames, to shed light on the investment timing. As shown in Table 3.2, we calculate the absolute number of investments (N) per quintile as well as the average investment amount per investor per startup (mean). We find a large number of crowdinvestors investing shortly before the end of the investment period. On average, the investment amount decreases in the mid of an investment period. Both findings suggest including fixed effects for investment timing into our further analyses to avoid biased results.

***Figure 3.6:*** *Example on the Distribution of Investment Periods*

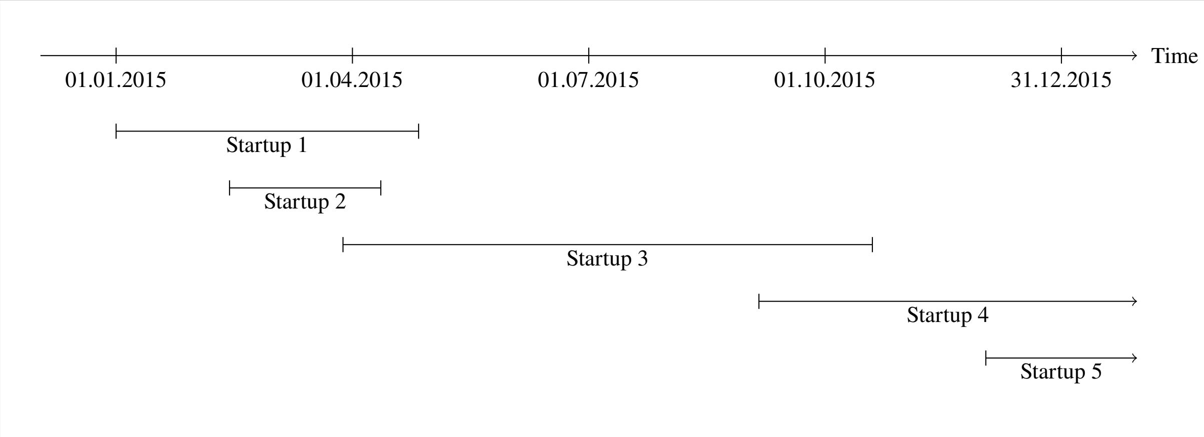



***Table 3.2:*** *Distribution of Investments per Startup Investment Period.*

| | N | Mean | StdDev | Min | p25 | p50 | p75 | Max |
|---|---|---|---|---|---|---|---|---|
| $1^{st}$ Quintile | 2,102 | 21,409 | 55,828 | 300 | 5,100 | 9,900 | 18,000 | 1,000,000 |
| $2^{nd}$ Quintile | 1,689 | 17,332 | 38,292 | 300 | 5,000 | 9,900 | 18,000 | 500,000 |
| $3^{rd}$ Quintile | 1,835 | 12,973 | 25,759 | 300 | 4,700 | 7,500 | 11,100 | 500,000 |
| $4^{th}$ Quintile | 2,102 | 12,690 | 26,549 | 300 | 4,000 | 7,500 | 10,200 | 600,000 |
| $5^{th}$ Quintile | 3,326 | 16,086 | 51,281 | 160 | 4,000 | 7,500 | 12,500 | 1,490,000 |
| Total | 11,054 | 16,126 | 43,139 | 160 | 4,500 | 7,500 | 15,000 | 1,490,000 |

*This table displays the distribution of the investment amount per quintile per startup investment period. The investment period is split into five equivalently long time frames. The sample covers all observations for which an investment has taken place.*

## 3.4.5 Research Design

### 3.4.5.1 Geographical Distance in Investment (Size) Decisions

Hypothesis 1a and Hypothesis 1b focus on the geographical distance between startups and (potential) investors. We expect that a greater distance between investors and startups negatively correlates with investors' decision to invest (Hypothesis 1a) as well as the investment amount (Hypothesis 1b). With these ideas, we follow several researchers in the field of finance and startup investments (Coval and Moskowitz 1999, Lutz et al. 2013; Niemand et al. 2018). For analyzing this relationship, we employ a logit regression followed by an OLS regression. We specify the logit regression as follows:

$$InvestmentDummy_{s,i} = \alpha + \beta_1 ln(Distance_{s,i}) + \rho_k + \tau_l + \epsilon_{s,i} \tag{3.1}$$

where, $InvestmentDummy_{s,i}$ in Equation (3.1) measures if an investor $i$ invests in startup $s$. In turn, we can derive the probability for startup investments via crowdinvesting given certain criteria. $ln(Distance_{s,i})$ is measured as the physical distance in kilometers between investor $i$ and startup $s$. We also include fixed effects for startup states $k$ ($\rho_k$) and investor states $l$ ($\tau_l$) to control for characteristics in the macroeconomic environment. Including fixed effects on a more fine-grained level, for example, the individual investor or startup level, would absorb the role of some explanatory variables (especially the moderating variables *Relationship* and *Likes* mentioned below) and cost heterogeneity, which is why we refrain from including these fixed effects in the model (Papke and Wooldridge 2023). Moreover, we would forfeit approximately a third of our observations in cases where only a single investment has occurred, thus diminishing the regression's power. Therefore, we specify the OLS regression as follows:

$$\begin{aligned} ln(InvestmentAmount_{s,i}) = \alpha &+ \beta_1 ln(Distance_{s,i}) \\ &+ \beta_2 ln(Experience_i) + \beta_3 Gender_i + \beta_4 ln(Goal_s) \\ &+ \sum_{j=1}^{6} \beta_{j+4} AgeGroup_{i,j} + \rho_k + \tau_l + \omega_m + \theta_t + \epsilon_{s,i} \end{aligned} \tag{3.2}$$

where, $ln(InvestmentAmount_{s,i})$ describes the investment amount of investor $i$ in startup $s$ in Mexican pesos. We add three investor characteristics to the regression equation to control for systematic differences in these regards. $ln(Experience_i)$ controls for the time investor $i$ has been registered on the crowdinvesting platform. $Gender_i$ takes the value of one for male investors and the value of zero for female investors. $AgeGroup_{i,j}$ measures the age of investor $i$ using six dummy variables for the different decades $j$ of age. The dummy variables are included for investors between 31 and 40, 41 and 50, 51 and 60, 61 and 70, 71 and 80, and over 80 years to account for biases due to differing individual financial needs. Investors under or equal to 30 years serve as the reference group. Moreover, we control for startups' financing goal using the variable $ln(Goals)$. In Equation (3.1), we do not include these control variables as we cannot calculate, e.g., investors' age at the time of investment, when no investment has taken place. We also include fixed effects for startup states $k$ ($\rho_k$) and investor states $l$ ($\tau_l$) to control for characteristics in the macroeconomic environment. In Equation (3.2), we add fixed effects for investment timing ($\omega_m$) and each specific month in our time period ($\theta_t$) to control for investment time-specific aspects.

#### 3.4.5.2 Investor-Startup-Relationship

The relationship between investors and startups constitutes a crucial factor with regard to the financial support provided. Several researchers find that social connections increase the probability of receiving external funding and, in turn, the chances of startup success (Angerer et al. 2017; Mollick 2014). Therefore, we expect in Hypothesis 2 that *an existing investor-startup relationship positively mitigates the role of geographical distance between investors and startups on the investment amount.* We analyze this by including relationship as a main coefficient and an interaction term between distance and relationship in our model.

#### 3.4.5.3 Startup Likes

We investigate the effects of signals, in this case, likes received on the crowdinvesting platform, on investment decisions. Startup likes are an observable signal by a third party about the perceived quality of a startup. Following Mochon et al. (2017) and Carr et al. (2018), we expect that *"the received likes positively mitigate the role of geographical distance between investors and startups on the investment amount"* (Hypothesis 3). Instead of the relationship, we now include a main coefficient for the received likes and an interaction term for likes and distance in our model.

## 3.5 Empirical Results

### 3.5.1 The Role of Geographical Distance

Column (1) of *Table 3.3* displays the regression results of a logit regression investigating the role of distance on the investment probability. Columns (2) and (3) depict the results of an OLS regression for distance and the investment amount. In line with our hypotheses and most studies in the research area of crowdfunding, we find a negative correlation between greater distance and the investment probability ($-0.1420$, p-value: $< 0.0001$) as well as the investment amount ($-0.0389$, p-value: 0.0003).

In contrast to researchers arguing that the rising use of the Internet diminishes the importance of geographic distance for investment decisions (Agrawal et al. 2015; Hornuf and Neuenkirch 2017), we find that distance still matters in crowdinvesting. A 10% increase in distance corresponds to a decrease in the investment probability by 1.42 pp and the investment amount by

***Table 3.3:*** *Link between Distance and Investment Decisions.*

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent Variable: | InvestmentDummy | ln(InvestmentAmount) | |
| ln(Distance) | -0.1420*** | -0.0433*** | -0.0389*** |
| | (0.0000) | (0.0001) | (0.0003) |
| ln(Experience) | | | -0.0189*** |
| | | | (0.0009) |
| Gender | | | 0.1980*** |
| | | | (<0.0001) |
| ln(Goal) | | | 0.4798*** |
| | | | (<0.0001) |
| AgeGroup 31-40 | | | 0.1736*** |
| | | | (<0.0001) |
| AgeGroup 41-50 | | | 0.3516*** |
| | | | (<0.0001) |
| AgeGroup 51-60 | | | 0.4728*** |
| | | | (<0.0001) |
| AgeGroup 61-70 | | | 0.7619*** |
| | | | (<0.0001) |
| AgeGroup 71-80 | | | 0.4795** |
| | | | (0.0375) |
| AgeGroup 81+ | | | 0.2332 |
| | | | (0.6718) |
| Constant | Yes | Yes | Yes |
| Investor-State FE | Yes | Yes | Yes |
| Startup-State FE | Yes | Yes | Yes |
| Investment-Timing FE | No | Yes | Yes |
| Time FE | No | Yes | Yes |
| Observations | 407,497 | 11,054 | 11,054 |
| (Pseudo) $R^2$ | 0.0241 | 0.3168 | 0.3899 |

*This table displays in Column (1) the results of a logit regression including all observations. Columns (2) and (3) display the results of OLS regressions which include only those observations in which an investment has taken place. Robust standard errors are clustered at the startup-state level. P-values are displayed in parentheses. ***, **, and *indicate significance at the 1%, 5%, and 10% level, respectively. Time fixed effects are included on a monthly level. Table A.3.7 of the Appendix displays the definitions of all variables.*

0.39%.[18] Both results are statistically significant at the 1% level. This result can be explained with two arguments. First, greater distance might lead to higher information asymmetries which result in higher uncertainty for investors. Investors with a larger distance to a startup are neither able to easily observe the startup's progress nor can they directly gather soft personal information on or by the startup's management. Therefore, investors further apart have to rely on hard information and informational reports created by the startup and published on the crowdinvesting platform. This hard information does not seem to have the same informative value for investors as soft information transferred with lower distance. The finding might be explained by the riskiness of startup investments which cannot easily be measured by hard facts (Carpenter and Petersen 2002). Second, since the invested amounts are relatively small on average and the cognitive effort might not be as high as for larger investments, investors might (unconsciously) engage in a cognitive bias, leading them to prefer startups located closer to themselves. A more positive attitude, triggered through the mere exposure effect, to the closer area could be transferred to the startups in this area and lead to a feeling of being more connected to them, supporting investors'

[18] Calculation of the investment probability: Marginal effects as displayed in *Derivation A.3.1* of the Appendix.

desire to invest. However, both explanations seem reasonable, and to get more clarity, the use of, for example, investor interviews is needed for more insights.

A longer time being registered on the crowdinvesting platform is significantly negatively correlated with the investment amount. We explain this finding by the idea that more experienced investors show lower risk appetite, are less overconfident, and diversify their portfolio more compared to less experienced investors (Goetzmann and Kumar 2008; Morin and Suarez 1983). These more experienced investors might be more cautious due to previous (unsuccessful) crowdinvesting activities and thus prefer to conduct a large number of small investments in different projects. Furthermore, we find that men invest, on average, 21.90% more compared to women.[19] As startup investments, and especially crowdinvesting activities, are usually characterized as a risky investment alternative, this is in line with Rieger et al. (2015), who state that men are less risk-averse than their female counterparts.

In addition, we find that the investment amount is higher if the startup has published a larger financing goal. A 10% increase in the financing goal is associated with a 4.80% increase in the individual investment amount. Moreover, the results show an inverted U-shape for the link between investors' age and the investment amount. This finding appears reasonable due to different stages in life with different financial bases and financial needs.

### 3.5.2 The Role of the Investor-Startup-Relationship

In the following step, we analyze the link of an existing relation between investors and startups on the actual amount provided by investors. As for the whole study, we focus on the first investment of each bilateral investor-startup combination. We distinguish between investors who are complete strangers to the startup, and investors who show some kind of relationship to the startup, that is, who are characterized as family/friends, customers, or who know the startup through an event or maker network in our dataset.

Hypothesis 2 suggests that an existing investor-startup relationship positively mitigates the role of distance between investors and startups on the investment amount. Table 3.4 displays the respective estimation results, depicting the coefficients from an OLS regression with and without the interaction of distance and an investor-startup relationship. As above, we find a significantly negative coefficient for the distance between investors and startups (−0.0349, p-value: 0.0007). The coefficient is of similar size as in the previous analyses. As expected, we find a positive main coefficient of relationships on the investment amount (0.0956, p-value: 0.0456). Investor-startup combinations with a relationship show an investment amount which is, on average, 10.03% higher compared to those without a relationship.[20] This result seems plausible as one can assume that investors with a relationship to a startup experience lower

[19] Calculation: $(e^{0.1980} - 1) * 100$.
[20] Calculation: $(e^{0.0956} - 1) * 100$.

***Table 3.4:*** *Link between Distance, the Investor-Startup-Relationship, and Investment Size.*

| Dependent Variable: ln(InvestmentAmount) | (1) | (2) |
|---|---|---|
| ln(Distance) | -0.0354*** | -0.0349*** |
| | (0.0005) | (0.0007) |
| Relationship | 0.0908*** | 0.0956** |
| | (0.0029) | (0.0456) |
| ln(Distance) × Relationship | | -0.0014 |
| | | (0.8886) |
| ln(Experience) | -0.0166*** | -0.0166*** |
| | (0.020) | (0.0018) |
| Gender | 0.2037*** | 0.2037*** |
| | (<0.0001) | (<0.0001) |
| ln(Goal) | 0.4779*** | 0.4779*** |
| | (<0.0001) | (<0.0001) |
| Age 31-40 | 0.1743*** | 0.1743*** |
| | (<0.0001) | (<0.0001) |
| Age 41-50 | 0.3542*** | 0.3544*** |
| | (<0.0001) | (<0.0001) |
| Age 51-60 | 0.4717*** | 0.4714*** |
| | (<0.0001) | (<0.0001) |
| Age 61-70 | 0.7353*** | 0.7351*** |
| | (<0.0001) | (<0.0001) |
| Age 71-80 | 0.4229** | 0.4233** |
| | (0.0393) | (0.0370) |
| Age 81+ | 0.2003 | 0.1999 |
| | (0.7116) | (0.7138) |
| Constant | Yes | Yes |
| Investor-State FE | Yes | Yes |
| Startup-State FE | Yes | Yes |
| Investment-Timing FE | Yes | Yes |
| Time FE | Yes | Yes |
| Observations | 11,054 | 11,054 |
| $R^2$ | 0.3910 | 0.3910 |

*This table displays the results of OLS regressions including observations in which an investment has taken place. Robust standard errors are clustered at the startup-state level. P-values are displayed in parentheses. ***, **, and *indicate significance at the 1%, 5%, and 10% level, respectively. Time fixed effects are included on a monthly level. Table A.3.7 of the Appendix displays the definitions of all variables.*

uncertainty due to previous interactions with the startup. Overall, our findings display that both distance and the relationship between startups and investors matter for the investment amount.

In contrast to our expectations, the interaction coefficient of distance and relationship appears nonsignificant (−0.0014, p-value: 0.8886). Thus, we can not say that investor-startup relationships counteract the negative main coefficient of distance. Even though it might seem counterintuitive at first glance, this finding is in line with the idea by Lee and Persson (2016). The authors claim that startups should not solely use funding by family and friends as, in times of trouble, “using family finance as risk capital undermines the preexisting familial insurance arrangement” (p. 2344). Hernandez-Trillo et al. (2005) and Hill et al. (2021) underscore this idea. They show that family and friends serve as a “type of insurance” or “safety net”, respectively, in financially difficult times.

We also find statistically significant results regarding investors' gender, experience, and all age groups except for the one with people aged 80+ years of age. The coefficients remain qualitatively similar as in the analyses above.

### 3.5.3 The Role of Startup Likes

In the next step, we analyze the role of signals on investment decisions via crowdinvesting. As one specific measurable signal, we employ the number of likes received, which is published on the crowdinvesting website. Block et al. (2018) show that likes influence funding decisions in startups. We assume a higher number of likes to be a positive signal for startups' future evolution and thus to reduce the negative role of physical distance on investment decisions. Consequently, Hypothesis 3 suggests that "the received likes positively mitigate the role of geographical distance between investors and startups on the investment amount".

Table 3.5 displays the respective estimation results. In columns (3) and (4), we find negative coefficients for a higher distance between investor and startup, and the investment amount. In contrast to our expectations that more likes lead to a higher investment amount, we find a statistically significant negative main coefficient (−0.0818, p-value: 0.0136). This is surprising at first sight. We explain this finding with the idea that projects with a high number of likes are seen as comparatively safe to be financed in the short run. Therefore, investors hop on to these projects with a small amount of money to benefit from investments by other investors. At the same time, these investors seem to keep risk diversification in mind. Thereby, they exploit the indirect support from other investors and can still split their total investment amount on several projects, reducing their individual risk.

The interaction coefficient between distance and likes counteracts the negative main coefficients of these variables. With a value of 0.0078, the interaction coefficient is small but statistically significant (p-value: 0.0451). However, it is still of economic significance. For the mean distance in our sample, a 10% increase in the number of likes is associated with a decrease in the investment amount of approximately 0.5%.[21] As in our previous analyses, the results with regard to investors' gender, age, and experience are still statistically significant and of similar size as above.

Columns (1) and (2) support our idea of an existing link between likes and investment decisions. We find a substantially larger main coefficient for likes (0.8093, p-value: <0.0001) compared to distance (−0.1457, p-value: $< 0.0001$). Thus, a higher number of likes positively correlate with the investment probability. This indicates that investors rely on other investors' assessments in their decision-making. This is in line with the literature (Coursaris et al. 2016). The positive main coefficient of likes persists when including the interaction between distance and likes. In contrast, the negative main coefficient of distance turns nonsignificant (−0.1137,

[21] Calculation: $[-0.0818 + 0.0078 * 3.9733)] * ln(1.1)$.

p-value: 0.2030) which displays that the interaction term captures part of the initial main effect. However, the interaction does not seem to explain sufficiently much of the effect on our dependent variable as the coefficient of this interaction term is also nonsignificant (0.0052, p-value: 0.7262). Summing up, our results suggest a link between physical distance, likes, and investment decisions.

***Table 3.5:*** *Link between Distance, Startup Likes, and Investment Decisions.*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable: | InvestmentDummy | | ln(InvestmentAmount) | |
| ln(Distance) | -0.1457*** | -0.1137 | -0.0384*** | -0.0864*** |
| | (<0.0001) | (0.2030) | (0.0003) | (0.0054) |
| ln(Likes) | 0.8093*** | 0.8284*** | -0.0543** | -0.0818** |
| | (<0.0001) | (<0.0001) | (0.0472) | (0.0136) |
| ln(Distance) × ln(Likes) | | -0.0052 | | 0.0078** |
| | | (0.7262) | | (0.0451) |
| ln(Experience) | | | -0.0186*** | -0.0187*** |
| | | | (0.0011) | (0.0009) |
| Gender | | | 0.1977*** | 0.1985*** |
| | | | (<0.0001) | (<0.0001) |
| ln(Goal) | | | 0.5271*** | 0.5254*** |
| | | | (<0.0001) | (<0.0001) |
| Age 31-40 | | | 0.1723*** | 0.1719*** |
| | | | (<0.0001) | (<0.0001) |
| Age 41-50 | | | 0.3498*** | 0.3477*** |
| | | | (<0.0001) | (<0.0001) |
| Age 51-60 | | | 0.4693*** | 0.4676*** |
| | | | (<0.0001) | (<0.0001) |
| Age 61-70 | | | 0.7605*** | 0.7567*** |
| | | | (<0.0001) | (<0.0001) |
| Age 71-80 | | | 0.4494** | 0.4397** |
| | | | (0.0137) | (0.0157) |
| Age 81+ | | | 0.2313 | 0.2239 |
| | | | (0.6729) | (0.6940) |
| Constant | Yes | Yes | Yes | Yes |
| Investor-State FE | Yes | Yes | Yes | Yes |
| Startup-State FE | Yes | Yes | Yes | Yes |
| Investment-Timing FE | No | No | Yes | Yes |
| Time FE | No | No | Yes | Yes |
| Observations | 407,497 | 407,497 | 11,054 | 11,054 |
| (Pseudo) $R^2$ | 0.0954 | 0.0954 | 0.3904 | 0.3909 |

*This table displays in Columns (1) and (2) the results of logit regressions including all observations. Columns (3) and (4) display the results of OLS regressions including only those observations in which an investment has taken place. Robust standard errors are clustered at the startup-state level. P-values are displayed in parentheses. ***, **, and *indicate significance at the 1%, 5%, and 10% level, respectively. Table A.3.7 of the Appendix displays the definitions of all variables.*

## 3.6 Robustness Tests

We first repeat the analyses from above only for observations with investors from outside Mexico City, which is by far the largest city of Mexico with approximately 22 million inhabitants.[22]

[22] The second and third largest cities of Mexico, that is, Guadalajara (Jalisco) and Monterrey (Nuevo Léon), show a population of approximately one quarter of the population of Mexico City each.

66.25% of our investors are from outside the capital city. Thereby, we account for differences in the general mindset of investors living in urban versus more rural areas. Bhayani et al. (2019, p. 3775) confirm that "there is considerable difference observed in investment behavior of urban and rural investors, with respect to the choice of investment." While the authors observe this fact within India, also for American investors it appears to exist (Copeland 2022). Sachan and Chugan (2020) go even further and explain that information processing is more limited for rural investors. Moreover, Chakravarty and Ahsan (2023) find that campaigns from global cities are more likely to be successful and receive more funding. Thus, it is of main importance to further investigate the differences in investment decisions between rural and urban investors, leading us to our first robustness test.

***Table 3.6:*** *Results Excluding Observations with Investors from Mexico-City.*

| Dependent Variable: | (1) Investment Dummy | (2) ln(Investment Amount) | (3) ln(Investment Amount) | (4) Investment Dummy | (5) ln(Investment Amount) |
|---|---|---|---|---|---|
| ln(Distance) | -0.2079*** | -0.0584*** | -0.0463** | -0.0983 | -0.1094*** |
| | (<0.0001) | (0.0003) | (0.0360) | (0.5942) | (0.0029) |
| Relationship | | | 0.1780 | | |
| | | | (0.2905) | | |
| ln(Likes) | | | | 0.8963*** | -0.1075*** |
| | | | | (<0.0001) | (0.0085) |
| ln(Distance) × Relationship | | | -0.0118 | | |
| | | | (0.6475) | | |
| ln(Distance) × ln(Likes) | | | | -0.0197 | 0.0088 |
| | | | | (0.5266) | (0.1373) |
| Investor Controls | No | Yes | Yes | No | Yes |
| Startup Controls | No | Yes | Yes | No | Yes |
| Constant | Yes | Yes | Yes | Yes | Yes |
| Investor-State FE | Yes | Yes | Yes | Yes | Yes |
| Startup-State FE | Yes | Yes | Yes | Yes | Yes |
| Investment-Timing FE | No | Yes | Yes | No | Yes |
| Time FE | No | Yes | Yes | No | Yes |
| Observations | 268,747 | 7,385 | 7,385 | 268,747 | 7,385 |
| (Pseudo) $R^2$ | 0.0265 | 0.4137 | 0.4154 | 0.0932 | 0.4146 |

*This table includes only observations from investors outside of Mexico-City. Columns (1) and (4) display the results of a logit regression with the investment probability as the dependent variable. Columns (2), (3) and (5) show the results of an OLS regression including only those observations where an investment has taken place. The dependent variable for these columns is the exact investment amount. The independent variable distance serves as the proxy for information asymmetries, and likes measure the number of likes a startup has received on the platform. Macroeconomic differences are captured by startup- and investor-state fixed effects, differences in investment decisions due to time by monthly and investment timing fixed effects. Robust standard errors are clustered at the startup state level. P-values are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and the 10% level, respectively. Table A.3.7 of the Appendix displays the definitions of all variables.*

As above, we find negative coefficients for distance in Table 3.6, showing that a higher distance decreases the investment probability as well as the investment amount. In contrast to the results above, we find a different importance of relationships in this sample. Whereas in the case when including all investors from inside and outside Mexico City the main coefficient of the dummy variable for relationship is significantly positive, it turns nonsignificant when reconducting the analyses for investors only from outside Mexico City. Several studies show that people living in rural areas give higher weight to the bonding to family and friends (House et al. 1988; Sørensen

2016). We have in mind that it is recommendable for early stage firms to keep family and friends to the greatest extent possible as fallback option when following risky business ideas (Hernandez-Trillo et al. 2005). While we find that investors from rural areas seem to be aware of and consider this concept in their decision on the investment amount, investors from urban areas seem to ignore this argument. The main coefficients for likes on the investment probability and investment amount, are similar to above. We again find a significantly positive coefficient for likes on the investment probability and a significantly negative coefficient for the investment amount. The interaction term in column (5) of Table 3.6 turns nonsignificant but remains of same size and direction. This finding has to be interpreted with caution but hints toward the idea that likes are more important for urban investors compared to investors living in rural areas.

In the following, we further investigate the role of a higher distance between investors and startups. For this purpose, we exclude investments within the same state.[23] In Table A.3.4, we find more negative coefficients for distance compared to the analyses above. This seems reasonable when excluding observations with zero distance between startup and investor. Interestingly, the main coefficient for relationship turns from significantly positive to significantly negative. For investors who are not located in the same state as the investee, having a relationship to the startup negatively correlates with the investment amount. This finding underscores the idea that these investors are aware of the thought that related parties should not be the exclusive supporters of an affiliated startup, but rather be kept as fallback options as far as possible (Hernandez-Trillo et al. 2005). In contrast, investors from the same state as the startup, do not seem to keep this idea in mind. The interaction term of distance and relationship changes from nonsignificantly negative to significantly positive. This change seems reasonable for information purposes and uncertainty reduction. If both parties are located in states which are far apart from each other, the information flow is not as easy as in the case of spatial proximity. A relationship could decrease information asymmetries or uncertainty regarding the startup's quality and allows a smoother information flow. In particular, relationships offset opportunity costs for seeking information (Agrawal et al. 2008). The main coefficients for likes are similar as in the main analyses. The coefficient of the interaction term turns again nonsignificant, but remains of the same direction as above.

Instead of including the natural logarithm of the investor-startup-distance into the regression analyses, we now include dummy variables for each quintile as independent variables. The lowest quintile, which serves as reference group in the analysis, contains only zero-distance observations.[24] The highest quintile comprises all investor-startup combinations with a distance of equal to or more than 897 km. Thereby, we strive to look at which distances exactly matter in investment decisions.[25]

---

[23] Unreported evidence shows that 87.63% of these zero-distance-observations describe investments within Mexico City. This equals 95.04% of the observations in our larger sample for the analysis of Hypothesis 1a.

[24] It is alike in the smaller sample in Hypothesis 1b, Hypothesis 2, and Hypothesis 3 with 3,387 observations in the lowest quintile.

[25] Table A.3.2 of the Appendix displays further details on the descriptive statistics.

Table A.3.5 of the Appendix confirms that a greater distance is associated with a decrease in the investment probability and investment amount. Compared to investor-startup combinations from the lowest-distance quintile, the probability to invest decreases by 1.15 pp[26] and the investment amount decreases by 44.83%[27] for investors with the largest distance to their investees. The results show that there is no threshold above which an increase in distance does not influence the investment decision more negatively anymore. This contradicts Zook (2002) who states that investors do not invest if they cannot drive to the startup. They find this limit in the investor-startup distance to be approximately 100 km. However, we cannot confirm this idea as coefficients become more negative with increasing quintiles. Compared to the main analyses, the coefficients measuring the role of distance appear relatively large. This can be explained by the fact that the dummy variables in this part of the analyses measure the correlation compared to the base group of investor-startup pairs with a very low distance.

Moreover, likes received on a platform and relationship might be endogenous variables. To alleviate this concern, we conduct an instrumental variable-free approach, namely the Gaussian copula method, since this approach offers advantageous features (Hult et al. 2018). Following Papies et al. (2017), this approach allows for a statistical test for the existence of endogeneity problems. One key requirement is that the potentially endogenous variable has to pass a non-normality test (Park and Gupta 2012). We use the Shapiro–Wilk to test for this nonnormality which is in line with previous literature and the most commonly test used by researchers in this regard (Becker et al. 2022). Unreported results show that Likes is nonnormally distributed and the Gaussian copula approach can be applied. For Relationship as a binary variable, the distribution is also nonnormal. The inclusion of the Gaussian copula in both regressions leads to a nonsignificant coefficient of the copula term, which indicates that an endogeneity problem should not exist (Papies et al. 2017). The main coefficients of the relationship and likes remain of similar size and significance as before, strengthening our main results.

Last, we investigate if the level of social proximity within a relationship matters for investment decisions. We now qualify only those investors as having a relationship to a startup when they are marked as “family/friends” in the platform’s managerial system. Investors knowing the startup as customers, from a fair, or through a maker network are now considered as having no relationship, that is, being strangers to the startup. Crowdinvesting describes a form of financing which startups usually choose in a very early development stage. In this stage, they often use mainly family and friends as financing source (Angerer et al. 2017) even though it might not be optimal (Lee and Persson 2016). As Table A.3.6 displays, we find a negative main coefficient for distance and a positive main coefficient for relationship. While the size and significance of the coefficient of distance stay similar as in the main analyses above, the main coefficient for the relationship strongly increases, indicating that a close relationship correlates with a high individual investment amount.

[26] Calculation: Marginal effects as displayed in *Derivation A.3.1* of the Appendix.

[27] Calculation: $(e^{-0.5948} - 1) * 100$.

## 3.7 Possible Limitations and Implications for Future Research

Several aspects constitute possible limitations to our analyses and call for future research. There are limitations regarding the dataset. The analyses are based on data stemming from one individual crowdinvesting platform which depicts the largest one in Mexico. This might lead to a self-selection bias with regard to the fact that this specific platform might attract certain participants of the financial sector more than others. In addition, our sample, by construction, only constitutes of investors having access to the Internet and being open-minded regarding innovative investment opportunities. However, the large size of our platform enables us to exploit a large heterogeneity in our variables of interest. Furthermore, we control for investor-specific and startup-specific characteristics.

Moreover, we know about potential endogeneity within our data. Endogeneity among the observations with regard to the investment probability, the investment amount as well as our independent variables might lead to biased results. Exemplarily, this might be the risk-return profile of a particular startup, the time horizon of startups' dividend payments, or state-specific regulation of the financial sector. We control for this issue by including startup-state and investor-state fixed effects, as well as time and investment timing fixed effects. In addition, by conducting the Gaussian copula approach, we empirically test that no endogeneity problem is present. The idea of measuring investor experience using investors' time being registered on the platform might also not be the ideal proxy. We argue that crowdinvesting activities are not comparable to any other equity investment alternative, which leaves space for future research, employing other proxies to measure investor experience.

One minor argument might be related to Goethner et al. (2021) who show that different factors and signals influence crowdinvestors differently. Hence, they claim that it is difficult to measure the impact of a certain signal on a heterogeneous group of investors. This concern is not of major importance in our study as our database contains in-depth information on investors as well as startups. However, this topic leaves space for future research following Piva and Rossi-Lamastra (2018) who call for deeper analyses with regard to the effect of signals on different types of investors.

## 3.8 Conclusion

In this study, we compile detailed information on investor and startup characteristics to reveal the factors influencing investment decisions in crowdinvesting. Specifically, we look at the effect of distance between investors and startups on the investment probability and investment amount. We investigate the moderating role of relationships and signals in this context. Recent studies show that the physical distance correlates with a lower investment probability and a lower investment amount (e.g., Bae et al. 2008; Dejean 2020). In crowdinvesting, information costs are comparatively high as individual investments are usually small and little contact between investors and startups precedes the investment. In those cases where investor and startup already have some kind of relationship (for example, through connections as family or friends, through meetings at a fair, and so on), uncertainty decreases, which might lead investors to support startups in a different way (Angerer et al. 2017; Mollick 2014). Furthermore, objectively observable signals which a startup has received on a platform, influence investors' decision-making (Courtney et al. 2017).

Our sample consists of 407,497 investor-startup combinations including 11,054 investments. We use data from the largest Mexican crowdinvesting platform and investigate the characteristics associated with investors' decision-making regarding their distance to the startup, relationships, and signals. First, we find that distance correlates negatively with the investment probability and the investment amount. A 10% increase in distance is associated with a decrease in the investment probability by 1.42 pp and the investment amount by 0.39%. In contrast, lower investor experience, male gender, and a higher financing goal correlate positively with the individual investment amount.

Second, Investors with a relationship to the startup, on average, invest more than those not having any kind of connection to the investee. However, relationships are not able to counteract the negative main role of higher distance in any case. While rural investors seem to be aware of the idea of startups keeping people with a relationship, especially family and friends, as fallback options for the case of a negative development of the startup (Hernandez-Trillo et al. 2005; Lee and Persson 2016), investors from urban areas do not seem to consider this idea. In the final part of this paper, we measure signals through likes on the crowdinvesting platform. We observe that signals positively correlate with the investment probability but have a significantly negative role on the investment amount. This is in line with the idea that private investors care about the opinion from other investors but also consider the concept of risk diversification.

Investigating the role of physical distance, relationships and signals on investment decisions is of main importance for researchers, politicians as well as all participants of the financial sector. Crowdinvesting constitutes a more and more important investment alternative for nonprofessional investors. With an expected increase in market transactions of 4–5% in the upcoming years in the U.S., Mexico, and Germany, crowdinvesting is currently experiencing a boom in its evolution. A few years ago, policymakers identified crowdfunding as a possibility to overcome geographic boundaries (Guenther et al. 2018). Given the currently rising inflation rates and still low interest rates on traditional investment opportunities, traditional equity investments and also nontraditional investment opportunities, for example, crowdinvesting, has become an almost inevitable approach for private investors to think about when deciding their investment alternatives. As equity investments occupy a more and more relevant part of the financial market, understanding its functionality and the factors influencing its development is of high interest for all participants of the financial system.

In addition, it is widely known that family and friends serve as first financial support in the very early stage of startups' existence (Angerer et al. 2017). However, in order to receive sufficient funding for future development, other financing sources should be taken into account (Lee and Persson 2016). Understanding the interplay between social and geographical distance and the investor-startup-relationship is thus of main importance for startups. Based on our findings, startups should be mindful of geographical factors that may influence their audience's behavior. It appears advantageous for startups to establish a presence in the minds of investors. Regular exposure to a startup can positively shape investors' perceptions and potentially increase their willingness to invest. If startups aim to expand and attract funding from investors in distant locations with differing cultural or social norms, employing an offensive advertising strategy in these regions could effectively introduce the startup and foster a favorable impression. This approach also helps mitigate information gaps by educating potential investors about the startup. Tailored advertising strategies, readily accessible through platforms like social media, are predestined for this purpose. Furthermore, startups that cultivate a sense of community may experience accelerated growth. When investors feel a strong connection to the brand and its supporters, demonstrated through engagement on social media or website interactions, it may attract new investors influenced by existing investors' behavior. Consequently, among other factors, the presentation of the startup itself, which is multidimensional (as seen in Ahsan et al. 2018), emerges as a crucial factor. However, further analyses and insights from disciplines such as marketing, psychology, and social science are required to fully understand the mechanics of this sense of "belonging."

# A Appendices

## Tables

***Table A.3.1:*** *Local Distribution of Investors and Startups.*

| | | Initial Sample | | | | Analyses | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Code acc. | Investors | | Startups | | Investors | | Startups | |
| State | ISO 3166-2 | Freq. | % | Freq. | % | Freq. | % | Freq. | % |
| Aguascalientes | AGU | 82 | 1.34 | 1 | 0.99 | 82 | 1.35 | 1 | 0.99 |
| Baja California | BCN | 37 | 0.60 | 0 | 0.00 | 37 | 0.61 | 0 | 0.00 |
| Baja California Sur | BCS | 25 | 0.41 | 0 | 0.00 | 25 | 0.41 | 0 | 0.00 |
| Campeche | CAM | 24 | 0.39 | 0 | 0.00 | 23 | 0.38 | 0 | 0.00 |
| Chiapas | CHP | 35 | 0.57 | 0 | 0.00 | 35 | 0.58 | 0 | 0.00 |
| Chihuahua | CHH | 110 | 1.80 | 0 | 0.00 | 110 | 1.81 | 0 | 0.00 |
| Coahuila de Zaragoza | COA | 70 | 1.14 | 0 | 0.00 | 70 | 1.15 | 0 | 0.00 |
| Colima | COL | 38 | 0.62 | 0 | 0.00 | 38 | 0.63 | 0 | 0.00 |
| Distrito Federal (Mexico City)[28] | CMX | 2,060 | 33.66 | 61 | 60.40 | 2,047 | 33.75 | 61 | 60.40 |
| Durango | DUR | 31 | 0.51 | 0 | 0.00 | 31 | 0.51 | 0 | 0.00 |
| Estado de México | MEX | 1,005 | 16.42 | 0 | 0.00 | 990 | 16.32 | 0 | 0.00 |
| Guanajuato | GUA | 185 | 3.02 | 9 | 8.91 | 183 | 3.02 | 9 | 8.91 |
| Guerrero | GRO | 36 | 0.59 | 0 | 0.00 | 36 | 0.59 | 0 | 0.00 |
| Hidalgo | HID | 84 | 1.37 | 0 | 0.00 | 82 | 1.35 | 0 | 0.00 |
| Jalisco | JAL | 478 | 7.81 | 6 | 5.94 | 473 | 7.80 | 6 | 5.94 |
| Michoacán de Ocampo | MIC | 88 | 1.44 | 1 | 0.99 | 88 | 1.45 | 1 | 0.99 |
| Morelos | MOR | 64 | 1.05 | 0 | 0.00 | 64 | 1.06 | 0 | 0.00 |
| Nayarit | NAY | 14 | 0.23 | 0 | 0.00 | 14 | 0.23 | 0 | 0.00 |
| Nuevo León | NLE | 458 | 7.48 | 5 | 4.95 | 454 | 7.48 | 5 | 4.95 |
| Oaxaca | OAX | 14 | 0.23 | 0 | 0.00 | 14 | 0.23 | 0 | 0.00 |
| Puebla | PUE | 182 | 2.97 | 3 | 2.97 | 181 | 2.98 | 3 | 2.97 |
| Querétaro de Arteaga | QUE | 241 | 3.94 | 1 | 0.99 | 240 | 3.96 | 1 | 0.99 |
| Quintana Roo | ROO | 92 | 1.50 | 6 | 5.94 | 90 | 1.48 | 6 | 5.94 |
| San Luis Potosí | SLP | 86 | 1.41 | 2 | 1.98 | 86 | 1.42 | 2 | 1.98 |
| Sinaloa | SIN | 65 | 1.06 | 1 | 0.99 | 65 | 1.07 | 1 | 0.99 |
| Sonora | SON | 97 | 1.58 | 0 | 0.00 | 97 | 1.60 | 0 | 0.00 |
| Tabasco | TAB | 29 | 0.47 | 1 | 0.99 | 29 | 0.48 | 1 | 0.99 |
| Tamaulipas | TAM | 73 | 1.19 | 0 | 0.00 | 72 | 1.19 | 0 | 0.00 |
| Tlaxcala | TLA | 35 | 0.57 | 0 | 0.00 | 33 | 0.54 | 0 | 0.00 |
| Veracruz | VER | 150 | 2.45 | 0 | 0.00 | 145 | 2.39 | 0 | 0.00 |
| Yucatán | YUC | 111 | 1.81 | 4 | 3.94 | 111 | 1.83 | 4 | 3.94 |
| Zacatecas | ZAC | 21 | 0.34 | 0 | 0.00 | 21 | 0.35 | 0 | 0.00 |
| Total | | 6,120 | 100.00 | 101 | 100.00 | 6,066 | 100.00 | 101 | 100.00 |

*This table displays the local distribution of startups and investors throughout the Mexican states for the initial sample (618,120 observations) and the sample used within the analyses (407,497 observations).*

[28] "Distrito Federal" is synonymous with the capital city of Mexico (Mexico City; Ciudad de México). It is an individual federal entity and does not belong to any of the 31 Mexican states. Thus, we treat it as an individual state.

***Table A.3.2:*** *Detailed Descriptive Statistics.*

| | N | Mean | StdDev | Min | p25 | p50 | p75 | Max |
|---|---|---|---|---|---|---|---|---|
| **Main Variables (H1a)** | | | | | | | | |
| ln(Distance) | 407,497 | 4.49 | 2.72 | 0.00 | 3.89 | 5.56 | 6.54 | 8.00 |
| InvestmentDummy | 407,497 | 0.03 | 0.16 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| **Main Variables (H1b, H2, H3)** | | | | | | | | |
| ln(Distance) | 11,054 | 3.97 | 2.81 | 0.00 | 0.00 | 4.90 | 6.50 | 8.00 |
| ln(InvestmentAmount) | 11,054 | 8.94 | 1.14 | 5.08 | 8.41 | 8.92 | 9.62 | 14.21 |
| ln(Likes) | 11,054 | 6.30 | 1.23 | 3.09 | 5.52 | 6.25 | 7.17 | 8.34 |
| Relationship | 11,054 | 0.72 | 0.45 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 |
| **Control Variables** | | | | | | | | |
| ln(Experience) | 11,054 | 3.57 | 2.37 | 0.00 | 1.10 | 4.16 | 5.58 | 7.55 |
| Gender | 11,054 | 0.84 | 0.37 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| ln(Goal) | 11,054 | 14.70 | 1.22 | 11.61 | 13.60 | 14.90 | 15.80 | 16.31 |
| InvestorAge | 11,054 | 34.25 | 8.72 | 16.00 | 28.00 | 33.00 | 39.00 | 89.00 |
| **Variables Included in the Analyses but without the Natural Logarithm** | | | | | | | | |
| Distance | 407,497 | 451.66 | 491.02 | 0.00 | 48.00 | 258.00 | 689.00 | 2987.00 |
| Distance | 11,054 | 355.26 | 441.21 | 0.00 | 0.00 | 133.00 | 663.00 | 2987.00 |
| InvestmentAmount | 11,054 | 16,126 | 43,139 | 160 | 4,500 | 7,500 | 15,000 | 1,490,000 |
| Likes | 11,054 | 1,081 | 1,295 | 21.00 | 249.00 | 515.00 | 1,305.00 | 4,194.00 |
| Experience | 11,054 | 195.05 | 292.05 | 0.00 | 2.00 | 63.00 | 264.00 | 1,898.00 |
| Goal | 11,054 | 4,163,737 | 3,859,619 | 109,800 | 800,000 | 3,000,000 | 7,000,000 | 12,100,000 |
| **Information on the Variable $AgeGroup_j$** | | | | | | | | |
| $AgeGroup_1$ (InvestorAge $\leq$ 30) | 4325 | 26.50 | 2.73 | 16.00 | 25.00 | 27.00 | 29.00 | 30.00 |
| $AgeGroup_2$ (31 < InvestorAge $\leq$ 40) | 4408 | 34.81 | 2.75 | 31.00 | 32.00 | 35.00 | 37.00 | 40.00 |
| $AgeGroup_3$ (41 < InvestorAge $\leq$ 50) | 1742 | 44.46 | 2.81 | 41.00 | 42.00 | 44.00 | 47.00 | 50.00 |
| $AgeGroup_4$ (51 < InvestorAge $\leq$ 60) | 467 | 54.32 | 2.78 | 51.00 | 52.00 | 54.00 | 56.00 | 60.00 |
| $AgeGroup_5$ (61 < InvestorAge $\leq$ 70) | 95 | 63.58 | 2.41 | 61.00 | 62.00 | 63.00 | 65.00 | 70.00 |
| $AgeGroup_6$ (71 < InvestorAge $\leq$ 80) | 9 | 76.22 | 3.38 | 71.00 | 74.00 | 77.00 | 79.00 | 79.00 |
| $AgeGroup_7$ (InvestorAge $\geq$ 81) | 8 | 85.88 | 1.55 | 84.00 | 85.00 | 85.00 | 86.00 | 89.00 |
| **Information on the quintiles of the Variable $Distance$ (H1a)** | | | | | | | | |
| $Distance_{Q1}$ | 99,450 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $Distance_{Q2}$ | 67,099 | 71.56 | 36.43 | 29.00 | 48.00 | 48.00 | 111.00 | 133.00 |
| $Distance_{Q3}$ | 96,918 | 345.60 | 123.17 | 141.00 | 250.00 | 320.00 | 495.00 | 495.00 |
| $Distance_{Q4}$ | 62,791 | 687.18 | 76.84 | 498.00 | 683.00 | 689.00 | 710.00 | 893.00 |
| $Distance_{Q5}$ | 81,239 | 1262.99 | 319.36 | 897.00 | 1,045.00 | 1,170.00 | 1,350.00 | 2,987.00 |
| **Information on the quintiles of the Variable $Distance$ (H1b, H2, H3)** | | | | | | | | |
| $Distance_{Q1}$ | 3,387 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $Distance_{Q2}$ | 2,153 | 68.03 | 34.37 | 48.00 | 48.00 | 48.00 | 66.00 | 133.00 |
| $Distance_{Q3}$ | 2,317 | 337.20 | 123.06 | 141.00 | 218.00 | 320.00 | 495.00 | 495.00 |
| $Distance_{Q4}$ | 1,786 | 682.44 | 72.92 | 508.00 | 671.00 | 689.00 | 689.00 | 893.00 |
| $Distance_{Q5}$ | 1,411 | 1261.87 | 323.67 | 897.00 | 1,045.00 | 1,170.00 | 1,350.00 | 2,987.00 |

*This table displays detailed descriptive statistics including all variables of the empirical analysis. N describes the number of observations. The variable $InvestorAge$ constitutes the base for creating the dummy variables $AgeGroup_j$ included in the empirical analysis.*

***Table A.3.3:*** *The Link between Distance and Investment Decisions.*

| Dependent Variable: ln(InvestmentAmount) | (1) |
|---|---|
| ln(Distance) | -0.0393*** |
| | (0.0002) |
| ln(Experience) | -0.0201*** |
| | (0.0007) |
| Gender | 0.1963*** |
| | (<0.0001) |
| ln(Goal) | 0.4769*** |
| | (<0.0001) |
| Age | 0.0369*** |
| | (<0.0001) |
| $Age^2$ | -0.0002*** |
| | (0.0109) |
| Constant | Yes |
| Investor-State FE | Yes |
| Startup-State FE | Yes |
| Investment-Timing FE | Yes |
| Time FE | Yes |
| Observations | 11,054 |
| $R^2$ | 0.3911 |

*This table displays the results including Age and Age-squared as control variables (instead of including investors' age as categorical variable through the different variables for $AgeGroup_j$). It displays the results of an OLS regression including observations in which an investment has taken place. Robust standard errors are clustered at the startup-state level. P-values are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. Time fixed effects are included on monthly level.*

***Table A.3.4:*** *Results Excluding Observations with Investors and Startups from the Same State.*

| Dependent Variable: | (1) Investment Dummy | (2) ln(Investment Amount) | (3) ln(Investment Amount) | (4) Investment Dummy | (5) ln(Investment Amount) |
|---|---|---|---|---|---|
| ln(Distance) | -0.2010*** | -0.0876** | -0.0974** | 0.1430 | -0.1315** |
| | (<0.0001) | (0.0179) | (0.0125) | (0.4788) | (0.0189) |
| Relationship | | | -0.2798*** | | |
| | | | (0.0058) | | |
| ln(Distance) × Relationship | | | 0.0605*** | | |
| | | | (0.0051) | | |
| ln(Likes) | | | | 1.1026*** | -0.0793* |
| | | | | (<0.0001) | (0.0975) |
| ln(Distance) × ln(Likes) | | | | -0.0550 | 0.0069 |
| | | | | (0.1074) | (0.3718) |
| Investor Controls | No | Yes | Yes | No | Yes |
| Startup Controls | No | Yes | Yes | No | Yes |
| Constant | Yes | Yes | Yes | Yes | Yes |
| Investor-State FE | Yes | Yes | Yes | Yes | Yes |
| Startup-State FE | Yes | Yes | Yes | Yes | Yes |
| Investment-Timing FE | No | Yes | Yes | No | Yes |
| Time FE | No | Yes | Yes | No | Yes |
| Observations | 308,047 | 7,667 | 7,667 | 308,047 | 7,667 |
| (Pseudo) $R^2$ | 0.0271 | 0.4068 | 0.4078 | 0.0924 | 0.4072 |

*This table includes only observations in which investors and startups are from different states. Columns (1) and (4) display the results of logit regressions including all observations meeting the criteria. Columns (2), (3), and (5) display the results of OLS regressions including only those observations in which an investment has taken place. Robust standard errors are clustered at the startup-state level. P-values are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. Time fixed effects are included on a monthly level.*

***Table A.3.5:*** *Results Dividing Distance Into 20%-Quantiles.*

| Dependent Variable: | (1) Investment Dummy | (2) ln(Investment Amount) | (3) ln(Investment Amount) | (4) Investment Dummy | (5) ln(Investment Amount) |
|---|---|---|---|---|---|
| $\ln(\text{Distance})_{Q2}$ | -0.3253*** | -0.0378 | 0.0125 | -0.8707*** | 0.0425 |
| | (0.0001) | (0.6451) | (0.8833) | (0.0054) | (0.7996) |
| $\ln(\text{Distance})_{Q3}$ | -0.8168*** | -0.1095 | -0.0671 | -0.3879 | -0.6094*** |
| | (<0.0001) | (0.1318) | (0.3992) | (0.1803) | (0.0002) |
| $\ln(\text{Distance})_{Q4}$ | -0.7958*** | -0.2252*** | -0.1866*** | -1.1310 | -0.5033*** |
| | (<0.0001) | (<0.0001) | (0.0003) | (0.3381) | (0.0009) |
| $\ln(\text{Distance})_{Q5}$ | -1.1528*** | -0.3789*** | -0.3463*** | -0.5027 | -0.5948*** |
| | (<0.0001) | (0.0021) | (0.0026) | (0.2228) | (0.0002) |
| Relationship | | | 0.1225** | | |
| | | | (0.0261) | | |
| $\ln(\text{Distance})_{Q2} \times$ Relationship | | | -0.1250** | | |
| | | | (0.0182) | | |
| $\ln(\text{Distance})_{Q3} \times$ Relationship | | | -0.0890 | | |
| | | | (0.3103) | | |
| $\ln(\text{Distance})_{Q4} \times$ Relationship | | | -0.0359 | | |
| | | | (0.7169) | | |
| $\ln(\text{Distance})_{Q5} \times$ Relationship | | | -0.0368 | | |
| | | | (0.7100) | | |
| ln(Likes) | | | | 0.8096*** | -0.0760*** |
| | | | | (<0.0001) | (0.0013) |
| $\ln(\text{Distance})_{Q2} \times$ ln(Likes) | | | | 0.0854** | -0.0102 |
| | | | | (0.0453) | (0.6523) |
| $\ln(\text{Distance})_{Q3} \times$ ln(Likes) | | | | -0.0742* | 0.0836*** |
| | | | | (0.0855) | (0.0006) |
| $\ln(\text{Distance})_{Q4} \times$ ln(Likes) | | | | 0.0499 | 0.0453** |
| | | | | (0.7908) | (0.0441) |
| $\ln(\text{Distance})_{Q5} \times$ ln(Likes) | | | | -0.1160* | 0.0376 |
| | | | | (0.0625) | (0.1242) |
| Investor Controls | No | Yes | Yes | No | Yes |
| Startup Controls | No | Yes | Yes | No | Yes |
| Constant | Yes | Yes | Yes | Yes | Yes |
| Investor-State FE | Yes | Yes | Yes | Yes | Yes |
| Startup-State FE | Yes | Yes | Yes | Yes | Yes |
| Investment-Timing FE | No | Yes | Yes | No | Yes |
| Time FE | No | Yes | Yes | No | Yes |
| Observations | 407,497 | 11,054 | 11,054 | 407,497 | 11,054 |
| (Pseudo) $R^2$ | 0.0244 | 0.3909 | 0.3923 | 0.0962 | 0.3924 |

*This table includes the variable for measuring information asymmetries using quintiles instead of a discrete variable. Columns (1) and (4) display the results of logit regressions including all observations. Columns (2), (3), and (5) display the results of OLS regressions including only those observations in which an investment has taken place. Robust standard errors are clustered at the startup-state level. P-values are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. Time fixed effects are included on monthly level.*

***Table A.3.6:*** *Results Only Qualifying Family as not Being Strangers to the Startup.*

| Dependent Variable: ln(InvestmentAmount) | (1) |
|---|---|
| ln(Distance) | -0.0300*** |
| | (0.0008) |
| Relationship | 0.3593*** |
| | (0.0009) |
| ln(Distance) × Relationship | 0.0115 |
| | (0.4800) |
| Investor Controls | Yes |
| Startup Controls | Yes |
| Constant | Yes |
| Investor-State FE | Yes |
| Startup-State FE | Yes |
| Investment-Timing FE | Yes |
| Time FE | Yes |
| Observations | 11,054 |
| $R^2$ | 0.3966 |

*This table characterizes only family and friends as not being strangers to the startup. It displays the results of an OLS regression including observations in which an investment has taken place. Robust standard errors are clustered at the startup-state level. P-values are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. Time fixed effects are included on monthly level.*

## Derivations

***Derivation A.3.1:*** *Marginal Effects*

*This part displays the mathematical approach on the calculation of marginal effects. As the dependent variable in logit regressions is a dummy variable, one must interpret the coefficients using marginal effects.*

As a starting point, we use the general formulation of a logistic regression with x being the log-specified regressor of interest and **z** being a vector of covariates:

$$Pr[y=1 \mid x,\mathbf{z}] = p = \frac{e^{\alpha+\beta\ln x+\mathbf{yz}}}{1+e^{\alpha+\beta\ln x+\mathbf{yz}}}$$

Taking the partial derivative with respect to x, we obtain

$$\begin{aligned}
\frac{\partial Pr[y=1 \mid x,\mathbf{z}]}{\partial x} &= \frac{\frac{\beta}{x} * e^{\alpha+\beta\ln x+\mathbf{yz}} * (1+e^{\alpha+\beta\ln x+\mathbf{yz}}) - e^{\alpha+\beta\ln x+\mathbf{yz}} * \frac{\beta}{x} * e^{\alpha+\beta\ln x+\mathbf{yz}}}{(1+e^{\alpha+\beta\ln x+\mathbf{yz}})^2} \\
&= \frac{\frac{\beta}{x} * e^{\alpha+\beta\ln x+\mathbf{yz}} * (1+e^{\alpha+\beta\ln x+\mathbf{yz}})}{1+e^{\alpha+\beta\ln x+\mathbf{yz}})^2} - \frac{e^{\alpha+\beta\ln x+\mathbf{yz}} * \frac{\beta}{x} * e^{\alpha+\beta\ln x+\mathbf{yz}}}{(1+e^{\alpha+\beta\ln x+\mathbf{yz}})^2} \\
&= \frac{\beta}{x} * \frac{e^{\alpha+\beta\ln x+\mathbf{yz}}}{(1+e^{\alpha+\beta\ln x+\mathbf{yz}})} - \frac{\beta}{x} * \frac{(e^{\alpha+\beta\ln x+\mathbf{yz}})^2}{(1+e^{\alpha+\beta\ln x+\mathbf{yz}})^2} \\
&= \frac{\beta}{x} * p - \frac{\beta}{x} * p^2 \\
&= \frac{\beta}{x} * p * (1-p).
\end{aligned}$$

This is approximately equivalent to

$$\frac{\Delta p}{\Delta x} = \frac{\beta}{x} * p * (1-p).$$

Thus, we obtain

$$\frac{\Delta p}{100\frac{\Delta x}{x}} = \frac{\beta}{100} * p * (1-p).$$

The average marginal effect thus reads

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\beta}{100} * p_i * (1-p_i).$$

## Variable Definitions

***Table A.3.7:*** *Variable Definitions.*

*All "explanatory and dependent variables" and the "control variables: investor and investment characteristics" are based on information retrieved from the internal management system of the crowdinvesting platform.*

**Explanatory and Dependent Variables**

| *Variable* | *Definition* |
|---|---|
| ln(Distance) | Variable for measuring information asymmetries between (potential) investors and startups. The proxy is measured as the natural logarithm of the physical distance between the investor and startup. It is measured between the midpoints of the investor's and the startup's state, and rounded to full kilometers. Higher values refer to higher information asymmetries. |
| ln(InvestmentAmount) | Variable for measuring the investment amount per investor per startup. It is measured in the natural logarithm of the investment amount in Mexican Pesos (MXN). Higher values refer to a higher investment amount. |
| InvestmentDummy | Dummy variable for investments having taken place. It takes the value of one if an investment has taken place, and zero otherwise. It serves as the base for the calculation of the investment probability. |
| ln(Likes) | Variable for measuring signaling effects from former users of the crowdinvesting platform to (potential) future investors. The proxy is measured as the natural logarithm of the number of likes which a startup has received on the platform. |
| Relationship | Dummy variable for an existing investor-startup-relationship. It takes the value of one if the investor and startup show any kind of relationship, e.g., as family, friends, or a previous meeting at a fair. It takes a value of zero otherwise. |

**Control Variables: Investor and Investment Characteristics**

| *Variable* | *Definition* |
|---|---|
| AgeGroup | Dummy variables for investors' age. One dummy variable is created per decade, i.e., for investors under or equal to 30, between 31 and 40, between 41 and 50, between 51 and 60, between 61 and 70, between 71 and 80, and above 80 years of age (see variable $InvestorAge$). It takes a value of one if an investor falls into a certain age group. |
| ln(Experience) | Variable for measuring investors' experience. The proxy is measured as the natural logarithm of the time span between the registration on the platform and the investment, and measured in days. |
| Gender | Dummy variable for measuring investors' risk preference. The proxy is measured as investors' gender, as men usually invest more risky than women. It takes the value of one for male investors, and the value for zero otherwise. |
| ln(Goal) | Variable for measuring the minimum financing goal set by startups within a predefined investment period. It is measured in the natural logarithm of the startups' goal in Mexican Pesos (MXN). Higher values refer to higher financing goals. |
| InvestorAge | Variable for measuring investors' age. It is measured in years between the year of birth and the year of investment. Higher values refer to a higher age. In the empirical analysis of *Table A.3.3* of the Appendix, it is denoted as the variable *Age*. |

# Chapter 4
# The Role of Societal Sentiment on Commercial Banks' Loan Portfolio

**Abstract** Motivated by prior sentiment analyses combined with financial questions and better data availability, we investigate whether the happiness of consumed music is correlated with banks' loan portfolios. More specifically, we focus on changes in societal sentiment, which is proxied by music consumption, and its association with changes in banks' loan volume, loan loss provisions, overdue loans, and finally loan charge-offs. Using a detailed dataset including information on streamed music from the world's leading music streaming platform and call report data on U.S. commercial banks, we show that an increase in societal sentiment is associated with an increase in banks' loan volume and banks' loan loss provisions. We also show that a positive change in sentiment is correlated with higher non-performing loans and charge-offs. Additionally, our results suggest a time dependency between changes in sentiment and the observed changes in banks' loan portfolio, that needs more attention in further research.

## 4.1 Introduction

The increasing digitalization of our daily lives has fundamentally changed how consumers interact with businesses. Online platforms, initially designed for purposes such as music streaming, are more and more important sources of data that extend beyond their original scope. For example, the information created through the specific use of such platforms has emerged as a valuable indicator of consumer preferences, behavior, and overall market trends. For example, sentiment derived from music streaming platforms, based on user activity and listening choices, can be used to proxy societal sentiments and emotional states. This information is also important for the financial sector, where sentiment analysis can play an important role in shaping business strategies and decision-making processes. Information intensive companies like banks, traditionally reliant on financial metrics and direct customer interactions, can add this additional information, expressed indirectly through external platforms, in their business decisions and potentially gain actionable insights. Such additional data can help financial institutions better understand consumers' mood, identify emerging (loan) risks, and understand emotional and behavioral patterns of their customers.

Many studies focus already on sentiment derived from music and its link to people's current emotions or induced behavior (Juslin and Västfjäll 2008; Saarikallio and Erkkilä 2007; Benetos et al. 2022). These studies do not have their origin in the field of economics but rather in the field of psychology. Researchers find that music is oftentimes used to validate current moods and feelings, which makes researchers agree that musical sentiment can function as a reflection of people's current state of mind. Moreover, many researchers already connected the field of psychology and economics by analyzing sentiment changes and economic outcomes. Jacob et al. (2009), for example, find a correlation between the type of music played in retail stores and the individual spending patterns of customers. Therefore, the authors provide empirical evidence that people's (spending) behavior might indeed be influenced by musical sentiment, which is not surprising since everybody might have experienced that very activating music, like techno for example, affects the mind differently than slow and steady piano music. However, studies on sentiment changes and the direct relation to questions and concerns of financial institutions have still not answered every open question. Many studies in this area focus more on macroeconomic data and gather insights on a broader scale like, e.g., associated changes in a country's consumer loan volume (e.g., Gric et al. 2022). For individual banks, this data aggregation makes it hard to derive any recommendations for them. Moreover, there is still a lack of an extensive analysis of a change in societal sentiment and its direct relation to a bank's loan business.

Therefore, in this study, we investigate the link between societal sentiment and key indicators for banks' loan business activities. Using data on the music listening behavior of the U.S. population, we conduct OLS regressions with time (quarter and year) and bank fixed effects, with 12,230 observations including data from the largest online streaming platform in the world

(Spotify.com). By integrating sentiment data from Spotify with banking performance metrics, we aim to assess the extent to which societal sentiment derived from music consumption correlates with banks' loan portfolio measures. For banks, the change in sentiment could affect their business in two ways. On the one hand, banks' loan officers themselves could be prone to sentiment changes and therefore tackle specific situations differently compared to situations in which this change in sentiment did not happen. On the other hand, borrowers' (repayment) behavior could change as well. When borrowers' priorities shift due to more positive or negative outlooks on the future, the repayment of an outstanding loan could rise or fall in the borrowers' list of priorities.

Our study helps deepen the understanding of the link between consumers' sentiment and the associated effects on banks' business. With the given dataset containing information on listening behavior, we provide insights into how observable and publicly available information potentially influence the financial market, or more specifically, banks in their core business tasks. First, we find that a positive change in sentiment in previous quarters is associated with a positive change in a bank's loan volume. Second, we provide evidence that changes in sentiment are also associated with changes in banks' provisioning behavior. With some kind of delayed reaction, banks react with higher loan loss provisions to positive changes in societal sentiment. Third, the repayment of loans is negatively correlated with a positive change in sentiment, since loans that are overdue, non-performing loans, and also charge-offs increase with positive sentiment changes in previous quarters. Thereby, this study confirms findings of previous studies and extends them by providing new insights into the link between banks' loan portfolios and their vulnerability to societal sentiment changes.

Our study contributes to the literature in different ways. First, this study adds to the strand of literature on the effect of musical sentiment on consumers' behavior or perception (e.g., Benetos et al. 2022; Hunter et al. 2011; Saarikallio and Erkkilä 2007; North and Hargreaves 1996). Furthermore, we add to the literature on the effects of sentiment on financial questions or decision-making in general. In contrast to previous studies (e.g., Delis et al. (2014)), we examine banks' loan portfolios on a broader scale by not only analyzing changes in loan volume but also incorporating loan provisioning, non-performing loans, and charge-offs. By establishing a direct link between sentiment changes and banks' annual reports, we provide new empirical evidence and insights into the connection between these two. By doing this, we make use of a publicly accessible and detailed dataset on consumed music and take the opportunity to create a sentiment measure based on observable music information. Therefore, this study does not rely on survey responses and is therefore not at someone's mercy to respond to these surveys, eliminating selection or nonresponse biases. Overall, since online activities in all areas have become of higher focus, our study sheds light on how banks can leverage and use non-traditional data to remain competitive in an increasingly data-driven world.

The paper proceeds as follows: Section 4.2 reviews the literature and develops our hypotheses. Section 4.3 describes the sample, methodology, descriptive statistics, and the research design. In Section 4.4, we present and discuss the empirical results. Section 4.5 provides additional tests and Section 4.6 describes limitations of this study. Section 4.7 concludes with a summary of the main findings.

## 4.2 Literature Review and Hypotheses Development

Sentiment refers to the collective mood, confidence, or perception of individuals or groups in the society.[29] It plays a significant role in a financial context as prior research shows that it influences decision-making processes and market outcomes beyond fundamental analysis. Prior literature already gives a broad insight into sentiments of specific individuals or groups and their association with financial questions as outlined in further detail below. Literature typically distinguishes between the sentiment of three main groups when discussing its association with financial questions.

First a large strand of literature focuses on managers' sentiment and its correlation with their decision making (e.g., Berger et al. 2024; Brückbauer and Cézanne 2022; Hribar et al. 2017). Understanding managerial sentiment is essential, as it directly influences and has consequences on key corporate decisions like investment strategies, credit allocation, and risk management. In contrast to investor sentiment, managerial sentiment has a direct impact on real firms' activities and, therefore, firms' performance. Decisions made based on (economic) conditions rather than economic fundamentals could therefore lead to suboptimal results.

Moreover, and building on the understanding of managerial sentiment, investor sentiment offers another critical perspective in financial markets. Focusing on the collective emotions and perceptions of market participants, investor sentiment, distinct from managerial decision-making, can drive decisions indirectly by influencing asset prices beyond their fundamental values. Its effects are particularly pronounced in speculative environments, where optimism or pessimism can amplify market volatility, mispricing, and systemic risks. Studying investor sentiment provides a deeper insight into human behavior and consequences coming with these types of psychological factors (e.g., Saunders 1993; Hirshleifer and Shumway 2003; Baker and Wurgler 2006).

Last but not least, the understanding of consumers' sentiment is also relevant for firms (e.g., Benhabib and Spiegel 2019; Delis et al. 2014; Gric et al. 2022; Caglayan and Xu 2016). Consumer sentiment could influence individual spending patterns, which in turn affect corporate revenues and broader economic cycles. Additionally, an overly optimistic consumer sentiment could lead

[29] In this paper, we use the terms "sentiment" and "mood" interchangeably. Some authors use "sentiment" as a broader term which then not only includes mood but also belief and preference changes from factors that are not mood-driven like, e.g., investor attention (Hirshleifer et al. 2020).

to excessive lending by banks and risk-taking by the consumers (and banks) since the future is seen as relatively positive, while negative sentiment could result in liquidity hoarding and reduced risk-taking due to uncertainty about future developments in politics or financial regards. Compared to other sectors, the interconnected nature of banks makes the banking sector even more vulnerable to changes in sentiment, as sentiment-driven behaviors could have systemic consequences in the worst case. Due to the central position banks hold in the financial system, the importance of understanding sentiment and its consequences from the involved counterparties is obvious.

The way sentiment is measured in prior literature diverges. Since a multitude of factors can potentially influence the sentiment of individuals or groups, the measurement and collection of data for calculating sentiment is far from straightforward. Many authors use exogenous shocks that are supposed to affect a group's sentiment in a specific way. For example, researchers use sports results or aviation accidents as a proxy for (national) sentiment (Edmans et al. 2007; Boido et al. 2023; Kaplanski and Levy 2010). However, since the interest in sports results or the awareness of aviation accidents in individuals (and the region they live in) varies, there might be room for an imprecise reflection of sentiment.

In the field of psychology, sentiment analysis is a very actively researched field of interest. Researchers come to the result that music affects thoughts, feelings, and can also change actions of a person. "Emotional contagion" describes the phenomenon that individuals reflect their mood in the choice of music. Emotional states such as happiness, sadness, anger, and calmness often find their way into melodies, harmonies, rhythms, and lyrics of music. This connection resonates with the theory of emotional contagion, which suggests that people can experience emotions indirectly through art forms like music (Juslin and Västfjäll 2008).

Consequently, individuals can use music to validate their emotions, and the music listened by a society could serve as a collective emotional snapshot, mirroring the community's shared mood. In line with this idea, Benetos et al. (2022) explore the use of music as an indicator for societal sentiment. They developed a national valence index by analyzing the emotional content of popular songs in the United Kingdom and show a significant correlation with survey-based measures of life satisfaction, suggesting that music can indeed serve as a reliable indicator of a nation's collective mood. Moreover, Saarikallio and Erkkilä (2007) show that adolescents often listen to sad music to express their emotions or find a sense of closure. North and Hargreaves (1996) find that an individual's music preference is closely connected to their current emotional state. Additionally, Hunter et al. (2011) observe in an experimental setting that upbeat music is less preferred after a sad mood was induced. Prior literature also shows that these changes in sentiment are connected to a person's economic behavior or beliefs that can drive and change an individual's behavior. For example, Jacob et al. (2009) find that the characteristics of music played in a retail store affect the spending pattern of customers. Another significant contribution is from Edmans et al. (2022) who explore the connection of sentiment and stock market returns, positing that the emotional tone of consumed music reflects broader societal optimism or pessimism. The findings reveal a positive association between sentiment induced by music and contemporaneous stock returns, even after controlling for external economic factors. The authors conclude that music sentiment can serve as a leading indicator of consumer and investor confidence.

Closer to our paper, Brückbauer and Cézanne (2022) show that when managerial sentiment, observed through the lens of textual sentiment, is too optimistic, lending growth can be high while equity investors' perception of the bank's risk declines. They find that excessive optimism largely disconnects from current economic data, which eventually leads to financial instability, highlighting how behavioral factors can influence credit supply and (systemic) risk in the financial sector. Examining loan growth and sentiment for different countries in Europe, Gric et al. (2022) measure household sentiment by using six survey questions from harmonized consumer surveys that are conducted to get information on private consumption growth in European countries. The authors surprisingly find that, on average, higher (irrational) sentiment of households leads to consumer loan growth. Caglayan and Xu (2016) analyze bank lending regarding business and consumer sentiment. The authors find that sentiment fluctuation has a negative effect on bank lending. Cortés et al. (2016) explore the impact of sentiment on lending decisions by using local weather patterns as an exogenous proxy for sentiment. They observe that the positive sentiment induced by sunny days increases bank lending - mostly on riskier applications that require greater discretionary judgment - and cloudy days have the opposite effect, leading to more conservative lending decisions. In addition, Delis et al. (2014) analyze anxious periods, evaluated by results from consumer surveys, in which economic agents hold a pessimistic outlook despite a lack of recessionary conditions. This study shows that in such anxious periods, lending decreases, especially when consumer and analyst confidence wanes. The authors argue that this mood-driven lending behavior can amplify economic downturns, particularly in banks with high credit risk. Cubillas et al. (2021) provide evidence that high investor sentiment drives up loan volumes, increasing risk-taking in banks. This sentiment-driven lending reduces stability, particularly in systems with weaker creditor protections, underscoring the systemic risks associated with overconfidence during optimistic phases. Moreover, by using a text-based measure, Agoraki et al. (2022) find that the change in outstanding loans is positive if investor sentiment is also positive. Unfortunately, the studies mentioned above do not disentangle whether the change in loan volume is supply or demand driven. However, and in line with their results, we hypothesize that in times of higher sentiment, banks extend their lending activities in terms of loan growth:

*Hypothesis 1: With increasing societal sentiment, the change in banks' loan volume is positive.*

Hribar et al. (2017) demonstrate that managerial sentiment biases accrual estimates, particularly in provisions for loan losses. Managers' optimism or pessimism, beyond justifiable economic indicators, affects loan risk assessments, resulting in potential misrepresentations of a bank's stability. This sentiment-induced bias suggests a need for objective control mechanisms in managerial decisions. In connection with *Hypothesis 1*, if banks are aware of an increasing change in loans (either through more demand or a less harsh process of originating a loan), they should carefully assess the risk they are taking on in their books. Gennaioli et al. (2015) state that financial crises result from a neglect of risks during economic booms. The authors find that investors overvalue recent good news and overlook bad outcomes. Thus, investors suffer from a cognitive bias that is rooted in the psychological principle of representativeness. And as the authors indeed find, during financial booms investors systematically underweight negative outcomes, showing why recurring crises cannot be perfectly predicted by conventional rational models. The 2008 financial crisis serves as a perfect example, in which the mispricing of mortgage-backed securities was based on an underestimation of the probability of mortgage default. In contrast, Bergman and Roychowdhury (2008) analyze corporate disclosure strategies as a function of investor sentiment and conclude that when sentiment is high, managers lower their forecasts for long-term future earnings, and when sentiment is low, managers increase disclosure to counteract the most pessimistic market expectations. Using the Michigan Consumer Confidence Index as one of the sentiment indicators, the study provides evidence of how companies improve communication with their investors and how their adjustment efforts affect the market in the hope of providing a favorable environment for their stocks. Therefore, Bergman and Roychowdhury (2008) state that managers are aware of investors' sentiment and could prepare for potential risks in the future. By using a news sentiment, Smales (2016) provides evidence that credit risk, measured through CDS spreads, decreases with an increase in sentiment. As a consequence, Smales (2016) suggests including news into banks' evaluation of credit risk. Cortés et al. (2016) do not only show that loan volume increases, but also riskier loans are granted. The authors find that loans approved in high sentiment periods are more likely to default, showing that the real economic consequence of sentiment in credit markets is hard to ignore. Thereby, the study provides evidence that sentiment changes subjective judgment while objective credit metrics are unaffected and illustrates the behavioral mechanisms underlying financial decisions. Agarwal et al. (2012) analyze the effect of loan officers' sentiment on their mortgage approval decision, using proxy variables for sentiment like sporting events and holidays. They show that positive sentiment raises loan approval rates by 4.5% for low-quality applications, for which officers have more discretion. These additional approvals, however, also contribute to a higher loan default rate among originated loans, which is then costly for banks. The authors point out that sentiment swings can lead to suboptimal decisions, underscoring the economic impact of sentiment on lending. Considering the studies mentioned above, two plausible hypotheses are possible. First, since biases made under the influence of positive sentiment are not done intentionally, and banks may be unaware of potential higher loan risks, an unchanged (or in this case inadequate) loan loss provisioning behavior could be observed. Second, banks could indeed be aware of the higher sentiment and "optimism" (as in, e.g., Bergman and Roychowdhury 2008) and could therefore build up higher provisions upfront to counteract potential downside risks in the future. All in all, we hypothesize that banks do change their loan provisioning behavior based on changes in (societal) sentiment:

*Hypothesis 2: An increase in societal sentiment goes along with a positive change in banks' LLP.*

As shown earlier, loan decisions are not always based on economic facts only (e.g., Cortés et al. 2016). Thereby, banks might end up taking on loan risks they objectively did not intend to accept in the first place, without being affected by sentiment, ultimately leading to an overall riskier loan portfolio than originally planned. In such a portfolio, banks are then more likely to face higher loan default probabilities or at least payment difficulties, affecting a bank's financial performance. Even though banks' loan officers are experts in their field, they might still be prone to engage in some form of bias due to the current societal sentiment. For example, Walther and Willis (2013) investigate how investor sentiment influences the accuracy of analysts' quarterly earnings forecasts by using the Michigan Index of Consumer Expectations as a sentiment proxy. They find that analysts are more optimistic and less accurate during periods of high sentiment. Their findings highlight the significance of sentiment for financial decision-making and they provide evidence that even experts are affected by sentiment induced biases. If we apply this framework to the loan origination process of banks, loan officers might underestimate the riskiness of the granted loans, likely leading to higher loan payment defaults in the future.

Besides potential riskier loan originations by banks on the supply side and therefore more risk of overdue or even non-performing loans and, in the worst case, write-offs in the future, consumers or the demand side could also influence a bank's (non-performing) loan portfolio by changing their (payment) behavior. Lindqvist (1981), for example, finds that debt repayments are negatively affected by higher economic satisfaction (evaluated with a measure deployed by Katona (1975)). This supports the theory of Katona (1975) stating that if people feel pessimistic about the future economy, they try to build up resources and buy durable goods. Therefore, people are more motivated to buy on instalments or with credit. In more optimistic times, spending priorities could change, and the mentioned behavior could plausibly decrease. Additionally, research by Souleles (2004) emphasizes the role of individual expectations and forecast errors in shaping consumption behavior, highlighting the heterogeneity in how consumers interpret and react to economic conditions. While this study does not directly address loan repayment, it points to the fact that subjective perceptions and future expectations, rather than purely objective economic indicators, significantly influence economic decisions. Therefore, while economic sentiment might have an indirect influence, the psychological factors of perceived control, attitudes towards debt, and individual expectations should be central to an understanding of credit behavior, with a potential feedback relationship between the ability to repay and the perception of control over debt. Additionally, if banks grant loans to companies, these companies face the risk of not being able to repay their debt when customers are not willing to buy their products or services, ultimately leading to problems in a bank's loan portfolio. Therefore, if there is a change in customers' (payment) behavior, it is highly informative to understand for banks to prepare for possible defaults in repayments. While Lindqvist (1981) suggests a direct link between consumer sentiment and loan repayment behavior, Livingstone and Lunt (1992) find no simple correlation between the two. Instead, their research indicates that individuals who are more diligent in repaying their loans tend to have a stronger sense of control over their debt.

In conclusion, we expect an underestimation of risks during periods of improved sentiment and, in addition, follow Lindqvist (1981) when deriving the following hypotheses:

***Hypothesis 3****: With increasing societal sentiment, future overdue and non-performing loans will increase.*

***Hypothesis 4****: With an increase in sentiment, future charge-offs will increase.*

## 4.3 Data and Research Design

### 4.3.1 Data

The dataset for this study comprises data from the largest music streaming platform worldwide, namely Spotify. For the U.S., 84% of recorded music revenues in 2023 were made by streamed music (RIAA 2024), making clear that the main market for music nowadays is online and making us confident that the use of this dataset is worthwhile. Spotify offers information on their songs through their API. Besides the song's key or tempo, the information includes a valence score, in other words "the musical positiveness", of a song (Spotify 2025). In addition, Spotify provides the number of streams per song listened to on a specific day for the Spotify Top 200 charts in the U.S. This allows for the calculation of a daily stream-weighted average of specific song characteristics (valence in our case). With this valence score, we proxy societal mood, traditionally done by surveys. To fit the sentiment derived from Spotify to banks' accounting data, we calculate the average valence score for each year's quarters.

Using data from Spotify offers distinct advantages over traditional surveys conducted by polling institutions or indices published by organizations for capturing societal sentiment. First, Spotify provides a significantly larger and more diverse dataset, as it encompasses listening behaviors and preferences of millions of users in the U.S. In contrast to surveys, which often focus on financial or political opinions, sentiment taken from Spotify reflects societal mood in general, capturing emotional states that are not limited to specific topics. In addition, the platform's user base is internet-savvy and most likely younger, which fits well with demographics that are likely to need credit or make financial decisions, compared to older demographics that may not use Spotify. Finally, Spotify's data eliminates potential biases that occur in surveys where users deliberately participate in sentiment analysis. As a result, selection bias and the influence of social desirability on responses can be avoided by construction. This passive data collection therefore leads to a more authentic, unfiltered representation of social sentiment.

For banks' accounting data, we use quarterly call reports published by the Federal Financial Institutions Examination Council (FFIEC) for all U.S. commercial banks in the period from 2017 to 2023. We drop bank observations with missing or negative values on core balance sheet items, such as equity or total assets. Additionally, we eliminate financial service firms and investment banks with a loan to asset ratio or deposits to total assets ratio below 50%. we drop small banks with total assets below 100,000 USD because small banks operate in a different regulatory environment. Furthermore, we delete observations characterized by extensive merger activities and financial distress to limit any confounding effects from corporate restructuring on the bank-level outcomes. After cleaning up the sample, we have 12,230 observations from 1,144 banks with data on all variables left for our regressions. The remaining banks in the sample are typical commercial banks. To further account for outliers, we winsorize all variables at the 1% level.

## 4.3.2 Variables

### 4.3.2.1 Main Variables

Within our analyses and throughout all our hypotheses, the main variable of interest is the societal sentiment. We capture societal mood by using data from Spotify. Spotify provides information on the 200 most streamed songs by a country per day. The information includes the name of the song, the artist, and the actual number of streams. Additionally, Spotify provides information, e.g., on the sentiment of the song called valence. Using an algorithm considering text, key, tempo, etc., Spotify determines whether a song is perceived more or less happy (with the value 1 for the happiest and 0 for the saddest possible song). Finally, we make use of this information and calculate the change of a stream-weighted daily average of valence per quarter ($\Delta SENT$).

In *Hypothesis 1*, we investigate the role of sentiment change on banks' change in loan volume ($\Delta LOANS$). In *Hypothesis 2*, we investigate how societal sentiment is correlated with banks' loan loss provisioning. Loan loss provisions ($LLP$) are scaled by a bank's total outstanding loans. Thereby, we can see whether banks are more or less cautious in a single period. For *Hypothesis 3*, non-performing loans ($NPL$) are measured by the share of loans that are overdue to total outstanding loans. Loans are categorized as non-performing once they are due for more than 90 days and the payment is not expected anymore (FFIEC 2024). The variable captures the share of non-performing loans to a bank's total loans. Additionally, banks must classify loans as overdue if payments are overdue for 30 to 89 days (FFIEC 2024). Therefore, we use the variable $PreNPL$ to capture these early stages of $NPL$. We also measure $PreNPL$ as the share of overdue loans to a bank's total loans. For *Hypothesis 4*, we analyze the relationship between societal sentiment and banks' loan charge offs ($CO$). To assess the proportion of loans written off relative to a bank's loan loss allowances, we calculate the ratio of charge-offs to the banks' overall provisioned amount. Loan loss allowances are therefore a cumulative reflection of a bank's loan loss provisions, representing the total reserve set aside to absorb potential credit losses over time.

#### 4.3.2.2 Control Variables

For the control variables, we follow several researchers (e.g., Caglayan and Xu 2016; Hribar et al. 2017) and include typical bank-business specific variables. Most of the control variables are included in lagged terms, helping to eliminate potential endogeneity concerns (Gric et al. 2022). First, we include the change in loan volume of the previous period ($\Delta LOANS$). The variable allows to capture previous demand- and supply-side effects in loan origination which, in turn, might correlate with the loan origination dynamics in the current period. We also add banks' loan-to-asset ratio ($LTA$). Thereby, differences in a bank's business model and (loan) risk structure can be captured. Additionally, we account for banks' currently perceived loan risks. We include these in two ways. First, we add the variable for banks' previous period's loan loss provisions ($LLP$), capturing their provisioning behavior for outstanding loans. The presence of troubled loans in banks' portfolios may influence their future decisions regarding loan origination. Therefore, accounting for their provisioning behavior makes sense. Second, we control for banks overall loan loss allowances ($LLA$) of the previous period. This might reflect the banks' attitude towards loan risks in general. Higher loan loss allowances might reflect a more cautious bank, while lower loan loss allowances could indicate more trust or confidence in a bank's loan portfolio.

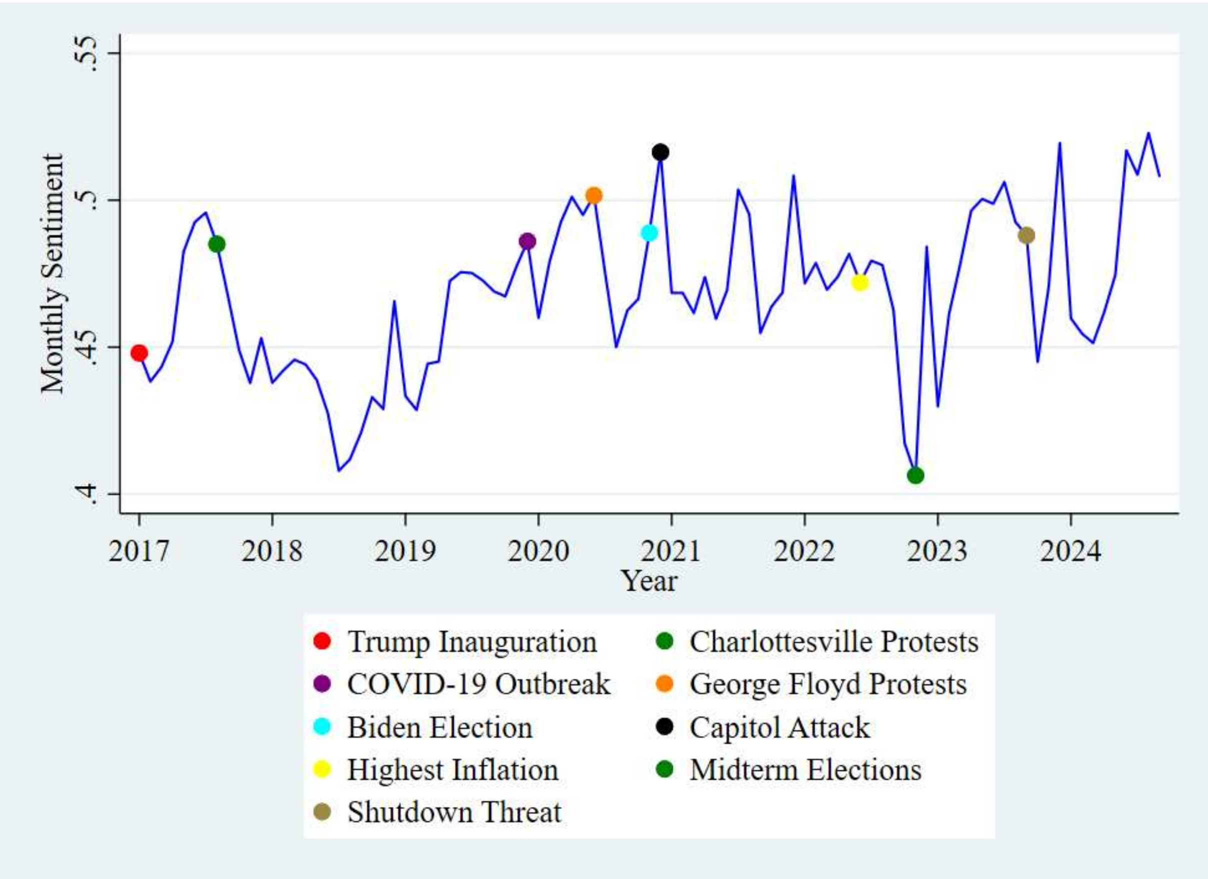
Moreover, we include the change in previous period's non-performing loans into the regression model ($\Delta NPL$). Here, the same argument as above holds: if banks have already originated problematic loans in the past, it could influence their decision making when originating new loans. Also, maybe banks have a structural problem in their loan processes. In this case, a previous increase in non-performing loans could be kind of an advance warning for an increase of non-performing loans in the current period. Furthermore, and closely related to non-performing loans, we add charge-offs from the previous period to account for the worst case in this scenario where banks have to write off some of their loans ($CO$).

In addition, we include banks' earnings before loan loss provisions ($EBLLP$) and their regulatory capital ($REGCAP$), as both variables can be influenced by managerial discretion. Specifically, banks may adjust their loan loss provisions in response to regulatory capital requirements, a phenomenon discussed in detail by Beatty and Liao (2014). Similarly, earnings management considerations are relevant, which is why we include earnings before loan loss provisions are deducted. Including both variables enables us to account for potential discretionary adjustments and enhances the reliability of our estimations for *Hypotheses 2* to *4*.

To account for macroeconomic conditions, we follow many researchers (e.g., Hribar et al. 2017) and include the change in current GDP rate ($\Delta GDP$), the change of the current unemployment rate ($\Delta UNEMP$), and lastly the return of the S&P/Case-Shiller U.S. National Home Price Index ($CSRET$).

### 4.3.3 Descriptive Statistics

***Figure 4.1:*** *Course of the Monthly Stream-Weighted Sentiment Score and Specific Events of U.S. Public Interest.*



Our primary (explanatory) variable of interest is the change of valence in the U.S. society. The average monthly change for the covered period is shown in Figure 4.1. As a plausibility check, we marked several events that potentially drove the collective mood in the U.S. Donald Trump's inauguration in 2017 is the first event. In the direct aftermath of the election, a small dip in sentiment can be observed, possibly reflecting public concern or political tension. From this point, the valence score continues to rise and the recovery in sentiment most likely reflects that the new administration was settling down and uncertainty slowly faded. The next marker stands for the violent clashes in Charlottesville. They sparked a nationwide discussion about racism and extremism after white nationalist rallies were held there. This event is observed alongside a significant drop in the valence score, indicating that such societal discord might have contributed to a lower collective mood. Next, the outbreak and escalation of the COVID-19 pandemic into a global health crisis coincides with a significant decline in the valence score, even though the decline is surprisingly small. However, this (slight) drop may reflect widespread fear, uncertainty, and social and economic disruption caused by the early stages of the pandemic. Last, there is a significant drop in the valence score after the murder of George Floyd and mass protests against racial injustice. The attack on George Floyd most likely caused anxiety and reduced trust in political stability, as reflected in the valence score's decline. The election of Joe Biden and the associated political shift appear to come with a stabilization of the valence score. While the immediate post-election period shows no dramatic changes, it goes along with a slight upward trend. Potentially, this reflects reduced political uncertainty. However, this stabilization was soon disrupted by the attack on the U.S. Capitol in January 2021. This event marked a serious challenge to democratic institutions and was widely perceived as a national crisis. This incident coincides with a noticeable dip in the valence score. Afterwards, inflation rose to the highest level in the U.S. since 1981 in the mid of 2022. Concerns about increased costs of living and economic uncertainty could therefore explain the decline in valence score. A sharp drop in the valence score can be observed in November 2022. Around the 2022 midterm elections, this decline in the valence score likely reflects political uncertainty and polarization during the election period. However, following the elections, the valence score rebounded, suggesting a restoration of public confidence and a reduction in political tension. In October 2023, the U.S. government faced a shutdown threat due to unresolved budget debates. This situation seems to have contributed to a drop in sentiment. Such substantial events commonly raise fears about economic stability, which likely influenced the observed decrease in the valence score.

In summary, a variation of the valence score around these key events is observable, indicating that the valence score captures aspects of collective societal sentiment. The timing of these events in relation to the sentiment trends offers plausible associations and strengthens the arguments for using a sentiment measure derived by music, even though a causal relationship can, of course, not be established.

***Figure 4.2:*** *Change in the Quarterly Sentiment Score and Michigan Consumer Sentiment Index.*

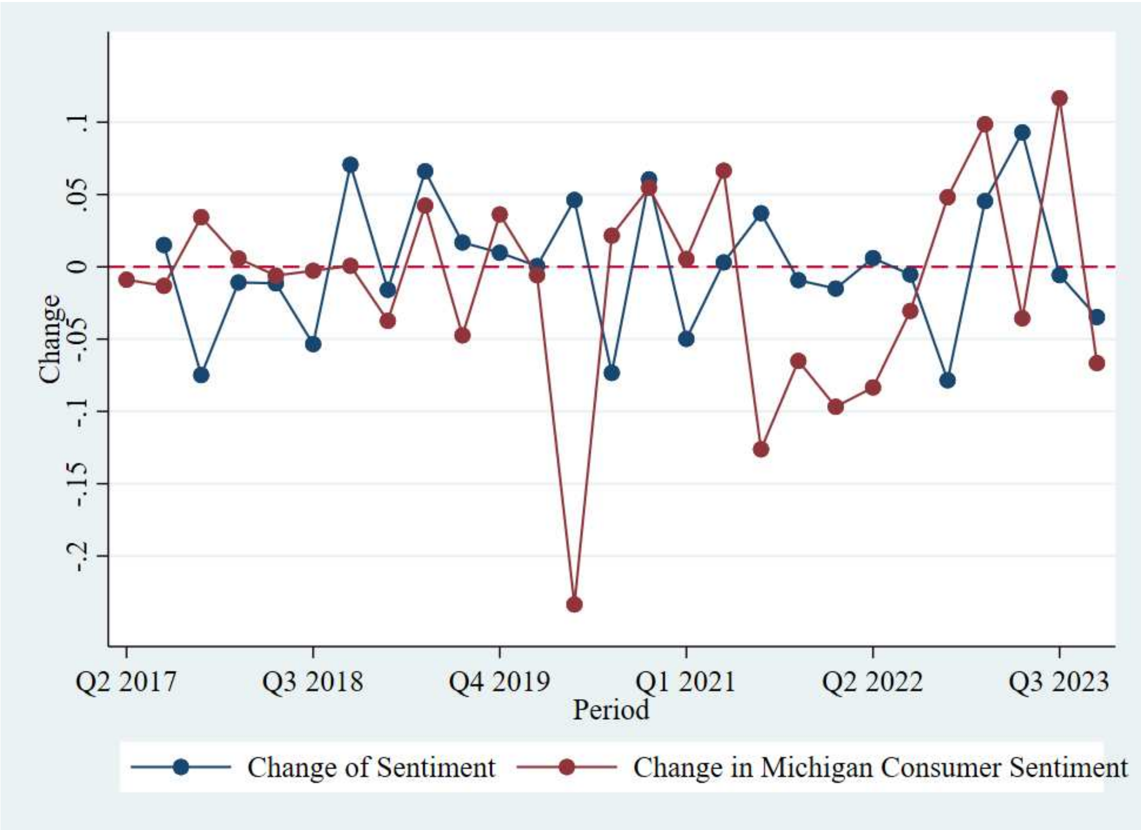
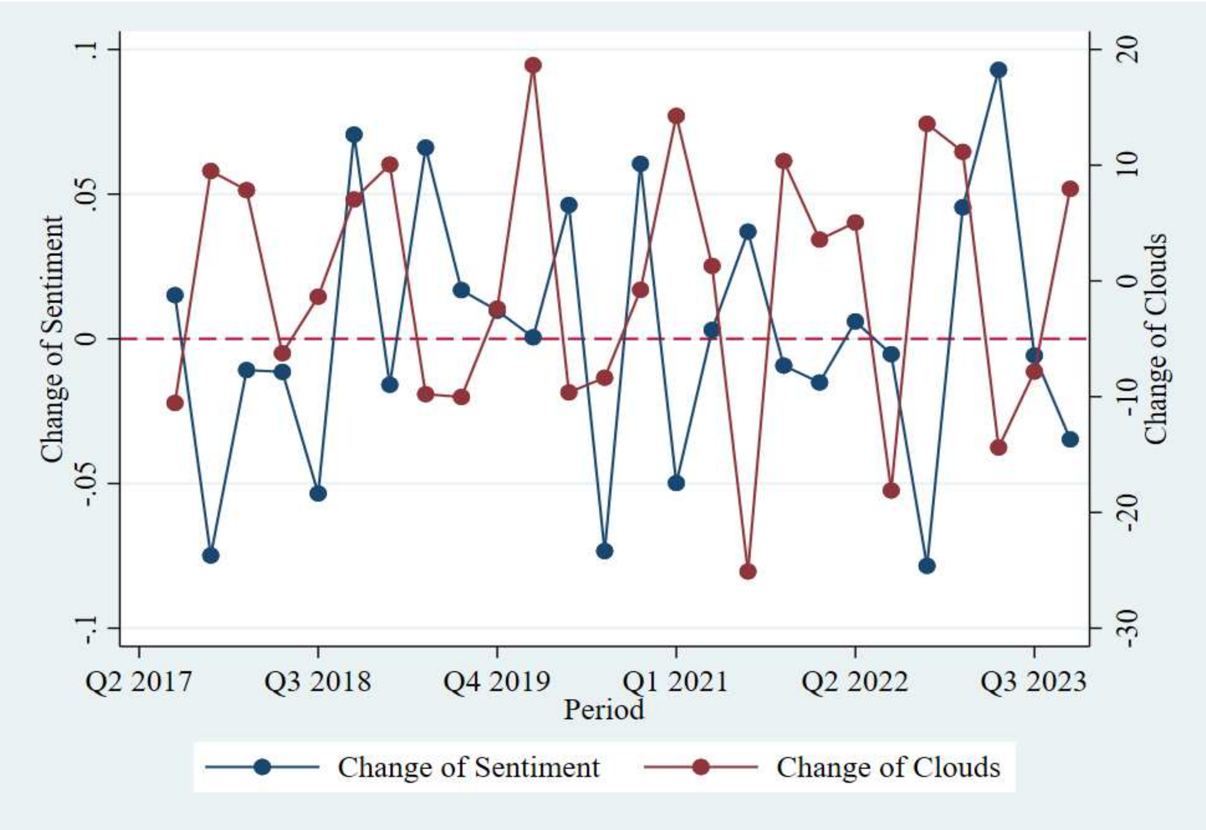



Figure 4.2 provides an additional graphical representation. It describes the course of the quarterly sentiment measure changes based on the stream-weighted average per quarter as used in the regression analyses. As a plausibility check, we added the quarterly change of the Michigan Consumer Sentiment Index in a specific quarter since several studies use it as their explanatory variable (e.g., Berger et al. 2024; Walther and Willis 2013). As shown, there is no perfectly positive correlation between these two sentiment measures. They show a similar shape, but the Michigan Consumer Sentiment Index seems to react stronger, especially in a sharp drop during the COVID-19 pandemic. Overall, these sentiment measures don't necessarily need to be very similar. The Michigan Consumer Sentiment Index captures answers from people who actively participated in the survey. The music sentiment of our study captures people's listening behavior without them being actively engaged in a survey. However, the sentiment measures seem to reflect U.S. sentiment, but at the same time capture somehow different sentiment information.

Figure 4.3 shows the course of the quarterly sentiment measure and the quarterly change of clouds in the U.S. Several studies use this weather proxy as their main variable (e.g., Cortés et al. 2016). As shown, there is no perfectly negative correlation between our sentiment and clouds, but on average an increase in sentiment is associated with a lower share of clouds ($\rho$ = -0.3885). Since the proportion of clouds might not capture all the people's sentiment and even a very good mood can occur when it is raining for example, our sentiment measure seems to be correlating

***Figure 4.3:*** *Change in the Quarterly Sentiment Score and U.S. Cloudiness.*




with cloudiness but, again, offers different information compared to the weather data. Therefore, we are convinced that our analyses and the use of this music-derived sentiment measure offer new and valuable insights.

Table 4.1 provides the descriptive statistics of the variables. On average, a 1.3 percent change in valence score can be observed on a quarterly basis, while the average value of valence is 47%. Additionally, the variable is slightly skewed to the left with a value of -0.47, meaning that the data points are more spread out at the lower end. The lowest value of a period's average valence is 43.57%, and the maximum reaches 49.92%.[30] Banks in the sample see, on average, an increase in loan growth of approximately 1.5% and have earnings before loan loss provisions of 1.7%. Additionally, banks in the sample have a loan to assets ratio of 72%, indicating a typical commercial bank balance sheet structure. The change in non-performing loans is slightly negative over the observation period, but the magnitude is so small that it does not indicate a systematic improvement of banks' loan portfolios in general.

Table 4.1 also displays the pairwise correlations of the valence score and other bank characteristics. The correlations do not indicate any issues related to multi-collinearity and are all in line with our expectations. We find a positive correlation between valence and change in loans that indicates that *Hypothesis 1* might not be rejected easily. Higher charge-offs tend to match

[30] Some of these information do not occur in Table 4.1 and are only reported in the text for informational purposes.

higher valence scores, suggesting that borrowers might not pay back their loans in times of a mood increase, supporting *Hypothesis 4*. However, correlations should be interpreted cautiously, since there is no control variables implemented, and a timing aspect is neglected. For example, the negative correlation of the change in valence and GDP change intuitively seems to head in the wrong direction.

***Table 4.1:*** *Descriptive Statistics and Correlation Matrix.*

| | Mean | Std. Dev | p50 | p25 | p75 | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) $\Delta SENT$ | 0.013 | 0.044 | 0.006 | -0.009 | 0.046 | | | | | | | | | | | |
| (2) $\Delta LOANS$ | 0.015 | 0.029 | 0.011 | -0.002 | 0.026 | 0.038 | | | | | | | | | | |
| (3) $LTA$ | 0.725 | 0.096 | 0.728 | 0.655 | 0.796 | 0.077 | 0.386 | | | | | | | | | |
| (4) $LLP$ | 0.001 | 0.001 | 0.000 | 0.000 | 0.001 | 0.023 | 0.146 | 0.121 | | | | | | | | |
| (5) $LLA$ | 0.012 | 0.005 | 0.012 | 0.009 | 0.014 | -0.038 | -0.130 | -0.236 | 0.310 | | | | | | | |
| (6) $\Delta NPL$ | -0.000 | 0.004 | -0.000 | -0.001 | 0.001 | 0.011 | 0.008 | 0.083 | 0.124 | 0.003 | | | | | | |
| (7) $CO$ | 0.030 | 0.056 | 0.009 | 0.001 | 0.031 | 0.060 | -0.042 | 0.040 | 0.447 | 0.072 | 0.032 | | | | | |
| (8) $EBLLP$ | 0.017 | 0.017 | 0.018 | 0.012 | 0.024 | -0.036 | -0.031 | -0.150 | -0.501 | -0.037 | -0.096 | -0.231 | | | | |
| (9) $REGCAP$ | 0.139 | 0.036 | 0.130 | 0.117 | 0.150 | 0.005 | -0.138 | -0.222 | -0.042 | 0.103 | -0.016 | -0.040 | 0.069 | | | |
| (10) $\Delta GDP$ | 0.014 | 0.028 | 0.016 | 0.009 | 0.023 | -0.436 | -0.311 | -0.166 | -0.174 | 0.042 | -0.031 | -0.038 | 0.104 | 0.009 | | |
| (11) $\Delta UNEMP$ | 4.566 | 2.404 | 3.867 | 3.200 | 4.733 | -0.053 | 0.180 | 0.066 | 0.179 | -0.011 | 0.084 | -0.000 | -0.067 | 0.022 | -0.300 | |
| (12) $CSRET$ | 0.019 | 0.023 | 0.018 | 0.002 | 0.030 | 0.138 | -0.035 | -0.177 | -0.083 | 0.088 | -0.034 | -0.067 | 0.087 | 0.013 | 0.168 | 0.180 |

Table 4.1 provides the summary statistics on the left-hand side of the table - columns (1) – (11) state the Pearson correlation coefficients between the variables. Continuous variables are winsorized at the 1% level. In Table A.4.1 in the Appendix, all variables are defined.

To strengthen the findings of the correlation analysis, we calculate variance inflation factors (VIF) to assure that the variables used in the regression analyses are most likely not affected by multicollinearity. Unfortunately, there is no critical threshold for VIF in general that indicates multicollinearity. However, a VIF of greater than 10 is likely a sign for multicollinearity. Since this is not the case for the main variables in our regression (maximum of 3.24), the VIF analysis is in line with the results of the correlation analysis.

### 4.3.4 Research Design

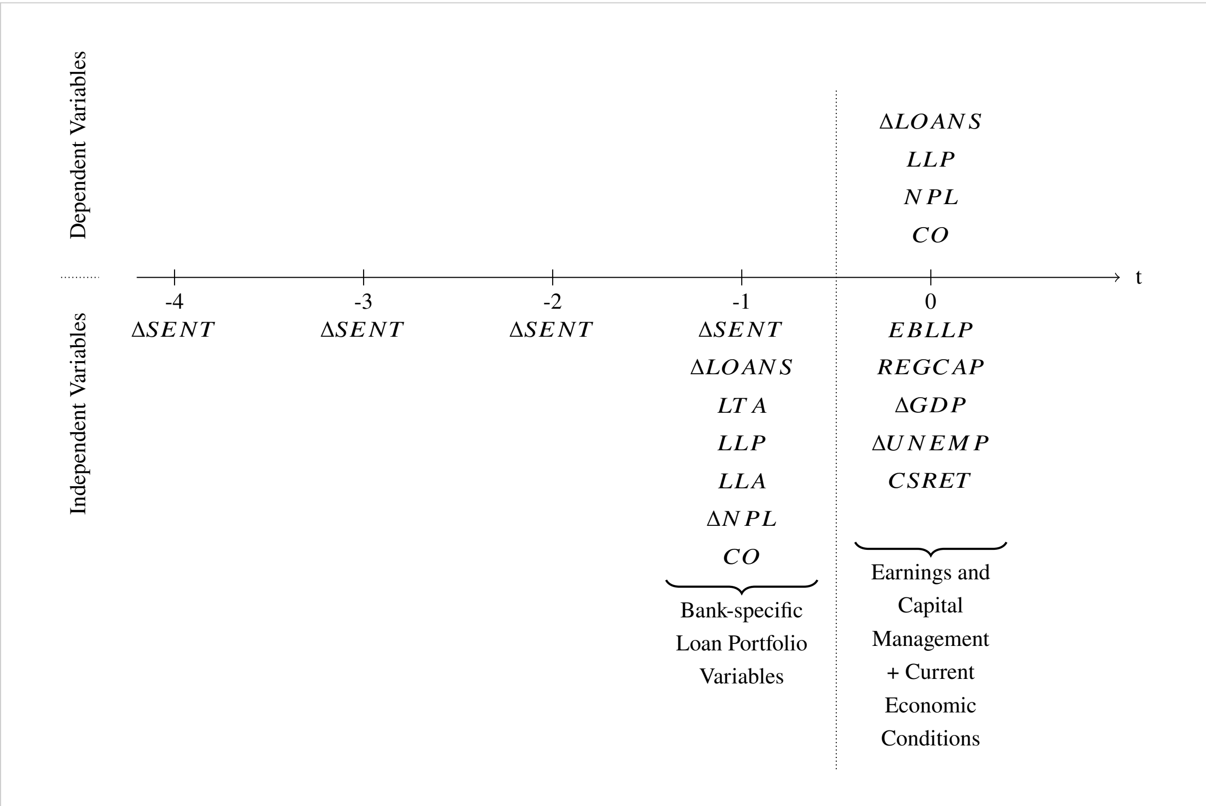
Our hypotheses focus on the influence of change in societal sentiment and its correlation with banks' loan business. For analyzing the relationship, we employ a fixed effects regression, closely related to prior studies (e.g., Hribar et al. 2017; Delis et al. 2014), with quarter and year fixed effects, and cluster standard errors on bank-level. Therefore, we specify the regression as follows:

$$\begin{aligned} Bank-specificvariable_{i,k,t} = {} & \alpha + \beta_1 * \Delta SENT_{k,t-1} + \beta_2 * \Delta SENT_{k,t-2} \\ & + \beta_3 * \Delta SENT_{k,t-3} + \beta_4 * \Delta SENT_{k,t-4} \\ & + \beta_5 * \Delta LOANS_{i,k,t-1} + \beta_6 * LTA_{i,k,t-1} \\ & + \beta_7 * LLP_{i,k,t-1} + \beta_8 * LLA_{i,k,t-1} \\ & + \beta_9 * \Delta NPL_{i,k,t-1} + \beta_{10} * CO_{i,k,t-1} \\ & + \beta_{11} * EBLLP_{i,k,t} + \beta_{12} * REGCAP_{i,k,t} \\ & + \beta_{13} * \Delta GDP_{k,t} + \beta_{14} * \Delta UNEMP_{k,t} \\ & + \beta_{15} * CSRET_{k,t} + \omega_i + \varphi_k + \tau_t + \varepsilon_{i,k,t} \end{aligned} \tag{4.1}$$

where $Bank-specificvariable_{i,k,t}$ takes on different values. For *Hypothesis 1*, it is the change in loan volume in the current period. For *Hypothesis 2*, the dependent variable is the current period's loan loss provisions ($LLP$) by bank $i$. Analyzing *Hypothesis 3*, the relationship between non-performing loans ($NPL$) and the explanatory variables is examined, while for *Hypothesis 4* charge-offs ($CO$) is the dependent variable. $\Delta SENT_{k,t}$ measures the change in societal sentiment from period t-1 to period t. $LLP_{i,k,t-1}$ stands for the loan loss provisions of period t-1. With $\Delta LOANS_{i,k,t-1}$, we include the change in loan volume from period t-1 to account for loan origination behavior and demand changes in the past. $\Delta NPL_{i,k,t}$ controls for the change of non-performing loans in period t. $EBLLP_{i,k,t}$ measures the earnings before loan loss provisions were deducted. To account for regulatory capital requirements of the individual banks, $REGCAP_{i,k,t}$ is included in the regression analysis. $CO_{i,k,t}$ measures the charge-offs made by bank i in period t, while $LLA_{i,k,t}$ control for banks' loan loss allowances employed in period t. We also include macroeconomic variables, namely the change in the current unemployment rate $\Delta UNEMP_{k,t}$, the return of the Case-Shiller Housing Price Index ($CSRET_{k,t}$), and the change in the U.S. GDP ($\Delta GDP_{k,t}$). Additionally, fixed effects for bank $i$ ($\omega_i$), year k ($\varphi_k$), and quarters t ($\tau_t$) are included to control for bank- and time-specific aspects and to capture potential seasonality effects.

Additionally, Figure 4.4 illustrates the regression model graphically. Most of the bank-specific variables are included with a lag of one quarter. Only $EBLLP_{k,t}$ and $REGCAP_{k,t}$ are included in the current quarter, since these variables might drive strategic bank decisions as explained before. Lastly, the macroeconomic variables are included in the current quarter without a quarterly lag.

***Figure 4.4:*** *Illustration of the Time Structure in the Regression Model.*




## 4.4 Empirical Results

Column (1) and (3) of Table 4.2 display the regression results of an OLS regression investigating the role of sentiment on the change in banks' loan volume and banks' loan loss provisions. In Columns (2) and (4), control variables are added. In line with our hypotheses, we find a positive correlation between higher sentiment and a change in loans. Additionally, higher sentiment is also associated with higher loan loss provisions.

We provide new evidence that loan volume rises when sentiment is higher. A one standard deviation increase in the sentiment measure in the previous quarter corresponds to an increase

***Table 4.2:*** *The Link between Sentiment Changes and Changes in Loan Volume and Loan Loss Provisioning.*

| Dependent Variable: | (1)<br>$\Delta LOANS_{i,k,t}$ | (2)<br>$\Delta LOANS_{i,k,t}$ | (3)<br>$LLP_{i,k,t}$ | (4)<br>$LLP_{i,k,t}$ |
|---|---|---|---|---|
| $\Delta SENT_{k,t-1}$ | 0.0696*** | 0.0257*** | 0.0142*** | 0.0000 |
| | (0.00746) | (0.00632) | (0.00023) | (0.00021) |
| $\Delta SENT_{k,t-2}$ | 0.00669 | -0.00524 | 0.00125*** | 0.000129 |
| | (0.00832) | (0.00644) | (0.000270) | (0.000229) |
| $\Delta SENT_{k,t-3}$ | 0.0697*** | 0.0360*** | 0.00181*** | 0.000643*** |
| | (0.00954) | (0.00753) | (0.000286) | (0.000230) |
| $\Delta SENT_{k,t-4}$ | 0.0918*** | 0.0644*** | 0.00107*** | 0.000506** |
| | (0.00953) | (0.00732) | (0.000233) | (0.000204) |
| $\Delta LOANS_{i,k,t-1}$ | | -0.0430*** | | 1.47e-05 |
| | | (0.0109) | | (0.000319) |
| $LTA_{i,k,t-1}$ | | 0.3100*** | | 0.000826*** |
| | | (0.0120) | | (0.000214) |
| $LLP_{i,k,t-1}$ | | -1.7390*** | | 0.164*** |
| | | (0.386) | | (0.0202) |
| $LLA_{i,k,t-1}$ | | 1.2750*** | | -0.0587*** |
| | | (0.225) | | (0.00971) |
| $\Delta NPL_{i,k,t-1}$ | | -0.3080*** | | 0.00913*** |
| | | (0.0646) | | (0.00190) |
| $CO_{i,k,t-1}$ | | 0.00531 | | -0.000375 |
| | | (0.00571) | | (0.000237) |
| $EBLLP_{i,k,t}$ | | -0.0157 | | -0.0258*** |
| | | (0.0239) | | (0.00279) |
| $REGCAP_{i,k,t}$ | | -0.0620 | | -0.000174 |
| | | (0.0428) | | (0.00115) |
| $\Delta GDP_{k,t}$ | -0.1030*** | 0.0286 | -0.0025*** | -0.0021*** |
| | (0.0289) | (0.0242) | (0.0010) | (0.0008) |
| $\Delta UNEMP_{k,t}$ | 0.0021*** | 0.0023*** | 0.0000 | 0.0000 |
| | (0.0004) | (0.0003) | (0.0000) | (0.0000) |
| $CSRET_{k,t-1}$ | -0.2090*** | -0.0156 | -0.0046*** | -0.0017*** |
| | (0.0189) | (0.0178) | (0.0006) | (0.0005) |
| Bank & Time FE | Yes | Yes | Yes | Yes |
| Observations | 12,230 | 12,230 | 12,230 | 12,230 |
| R-squared | 0.2210 | 0.4420 | 0.1820 | 0.4330 |

*This table displays in Column (1) and (3) the results of a fixed effects regression including only the explanatory variable, macroeconomic control variables and fixed effects. Columns (2) and (4) display the results of OLS regressions which include bank specific control variables. Robust standard errors are clustered at the bank level. Standard errors are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. Time fixed effects are added on a quarter and year level.*

in the loan volume of 0.0011 in the current quarter which can be translated into an increase of approximately 7.8% compared to the average level of change in loans.[31] The result is statistically significant at the 1% level. The same holds true for most of the lagged versions of the sentiment measure, even though the coefficient for the second lag variable is not significant. The results are plausible. Higher sentiment potentially reflects consumers optimism about the future. Therefore, people are more likely to spend money, resulting in higher needs for money, at least partially financed by credit. The control variables are rather inconspicuous. A positive change in the unemployment rate is positively correlated with a bank's change in loan volume. A higher unemployment rate might heighten the need for money and therefore credit. Of course, achieving a loan approval for individuals might be harder to achieve when they are unemployed, but since

[31] Marginal Effect = 0.0256973 (coefficient) * 0.04426118 (s.d. of $\Delta SENT$) = 0.0011377. The average of banks' quarterly loan volume change is 0.0146188.

we have no information on the loan characteristics (interest rate, time horizon, etc.), it is still a plausible explanation. Additionally, we observe a negative coefficient for the Case-Shiller Index. At first sight, this finding looks counterintuitive. But if the real estate market increases in value, housing prices rise, and consumers might shy away from getting a relatively large loan. The bank business specific control variables are also plausible. Banks with higher changes in loans in the previous quarter have lower changes in loan volume in the current quarter. This makes sense, as either the demand side is saturated, or the bank does not want to expand its loan portfolio if the loan volume has already increased in the previous quarter. If banks have a higher loan to asset ratio, they are more likely to have a higher change in loans, which can be explained by specialization advantages. Higher loan loss provisions in the previous quarter might indicate problems in the loan portfolio, therefore, the negative coefficient could reflect a more cautious loan origination behavior. In contrast, if banks have higher loan loss allowances, they might grant more loans, since the bank already has a higher risk buffer, explaining the positive coefficient of loan loss allowances. An increase in previous $NPL$ is correlated with a negative change in loan volume, reflecting a potentially more conservative loan origination process. Overall, there are no surprising coefficients for the control variables.

Column (3) and (4) of Table 4.2 present the relation between banks' loan loss provisions and changes in sentiment. The size of the sentiment coefficients is interesting in this case when control variables are included. The coefficients for the third and fourth period lag of sentiment changes are much higher than the ones for lag one and two and statistically significant on the 1% and 5% level. This shows some kind of delayed provisioning behavior in banks, meaning that granting loans and building up provisions most likely occur in a delayed fashion. On the one hand, this makes sense since it is unlikely to have a loan default in the first quarter of a freshly originated loan. On the other hand, the loan portfolio does not consist of newly originated loans only. If the (repayment) behavior of the borrowers' changes, this delay in provisioning might play a role for banks' stability, depending on how fast a potential behavioral change in borrowers is triggered by changes in sentiment. However, banks seem to consider societal sentiment in their decision-making processes which is positive. Unfortunately, whether it is driven by expectations on a changed behavior of borrower or a reaction based on previous lending decisions is unclear and cannot be disentangled with the use of our data. Therefore, the findings are intriguing in themselves but require further investigation to establish a causal relationship. Additionally, the economic effect seems to be rather small, but noticeable. A one standard deviation increase in three period lagged sentiment change is associated with an increase in banks' $LLP$ of approximately 5.4% relative to the average level of $LLP$.[32] The control variables are again unobtrusive. A higher $LTA$, as well as higher previous period's $LLP$, and $\Delta NPL$ are correlated with a higher provisioning rate, while higher $LLA$, and $EBLLP$ are associated with a decrease in $LLP$. Moreover, an increase in the GDP rate is associated with a decrease in $LLP$, which makes

[32] Marginal Effect = 0.0006427 (coefficient) * 0.04426118 (s.d. of $\Delta SENT$) = 0.00002844. The average of banks' quarterly LLP is 0.0005.

sense since borrowers are more likely capable of paying back outstanding debt. An increase in $CSRET$ is negatively correlated with $LLP$, most likely reflecting banks' trust in loans' collateral.

***Table 4.3:*** *The Link between the Change in Sentiment, Overdue Loans, Non-Performing Loans, and Charge Offs.*

| Dependent Variable: | (1) $PreNPL_{i,k,t}$ | (2) $PreNPL_{i,k,t}$ | (3) $NPL_{i,k,t}$ | (4) $NPL_{i,k,t}$ | (5) $CO_{i,k,t}$ | (6) $CO_{i,k,t}$ |
|---|---|---|---|---|---|---|
| $\Delta SENT_{k,t-1}$ | 0.00723*** | 0.00657*** | 0.00218** | 0.00160** | 0.0513*** | 0.0385*** |
| | (0.000888) | (0.000906) | (0.000862) | (0.000805) | (0.0124) | (0.0127) |
| $\Delta SENT_{k,t-2}$ | 0.00689*** | 0.00663*** | 0.00367*** | 0.00324*** | 0.0646*** | 0.0592*** |
| | (0.00130) | (0.00133) | (0.00114) | (0.00102) | (0.0159) | (0.0163) |
| $\Delta SENT_{k,t-3}$ | 0.00494*** | 0.00470*** | 0.00204 | 0.00219** | 0.0465*** | 0.0430** |
| | (0.00152) | (0.00155) | (0.00125) | (0.00109) | (0.0173) | (0.0174) |
| $\Delta SENT_{k,t-4}$ | 0.00104 | 0.001000 | 0.00123 | 0.00122 | 0.0270* | 0.0209 |
| | (0.00110) | (0.00112) | (0.000938) | (0.000919) | (0.0163) | (0.0161) |
| $\Delta LOANS_{i,k,t-1}$ | | -0.00319** | | -0.00318*** | | -0.000121 |
| | | (0.00144) | | (0.00123) | | (0.0187) |
| $LTA_{i,k,t-1}$ | | 0.000244 | | -0.00183 | | -0.0144 |
| | | (0.00105) | | (0.00140) | | (0.0144) |
| $LLP_{i,k,t-1}$ | | 0.150*** | | 0.0631 | | 3.396*** |
| | | (0.0578) | | (0.0476) | | (1.138) |
| $LLA_{i,k,t-1}$ | | 0.0514 | | 0.209*** | | 1.604*** |
| | | (0.0317) | | (0.0375) | | (0.558) |
| $\Delta NPL_{i,k,t-1}$ | | 0.00604 | | 0.366*** | | 1.175*** |
| | | (0.0135) | | (0.0141) | | (0.180) |
| $CO_{i,k,t-1}$ | | 0.00313*** | | 0.00240*** | | -0.0226 |
| | | (0.00116) | | (0.000905) | | (0.0211) |
| $EBLLP_{i,k,t}$ | | -0.0128*** | | -0.0164*** | | -0.491*** |
| | | (0.00411) | | (0.00338) | | (0.0958) |
| $REGCAP_{i,k,t}$ | | 0.00610 | | 0.0129 | | -0.0590 |
| | | (0.00552) | | (0.00886) | | (0.0659) |
| $\Delta GDP_{k,t}$ | -0.0187*** | -0.0182*** | -0.00266 | -0.00348 | -0.0305 | -0.0559 |
| | (0.00346) | (0.00363) | (0.00231) | (0.00234) | (0.0443) | (0.0465) |
| $\Delta UNEMP_{k,t}$ | -0.00025*** | -0.00025*** | -5.34e-05* | -6.13e-05** | -0.000227 | -0.000543 |
| | (4.48e-05) | (4.61e-05) | (2.97e-05) | (2.90e-05) | (0.000538) | (0.000558) |
| $CSRET_{k,t-1}$ | -0.00976*** | -0.0108*** | 0.00730*** | 0.00195 | -0.0152 | -0.0226 |
| | (0.00198) | (0.00217) | (0.00196) | (0.00205) | (0.0285) | (0.0309) |
| Bank & Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 12,230 | 12,230 | 12,230 | 12,230 | 12,230 | 12,230 |
| R-squared | 0.0763 | 0.0852 | 0.0736 | 0.3070 | 0.0447 | 0.0907 |

*This table displays in Column (1), (3), and (5) the results of a fixed effects regression including only the explanatory variable, macroeconomic control variables and fixed effects. Columns (2), (4), and (6) display the results of OLS regressions which include bank specific control variables. Robust standard errors are clustered at the bank level. Standard errors are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. Time fixed effects are added on a quarter and year level.*

In column (2), (4), and (6) of Table 4.3, the regression results of the relation between sentiment change, overdue loans, non-performing loans, and charge-offs are presented. For loans that are overdue, the coefficients of sentiment change for the first and second lag are almost the same. Both are positive and statistically significant on the 1% level. For the third lag, the significance level is still the same, but the absolute size of the coefficient is smaller. The fourth lag has no significant coefficient at all. More precisely, an increase of one standard deviation in the sentiment change one or two periods earlier is associated with an increase in $PreNPL$ of approximately 6.8% relative to the average level of $PreNPL$ in the current quarter. When looking at the results for $NPL$, it is obvious that the coefficients are smaller than for the loans that are "only" overdue.

Additionally, the highest coefficient is seen in the two-lag period while the smallest statistically significant coefficient is observed in the first lag period. These results make sense since for a loan to be classified as non-performing, it must go through being a loan that is classified as overdue. Moreover, since not all loans that are overdue end up being non-performing, the size of the coefficients are plausible as well. In this case, a one standard deviation increase in sentiment change of the two lag period is associated with an increase in $NPL$ of approximately 2.8%. For $CO$ in column (6), the comparison in coefficients' size does not make sense since charge-offs are scaled by the banks' current $LLA$ and not total loans.[33] However, a one standard deviation increase in change of sentiment lagged for two periods is associated with an increase in charge-offs of approximately 8.8% relative to its average level which is indeed remarkable. Interesting is the timing aspect of sentiment changes and charge-offs. If we compare the coefficients between the second and third lag, there are higher charge-offs associated with the (positive) sentiment change two lags before compared to a (positive) sentiment change three periods before. This is interesting and similar to the findings from the $NPL$ regression, why we assume banks to react quite fast in the conversion from non-performing loans to write (parts of) them off. For the control variables, two coefficients stand out. First, the coefficient of the previous quarter's $LLP$ is huge. Since both variables are scaled differently, this is not worrisome. However, the association between $LLP$ and $CO$ is interesting. The results suggest that banks do anticipate charge-offs in the quarter before the charge-offs actually take place and thereby build up their loan loss allowances. Second, the change in $NPL$ from the previous quarter is also directly related to higher charge-offs. In line with the argumentation above (and per definition), higher non-performing loans are in general a prerequisite for charge-offs, naturally increasing the chance of higher future charge-offs.

Overall, we find positive correlations between sentiment changes and banks' overdue and non-performing loans as well as banks' charge-offs. The results are in line with our hypotheses 3 to 5 stating that for more positive sentiments, the loan repayments are more likely lower. Unfortunately, unbundling whether this effect is driven by the supply side (granting riskier loans) or demand driven (worse repayment behavior by borrowers) is not possible with the data at hand. However, if banks observe this relation between sentiment and repayment behavior in their loan portfolio while using very standardized loan processes (for example a highly automated loan origination process with low discretionary space for loan officers), it might strengthen the assumption that this phenomenon is indeed behavior or at least borrower driven. In this case, banks could confidently and timely include changes in sentiment into their provisioning decisions and prevent future surprises in defaulted repayments. Summing up, our results show a consistent positive association of positive sentiment changes with higher loan volume, higher (but somehow delayed) loan loss provisions, lower loan repayments, and lastly higher charge-offs.

[33] Unreported results show that if we scale CO by total loans, the size of the relevant coefficients is smaller than the ones from column (3). This is in line with our previous argumentation since not all non-performing loans end up being written off by a bank.

## 4.5 Additional Tests

We first repeat the analyses above with a different sentiment measure. We use the change in cloud cover in the U.S. for a specific quarter as the new sentiment variable. Since higher cloudiness is supposed to be associated with lower mood, the coefficients should point in the opposite direction as before. By using different sentiment measures, we account for potential problems coming up with the use of only one music streaming platform, even though it's the largest one in the world.

***Table 4.4:*** *The Link between Sentiment Measured by Cloudiness in the U.S. and Banks' Loan Business Variables.*

| Dependent Variable: | (1) $\Delta LOANS_{i,k,t}$ | (2) $LLP_{i,k,t}$ | (3) $PreNPL_{i,k,t}$ | (4) $NPL_{i,k,t}$ | (5) $CO_{i,k,t}$ |
|---|---|---|---|---|---|
| $\Delta CLOUDS_{k,t-1}$ | -0.213*** | -0.000770 | -0.0573*** | -0.0107* | -0.272*** |
| | (0.0408) | (0.00147) | (0.00680) | (0.00545) | (0.0926) |
| $\Delta CLOUDS_{k,t-2}$ | -0.0644 | -0.00258 | -0.0387*** | -0.0200** | -0.416*** |
| | (0.0525) | (0.00161) | (0.00931) | (0.00827) | (0.134) |
| $\Delta CLOUDS_{k,t-3}$ | -0.242*** | -0.00448** | -0.0718*** | -0.0103 | -0.319** |
| | (0.0574) | (0.00184) | (0.00972) | (0.00818) | (0.144) |
| $\Delta CLOUDS_{k,t-4}$ | -0.217*** | -0.00147 | -0.0553*** | -0.00171 | -0.112 |
| | (0.0402) | (0.00131) | (0.00733) | (0.00548) | (0.0943) |
| $\Delta LOANS_{i,k,t-1}$ | -0.0390*** | -5.78e-06 | -0.00337** | -0.00325*** | -0.00186 |
| | (0.0110) | (0.000319) | (0.00143) | (0.00123) | (0.0188) |
| $LTA_{i,k,t-1}$ | 0.313*** | 0.000835*** | 2.25e-05 | -0.00179 | -0.0140 |
| | (0.0119) | (0.000211) | (0.00104) | (0.00139) | (0.0143) |
| $LLP_{i,k,t-1}$ | -1.403*** | 0.165*** | 0.172*** | 0.0607 | 3.412*** |
| | (0.384) | (0.0204) | (0.0576) | (0.0478) | (1.138) |
| $LLA_{i,k,t-1}$ | 1.289*** | -0.0587*** | 0.0567* | 0.208*** | 1.603*** |
| | (0.227) | (0.00969) | (0.0315) | (0.0375) | (0.558) |
| $\Delta NPL_{i,k,t-1}$ | -0.303*** | 0.00913*** | 0.00756 | 0.366*** | 1.175*** |
| | (0.0642) | (0.00189) | (0.0135) | (0.0142) | (0.179) |
| $CO_{i,k,t-1}$ | 0.00431 | -0.000385 | 0.00296** | 0.00243*** | -0.0225 |
| | (0.00574) | (0.000238) | (0.00115) | (0.000905) | (0.0210) |
| $EBLLP_{i,k,t}$ | -0.0112 | -0.0258*** | -0.0120*** | -0.0166*** | -0.492*** |
| | (0.0236) | (0.00279) | (0.00408) | (0.00339) | (0.0961) |
| $REGCAP_{i,k,t}$ | -0.0543 | -0.000161 | 0.00714 | 0.0128 | -0.0586 |
| | (0.0428) | (0.00115) | (0.00546) | (0.00889) | (0.0659) |
| $\Delta GDP_{k,t}$ | 0.0321 | -0.00233*** | -0.0226*** | -0.00282 | -0.0529 |
| | (0.0236) | (0.000836) | (0.00371) | (0.00247) | (0.0482) |
| $\Delta UNEMP_{k,t}$ | 0.00287*** | 1.17e-06 | -0.000267*** | -5.61e-05* | -0.000489 |
| | (0.000298) | (1.01e-05) | (4.49e-05) | (2.88e-05) | (0.000545) |
| $CSRET_{k,t-1}$ | -0.00892 | -0.00181*** | -0.00990*** | 0.00101 | -0.0405 |
| | (0.0182) | (0.000501) | (0.00231) | (0.00212) | (0.0319) |
| Bank & Time FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 12,230 | 12,230 | 12,230 | 12,230 | 12,230 |
| R-squared | 0.4390 | 0.4330 | 0.0904 | 0.3070 | 0.0906 |

*This table displays the results of fixed effects regressions with the same dependent variables as in the previous regression analyses including bank specific control variables. As a robustness test, the sentiment variable is now switched to a variable that measures the change in cloudiness in the U.S. over the respective quarter. Robust standard errors are clustered at the bank level. Standard errors are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. For representational purposes, we scaled the cloud variable by dividing with the factor* $10^3$. *Time fixed effects are added on a quarter and year level.*

The findings are in line with the previous results. A decrease in the cloud sentiment measure is associated with a better mood, therefore the sign of the sentiment coefficients is negative for the different regression specifications. As presented in Table 4.4, the coefficients are most of the time statistically significant, too, supporting the findings from the previous analyses. Interesting to see is that the sentiment coefficients for loans that are overdue, but not non-performing yet, are consistently significantly negative on the 1% level. However, the sentiment coefficients for $NPL$ and $CO$ are only significantly negative for the first two or three period lags, possibly indicating that being "only" overdue with loan payments seems to be less scary for borrowers than being classified as non-performing or being written-off. The role of the control variables is similar to the main analyses.

***Table 4.5:*** *The Link between Sentiment Measured Through Survey Results and Banks' Loan Business Variables.*

| Dependent Variable: | (1) $\Delta LOANS_{i,k,t}$ | (2) $LLP_{i,k,t}$ | (3) $PreNPL_{i,k,t}$ | (4) $NPL_{i,k,t}$ | (5) $CO_{i,k,t}$ |
|---|---|---|---|---|---|
| $\Delta MSENT_{k,t-1}$ | -0.0792*** | -0.000101 | 0.00603*** | -0.000318 | 0.0168 |
| | (0.00540) | (0.000183) | (0.000870) | (0.000636) | (0.0119) |
| $\Delta MSENT_{k,t-2}$ | -0.00263 | 0.000226 | 0.00872*** | 0.00119 | 0.0391*** |
| | (0.00536) | (0.000208) | (0.000943) | (0.000792) | (0.0130) |
| $\Delta MSENT_{k,t-3}$ | -0.0110*** | -6.85e-05 | 0.00188*** | -0.000509 | 0.00320 |
| | (0.00373) | (0.000101) | (0.000508) | (0.000478) | (0.00675) |
| $\Delta MSENT_{k,t-4}$ | 0.0497*** | 0.000195* | -0.000804 | -0.000461 | -0.0162* |
| | (0.00385) | (0.000108) | (0.000552) | (0.000491) | (0.00864) |
| $\Delta LOANS_{i,k,t-1}$ | -0.0430*** | 2.97e-05 | -0.00388*** | -0.00333*** | -0.00522 |
| | (0.0108) | (0.000320) | (0.00144) | (0.00124) | (0.0188) |
| $LTA_{i,k,t-1}$ | 0.310*** | 0.000816*** | 0.000606 | -0.00190 | -0.0124 |
| | (0.0120) | (0.000218) | (0.00105) | (0.00142) | (0.0145) |
| $LLP_{i,k,t-1}$ | -1.906*** | 0.163*** | 0.182*** | 0.0536 | 3.469*** |
| | (0.382) | (0.0205) | (0.0592) | (0.0481) | (1.156) |
| $LLA_{i,k,t-1}$ | 1.234*** | -0.0590*** | 0.0535* | 0.207*** | 1.583*** |
| | (0.219) | (0.00969) | (0.0313) | (0.0373) | (0.558) |
| $\Delta NPL_{i,k,t-1}$ | -0.258*** | 0.00929*** | 0.00600 | 0.366*** | 1.175*** |
| | (0.0626) | (0.00189) | (0.0134) | (0.0141) | (0.179) |
| $CO_{i,k,t-1}$ | 0.00679 | -0.000376 | 0.00292** | 0.00243*** | -0.0232 |
| | (0.00564) | (0.000237) | (0.00115) | (0.000908) | (0.0211) |
| $EBLLP_{i,k,t}$ | -0.0133 | -0.0258*** | -0.0121*** | -0.0167*** | -0.492*** |
| | (0.0239) | (0.00280) | (0.00411) | (0.00340) | (0.0961) |
| $REGCAP_{i,k,t}$ | -0.0593 | -0.000151 | 0.00637 | 0.0129 | -0.0570 |
| | (0.0427) | (0.00115) | (0.00552) | (0.00885) | (0.0658) |
| $\Delta GDP_{k,t}$ | -0.230*** | -0.00293*** | -0.00521 | -0.00550** | -0.0272 |
| | (0.0307) | (0.00101) | (0.00443) | (0.00274) | (0.0608) |
| $\Delta UNEMP_{k,t}$ | 0.000134 | -6.83e-06 | -0.000190*** | -7.42e-05** | -0.000263 |
| | (0.000333) | (1.13e-05) | (4.99e-05) | (3.09e-05) | (0.000646) |
| $CSRET_{k,t-1}$ | -0.122*** | -0.00178*** | 0.000192 | 0.00559** | 0.0709* |
| | (0.0194) | (0.000505) | (0.00242) | (0.00252) | (0.0390) |
| Bank & Time FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 12,230 | 12,230 | 12,230 | 12,230 | 12,230 |
| R-squared | 0.4630 | 0.4330 | 0.0894 | 0.3070 | 0.0906 |

*This table displays the results of fixed effects regressions with the same dependent variables as in the previous regression analyses including bank specific control variables. As a robustness test, the sentiment variable is now switched to a variable that measures the change in U.S. sentiment by using the results of the Michigan Consumer Sentiment Index over the respective quarter. Robust standard errors are clustered at the bank level. Standard errors are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. Time fixed effects are included on a quarter and year level.*

For further robustness tests, we conduct the regressions again with a different sentiment variable, this time gathered from the Michigan Consumer Sentiment Index and presented in

Table 4.5. Interestingly, the regression results for the change in banks' loan volume are mixed. Other than for our main sentiment measure and the cloudiness score, it seems that with the Michigan Consumer Sentiment Index there is some kind of cyclical pattern. At least, the sign of the coefficients differs and the sentiment coefficients are (apart from the second lag) statistically significant on the 1% level. Compared to the results from the main analysis and the first robustness test, there is no straightforward explanation for this. However, for the $LLP$, the findings are weaker but in line with the main findings of this paper. A positive change in sentiment correlates positively with an increase in $LLP$, supporting *Hypothesis 2*. Again, an interesting finding can be observed when looking at $NPL$ and loans that are overdue. For $NPL$, the sentiment measure derived from the consumer poll does not provide statistically significant insights. But looking at loans, for which payments are overdue, there are highly significant and positive coefficients calculated through the regression. Therefore, the previous findings are mostly supported. Moreover, the main findings suggest that not all loans that are overdue end up being non-performing. This argumentation can be applied here as well, even though there are indeed no statistically significant findings for $NPL$ at all when using the sentiment measure constructed by the results from the consumer survey. It is even more surprising that there are statistically significant findings for charge-offs even though non-performing loans seem to be uncorrelated. For charge-offs, there is also some kind of a timely pattern observed. While there is one statistically significant positive coefficient calculated for the two period lag, the coefficient for the fourth lag turns its sign and gets significantly negative (only on the 10% level). Again, the role of the control variables is similar to the main analyses. In conclusion, there might be more research needed to find out what is exactly measured by these different sentiment measures. Even though the regression results head in the same direction, it is still the case that they most likely carry or provide different information. Especially the regression results using the sentiment from the consumer survey, where people must actively respond, seem to generate weaker and less clear results. If this is driven by some kind of selection bias or the influence of social desirability on consumers' answers, as outlined previously, this study provides evidence for using a new sentiment measure derived from a large music dataset that is not subject to these biases.

Besides these two robustness tests, further insights are valuable for banks. As of now, regression results show that a positive sentiment change is associated with a positive change in loans, higher (but somehow delayed) loan loss provisions, higher overdue and non-performing loans, and also higher charge-offs. A relevant question for banks is whether these findings are problematic. If banks can generate profits overall, they might be fine with shifts in their loan portfolio. To test the profitability of banks, we conduct one more regression analysis. With respect to the controls, we follow Menicucci and Paolucci (2016) and specify the following regression model:

$$Bankprofitability_{i,k,t} = \alpha + \beta_1 * \Delta SENT_{k,t-1} + \beta_2 * \Delta SENT_{k,t-2} + \beta_3 * \Delta SENT_{k,t-3} + \beta_4 * \Delta SENT_{k,t-4} + \beta_5 * SIZE_{i,k,t} + \beta_6 * REGCAPi,k,t + \beta_7 * LTA_{i,k,t} + \beta_8 * DEPOSITS_{i,k,t} + \beta_9 * LLP_{i,k,t} + \omega_i + \rho_k + \tau_t + \varepsilon_{i,k,t} \tag{4.2}$$

***Table 4.6:*** *Variable Descriptions for the Additional Test.*

| Variable | Description | Measurement |
|---|---|---|
| *Dependent Variable* | | |
| $ROA_{i,k,t}$ | Return on Assets | Income (loss) before applicable income taxes to total loans |
| $NII_{i,k,t}$ | Net Interest Income | Net interest income to total loans |
| *Independent Variables* | | |
| $SIZE_{i,k,t}$ | Bank Size | Natural logarithm of banks' total assets |
| $REGCAP_{i,k,t}$ | Regulatory Capital | Tier 1 Capital divided by the sum of risk-weighted assets |
| $LTA_{i,k,t}$ | Loan Ratio | Banks' loans divided by total assets |
| $DEPOSITS_{i,k,t}$ | Deposits Ratio | Banks' deposits divided by total assets |
| $LLP_{i,k,t}$ | Loan Loss Provisions | Banks' loan loss provisions divided by total loans |
| $\omega_i$ | Bank Fixed Effect | |
| $\rho_k$ | Year Fixed Effect | |
| $\tau_t$ | Quarter Fixed Effect | |

*This table shows the variable descriptions for the additional test on the link of sentiment changes and banks' profitability measures.*

The variables for this additional test are described in Table 4.6. As before, we add bank and time fixed effects. The regression results are presented in Table 4.7, and they are indeed interesting. For both profitability measures (return on assets and net interest income), an increase in sentiment is associated with a decrease in profitability. For the return on assets, this correlation is true for the first two lags. For lags three and four, the sign of the coefficient turns positive. Nevertheless, for the first lag, banks face a shrinkage of about 1.3% relative to the average return on assets in the sample if the change in sentiment increases one standard deviation. The net interest income, however, "only" decreases about 0.8% relative to its average value in the sample.[34] These findings suggest that banks should incorporate a sentiment measure in their loan decision making process. The most important point in time for incorporating sentiment is likely when the loan granting decision is made. If the loan has been granted, banks (in general) have a potential risk of delayed or defaulted payments in their loan portfolio. However, even if the loan is already granted, banks can incorporate sentiment measures in their risk assessment and, e.g., make loan loss provisions timelier and (hopefully) more accurate. Again, since we do not have information on the specific loan contracts (e.g., interest rate, borrower information) and no statistical identification, further studies need to evaluate more precisely what the drivers of

[34] Since this is only a small additional test, the descriptive statistics are not reported in a table within the text. For transparency purposes, the average value of the return on assets in the sample is 0.0218. The average value for the net interest income is 0.0473.

these preliminary results are. Nevertheless, if these results are robust, the need for incorporating sentiment measures in the loan decision-making process is clear.

***Table 4.7:*** *The Link between Changes in Sentiment and Banks' Profitability.*

| Dependent Variable: | (1) $ROA_{i,k,t}$ | (2) $NII_{i,k,t}$ |
|---|---|---|
| $\Delta SENT_{k,t-1}$ | -0.00658*** | -0.00919*** |
| | (0.00115) | (0.000711) |
| $\Delta SENT_{k,t-2}$ | -0.00734*** | -0.0109*** |
| | (0.00178) | (0.000998) |
| $\Delta SENT_{k,t-3}$ | 0.00590*** | -0.0114*** |
| | (0.00212) | (0.00136) |
| $\Delta SENT_{k,t-4}$ | 0.00587*** | -0.00950*** |
| | (0.00197) | (0.00131) |
| $SIZE_{i,k,t}$ | 0.000655 | -0.00376*** |
| | (0.00111) | (0.00110) |
| $REGCAP_{i,k,t}$ | 0.0839*** | 0.0592*** |
| | (0.0159) | (0.0111) |
| $LTA_{i,k,t}$ | -0.00485** | -0.0179*** |
| | (0.00233) | (0.00195) |
| $DEPOSITS_{i,k,t}$ | 0.00473* | 0.00671*** |
| | (0.00242) | (0.00175) |
| $LLP_{i,k,t}$ | 0.767*** | 0.341*** |
| | (0.157) | (0.120) |
| Bank & Time FE | Yes | Yes |
| Observations | 12,230 | 12,230 |
| R-squared | 0.0405 | 0.1320 |

*This table displays the results of fixed effects regressions with the bank profitability variables as dependent variables including bank specific control variables. Robust standard errors are clustered at the bank level. Standard errors are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. Time fixed effects are included on a quarter and year level.*

## 4.6 Limitations and Implications for Future Research

Several aspects constitute possible limitations to our analyses and call for future research. First, there are limitations regarding the dataset. The analyses are based on data stemming from an Internet platform which, by construction, only constitutes people having access to the Internet and being active listeners on the bespoken platform. However, the mere size of the platform enables us to exploit a large heterogeneity and construct a general societal sentiment measure. Additionally, in times where technological progress is the order of the day, people get more and more confident using digital media for entertainment purposes where music is a big part of and continues to increase its role (IFPI 2023).

Furthermore, we are aware of limitations regarding long-term effects and time horizons of sentiment swings within our data. This means that we are unable to disentangle, e.g., how long a temporary change in people's mood potentially affect their actual behavior. Additionally, we are not able to determine an exact point in time when sentiment changes start to affect human behavior. Since we observe societal sentiment swings aggregated on a quarter level and include multiple lagged quarters, we are confident to control for delayed reactions on sentiment changes. However, it remains unclear to what extent possible sentiment changes overlap, especially in longer periods of good or bad sentiment. As mentioned, the duration and impact of sentiment changes cannot be determined in this study and need further investigations.

Unfortunately, we do not have information on specific loan conditions. The use of such information could provide further insights, particularly on the demand side. Currently, we are mostly limited to the use of information from banks' call reports, therefore, from the supply side (apart from the sentiment measure itself). However, since we do not have data on loan conditions and information, future research could provide evidence on this connection.

Additional analyses with different musical characteristics could produce interesting insights, too. Since Spotify offers individual song information like key or tempo, it would be exciting to see whether there is a correlation to customers' behavior. Therefore, studies from the field of psychology could point to musical measures that may influence human behavior or perception.

Moreover, since a large battery of sentiment measures is used in prior literature, a detailed analysis on advantages or disadvantages of single sentiment measures might offer serious advantages for banks' business. We already briefly discussed that the different sentiment measures might deliver somehow different information in the section regarding the robustness checks. Whether this information is insightful or just reflects an attitude towards something that is only closely related to banks' business, is of high importance. To disentangle which measure captures societal sentiment the best, could therefore lead to more reliable and precise estimates for banks and ultimately improve banks' stability and their loan business. If there is a measure superior to others and easily accessible, there is no need for a crude measure constructed from, e.g., weather data which makes it easier for banks to handle societal sentiment and its effects on consumer behavior. However, this study leaves space for future research and calls for deeper analyses regarding the effects of societal sentiment swings and their relation to customer behavior.

## 4.7 Conclusion

In this study, we analyze the relation between societal sentiment changes and the loan portfolio of banks. More specifically, we investigate the interplay between sentiment and banks' changes in loan volume, loan loss provisions, loans that are overdue or non-performing, and loan charge-offs. By integrating sentiment analysis with bank accounting data, we conduct an extensive empirical analysis and provide new insights into the role of sentiment for banks' loan business.

Our dataset consists of a large dataset gathered from the leading online music streaming platform. This dataset enables us to calculate stream-weighted averages of the "musical positiveness" of a song (Spotify 2025), which proxies societal sentiment in the U.S. Moreover, we use quarterly call data from banks provided by the FFIEC for banks' accounting information. Overall, we conduct regression analyses with 12,230 bank-quarter observations from 1,144 banks. Our results indicate that a positive change in sentiment is associated with a positive change in banks' loan volume, loan loss provisions, overdue loans, and even charge-offs.

Understanding the connection between non-financial factors like sentiment, and banking activities is relevant for researchers, financial institutions, and also banking supervision. While digitalization reshapes consumers' behavior, banks can use this for their advantage and exploit publicly available sources of information for gaining an informational advantage. Of course, traditional financial indicators remain essential for banks, but the benefits of using more information are clear. Since banks themselves become more data-driven, the implementation of more data in the decision-making process should be manageable. Banking supervision could also benefit from analyzing sentiment data and incorporating it in their risk profile of banks. When sentiment does, as shown in this study, relate to changes in banks' loan business, the neglection of sentiment information would be imprudent.

Even though this study cannot establish a causal relationship between sentiment changes and banks' loan business, it contributes to prior literature on sentiment analysis in the financial sector. However, further research is needed due to multiple factors. First, the disentanglement of whether the findings stem from the supply or demand side would be very insightful. Moreover, the timely aspect of sentiment changes on human behavior is indeed relevant for banks as well as for banking supervision. Before this relation is analyzed, recommendations seem to be too early. Lastly, since a battery of sentiment measures is used in the literature, it would be desirable to analyze which sentiment measure captures the "correct" sentiment and at what cost.

# A Appendices

## Tables

***Table A.4.1:*** *Variable Definitions.*

| *Variable* | *Definition* |
|---|---|
| $\Delta SENT$ | Variable for measuring the change in sentiment in the U.S. society. The proxy is measured as quarterly change in the mean valence score gathered from the Spotify API. The underlying sentiment score is measured on a scale between 0 and 1, with 0 for the saddest song and 1 for the most happiest song. |
| $\Delta LOANS$ | Variable for measuring the changes in a bank's loan portfolio. It is measured as the quarterly change in a bank's loans scaled by a bank's total assets. |
| $LTA$ | Variable for measuring the level of the loan to asset ratio. The variable is calculated as the share of a bank's loans scaled by total assets. Higher values refer to a higher proportion of loans in a bank's balance sheet. |
| $LLP$ | Variable for measuring a bank's loan loss provisions. The variable is scaled by a bank's total loans. The specific values are gathered from the bank's specific quarterly call report. |
| $LLA$ | This variable measures a bank's loan loss allowances. Similar to the loan loss provisions, this variable is scaled by a bank's total loans and gathered from the specific quarterly call reports. |
| $\Delta NPL$ / $PreNPL$ | Variable for measuring the change in a bank's non-performing loans. Non-performing loans are scaled by a bank's total loans. Loans are classified as non-performing in the quarterly call reports. The same holds true for loans that are currently "only" overdue (PreNPL). |
| $CO$ | This variable measures a bank's loan charge-offs and is scaled by a bank's loan loss allowances. Therefore, this variable reflects the proportion of loans for which a bank does not expect a payment anymore in relation to the bank's accumulated loan provisions. |
| $EBLLP$ | This variable reflects a bank's earnings before loan loss provisions are deducted. This variable is also scaled by a bank's total loans at the beginning of the respective quarter. |
| $REGCAP$ | This variable reflects the proportion of Tier 1 capital divided by the sum of risk-weighted assets. |
| $\Delta GDP$ | Variable for measuring the change in the U.S. GDP for the respective quarter. |
| $\Delta UNEMP$ | Variable for measuring the change in the current unemployment rate in the U.S. for the specific quarter. |
| $CSRET$ | This variable reflects the return of the S&P/Case-Shiller U.S. National Home Price Index. |

# Chapter 5
# Conclusion and Outlook

The role of IT in financial markets has changed and grown significantly in recent decades. With the increasing availability of data and the rise of digital tools, market participants face both new opportunities and challenges. Moreover, financial decisions are no longer solely based on information stemming from traditional data streams. These developments raise important questions about how IT can be used to process and analyze data efficiently.

This thesis examines the role of IT in financial markets. By presenting empirical evidence on IT-related topics in specific financial market settings, the goal of this thesis is to provide new insights into the relationship between the role of IT and financial market dynamics.

The first part of the analyses (Chapter 2, *"The Benefits of Banks' IT Investments in Times of Trouble: Evidence from Loan Loss Accruals During the COVID-19 Pandemic"*) shows the importance of being technologically prepared for new and surprising economic situations. By using a detailed dataset containing call report information for U.S. commercial banks, the study provides empirical evidence that banks with higher IT investments are better able to estimate their loan loss provisions in times of high economic uncertainty. With the COVID-19 pandemic as an exogenous shock, a causal relation can be established which strengthens our empirical findings. This study extends existing literature on IT investments in banks by presenting strong empirical evidence on the relevance of IT in terms of banks' loan risk management. In this way, this study contributes to the discussion on the so-called "profitability paradox" as introduced by Beccalli (2007) and acts as a base for future research on IT investments in banks.

Chapter 3 (*"The Role of Geographical Distance for Investment Decisions in Crowdinvesting"*) examines the role of geographical distance in investment decisions in crowdinvesting. Studies show that informational differences play a major role in decision-making (Kleinert and Volkmann 2019). By using a proprietary dataset from the largest online crowdinvesting platform in Mexico, geographical distance between investors and startups still seems to play a role, even in an online environment. A larger distance between investors and startups is correlated with a decrease in investment probability and investment amount. Moreover, an investor-startup-relationship does not appear strong enough to fully compensate for this negative relation. However, observable, yet partly subjective, signals on the online platform in form of startup likes are positively correlated with investors' investment decisions. Surprisingly, the online distribution of information does not appear to fully provide the same informative value to all investors. Thus, IT may extend the access for more distant investors, but it does not overcome all informational disadvantages or certain trust-related issues in crowdinvesting. Despite providing robust empirical evidence, we cannot establish a causal relationship with the dataset at hand. Therefore, future research might extend this study and generate a causal relationship by exploiting an exogenous economic shock, using an even more detailed dataset, or applying more sophisticated statistical methods.

In Chapter 4 (*"The Role of Societal Sentiment on Commercial Banks' Loan Portfolio"*), we show that the mood in a society correlates with the loan business of commercial banks. More specifically, we find that an increase in a society's sentiment is associated with a positive change in banks' loan volume, loan loss provisions, overdue and non-performing loans, and even charge-offs. We thereby confirm findings of studies focusing on sentiment changes and banks' loan volume changes on the one hand (e.g., Agoraki et al. 2022; Cubillas et al. 2021). On the other hand, we extend previous studies by analyzing sentiment and banks' loan business in its entirety, from changes in loan volume to banks' charge-offs. This part of the analyses shows that the use of alternative data can be worthwhile for financial institutions, even though traditional financial indicators remain, of course, essential. Therefore, the combination of soft, music-derived, sentiment information and banks' hard call report information offers new and valuable insights for banks' loan business. Consequently, it is important to take such data combinations into consideration when aiming to improve data-driven decision-making in a financial market setting. While being unable to establish a causal relationship, our results shed light on the importance of being technologically capable and encourage financial institutions to use information (in our case, on customers) that is publicly available.

Taken together, the three empirical chapters illustrate the multifaceted nature of information and the diverse ways in which IT can be applied in a financial market setting. Across these settings, this thesis demonstrates that technological capabilities must be aligned with the informational environment to gather meaningful insights and enhance decision-making.

In conclusion, the interplay between information and technological capabilities has become a key element of the modern financial system. The ability to efficiently collect, process, and interpret data is now a major determinant for financial market participants (BIS 2024). Financial institutions use their IT infrastructure to better respond to economic uncertainty and adapt quickly to new market conditions. The use of non-traditional data sources allows for additional insights and demonstrates an evolving informational structure in the finance sector. Moreover, financial decisions are increasingly shifted to digital platforms with observable, but not entirely objective, signals from peers and therefore allowing for new data streams. The value of this information depends on market participants themselves and their technological capabilities to effectively utilize the information. Future research should continue to explore the interfaces between technology, data, and financial behavior, especially in the light of ongoing innovations such as, e.g., generative AI and decentralized finance. A key challenge will be to ensure that these technological advances contribute to efficiency, while keeping an eye on financial stability (GCEE 2024). Thus, IT capabilities are no longer a background tool in the financial sector. They are a core determinant of how risks are assessed, investments are made, and economic signals are interpreted. The findings of this thesis provide a differentiated perspective on when and how IT adds value in financial market settings - and where its limits lie.

# References


**Agarwal, Sumit, David Lucca, Amit Seru, and Francesco Trebbi.** 2014. Inconsistent regulators: evidence from banking. *Quarterly Journal of Economics*, 129(2): 889–938.

**Agarwal, Sumit, Ran Duchin, and Denis Sosyura.** 2012. In the mood for a loan: The causal effect of sentiment on credit origination. SSRN Electronic Journal SSRN 2141030.

**Agoraki, Maria-Eleni K., Nektarios Aslanidis, and Georgios P. Kouretas.** 2022. U.S. banks' lending, financial stability, and text-based sentiment analysis. *Journal of Economic Behavior & Organization*, 197: 73–90.

**Agrawal, Ajay, Christian Catalini, and Avi Goldfarb.** 2015. Crowdfunding: geography, social networks, and the timing of investment decisions. *Journal of Economics & Management Strategy*, 24(2): 253–274.

**Agrawal, Ajay, Devesh Kapur, and John McHale.** 2008. How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics*, 64(2): 258–269.

**Ahlers, Gerrit K.C., Douglas Cumming, Christina Günther, and Denis Schweizer.** 2015. Signaling in equity crowdfunding. *Entrepreneurship Theory and Practice*, 39(4): 955–980.

**Ahsan, Muhammad, Eveline F. I. Cornelis, and Alan Baker.** 2018. Understanding backers' interactions with crowdfunding campaigns. *Journal of Research in Marketing and Entrepreneurship*, 20(2): 252–272.

**Akerlof, George A.** 1970. The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3): 488–500.

**Angerer, Martin, Alexander Brem, Sascha Kraus, and Andreas Peter.** 2017. Start-up funding via equity crowdfunding in Germany: A qualitative analysis of success factors. *Journal of Entrepreneurial Finance*, 19(1): 1–34.

**Angerer, Martin, Thomas Niemand, Sascha Kraus, and Thies Ferdinand.** 2018. Risk-reducing options in crowdinvesting: An experimental study. *Journal of Small Business Strategy*, 28(3): 1–17.

**Bae, Kee-Hong, René M. Stulz, and Hongping Tan.** 2008. Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts. *Journal of Financial Economics*, 88(3): 581–606.

**Baker, Malcolm, and Jeffrey Wurgler.** 2006. Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4): 1645–1680.

**Bandura, Albert.** 1977. *Social learning theory.* Englewood Cliffs, NJ: Prentice Hall.

**Banerji, Devika, and Torsten Reimer.** 2019. Startup founders and their LinkedIn connections: Are well-connected entrepreneurs more successful? *Computers in Human Behavior*, 90: 46–52.

**Beatty, Anne, and Scott Liao.** 2014. Financial accounting in the banking industry: A review of the empirical literature. *Journal of Accounting and Economics*, 58(2-3): 339–383.

**Beccalli, Elena.** 2007. Does IT investment improve bank performance? Evidence from Europe. *Journal of Banking & Finance*, 31(7): 2205–2230.

**Becker, Jan-Michael, Dennis Proksch, and Christian M. Ringle.** 2022. Revisiting Gaussian copulas to handle endogenous regressors. *Journal of the Academy of Marketing Science*, 50(1): 46–66.

**Beck, Thorsten, Aslı Demirgüç-Kunt, and Vojislav Maksimovic.** 2006. The influence of financial and legal institutions on firm size. *Journal of Banking & Finance*, 30(11): 2995–3015.

**Benetos, Emmanouil, Alessandro Ragano, Daniel Sgroi, and Anthony Tuckwell.** 2022. Measuring national mood with music: using machine learning to construct a measure of national valence from audio data. *Behavior Research Methods*, 54(6): 3085–3092.

**Benhabib, Jess, and Mark M. Spiegel.** 2019. Sentiments and economic activity: Evidence from US States. *The Economic Journal*, 129(618): 715–733.

**Berger, Allen N.** 2003. The economic effects of technological progress: evidence from the banking industry. *Journal of Money, Credit, and Banking*, 35(2): 141–176.

**Berger, Allen N., and Gregory F. Udell.** 2002. Small business credit availability and relationship lending: The importance of bank organisational structure. *The Economic Journal*, 112(477): 32–53.

**Berger, Allen N., Hugh Hoikwang Kim, and Xiaonan "Flora" Ma.** 2024. Bank sentiment and liquidity hoarding. *Contemporary Accounting Research*, 41(3): 1513–1542.

**Bergman, Nittai K., and Suhata Roychowdhury.** 2008. Investor sentiment and corporate disclosure. *Journal of Accounting Research*, 46(5): 1057–1083.

**Berg, Tobias, Verena Burg, Aleksandar Gombović, and Manju Puri.** 2020. On the rise of FinTechs: credit scoring using digital footprints. *Review of Financial Studies*, 33(7): 2845–2897.

**Berre, Max, and Benjamin Le Pendeven.** 2021. Startups, growth, and the macroeconomic environment: Valuation and business cycles. *Working Paper*. http://entfin.org/wp-content/uploads/2021/06/Startups-growth-and-the-Macroeconomic-Environment-Valuation-and-Business-Cycles.pdf. Accessed on 2023-01-25.

**Bharadwaj, Anandhi S.** 2000. A resource-based perspective on information technology capability and firm performance: an empirical investigation. *MIS Quarterly*, 24(1): 169.

**Bhat, Gauri, Jacqueline A. Lee, and Stephen G. Ryan.** 2020. Using loan loss indicators by loan type to sharpen the evaluation of the determinants and implications of banks' loan loss accruals. SSRN Electronic Journal.

**Bhatt, Ganesh D., and Varun Grover.** 2005. Types of information technology capabilities and their role in competitive advantage: an empirical study. *Journal of Management Information Systems*, 22(2): 253–277.

**Bhayani, Mital, Priyanka Bobade, Vinod Malkar, and Pooja Kawale.** 2019. Disparity in investment behavior of urban and rural individual investors. *International Journal of Social Science and Economic Research*, 4(5): 3764–3776.

**Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch.** 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5): 992–1026.

**BIS.** 2024. Artificial intelligence and the economy: Implications for central banks. In *Annual Economic Report 2024*, Chapter III. Basel: Bank for International Settlements.

**Block, Jörn, Lars Hornuf, and Alexandra Moritz.** 2018. Which updates during an equity crowdfunding campaign increase crowd participation? *Small Business Economics*, 50(1): 3–27.

**Blum, Bernardo S., and Avi Goldfarb.** 2006. Does the internet defy the law of gravity? *Journal of International Economics*, 70(2): 384–405.

**Boido, Claudio, Mauro Aliano, and Giuseppe Galloppo.** 2023. Top-flight European football teams and stock returns: market reactions to sporting events. *Journal of Economics and Finance*, 47(4): 1041–1061.

**Bornstein, Robert F., and Catherine Craver-Lemley.** 2022. Mere exposure effect. In *Cognitive illusions*, ed. by Rüdiger F. Pohl, 241–258. Routledge.

**Brückbauer, Frank, and Thibault Cézanne.** 2022. Bank manager sentiment, loan growth and bank risk. ZEW – Leibniz Centre for European Economic Research ZEW Discussion Paper 22-066, Mannheim. https://ftp.zew.de/pub/zew-docs/dp/dp22066.pdf.

**Buchak, Gregor, Gregor Matvos, Tomasz Piskorski, and Amit Seru.** 2018. Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3): 453–483.

**Caglayan, Mustafa, and Bing Xu.** 2016. Sentiment volatility and bank lending behavior. *International Review of Financial Analysis*, 45: 107–120.

**Caliendo, Marco, and Sabine Kopeinig.** 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1): 31–72.

**Campbell, Mary.** 2012. What a difference a year makes: time lag effect of information technology investment on firm performance. *Journal of Organizational Computing and Electronic Commerce*, 22(3): 237–255.

**Carpenter, Robert, and Bruce Petersen.** 2002. Capital market imperfections, high-tech investment, and new equity financing. *The Economic Journal*, 112(477): 54–72.

**Carr, Caleb T., Rebecca A. Hayes, and Emma M. Sumner.** 2018. Predicting a threshold of perceived Facebook post success via likes and reactions: A test of explanatory mechanisms. *Communication Research Reports*, 35(2): 141–151.

**Chakravarty, D., and M. Ahsan.** 2023. The effect of global city founding location on crowdfunding success and internationalization. *Working Paper*, 1–36.

**Chen, Henry, Paul Gompers, Anna Kovner, and Josh Lerner.** 2010. Buy local? The geography of venture capital. *Journal of Urban Economics*, 67(1): 90–102.

**Copeland, Craig.** 2022. Understanding the finances of rural vs. urban americans. *EBRI Issue Brief*, 553: 1–16. https://www.ebri.org/content/understanding-the-finances-of-rural-vs.-urban-americans. Accessed on 2023-01-31.

**Cornelis, Elien F. I., Amanda M. Baker, and Muntasir Ahsan.** 2022. Mitigating a crisis of confidence: The effect of crisis response strategies on reward-based crowdfunding success. *Strategic Entrepreneurship Journal*, 16(1): 67–96.

**Cortés, Kristle, Ran Duchin, and Denis Sosyura.** 2016. Clouded judgment: The role of sentiment in credit origination. *Journal of Financial Economics*, 121(2): 392–413.

**Coursaris, Constantinos K., Wietske van Osch, and Brigitte A. Balogh.** 2016. Do facebook likes lead to shares or sales? Exploring the empirical links between social media content, brand equity, purchase intention, and engagement (49th HICSS 2016). https://ieeexplore.ieee.org/abstract/document/7427628. Accessed on 2023-02-26.

**Courtney, Christopher, Supradeep Dutta, and Yong Li.** 2017. Resolving information asymmetry: Signaling, endorsement, and crowdfunding success. *Entrepreneurship Theory and Practice*, 41(2): 265–290.

**Coval, Joshua D., and Tobias J. Moskowitz.** 1999. Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, 54(6): 2045–2073.

**Cubillas, Elena, Elena Ferrer, and Nuria Suárez.** 2021. Does investor sentiment affect bank stability? International evidence from lending behavior. *Journal of International Money and Finance*, 113: 102351.

**Cunningham, Scott.** 2021. *Causal inference: The mixtape.* New Haven, CT: Yale University Press.

**Dal Maso, Laura, Kiridaran Kanagaretnam, Gerald J. Lobo, and Stefano Terzani.** 2018. The influence of accounting enforcement on earnings quality of banks: Implications of bank regulation and the global financial crisis. *Journal of Accounting and Public Policy*, 37(5): 402–419.

**Davila, Antonio, George Foster, and Mahendra Gupta.** 2003. Venture capital financing and the growth of startup firms. *Journal of Business Venturing*, 18(6): 689–708.

**de Chaisemartin, Clément, and Xavier D'Haultfœuille.** 2018. Fuzzy differences-in-differences. *Review of Economic Studies*, 85(2): 999–1028.

**Degryse, H., and S. Ongena.** 2005. Distance, lending relationships, and competition. *The Journal of Finance*, 60(1): 231–266.

**Dejean, Sylvain.** 2020. The role of distance and social networks in the geography of crowdfunding: Evidence from France. *Regional Studies*, 54(3): 329–339.

**Delis, Manthos D., Georgios P. Kouretas, and Chris Tsoumas.** 2014. Anxious periods and bank lending. *Journal of Banking & Finance*, 38: 1–13.

**Demir, Firat.** 2009. Macroeconomic uncertainty and private investment in Argentina, Mexico and Turkey. *Applied Economics Letters*, 16(6): 567–571.

**DeYoung, Robert, William W. Lang, and Daniel L. Nolle.** 2007. How the Internet affects output and performance at community banks. *Journal of Banking and Finance*, 31(4): 1033–1060.

**Di Maggio, Marco, and Vincent Yao.** 2020. Fintech borrowers: Lax-screening or cream-skimming? National Bureau of Economic Research Working Paper 28021, Cambridge, MA.

**Di Pietro, Francesca, Luca Grilli, and Francesca Masciarelli.** 2020. Talking about a revolution? Costly and costless signals and the role of innovativeness in equity crowdfunding. *Journal of Small Business Management*, 61(2): 831–862.

**Duan, Yixin, Sadok El Ghoul, Omrane Guedhami, Hui Li, and Xiaofei Li.** 2021. Bank systemic risk around COVID-19: a cross-country analysis. *Journal of Banking and Finance*, 133: 106299.

**EBA.** 2022. EBA discussion paper on Machine Learning for IRB models. European Banking Authority.

**EBA.** 2025. Final report on amending guidelines on ICT and security risk management (EBA/GL/2019/04). European Banking Authority EBA/GL/2025/02.

**Edmans, Alex, Adrian Fernandez-Perez, Alexandre Garel, and Ivan Indriawan.** 2022. Music sentiment and stock returns around the world. *Journal of Financial Economics*, 145(2): 234–254.

**Edmans, Alex, Diego Garcia, and Øyvind Norli.** 2007. Sports sentiment and stock returns. *The Journal of Finance*, 62(4): 1967–1998.

**Enria, Andrea.** 2021. Digital innovation in the banking sector: pay-offs and perils. European Central Bank. https://www.bankingsupervision.europa.eu/press/speeches/date/2021/html/ssm.sp210921 78d06f2393.en.html. Accessed on 2022-03-28.

**Eranti, Veikko, and Markku Lonkila.** 2015. The social significance of the Facebook Like button. *First Monday*, 20(6).

**Fama, Eugene F.** 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2): 383–417.

**FFIEC.** 2024. Instructions for preparation of consolidated reports of condition and income. https://www.ffiec.gov/pdf/FFIEC_forms /FFIEC031_FFIEC041_202406_i.pdf. Accessed on 2025-01-20.

**Fortune Business Insights.** 2025. Big data analytics market size, share & industry analysis. Fortune Business Insights Industry Report. Accessed on 2025-05-18.

**FSB.** 2024. The financial stability implications of artificial intelligence. Financial Stability Board. Accessed on 2025-05-22.

**Fu, Jun, and Manish Mishra.** 2022. Fintech in the time of COVID 19: technological adoption during crises. *Journal of Financial Intermediation*, 50: 100945.

**Fuster, Andreas, Matthew Plosser, Philipp Schnabl, and James Vickery.** 2019. The role of technology in mortgage lending. *Review of Financial Studies*, 32(5): 1854–1899.

**GCEE.** 2024. Annual Report 2024/25: Addressing shortfalls, modernising resolutely. German Council of Economic Experts Annual Report. Accessed on 2025-05-18.

**Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny.** 2015. Neglected risks: The psychology of financial crises. *American Economic Review*, 105(5): 310–314.

**Goethner, Maximilian, Sebastian Luettig, and Tobias Regner.** 2021. Crowdinvesting in entrepreneurial projects: Disentangling patterns of investor behavior. *Small Business Economics*, 57(2): 905–926.

**Goetzmann, William N., and Alok Kumar.** 2008. Equity portfolio diversification. *Review of Finance*, 12(3): 433–463.

**Grasshoff, Günter, Marc Coppola, Thomas Pfuhler, Nicolas Gittfried, Sebastian Bochtler, Viktor Vonhoff, and Christian Wiegand.** 2019. Global risk 2019: creating a more digital, resilient bank. https://www.bcg.com/publications/2019/global-risk-creating-digital-resilient-bank. Accessed on 2021-04-16.

**Gric, Zuzana, Dominika Ehrenbergerova, and Martin Hodula.** 2022. The power of sentiment: Irrational beliefs of households and consumer loan dynamics. *Journal of Financial Stability*, 59: 100973.

**Grossman, Sanford J., and Joseph E. Stiglitz.** 1980. On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3): 393–408.

**Grunert, Jens, and Lars Norden.** 2012. Bargaining power and information in SME lending. *Small Business Economics*, 39(2): 401–417.

**Guenther, Christina, Sofia Johan, and Denis Schweizer.** 2018. Is the crowd sensitive to distance? How investment decisions differ by investor type. *Small Business Economics*, 50(2): 289–305.

**Hauswald, Robert, and Robert Marquez.** 2003. Information technology and financial services competition. *Review of Financial Studies*, 16(3): 921–948.

**Hernandez-Trillo, Fausto, Jose A. Pagan, and Julia Paxton.** 2005. Start-up capital, microenterprises and technical efficiency in Mexico. *Review of Development Economics*, 9(3): 434–447.

**Hervé, Fabrice, Elodie Manthé, Aurélie Sannajust, and Armin Schwienbacher.** 2019. Determinants of individual investment decisions in investment-based crowdfunding. *Journal of Business, Finance & Accounting*, 46(5-6): 762–783.

**Hill, Katherine, Donald Hirsch, and Abigail Davis.** 2021. The role of social support networks in helping low income families through uncertain times. *Social Policy & Society*, 20(1): 17–32.

**Hirshleifer, David, and Tyler Shumway.** 2003. Good day sunshine: Stock returns and the weather. *The Journal of Finance*, 58(3): 1009–1032.

**Hirshleifer, David, Danling Jiang, and Yuting Meng DiGiovanni.** 2020. Mood beta and seasonalities in stock returns. *Journal of Financial Economics*, 137(1): 272–295.

**Hornuf, Lars, and Matthias Neuenkirch.** 2017. Pricing shares in equity crowdfunding. *Small Business Economics*, 48(4): 795–811.

**Hornuf, Lars, Matthias Schmitt, and Eliza Stenzhorn.** 2022. The local bias in equity crowdfunding: Behavioral anomaly or rational preference? *Journal of Economics & Management Strategy*, 693–733.

**House, J. S., D. Umberson, and K. R. Landis.** 1988. Structures and processes of social support. *Annual Review of Sociology*, 14(1): 293–318.

**Hribar, Paul, Samuel J. Melessa, R. Christopher Small, and Jaron H. Wilde.** 2017. Does managerial sentiment affect accrual estimates? Evidence from the banking industry. *Journal of Accounting and Economics*, 63(1): 26–50.

**Hult, G. Tomas M., Joseph F. Hair, Dorian Proksch, Marko Sarstedt, Andreas Pinkwart, and Christian M. Ringle.** 2018. Addressing endogeneity in international marketing applications of partial least squares structural equation modeling. *Journal of International Marketing*, 26(3): 1–21.

**Hunter, Patrick G., E. Glenn Schellenberg, and Andrew T. Griffith.** 2011. Misery loves company: mood-congruent emotional responding to music. *Emotion (Washington, D.C.)*, 11(5): 1068–1072.

**IFPI.** 2023. IFPI's global study finds we're listening to more music in more ways than ever - IFPI. Updated on 12/11/2023, checked on 1/20/2025.

**Jacob, Céline, Nicolas Guéguen, Gaëlle Boulbry, and Selmi Sami.** 2009. 'Love is in the air': congruence between background music and goods in a florist. *The International Review of Retail, Distribution and Consumer Research*, 19(1): 75–79.

**Jin, Jingyao, Kiridaran Kanagaretnam, and Ying Liu.** 2018. Banks' funding structure and earnings quality. *International Review of Financial Analysis*, 59: 163–178.

**Jin, Justin, Kiridaran Kanagaretnam, Y. I. Liu, and Maoyong Cheng.** 2021. Does citizens' financial literacy relate to bank financial reporting transparency? *European Accounting Review*, 30(5): 887–912.

**Juslin, Patrik N., and Daniel Västfjäll.** 2008. Emotional responses to music: the need to consider underlying mechanisms. *The Behavioral and Brain Sciences*, 31(5): 559–575; discussion 575–621.

**Kanagaretnam, Kiridaran, Charles Y. Lim, and Gerald J. Lobo.** 2010*a*. Auditor reputation and earnings management: international evidence from the banking industry. *Journal of Banking and Finance*, 34(10): 2318–2327.

**Kanagaretnam, Kiridaran, Gopal V. Krishnan, and Gerald J. Lobo.** 2010*b*. An empirical analysis of auditor independence in the banking industry. *Accounting Review*, 85(6): 2011–2046.

**Kaplanski, Guy, and Haim Levy.** 2010. Sentiment and stock prices: The case of aviation disasters. *Journal of Financial Economics*, 95(2): 174–201.

**Katona, George.** 1975. *Psychological Economics.* New York, NY, Oxford, Amsterdam: Elsevier.

**Kilic, Emre, Gerald Lobo, Tharindra Ranasinghe, and Lu Yi.** 2021. Strategic usefulness of ignorance: evidence from income smoothing via retained interest of securitized loans. *Review of Quantitative Finance and Accounting*, 56(1): 245–272.

**Kleinert, Simon, and Christine Volkmann.** 2019. Equity crowdfunding and the role of investor discussion boards. *Venture Capital*, 21(4): 327–352.

**Kleinert, Simon, Christine Volkmann, and Marc Grünhagen.** 2020. Third-party signals in equity crowdfunding: The role of prior financing. *Small Business Economics*, 54(1): 341–365.

**Koetter, Michael, and Felix Noth.** 2013. IT use, productivity, and market power in banking. *Journal of Financial Stability*, 9(4): 695–704.

**Konovalova, Maria E., Olga Y. Kuzmina, and Sergey A. Zhironkin.** 2020. Digital technologies as a factor of expanding the investment opportunities of business entities. In *Digital Age: Chances, Challenges and Future*, Vol. 84 of *Lecture Notes in Networks and Systems*, ed. by Svetlana Igorevna Ashmarina, Marek Vochozka and Valentina Vyacheslavovna Mantulenko, 180–188. Cham: Springer International Publishing.

**Kriebel, Jens, and Jens Debener.** 2019. Measuring the effect of digitalization efforts on bank performance. *SSRN Journal*, 1–36. Working Paper.

**Kwan, Alan, Chen Lin, Vesa Pursiainen, and Mingzhu Tai.** 2020. Stress testing banks' digital capabilities: Evidence from the COVID-19 pandemic. *SSRN Electronic Journal*, 1–46.

**Kwan, Lai-Yee-Yee, Stephanie Yap, and Chi-Yue Chiu.** 2015. Mere exposure affects perceived descriptive norms: Implications for personal preferences and trust. *Organizational Behavior and Human Decision Processes*, 129: 48–58.

**Leboeuf, Gael, and Armin Schwienbacher.** 2018. Crowdfunding as a new financing tool. In *The Economics of Crowdfunding: Startups, Portals and Investor Behavior*, Vol. 1, ed. by Douglas Cumming and Lars. Hornuf, 11–28. Palgrave Macmillan, Cham: Springer Nature.

**Lee, Samuel, and Petra Persson.** 2016. Financing from family and friends. *Review of Financial Studies*, 29(9): 2341–2386.

**Levine, Ross.** 2005. Finance and growth: Theory and evidence. In *Handbook of Economic Growth*, Vol. 1 of *Handbooks in Economics*, ed. by Philippe Aghion and Steven N. Durlauf, 865–934. Amsterdam: Elsevier.

**Leyden, Dennis P., Albert N. Link, and Donald S. Siegel.** 2014. A theoretical analysis of the role of social networks in entrepreneurship. *Research Policy*, 43(7): 1157–1163.

**Liberti, Jose Maria, and Mitchell A. Petersen.** 2019. Information: Hard and Soft. *Review of Corporate Finance Studies*, 8(1): 1–41.

**Lindblom, Tobias, Taylan Mavruk, and Stellan Sjögren.** 2018. East or west, home is best: The birthplace bias of individual investors. *Journal of Banking and Finance*, 92: 323–339.

**Lindqvist, Alf.** 1981. A note on determinants of household saving behavior. *Journal of Economic Psychology*, 1(1): 39–57.

**Liu, Chih-Chih, and Stephen G. Ryan.** 1995. The effect of bank loan portfolio composition on the market reaction to and anticipation of loan loss provisions. *Journal of Accounting Research*, 33(1): 77–94.

**Liu, Chih-Chih, and Stephen G. Ryan.** 2006. Income smoothing over the business cycle: changes in banks' coordinated management of provisions for loan losses and loan charge-offs from the Pre-1990 bust to the 1990s boom. *The Accounting Review*, 81(2): 421–441.

**Livingstone, Sonia M., and Peter K. Lunt.** 1992. Predicting personal debt and debt repayment: Psychological, social and economic determinants. *Journal of Economic Psychology*, 13(1): 111–134.

**Li, Xinxin, and Lorin Hitt Wu.** 2013. Observational learning and social-network word-of-mouth: Evidence from Groupon. *SSRN Electronic Journal*.

**Lutz, Eva, Marko Bender, Ann-Kristin Achleitner, and Christoph Kaserer.** 2013. Importance of spatial proximity between venture capital investors and investees in Germany. *Journal of Business Research*, 66(11): 2346–2354.

**Malloy, Christopher J.** 2005. The geography of equity analysis. *The Journal of Finance*, 60(2): 719–755.

**Menicucci, Elisa, and Guido Paolucci.** 2016. The determinants of bank profitability: empirical evidence from European banking sector. *Journal of Financial Reporting and Accounting*, 14(1): 86–115.

**Merton, Robert C.** 1987. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3): 483–510.

**Mochon, Daniel, Katherine Johnson, Jonah Schwartz, and Dan Ariely.** 2017. What are likes worth? A Facebook page field experiment. *Journal of Marketing Research*, 54(2): 306–317.

**Mollick, Ethan.** 2014. The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, 29(1): 1–16.

**Monti, Matteo, Vittorio Pelligra, Laura Martignon, and Nathan Berg.** 2014. Retail investors and financial advisors: New evidence on trust and advice taking heuristics. *Journal of Business Research*, 67(8): 1749–1757.

**Morin, Robert-A., and Antonio F. Suarez.** 1983. Risk aversion revisited. *The Journal of Finance*, 38(4): 1201–1216.

**Moritz, Alexandra, and Joern Hendrich Block.** 2014. Crowdfunding und Crowdinvesting: State-of-the-Art der wissenschaftlichen Literatur (Crowdfunding and Crowdinvesting: A Review of the Literature). *Zeitschrift für KMU und Entrepreneurship*, 62(1): 57–89. 10.2139/ssrn.2274141.

**Niemand, Thomas, Martin Angerer, Ferdinand Thies, Sascha Kraus, and René Hebenstreit.** 2018. Equity crowdfunding across borders: A conjoint experiment. *International Journal of Entrepreneurial Behavior & Research*, 24(4): 911–932.

**Nitani, Miwako, Allan Riding, and Beichuan He.** 2019. On equity crowdfunding: Investor rationality and success factors. *Venture Capital*, 21(2-3): 243–272.

**North, Adrian C., and David J. Hargreaves.** 1996. Situational influences on reported musical preference. *Psychomusicology: A Journal of Research in Music Cognition*, 15(1-2): 30–45.

**OECD.** 2019. An introduction to online platforms and their role in the digital transformation. https://www.oecd.org/innovation/an-introduction-to-online-platforms-and-their-role-in-the-digital-transformation-53e5f593-en.htm. Accessed on 2022-06-14.

**OECD.** 2020. Taxing wages 2020. https://www.oecd.org/en/publications/taxing-wages/volume-/issue-_047072cd-en.html. Accessed on 2022-04-22.

**OECD.** 2022. OECD Economic surveys (Mexico, May 2019). https://www.oecd-ilibrary.org/economics/oecd-economic-surveys-mexico-2019_a536d00e-enhttp://www.oecd.org/economy/mexico-economic-snapshot/. Accessed on 2022-04-22.

**Papies, Dominik, Peter Ebbes, and Harald J. van Heerde.** 2017. Addressing endogeneity in marketing models. In *Advanced Methods for Modeling Markets*, *International Series in Quantitative Marketing*, ed. by Peter S. H. Leeflang, Jaap E. Wieringa, Tammo H. A. Bijmolt and Koen H. Pauwels, 581–627. Springer International Publishing.

**Papke, Leslie E., and Jeffrey M. Wooldridge.** 2023. A simple, robust test for choosing the level of fixed effects in linear panel data models. *Empirical Economics*, 64(6): 2683–2701.

**Park, Sangman, and Sunil Gupta.** 2012. Handling endogenous regressors by joint estimation using Copulas. *Marketing Science*, 31(4): 567–586.

**Petersen, Mitchell A., and Raghuram G. Rajan.** 1994. The benefits of lending relationships: Evidence from small business data. *The Journal of Finance*, 49(1): 3–37.

**Pierri, Nicola, and Yannick Timmer.** 2020. Tech in fin before FinTech. *IMF Working Papers*, 20(14): 1–41.

**Pierri, Nicola, and Yannick Timmer.** 2022. The importance of technology in banking during a crisis. *Journal of Monetary Economics*, 128: 88–104.

**Piva, Evila, and Cristina Rossi-Lamastra.** 2018. Human capital signals and entrepreneurs' success in equity crowdfunding. *Small Business Economics*, 51(3): 667–686.

**Reuters.** 2020. Banks roll out robots as pandemic shakes up IT plans. https://www.reuters.com/article/idUSL4N2GR4B3. Accessed on 2022-03-28.

**RIAA.** 2024. Year-end 2023 RIAA revenue statistics. Recording Industry Association of America. https://www.riaa.com/wp-content/uploads/2024/03/2023-Year-End-Revenue-Statistics.pdf. Accessed on 2025-01-20.

**Rieger, Marc Oliver, Mei Wang, and Thorsten Hens.** 2015. Risk preferences around the world. *Management Science*, 61(3): 637–648.

**Rosser, John Barkley.** 2003. A Nobel Prize for asymmetric information: The economic contributions of George Akerlof, Michael Spence and Joseph Stiglitz. *Review of Political Economy*, 15(1): 3–21.

**Saarikallio, Suvi, and Jaakko Erkkilä.** 2007. The role of music in adolescents' mood regulation. *Psychology of Music*, 35(1): 88–109.

**Sachan, Abhishek, and Pawan K. Chugan.** 2020. Availability bias of urban and rural investors: Relationship study of the Gujarat State of India. *Journal of Behavioural Economics, Finance, Entrepreneurship, Accounting and Transport*, 8(1): 1–6.

**Sah, Raaj Kumar, and Joseph E Stiglitz.** 1986. The architecture of economic systems: Hierarchies and polyarchies. *The American Economic Review*, 76(4): 716–727.

**Saunders, Edward M.** 1993. Stock prices and Wall Street weather. *The American Economic Review*, 83(5): 1337–1345.

**Saxton, Gregory D., and Richard D. Waters.** 2014. What do stakeholders like on Facebook? Examining public reactions to nonprofit organizations' informational, promotional, and community-building messages. *Journal of Public Relations Research*, 26(3): 280–299.

**Shipman, James E., Quinn T. Swanquist, and Robert L. Whited.** 2017. Propensity score matching in accounting research. *Accounting Review*, 92(1): 213–244.

**Smales, Lee A.** 2016. News sentiment and bank credit risk. *Journal of Empirical Finance*, 38: 37–61.

**Sørensen, Jens F. L.** 2016. Rural–urban differences in bonding and bridging social capital. *Regional Studies*, 50(3): 391–410.

**Sorenson, Olav, and Toby E. Stuart.** 2001. Syndication networks and the spatial distribution of venture capital investments. *The American Journal of Sociology*, 106(6): 1546–1588.

**Souleles, Nicholas S.** 2004. Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the michigan consumer sentiment surveys. *Journal of Money, Credit, and Banking*, 36(1): 39–72.

**Spotify.** 2025. Web API Reference | Spotify for Developers. https://developer.spotify.com/documentation/web-api/reference/get-audio-features. Accessed on 2025-01-20.

**Stafford, Tom, and Alex Grimes.** 2012. Memory enhances the mere exposure effect. *Psychology and Marketing*, 29(12): 995–1003.

**Statista.** 2022*a*. Crowdinvesting in Germany. https://www.statista.com/outlook/dmo/fintech/alternative-financing/crowdinvesting/germany?currency=usd. Accessed on 2022-04-26.

**Statista.** 2022*b*. Crowdinvesting in Mexico. https://www.statista.com/outlook/dmo/fintech/alternative-financing/crowdinvesting/mexico. Accessed on 2022-04-26.

**Statista.** 2022*c*. Crowdinvesting in the United States. https://www.statista.com/outlook/dmo/fintech/alternative-financing/crowdinvesting/united-states?currency=usd. Accessed on 2022-04-26.

**Sumner, Edward M., Rebecca A. Hayes, Caleb T. Carr, and Donghee Yvette Wohn.** 2020. Assessing the cognitive and communicative properties of Facebook reactions and likes as lightweight feedback cues. *First Monday*, 25(2).

**Varian, Hal R.** 2014. Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2): 3–28.

**Vismara, Silvio.** 2016. Equity retention and social network theory in equity crowdfunding. *Small Business Economics*, 46(4): 579–590.

**Wadhwa, Anil, and Suresh Kotha.** 2006. Knowledge creation through external venturing: Evidence from the telecommunications equipment manufacturing industry. *Academy of Management Journal*, 49(4): 819–835.

**Walther, Beverly R., and Richard H. Willis.** 2013. Do investor expectations affect sell-side analysts' forecast bias and forecast accuracy? *Review of Accounting Studies*, 18(1): 207–227.

**Wang, Ying, Hong Cai, Chenguang Li, Zhi Jiang, Lujun Wang, Jian Song, and Jing Xia.** 2013. Optimal caliper width for propensity score matching of three treatment groups: a Monte Carlo study. *PLoS ONE*, 8(12): e81045.

**Xin, Ming, and Vidyanand Choudhary.** 2019. IT investment under competition: the role of implementation failure. *Management Science*, 65(4): 1909–1925.

**Zajonc, Robert B.** 1968. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9(2): 1–27.

**Zbinden, Fabian, and Galia Kondova.** 2019. Economic development in Mexico and the role of blockchain. *Advances in Economics and Business*, 7(1): 55–64.

**Zook, Matthew A.** 2002. Grounded capital: venture financing and the geography of the Internet industry, 1994-2000. *Journal of Economic Geography*, 2(2): 151–177.