Three Essays on Technology and Competition in Bank Lending and Corporate Finance

Yunying Huang

A thesis submitted in fulfillment of the
requirements for the degree of Doctor of Philosophy

Department of Finance
The University of Sydney Business School
THE UNIVERSITY OF SYDNEY

2025

Declaration of Authorship

I declare that this thesis titled, 'Three Essays on Technology and Competition in Bank Lending and Corporate Finance' and the work presented in it are my own. I confirm that:

- This thesis has not been submitted for any degree or other purposes.
- The intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Signed: ______________________________

Yunying Huang

May 2025

Authorship Attribution Statement

This is to certify that I have not included any works (published or unpublished) that I have only made minor contribution to in the main body of this thesis.

I design the study, analyze the data, and write the original draft of the manuscript. The co-authors review and edit the writing.

Signature: ________________________________

Yunying Huang
May 2025

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Signature: ________________________________

Eliza Wu
May 2025

Artificial Intelligence Statement

Following the University of Sydney's requirements for thesis submission, I declare the following use of generative artificial intelligence tools in this thesis:

- **Large Language Models for Data Analysis**: In Chapter 4, I used OpenAI's GPT-4o via API as a large language model (LLM) textual analysis tool to classify Department of Justice antitrust lawsuits. The LLM was used to:
  - Identify procurement-related cases within the corpus of antitrust lawsuits
  - Extract violation types, defendant identities, and product market identifiers
  - Classify misconduct types and legal outcomes

  All LLM outputs are cross-checked by comprehensive human verification to ensure accuracy and prevent hallucinations.

I confirm that:

- All conceptual work, analysis, and interpretation are my own
- The thesis structure, arguments, and conclusions are original
- All AI-generated content was critically evaluated and verified
- The use of AI tools was limited to assistance and did not replace my intellectual contribution

Signed: ______________________________

Yunying Huang

May 2025

Acknowledgements

The completion of this Sydney University Business School doctoral dissertation has involved a long and challenging path punctuated by extensive intellectual development, personal challenges, and many enriching experiences. The transition from a question-driven student to an independent researcher has required not only methodological skill in research and analytical capacity, but also qualities such as grit, perseverance, and the ability to handle ambiguity.

First, I am grateful to my primary supervisor, Professor Eliza Wu, whose mentorship has been the cornerstone of my PhD journey. She provided the perfect balance of guidance and independence, seeing potential where I saw confusion, and celebrating my breakthroughs with genuine joy. Despite her demanding schedule as a senior academic leader, she always made time for my questions. I am equally grateful to Professor Buhui Qiu, whose expertise in empirical finance has been instrumental in shaping this thesis. His strict standards pushed me to continuously improve my work with his remarkable patience guided me through research. I am also grateful to Professor Junbin Gao, whose machine learning expertise provided valuable insights.

Beyond my supervision team, I owe a special debt of gratitude to Professor Iftekhar Hasan, who has been like family throughout this journey. His unwavering support extended far beyond academic guidance—he believed in my potential when I doubted myself and encouraged me during the most challenging moments. Without his care, I would not have found the confidence to pursue an academic career.

Throughout my PhD journey, I have benefited from presenting my research at international conferences, including the European Finance Association Annual Meeting 2023 and the SFS Cavalcade North America 2024, where I received valuable feedback from the global academic community. I have been particularly fortunate to learn from several distinguished scholars who generously shared their time and expertise, including Professor Steven Ongena, Professor Decio Coviello, Dr. Mingze Gao, and Dr Stefano Manfredonia. I am especially grateful to my Postgraduate Research Coordinators, Professor Peter Pham and Associate Professor Jing Yu, for their invaluable guidance and constructive feedback throughout my PhD journey. I also thank the members of the Discipline of Finance at the University of Sydney Business School for creating a supportive research environment. I gratefully acknowledge the financial support provided by the University of Sydney Business School Enhanced Business School Scholarship.

Finally, my deepest gratitude goes to my mom, my girlfriend Ziyue Li, and my dog. To my mother and Ziyue, your love, understanding, and constant support have been my strength during the ups and downs of this PhD journey. This thesis is dedicated to all three of you.

Abstract

This thesis consists of three empirical studies on bank lending and corporate finance, addressing two central policy-relevant issues for financial market dynamism and technology development. The first study explores whether and how banks develop specialized knowledge about corporate technology through their lending activities. We find that loans to firms sharing similar technologies with banks' prior borrowers obtain lower loan spreads. We can rule out product market competition, the value of their technology and ability to innovate, and/or numerous other firm characteristics as alternative explanations. By estimating a structural bank-borrower matching model and exploiting the consummation of bank mergers and acquisitions, we can show that shocks to banks' technology knowledge causally affect loan pricing.

In the second study, we develop a novel AI-driven Co-Lending Graph Neural Network (CoL-GNN) model to capture risk spillovers in syndicated lending markets. Our approach integrates bank characteristics, loan attributes, and network topology to generate a robust and novel spillover risk measure (CLN score) that predicts bank distress and profitability up to two years ahead. Leveraging natural experiments involving credit rating downgrades and the unexpected collapse of Lehman Brothers, we provide causal evidence of risk transmission across financial institutions. The CLN score significantly outperforms traditional risk metrics and standard network centrality measures, particularly for smaller, complex, and privately held banks. Further analysis reveals that revolving credit facilities primarily drive risk spillovers due to their intensive monitoring requirements. By uncovering previously hidden bank vulnerabilities, our deep learning framework offers regulators and market participants an effective early-warning tool for managing spillover risks in highly interconnected banking ecosystems.

In the third study, we swift to a corporate finance to investigate how antitrust lawsuits affect corporations. We focus on government procurement cases, which we identify using large language models. Using a difference-in-differences design, we document that non-defendant firms gain federal contracts, expand employment and sales, and experience positive abnormal stock returns following antitrust filings. We provide evidence that these effects arise from the effective exclusion of defendant firms from the market. While antitrust lawsuits enhance market competition and reduce concentration, we document that the benefits are not evenly distributed: larger, well-established firms benefit more than smaller, financially constrained businesses. We also find no significant impact on government acquisition costs.

Contents

List of Figures

List of Tables

Chapter 2 Tables

Chapter 3 Tables

Chapter 4 Tables

Appendix Tables

Chapter 1

Introduction

Financial markets worldwide face a period of exceptional change characterized by unprecedented technological advances, enhanced interconnectedness, and increasing concerns about market concentration. The widespread application of artificial intelligence, machine learning, and other technologies has profoundly transformed the methods used in finance to value counterparties and allocate resources. These innovations potentially mitigate certain information asymmetries while simultaneously creating new forms of specialized knowledge that remain unevenly distributed across market participants. Despite their growing importance, we have limited empirical evidence on how these technological transformations affect fundamental financial activities, lending, risk management, and market competition.

These technological developments have significant implications for the three critical aspects of modern financial markets. First, financial institutions increasingly accumulate specialized technology knowledge, creating information advantages that influence lending decisions and resource allocation. Second, the growing interconnectedness among financial institutions has created complex embedded networks for the rapid transmission of information and risk, often through channels poorly monitored by traditional systems. Third, competitive market forces emerged with concentrated knowledge and network structures, leading to important questions about the effectiveness and distributive impacts of interventions. Each of these facets reflects an underlying tension in the market that impacts economic outcomes; however, thorough empirical analysis has been hampered by challenges arising from the difficulties in identifying and measuring them.

Asymmetric information, technical specialization, and competitive pressures are the building blocks of lending choice, risk transmission, and firm market performance in modern financial markets. This thesis examines these issues through three empirical studies that investigate critical, yet underexplored, dimensions of banking and corporate finance. We examine how banks use specialized technological knowledge in loan contracting, how interconnection among lenders in syndicated lending networks transfers risk between financial institutions, and how enforcement policy in procurement markets by governments affects firm performance. Together, these three studies contribute novel empirical evidence to the literature on banking and corporate finance, while offering actionable insights for practitioners and regulatory authorities concerned with promoting technological innovation, strengthening financial stability, and ensuring effective competition in markets.

This thesis responds to policy and economic concerns regarding technological innovation, lending relationships, and competition across markets through three complementary empirical studies. In our first investigation, we contribute to long-standing concerns about how knowledge specialization affects financial intermediation by documenting how banks develop expertise by lending to technologically similar firms, revealing that technology similarity between the current borrower and a bank's prior borrowers significantly impacts loan pricing decisions. Our findings challenge the traditional focus on industry specialization by showing that technology-based information advantages operate through distinct channels and have independent economic value. In our second investigation, we advance our understanding of risk transmission in financial networks by developing a novel AI-driven deep learning framework, the Co-Lending Graph Neural Network (CoLGNN), which quantifies how risk propagates through co-lending relationships. Our model identifies directional risk spillovers between lead arrangers and participant banks and offers a comprehensive risk monitoring framework that accommodates public and private institutions. Our third research investigation makes ongoing contributions to the distributional effects of antitrust enforcement by using large language models to analyze procurement-related antitrust cases. In revealing that increased market competition from antitrust enforcement benefits established firms at the expense of entry by small firms, our research challenges conventional expectations on the democratizing impact of policy intervention and presents evidence of friction in market reallocation processes.

Specifically, in the first study, we investigate whether banks develop specialized knowledge about corporate technology through their lending activities, and how this accumulated expertise influences loan contracting. Current research on banks has identified industry-specific knowledge as a key factor in lending relationships (Blickle et al., 2023; Giometti & Pietrosanti, 2022); however, we show that technological expertise transcends industry boundaries and represents an increasingly important dimension of bank specialization. We measure the technology similarity between a borrower and a bank's prior borrowers using patent-based technology profiles, and extend it to quantify the accumulated technology knowledge of banks. Using three-decade U.S. syndicated loan data from 1990 to 2020, we find that firms whose technology profiles are similar to a bank's previous borrowers receive significantly lower loan spreads, translating to meaningful cost savings. A one-standard-deviation increase in technology similarity reduces loan spreads by approximately 4 basis points, equivalent to annual savings of $170,000 for the average loan in our sample. This cost savings is economically considerable, aligning with previous literature documenting the economic significance of similarly sized spread reductions (e.g. Hasan et al., 2014).

To establish causality, we employ two complementary identification strategies: first, a structural bank-borrower matching model demonstrating that the economic surplus is maximized when banks are matched with technologically similar borrowers; second, a difference-in-differences analysis exploiting bank mergers and acquisitions as a quasi-natural experiment that exogenously increases acquirer banks' technological expertise. Our extensive robustness tests rule out alternative explanations such as product market competition, patent value, or industry specialization. In particular, we find that technology similarity affects pricing primarily in relationship loans such as revolving credit facilities and does not affect term loans that are frequently traded in secondary markets, as predicted by the relationship-intensive nature of technological knowledge. Our mechanism tests reveal that technology similarity functions as a channel of information and reduces screening costs because firms having the same type of technology are similarly creditworthy at origination and in the long term. We also document how expertise in specific technologies directly improves the profitability and stability of banks and how banks with specialization in technology are more profitable and have a higher distance to default.

Furthermore, we show that the benefits of technological expertise are most pronounced when information asymmetries are high. Specifically, borrowers with lower technological obsolescence, less product disclosure, and those that are more opaque experience greater loan spread reductions. Similarly, the effect is stronger on less well-capitalized banks and liquidity-constrained banks with stronger resource constraints on screening and monitoring borrowers. These findings show how technology similarity reduces information frictions in lending markets and improves credit allocation efficiency. This study contributes to the growing literature on the economic value of technological innovation by developing a new measure of banks' technological expertise to capture knowledge spillovers between borrowers. These findings have significant policy implications, particularly for addressing financing challenges in emerging technological domains such as green innovation and artificial intelligence. Our evidence suggests that targeted government support during the initial adoption phases could be warranted, as this would allow banks to accumulate technological expertise, eventually reducing information frictions and enabling more efficient capital allocation without the need for ongoing government support for the financing of technology advancement.

In the second study, we introduce a novel AI-driven Co-Lending Graph Neural Network (CoL-GNN) model to capture the risk spillovers in syndicated lending markets. Building on research on interconnectedness among financial institutions (e.g. Acemoglu et al., 2015; Cabrales et al., 2017; Elliott et al., 2014) and the importance of information asymmetry in syndicated lending (e.g. Gopalan et al., 2011; Ivashina, 2009), we close a central gap in risk monitoring systems based predominantly on market data, omitting private banks due to lack of reliable data, and ignoring network transmission mechanisms due to complex topology. We construct a directed co-lending network accounting for the naturally asymmetric co-lending relations among lead arrangers and participant banks, where information asymmetries create unique channels for risk propagation. We employ recent developments in deep learning on graph-structured data to jointly model features of banks and loans as well as the network topology. Using comprehensive data from U.S. syndicated loans and bank holding company information from 1991-2020, we generate a co-lending network risk score (*CLN score*) that serves as an early warning indicator of bank risk, predicting future bank risk and performance up to eight quarters ahead. To improve model interpretability, we further apply the integrated gradient method to quantify the contribution of each input feature on the *CLN score*, thereby identifying key drivers of network-based bank risk.

To validate that our model captures genuine risk transmission rather than common exposure to distressed borrowers, we exploit two quasi-natural experiments: credit rating downgrades and the collapse of Lehman Brothers. Using a stacked-cohort difference-in-differences design, we show that banks directly connected to downgraded banks experience higher subsequent increases in their *CLN score* than control banks without such exposure. Similarly, following Lehman's collapse, a shock originating outside its corporate lending activities, banks with prior co-lending ties exhibit significant increases in their *CLN score*, supporting directional spillovers through the co-lending network. Importantly, the compared *CLN score* exhibits predictive power beyond common risk indices, particularly for private banks. We show that the *CLN score* significantly outperforms standard network centrality measures in predicting future bank risk, suggesting the limitations of conventional network analysis approaches that focus solely on relative node importance. Even when controlling for stock returns, our *CLN score* maintains its predictive power, implying that it captures unique risk information not reflected in market performance.

Our channel tests show that risk propagates primarily through monitoring-intensive revolving credit facilities rather than term loans that are frequently sold in secondary markets, highlighting the importance of ongoing lending relationships in risk transmissions. In addition, we find that more vulnerable banks—those with smaller size, weaker earnings, higher return volatility, and lower capital adequacy—are particularly susceptible to network-driven risk spillovers. The *CLN score* shows robust predictive performance across different sample periods, with significantly stronger effects during the global financial crisis when systemic risk and contagion intensified. This study extends the financial network literature by modeling the topological structure of co-lending relationships, advances bank risk measurement by developing an early-warning system that includes private institutions, and demonstrates the practical application of artificial intelligence in enhancing financial stability monitoring. The findings have important policy relevance for regulators seeking to identify potential channels of risk spillover before they materialize into systemic crises and for banks looking to better manage counterparty exposures in syndicated lending markets.

In the third study, we examine how antitrust enforcement affects corporate performance using government procurement markets as a natural laboratory. An increase in corporate market concentration has been documented in the economic and finance literature (e.g., Autor et al., 2020; Barkai, 2020). These dynamics have reignited interest in antitrust enforcement among policymakers, researchers, and the media, as more robust antitrust measures are being considered as potential solutions to limit corporate market power, protect competition, and promote market fairness. However, empirical evidence on the distributive effects of antitrust enforcement remains surprisingly limited and lacks consensus on whether these legal actions benefit or harm affected markets and their participants. The government procurement market provides an ideal setting to study these persistent questions, since it offers guaranteed demand that eliminates demand-side confounders, it is highly susceptible to supplier collusion, and it provides high-quality data on market participants and structure. We employ large language models to classify the Department of Justice antitrust cases on procurement between the period 2001-2021 and extract in-depth information on product markets, defendants, and misconduct types. We merge these data with granular data on government procurement contracts and government contractor establishment-level data.

Using a difference-in-differences design, we find that antitrust enforcement generates large redistribution effects accruing to non-defendant firms competing in the targeted product markets. Specifically, non-defendant contractors experience a 9% increase in government contract awards compared to those in the control markets, as well as large increases in employment and sales without adverse effects on financial health. This improved access to government contracts translates into higher stock market valuations for publicly traded non-defendant firms, with an average increase in shareholder wealth of $8.45 million. Our results are robust to a series of robustness checks, including controls for industry trends, product market-specific time trends, propensity score matching, and placebo tests with non-procurement establishments. Specifically, our findings indicate that the beneficial impact is stronger for antitrust action to deal with corruption, fraud, and bribery allegations, where defendant firms face formal debarment from government contracting, thereby producing a particularly clear market exclusion resulting in significant redistribution.

At the product market level, we record higher market participation and lower concentration after antitrust enforcement, in line with the theoretical aim of enhancing competition. However, we find no indication that such beneficial competitive effects are passed on in the form of lower government procurement costs—if anything, contract renegotiation becomes more common, which may well raise costs. Furthermore, against the common perception that antitrust enforcement primarily helps small firms and new entrants in the market by lowering the costs of entry, our findings indicate that larger and more established firms predominantly reap the advantages in the altered competitive environment. Small and financially weaker firms also have comparatively lower market shares after enforcement interventions, indicating that increased competition involves significant resources to be in a position to compete in newly opened markets. The evidence in this paper contradicts conventional wisdom about the democratizing impact of antitrust actions and exposes fundamental frictions in market reallocation processes that are more consistent with recent empirical evidence demonstrating how market leadership is reorganized among incumbent firms rather than being contested by new market entrants (Faccio & McConnell, 2020). The paper provides new evidence on the multifaceted distributional effects of antitrust enforcement with important implications for competition policy, procurement law, and policy on helping small businesses.

The remainder of this thesis is structured as follows. Following the introduction, Chapter 2 examines the role of technological expertise in bank lending decisions and its influence on loan pricing. Chapter 3 investigates how risk propagates through financial institutions' co-lending relationships using a novel graph neural network framework. Chapter 4 analyzes the impact of antitrust enforcement on corporate performance in government procurement markets. Chapter 5 concludes the thesis by synthesizing the findings and discussing broader implications for financial intermediation, risk management, and market competition.

Chapter 2

Borrower Technology Similarity and Bank Loan Contracting

2.1 Introduction

In spite of technological innovations being a key driver of economic growth, the role of borrowers' technology profiles in bank lending decisions remains an open issue. In the U.S., corporations are the main loci for technological innovations, which account for the bulk of R&D expenditures (Chava et al., 2017). The financing of innovative firms is therefore an important function of the financial sector in which banks play a critical role, for example through the syndicated loan market. However, the information asymmetry and adverse selection faced by banks increase the costs in loan screening and monitoring, raising the cost of bank financing for corporate borrowers and the likelihood of under-funding productive firms (Greenwood et al., 2010). The inherently risky high-tech nature of innovative firms exacerbates the under-funding problem due to their high opacity and hence the need for tech-specific knowledge to assess credit risk.

To alleviate these funding challenges, greater information disclosure for borrowers is an option (Saidi & Žaldokas, 2021). Alternatively, the bank's accumulated knowledge from past lending to firms sharing similar technologies could imply a source of value via cost savings in loan screening and monitoring. Such an information advantage may rationally lead to a degree of bank specialization in technologies that overrides banks' industry specialization because firms from different industries can share similar technologies. For example, Hyundai Steel Co. (a steel-making company), Deutsche Post AG (a mail services company), and Berkshire Grey Inc. (a robotics company) all have patents granted in the Cooperative Patent Classification (CPC) class B65G "Transport or Storage Devices".[1] As shown in Figure 2.1, the average technology similarity

[1] CPC is a patent classification system jointly developed by the United States Patent and Trademark Office (USPTO) and European Patent Office (EPO) which has replaced the United States Patent Classification (USPC) in 2013. See our later discussions for more details. The three companies Hyundai Steel Co., Deutsche Post Ag, and Berkshire Grey Inc. have distinct Standard Industrial Classification (SIC) codes 3312, 7389, and 3569, respectively, but all have some patents related to conveyors (group 15 under class B65G).

between a bank's borrower and the bank's prior borrowers has shown a marked increase since the 2007-09 Global Financial Crisis, reaching a historical high in 2020. This trend reflects banks' increasing technological specialization. Moreover, it underscores the value implications of technological specialization, as we expect and find that banks—especially medium and small-sized ones—specialize in a few technological classes to sharpen their competitive edge and boost performance.

[Insert Figure 2.1 about here]

Further, borrowing firms' technology diversity (Hsu et al., 2018) has been declining substantially since the 1990s, as shown in Figure 2.2. This indicates firms' patent portfolios have become increasingly concentrated in specific technological areas over time. This trend suggests that banks' accumulated technological knowledge from past lending experience should be particularly valuable for reducing information asymmetry and screening costs when evaluating new borrowers with similar technological profiles. In this paper, we examine whether this accumulated technological expertise translates into tangible benefits in loan contracting terms, specifically through lower loan spreads, and investigate whether such bank-borrower matching is economically optimal.

[Insert Figure 2.2 about here]

We rely on borrower patent information to measure, as in Jaffe (1986) and Bloom et al. (2013), the average pair-wise technology similarity between the prospective borrower and its bank's prior borrowers in recent years. While past information accumulation by the bank could lead to a lowering of spreads, the net impact, however, remains ex-ante unclear and warrants an empirical study for two reasons. First, borrowers with similar technologies may face greater product market competition as a result of the technologies being more likely to be applied in related product markets (e.g., Bereskin et al., 2023). A high technology similarity between the bank's current and past borrowers could also imply more industry segment concentration in the bank's loan portfolio, undermining potential diversification benefits (e.g. Boyd & Prescott, 1986; Diamond, 1984).[2] Second, banks may extract rents based on their accumulated information advantage (e.g., Rajan, 1992) instead of passing on to the borrower the cost savings from the reduced due diligence needed in assessing the borrower's technology profile. These possible channels could lead to a positive relationship between technology similarity and loan spreads, but are not supported by our empirical results. We find that increased industry segment concentration is positively associated with the loan spread, but a higher borrower technology similarity causally reduces the loan spread. Further, we show that offering lower loan spreads

[2] The technology similarity of a prospective borrower and the bank's prior borrowers is positively correlated with their industry segment similarity with a correlation coefficient of 0.21 in our sample.

to borrowers with a high technology similarity to the banks' prior borrowers is economically optimal as the total economic surplus for both banks and borrowers is expanded. We establish this finding using a structural bank-borrower matching model similar to Fox (2017, 2018) and Schwert (2018).

Specifically, at each loan's origination, we compute the technology similarity of the borrower and its bank's prior borrowers in the past five years using the average pair-wise technology similarity measure from Jaffe (1986) and Bloom et al. (2013). The pair-wise technology similarity is a cosine similarity of the firm's technology profiles measured by the proportions of patents granted in each of the Cooperative Patent Classification (CPC) technology classes. In this application our technology similarity measure does differ somewhat from the technology spillover measure deployed in Bloom et al. (2013) and Qiu and Wan (2015). These prior studies captured the firm-year technology similarity to the whole economy. Instead, our bank-firm-year level technology similarity measure represents a bank's time-varying technological expertise specific to each borrower.

Using a comprehensive sample of U.S. syndicated loans from January 1990 to December 2020, we show that loans to firms with a higher technology similarity with banks' prior borrowers have lower loan spreads and total costs. A one-standard-deviation increase in technology similarity is associated with approximately a 4 basis points (bps) reduction in loan spreads. This reduction is economically meaningful for a mean loan spread of 205 bps (175 bps), being equivalent to an annual loan cost saving of $170,000 ($66,000) for a mean (median) loan of $425.14 (166) million in our sample. Such effect remains even after controlling for borrower's product market rivalry (Hoberg & Phillips, 2016; Hoberg et al., 2014), product market segment similarity (Bereskin et al., 2023; Bloom et al., 2013), prior lending relationship (Bharath et al., 2011), and borrower's patent stock and value (Chava et al., 2017; Kogan et al., 2017).

Innovative borrowers such as green firms may take this opportunity to reduce their financing costs, which have been shown to receive favorable loan recommendations from bankers (Bu et al., 2023).

As for other loan characteristics beyond loan spread, we find that loan size, maturity, and secured status for example are not affected by technology similarity, possibly because these loan characteristics are more driven by corporate needs and assets, and are thereby pre-determined. Hence, the bad control problem that may exist when including these loan terms as independent variables may be less acute; yet, we confirm that their exclusion leaves the main estimates we report mostly unaltered.

In robustness tests, we rule out the possibility that bank industry specialization drives our results by controlling for lender times industry times year fixed effects. We show that our results are also robust to alternative industry classifications, alternative window sizes used for measuring firm patent classes, as well as alternative constructs of borrower technology similarity addressing potential attenuation bias. Further, even when we exclude firms without patents, loans originated by banks with few recent borrowers, or loans to major customers, we find the same inference. Our results remain largely unchanged after controlling for the participation of non-bank lenders, who may be invited to join loan syndicates due to their specialized technology knowledge. We additionally conduct a placebo test and find no relationship between loan spreads and the technology similarity of a borrowing firm with the bank's future borrowers.

To investigate how banks benefit from the accumulated technology knowledge, we show that lending to borrowers sharing technologies with banks' prior borrowers is associated with higher bank future return on assets (ROA) and distance to default. We then show that technology similarity is indeed informative about borrower's creditworthiness in both contemporaneous and predictive regressions. A higher technology similarity with prior borrowers is negatively associated with the absolute difference between their credit risks measured by the Altman Z-score, Merton (1974) default probability, and their debt service capabilities measured by profitability and cash holding. These results suggest that firms with similar technologies tend to have similar credit risks both at and post loan origination, which could lower banks' screening and monitoring costs, leading to reduced loan spreads.

However, establishing a causal link between a borrower's technology similarity with its bank's prior borrowers and loan spreads is empirically challenging. Given that our technology similarity measure is specific to each bank-borrower matched pair (at loan origination), it is difficult to find an exogenous shock directly to the technology similarity measure, which alters the borrower's technology profile but does not affect the bank-borrower matching or the borrower's fundamentals and future business prospects. An instrumental variable regression approach is also precluded as an instrumental variable that correlates with technology similarity but not with borrower characteristics that are important determinants of bank loan pricing is hard to come by. To the best of our knowledge, prior studies such as Bereskin et al. (2023), Lee et al. (2019), and McLemore et al. (2022) also encounter similar challenges. We therefore explore two alternative methods for our identification strategy.

We begin by estimating a structural matching model, building on the frameworks of Fox (2017, 2018) and Schwert (2018), to identify the key factors influencing bank-borrower matching decisions in the absence of unobservable counterfactual assignments. Specifically, our model assumes that both banks and borrowers seek to maximize their respective values. Loan contracts involve transfer payments, such as interest rates and fees. However, due to the absence of data on unobservable counterfactual loans, we establish an inequality condition for simultaneous value maximization that abstracts from transfer payments. Under the assumption that realized bank-borrower matches (i.e., observed loans) generate higher value than unobserved counterfactual matches, we apply a semi-parametric estimation approach to analyze the determinants of loan formation at the bank-borrower level (Schwert, 2018). Our results suggest that overall economic surplus for both banks and borrowers is maximized when banks engage with borrowers whose technological profiles align with those of their prior borrowers. Furthermore, borrowers with technology profiles similar to those of banks' previous clients tend to receive loans at lower spreads, a pattern that is economically optimal for both parties.

To complement the matching model, we then use difference-in-differences (DiD) estimations. While it is challenging to find an exogenous shock to borrower technology similarity, we alternatively exploit reasonably exogenous shifts to the stock of banks' accumulated technological knowledge using bank mergers and acquisitions (M&As), which increase acquirer banks' technology knowledge but do not affect their *extant* borrowers' creditworthiness. Using stacked cohorts of treated banks with M&A activities and control banks matched via propensity scores that have no M&A activities in the five-year window around the M&A event year, we find that loans extended to extant borrowers by acquirer banks after the M&A event are significantly cheaper than loans issued before, as compared to the loans originated by control banks.

Additionally, we show that the documented beneficial effect of borrower technology similarity on loan spreads is stronger for borrowers who are ex ante more opaque, especially in their innovation profiles, with lower technological obsolescence and lower product disclosure. These opaque borrowers require greater due diligence and screening efforts, enhancing the value of bank technological expertise in reducing information asymmetry. Banks are more willing to pass on the cost savings to less-risky borrowers with lower credit risk, leverage, or higher profitability. Further, we show that the effect is stronger for loans originated by smaller, less-capitalized, or less-liquid banks. These results are consistent with our conjecture that such banks have more limited resources to screen and monitor borrowers and hence may place higher value on their accumulated technology knowledge from past lending.

This study contributes to the literature in four directions. Firstly, we contribute to the burgeoning literature on the implications of technological innovation (e.g., He & Tian, 2020). For example, recent studies identify the impact of technology spillover on product market rivalry (Bloom et al., 2013), cash holdings (Qiu & Wan, 2015), technology styles (Byun et al., 2021) and innovation outputs (Matray, 2021). Importantly, to assess borrower's technological profiles We extend the technology measures developed by Bloom et al. (2013), Hall et al. (2001), and Kogan et al. (2017) in the context of bank lending by proposing a measure of a bank's time-varying technological expertise specific to each borrower. In doing so, we shed new light on the economic value of firms' technological innovation in financing decisions.

Secondly, we contribute to the literature on the interplay between financial intermediaries and firm technological innovation. On the one hand, many extant studies investigate how the banking industry affects firm innovation outputs (e.g., Chava et al., 2013; Cornaggia et al., 2015).

On the other hand, Chava et al. (2017) find that firms with more innovation outputs receive cheaper bank loans. Saidi and Žaldokas (2021) find that enhanced technology disclosure improves banking competition and helps reduce loan costs for borrowers. We extend the work of Mann (2018) on the intangible and collateral value of firms' technology profiles by showing that the similarity across different borrowers' technology profiles is informative for credit risk assessment.

Thirdly, our study contributes to the relationship lending and bank loan contracting literature. We highlight the value of accumulated technology knowledge from banks' past lending, thereby extending the literature on relationship lending (e.g. Bharath et al., 2011; Demiroglu & James, 2010; Ioannidou & Ongena, 2010; Murfin, 2012). Our study provides further evidence for the crucial role of information asymmetry on bank loan contracting (e.g. Demiroglu et al., 2021; Gustafson et al., 2021; Ivashina, 2009; Sufi, 2007), and sheds new light on the relation between bank's private information advantage and rent extraction (e.g. Schenone, 2010). We extend the prior literature on various alternative determinants of bank loan contracting (e.g. Campello & Gao, 2017; Carvalho et al., 2023; Hasan et al., 2014) by showing that bank loan costs are dependent on the shared technological knowledge across banks' borrowers.

Lastly, our study contributes to the literature on bank lending specialization. Extant studies focus mostly on the banks' specialization in borrower industries and find that banks' industry-specific knowledge leads to laxer contract terms for borrowers in the industry (Giometti & Pietrosanti, 2022), reduces loan spreads and defaults (Blickle et al., 2023, 2024), and reduces both individual and systemic bank risk (Beck et al., 2022). Gopal (2021) shows that banks increase credit supply to borrowers whose collateral they have specialization in. We document that banks' accumulated technology knowledge reduces loan costs and that matching borrowers to banks with a clear specialization in the borrower's technology increases total economic surplus. More importantly, even when accounting comprehensively for bank industry specialization, bank technology specialization continues to explain salient loan conditions.

Our findings also have significant policy implications, particularly for the financing of innovative technologies. For instance, the green transition is clearly a juncture in time when innovative technology is urgently needed but is not yet very familiar to banks. Due to the high information asymmetry involved, banks may be hesitant to finance technologies they have not previously encountered, leading to an under-funding problem (Greenwood et al., 2010). To address this, government support may be introduced to encourage bank financing for technological innovation by, for example, subsidizing initial loans. As banks gain experience and knowledge, loan rates are likely to decrease, allowing for the phasing out of government subsidies. This approach, akin to policies aimed at small business financing, could be particularly beneficial for new technology adoption.

The rest of the paper proceeds as follows. Section 2.2 develops the hypotheses. Section 2.3 discusses our data and key measurements. Section 2.4 presents our baseline results. Section 2.5 discusses the identification challenges and investigates the economic mechanisms. Section 2.6 presents some additional results. Section 2.7 concludes.

2.2 Hypothesis development

Information asymmetry between banks and borrowing firms plays a major role in bank loan contracting (e.g., Ivashina, 2009; Sufi, 2007). Banks invest substantial resources in loan screening and monitoring to collect and assess information relevant to prospective borrowers' creditworthiness (e.g., Agarwal & Hauswald, 2010; Botsch & Vanasco, 2019; Gustafson et al., 2021; Rajan et al., 2015; Schenone, 2010; Sufi, 2007). Beyond the borrowing firm's fundamental financial metrics such as leverage, profitability, and so on, non-financial firm characteristics have also been receiving increased attention. One strand of banking literature, for example, focuses on intangible capital, including technology capital, and its impact on bank loan contracting (e.g. Agarwal & Ben-David, 2018; Hasan et al., 2017; Hollander & Verriest, 2016; Karolyi, 2018). Specifically, firm technology capital is related to its cash holding (Qiu & Wan, 2015), governance structure (Frydman & Papanikolaou, 2018), creditworthiness (Dannhauser, 2017), competitive scope and long-term growth (e.g. Glasso & Schankerman, 2013; Romer, 1990). Chava et al. (2017) find that an exogenous enhancement of intellectual property protection and patent value results in lower bank loan costs. Mann (2018) identifies that improved pledgeability of patents contributes to the use of debt financing. Saidi and Žaldokas (2021) show that increased information disclosure on borrowers' technology profiles reduces the cost of switching lenders and results in a more competitive loan market and lowers the cost of bank financing.

However, firm technology capital is inherently difficult to evaluate due to its opacity and limited redeployability, exhibiting higher knowledge barriers compared to fundamental or other soft information sources (e.g., Hall & Lerner, 2010; He & Tian, 2013). Therefore, the accumulated knowledge from banks' prior experience in lending to firms with certain technology profiles is arguably valuable and relevant for future bank financing.

To the extent that such accumulated technology knowledge reduces adverse selection and information asymmetry, we expect that banks are more likely to lend to borrowers sharing similar technologies with the banks' prior borrowers and they pass on the cost savings from reduced screening and monitoring to their borrowers (Bharath et al., 2011; Blickle et al., 2024). We therefore develop our Hypothesis 1 as follows:

Hypothesis 1. Banks charge lower loan spreads for borrowing firms with a higher technology similarity with the banks' prior borrowers.

Nevertheless, we acknowledge that Hypothesis 1 is not clear-cut for two reasons. First, borrowers sharing similar technology profiles may face greater product market competition, which is known to cause higher bank loan costs (e.g., Valta, 2012). It could also imply an increased industry concentration of bank loan portfolios as borrowers sharing similar technologies are likely to compete in the same or related industries, undermining the potential diversification benefits to the bank. Second, banks may use their information advantage to extract rents (Rajan, 1992), since the borrowing firm could face worse outside options as it represents greater information asymmetry to other lenders who have less experience in lending to firms with similar technologies. Therefore, it remains an empirical question whether a borrower with a higher technology similarity as the bank's prior borrowers receives lower loan spreads.

Additionally, we consider the role of technology similarity from both the borrower and bank sides. From a borrower's perspective, we expect borrowers who are ex ante more opaque to receive lower loan spreads when they share similar technologies with banks' prior borrowers. This is because the accumulated technology knowledge from banks' prior lending activities reduces adverse selection and information asymmetry, which arguably should be more valuable when borrowers are more opaque, especially in their technology profiles rather than in other aspects. For example, the importance of banks' technological expertise increases when borrowers' technologies are cutting-edge and up-to-date. More advanced technologies are typically harder to fully comprehend and of higher value, demanding greater screening efforts from lenders. In this case, banks' knowledge of the borrower's technology plays a more important role in reducing due diligence costs and hence leads to relatively lower loan spreads. Firms with a high technological obsolescence are shown to have lower growth and productivity (Ma, 2021), and hence may not receive reduced loan costs when borrowing even if they share similar technologies with banks' prior borrowers. Similarly, Cao et al. (2018) show that firm product disclosure contains proprietary information and is related to technological peer pressure. As a result, more opaque borrowers that have lower product disclosure but share similar technologies with the banks' prior borrowers are more likely to receive reduced loan spreads. Therefore, we form our second hypothesis as below:

Hypothesis 2. Banks charge lower loan spreads for borrowers that share similar technologies with the banks' prior borrowers when borrowers are more opaque with lower levels of technological obsolescence or product disclosure.

Further, we expect less-risky borrowers with lower credit risk, lower leverage and higher profitability to receive lower loan spreads from banks when they have a higher technology similarity with the banks' prior borrowers. These borrowers have better capabilities in servicing debt, to whom banks may be more willing to pass on the cost savings due to the reduced due diligence needed. This helps with retaining and attracting good-quality borrowers. On the other hand, while banks' accumulated technological knowledge is arguably valuable when screening lower-quality borrowers with higher risks, it is ambiguous whether the cost savings due to such knowledge will be passed on to these borrowers more than to the higher-quality borrowers. If more risky firms receive greater benefits when borrowing from banks whose prior borrowers share similar technologies, an adverse selection problem seems imminent. Therefore, we form our third hypothesis as below:

Hypothesis 3. Banks charge lower loan spreads for less risky, less levered, and more profitable borrowers that share similar technologies with the banks' prior borrowers.

For banks, because the accumulated technology knowledge reduces the expensive resources required for evaluating borrower technology profiles, we conjecture that more constrained banks are likely to benefit more from such accumulated knowledge. Specifically, we expect that smaller banks with less technology expertise would value borrower technology similarity relatively more. Figure 2.1 has provided some initial evidence. Similarly, less-capitalized banks may be more risk-averse and value more the certainty from borrowers of higher technology similarity to their prior borrowers. Less-liquid banks are likewise more constrained by resources and may value more the cost savings from the accumulated technology knowledge. As Gustafson et al. (2021) show that the value of information obtained is negatively related to loan spreads, we expect that smaller, less-capitalized, or less-liquid banks are more willing to reduce loan spreads for borrowers with a higher technology similarity to their existing borrowers as this allows banks to capitalize more on their accumulated knowledge. It is possible, however, that more constrained banks do not pass on the cost savings to borrowers. The banks may trade off the short-term gains from retaining the surplus and the potential losses due to their customer switching banks. Given that less-bank-dependent firms (e.g., with bond market access) are more likely to borrow from less-capitalized banks (Schwert, 2018), we expect that the more constrained banks, due to concerns of customer retention, will charge lower loan spreads when borrowing firms have a higher technology similarity with their prior borrowers. We thus have the following fourth hypothesis:

Hypothesis 4. Smaller, less-capitalized, or less-liquid banks charge lower loan spreads for borrowing firms that share similar technologies with the banks' prior borrowers.

2.3 Sample and variable construction

2.3.1 Measuring technology similarity

To capture banks' time-varying knowledge of borrowers' technology profiles, we require a measure that varies by bank-firm pair and over time. Therefore, we use the technology similarity of a borrower with the bank's recent borrowers to gauge the bank's knowledge of this particular borrower's technology profile at loan origination. Because this technology similarity measure is computed for each loan, it is naturally specific to the pair of bank and borrower at loan origination and offers some important features. It allows for not only a bank having different levels of knowledge in different technology classes at a given time, but also a bank with time-varying knowledge in a given technology class. Specifically, we compute the technology similarity measure in two steps. First, at each loan's origination, we compute the pairwise technology similarity between the borrower and each of the bank's recent borrowers. Second, we aggregate the pairwise similarities to the bank-borrower level as our measure of the bank's knowledge of the borrower's technology at the time of loan origination.

In the first step, we compute the pairwise technology similarity, consistent with Jaffe (1986), as the spatial proximity of two firms' technology profiles measured by patents granted and their technology classifications. Using this method, for example, Bloom et al. (2013), Byun et al. (2021), and Qiu and Wan (2015) obtain the technology similarity between each firm and the whole economy to measure technology spillover. Lee et al. (2019) and McLemore et al. (2022) sum over the technology similarity within pre-specified firm sets to measure the technology linkage between firms.

Empirically, we collect all firms' patents granted and their technology classifications using the Cooperative Patent Classification (CPC), a classification system jointly developed by the United States Patent and Trademark Office (USPTO) and European Patent Office (EPO).[3] The pairwise technology similarity between a borrowing firm i and a bank's prior borrower j, as at the origination time t, is the normalized uncentered cosine similarity between the patent portfolio of firm i at time t and the portfolio of firm j at its prior borrowing time τ:

$$\textit{Pairwise Technology Similarity}_{ijt\tau} = \frac{(\boldsymbol{T}_{it}\boldsymbol{T}'_{j\tau})}{(\boldsymbol{T}_{it}\boldsymbol{T}'_{it})^{0.5}(\boldsymbol{T}_{j\tau}\boldsymbol{T}'_{j\tau})^{0.5}} \tag{2.1}$$

where T_{it} is a k-dimensional vector of firm i's proportions of patents granted in each of the k technology classes over the past five years,[4] where the value of each element in T_{it} is strictly between zero and one. We assume that a bank learns the most about a borrower's technology profile at loan origination, and hence we use the patent portfolio of prior borrowers at their respective borrowing time τ in the five-year window, i.e., $\tau \in [t-5, t-1]$, instead of time t, in computing the pairwise technology similarity. Figure 2.A1 in the Appendix provides a graphical illustration.

[3] In the later section of data sources we explain in detail the technology classifications.

[4] The total number of technology classifications, k, varies with time. We use a k of 660. We follow the standard innovation literature to use a five-year window to allow for some accumulation of technology stock (e.g. Bloom et al., 2013; Lee et al., 2019). Our results are robust to alternative window sizes such as 1-year, 3-year, 7-year and all-history windows as shown in Table 2.A1 in the Appendix. All of our empirical results still hold if we use the USPTO classification system using data before 2013 (the USPTO technology class system was replaced by the newer CPC technology class system in 2013).

In the second step, our key variable of interest, the technology similarity between the borrower firm i and the bank b's recent borrowers is the average pairwise similarity:[5]

$$Technology\ Similarity_{ibt} = \frac{1}{N}\sum_{j=1}^{N} Pairwise\ Technology\ Similarity_{ijt\tau} \tag{2.2}$$

where N is the total number of the loans that bank b serves as the lead bank in the five-years leading up to time t. Note that we do not exclude the borrowing firm from the sample of the bank's recent borrowers because a firm's technology profile varies over time, i.e., we allow T_{it} and $T_{j\tau}$ in Equation 2.1 to represent the patent portfolio of the same firm at different times. To a certain extent, this could cause a mechanical correlation between technology similarity and relationship lending should the firm's technology profile be stable over time. However, there are two reasons why it is less of a concern. First, the technology profile of a firm changes as measured by recently granted patents, and lending banks still face increased screening and monitoring costs if the same borrower experience a change in its technology profile. Second, our measure is averaged across all pairs of recent borrowers and the focal firm. As long as the bank lends to more than one firm in the past five years, the concern of a mechanical correlation between technology similarity and relationship lending due to the inclusion of the focal firm in the group of recent borrowers is mitigated.[6]

Our technology similarity measure differs from the well-known technology spillover measure (Bloom et al., 2013; Qiu & Wan, 2015) which captures a firm's technology similarity to the whole economy. Our measure also differs from firm-level technological obsolescence measure by Ma (2021) based on patent citations. While a firm may have highly-valued frontier technologies, they are not necessary familiar to the bank. As discussed earlier, we calculate a borrower technology similarity measure specific to each bank-borrower pair at loan origination to capture each bank's time-varying technological expertise on a specific borrower. In the later section of robustness tests, we present and discuss alternative constructs of technology similarity measure.

2.3.2 Sample and summary statistics

We collect the patent data from the United States Patent and Trademark Office (USPTO) for the period from 1985 to 2020. We use a five-year window for computing technology similarity. To match patent assignee names to Compustat firms, we rely on the Kogan et al. (2017) (KPSS) linking map. We further cross-check the linkage using other mappings such as Stoffman et

[5] All our results are robust to the alternative use of technology similarity weighted by loan amount.

[6] Further, we control for past lending relationships, following the relationship lending literature (Bharath et al., 2011), in our regression analysis. Our results are robust to the use of a 3-year or 5-year relationship window and to alternative relationship strength measures based on the prior number of loans or total loan amount.

al. (2020).[7] Our final sample of patents matched with the Center for Research in Security Prices (CRSP)/Compustat firms consists of 2,331,801 unique observations. Notably, we use the patent grant date when identifying a firm's patent portfolio.[8] We obtain patents' technology class classification data directly from the USPTO PatentsView, which regularly updates patent information including classifications, inventors, and organizations.[9] Prior studies have relied on the United States Patent Classification (USPC) made available by the USPTO (e.g., Bloom et al., 2013; Byun et al., 2021; Hsu et al., 2014; Lee et al., 2019; Qiu & Wan, 2015). However, since the USPTO officially moved to the Cooperative Patent Classification (CPC) system on January 1, 2013, most studies are based on limited sample periods up to 2012. We use the CPC classification to incorporate more recent patent information enabling the expanded identification of 660 technology classes.[10]

Our bank loan sample is sourced from the Thomson Reuters Loan Pricing Corporation (LPC) DealScan database for the same sample period. Specifically, we include all U.S. dollar-denominated loan facilities to U.S. borrowers that can be linked to Compustat using the DealScan-Compustat link table by Chava and Roberts (2008). We use Schwert (2018)'s updated DealScan lender link table to obtain the lender's Compustat identification. We remove utility and financial firms and loans with missing observations on all-in-drawn spread, loan maturity, loan amount, and other necessary loan information. Following Ivashina (2009), a bank in the loan syndicate is classified as the lead bank if it is the administrative agent (if defined), or if it acts as the agent, arranger, bookrunner, lead arranger, lead bank, or lead manager. If a loan has multiple lead banks identified, we assign the one with the highest technology similarity as the lead bank.[11] We collect borrower firms' financial information from Compustat, industry-specific sales data from Compustat Segment, and market data from CRSP. We obtain lender banks' information from Compustat Bank.

[7] Several firm-patent mapping tables are available. For example, the National Bureau of Economic Research (NBER) patent database by Hall et al. (2001) is used in He and Tian (2013) and Tian and Wang (2014), but ends in 2005. Kogan et al. (2017) (KPSS) provide an updated mapping table to 2020, which is another well-known concordance file. Stoffman et al. (2020) (SWY) publish a similar linkage dataset updated to 2020. Given the challenge of fuzzy matching patent assignee names and firm names, we rely on the KPSS dataset primarily and use SWY as a cross-validation and to fill missing mappings wherever possible to ensure maximum accuracy.

[8] The American Inventor Protection Act (AIPA) enacted in 1999 mandates that patent information becomes public at either grant date or 18 months after the patent application date, which significantly affects the banking relationship of innovative firms (Saidi & Žaldokas, 2021). Nevertheless, Lee et al. (2019) argue that a significant proportion of patents might eventually fail to be issued, resulting in actually no innovation outputs for firms. Lee et al. (2019) suggest that using patent grant date would be a conservative choice to assess firm technology profiles.

[9] See, https://www.uspto.gov/ip-policy/economic-research/patentsview.

[10] We identify 660 technology classes similar to McLemore et al. (2022) who identify 642 technology classes. The difference is due to new classifications added by the CPC over time, which does not have any material impact on the technology similarity measure.

[11] Bharath et al. (2011) use a similar approach. In studying the lending relationship and loan contract terms, they choose from the multiple lead banks the one that yields the strongest lending relationship with the borrower and assign it to the loan.

Our final sample consists of 36,166 loan facilities originated by 110 bank holding companies (banks, hereafter) identified by Compustat Bank to 5,522 distinct firms from 1990 to 2020.[12] Given that technology similarity may correlate with product market competition, we control for borrowers' product market rivalry measured by Hoberg and Phillips (2016) and Hoberg et al. (2014). Additionally, we control for product market similarity using a segment similarity measure as in Bereskin et al. (2023) and Bloom et al. (2013), defined as $\frac{1}{N}\sum_{j=1}^{N}\frac{(S_{it}S'_{j\tau})}{(S_{it}S'_{it})^{0.5}(S_{j\tau}S'_{j\tau})^{0.5}}$, where S_{it} is the vector of firm i's proportions of sales in each industry segment and other symbols follow previous notations in Equation 2.1. To an extent, this segment similarity measure also captures the contribution of the borrower to the bank's loan portfolio industry segment concentration. For other borrower characteristics, we include borrower size, leverage, cash holding, profitability, market-to-book ratio, Altman Z-score, and a dummy variable for whether the borrower has a credit rating (e.g. Bharath et al., 2011; Carvalho et al., 2023; Hasan et al., 2014). We control for the relationship lending as in Bharath et al. (2011). For bank loan characteristics, we include the loan size, maturity, and a dummy variable for whether the loan is secured. The merged sample requires all firm, loan, and bank characteristics to be non-missing. Table 2.1 reports the summary statistics. Definitions of the variables and data sources are provided in Table 2.A1 in the Appendix. We winsorize all continuous variables used in the analyses by year at the 1$^{\text{st}}$ and 99$^{\text{th}}$ percentiles.

[Insert Table 2.1 about here]

Table 2.1 reports the summary statistics of our sample. Our key variable, technology similarity, ranges from 0 to 1 by construction, with a value of 0 indicating no similarity and 1 the same technology profile. The technology similarity in our sample has a mean of 3.9% and a standard deviation of 6%, with a skewed distribution due to most borrowers sharing no similar technologies. The distribution statistics of our technology similarity measure are comparable to Bereskin et al. (2023), who document a mean of 4.3% and a standard deviation of 11% using a firm-by-firm pairwise sample. The segment similarity variable exhibits a relatively similar distribution to technology similarity with a mean (median) of 6.3% (5.2%). This is consistent with prior studies documenting that the scale of technology similarity and segment similarity should be consistent (Bloom et al., 2013).

The key dependent variable is the cost of bank loans measured by the all-in-drawn loan spreads. The mean (median) of loan spreads is 205.31 bps. The average (median) loan size is \$425.14 (166) million U.S. dollars. The average (median) maturity is 48.19 (60.00) months. The loan characteristics are consistent with prior literature (e.g. Campello & Gao, 2017; Carvalho et

[12] We aggregate 533 unique lenders from DealScan to 110 bank holding companies. The DealScan database starts from 1984. We restrict our sample period to match the product market competition data from Hoberg et al. (2014) and Hoberg and Phillips (2016). Nevertheless, our results are robust if our sample starts from 1984, removing Hoberg and Phillips (2016) product market competition measures.

al., 2023; Hasan et al., 2014; Hollander & Verriest, 2016). For example, Campello and Gao (2017) reports average (median) loan spreads of 179.16 (175.00) bps and average (median) loan maturity of 46 (48) months. Hasan et al. (2014) reports average (median) loan spreads of 167 (150) bps and average (median) loan size $487 (150) million U.S. dollars.

In terms of borrower characteristics, the summary statistics show that we have selected a comparable sample of borrowers to those examined in the literature. For example, the average borrower firm has a book value of total assets of $6.566 billion dollars in our sample. The average (median) natural logarithm of the total asset size of borrowers in our sample is 7.074 (7.101), and the average (median) leverage ratio is 0.314 (0.297). Similarly, Carvalho et al. (2023) reports the mean (median) borrower size of 8.04 (7.99) and the mean (median) borrower leverage of 0.37 (0.32). Our bank characteristics are also comparable to previous studies (e.g. Acharya & Mora, 2015; Schwert, 2018).

2.4 Main results

2.4.1 Baseline model and results

Our first hypothesis postulates that banks charge lower loan spreads for borrowing firms with a higher technology similarity with the banks' prior borrowers due to cost savings from accumulated technology knowledge. To empirically test whether bank's technology knowledge of a borrower, measured by the borrowers' technology similarity with the bank's prior borrowers at loan origination, reduces loan costs, we start by estimating the following baseline regression:

$$\begin{aligned}\ln(\mathit{Loan\ Spread}_{i,l,t}) = \beta_1\, \mathit{Technology\ Similarity}_{i,l,t} + \beta_2 X_{i,t-1} + \beta_3 \Gamma_{l,t} \\ + \text{Fixed Effects} + \varepsilon_{i,l,t}\end{aligned} \tag{2.3}$$

where $\mathit{Loan\ Spread}_{i,l,t}$ is the all-in-drawn spread of loan l for the firm i at year t. $X_{i,t-1}$ represents the vector of borrower firm characteristics as at year $t-1$ and $\Gamma_{l,t}$ the vector of loan characteristics. Specifically, for borrower characteristics, we control for firm size, leverage, credit risk measured by the Altman Z-score, profitability, market-to-book ratio, cash holding, and whether the borrower has received a credit rating. For loan characteristics, we control for loan size, maturity, and whether it is secured. We control for the bank-borrower prior lending relationships following (Bharath et al., 2011), as relationship lending can also reduce the screening and monitoring costs and hence lower loan spreads. We include borrower industry times year fixed effects to capture unobservable time-varying borrower industry heterogeneity, given that technology innovation and adoption could pertain to industry sectors and possibly cluster by time.[13] Additionally, we control for bank fixed effects, loan type fixed effects, and loan purpose fixed effects. Heteroskedasticity-robust standard errors are clustered by borrower firm.[14]

[Insert Table 2.2 about here]

Table 2.2 presents the results for our baseline model. Consistent with our Hypothesis 1, we find that loans to borrowers with a higher technology similarity with the bank's prior borrowers have lower spreads. Specifically, column (1) shows that our borrower technology similarity measure is negatively associated with loan spreads and is statistically significant at the 1% level. A one-standard-deviation increase in the borrower technology similarity reduces the loan spread by 4 bps.[15] Economically, given a sample mean loan size of $425 million, it translates to a sizable annual loan cost saving of $170,000.

Next, we perform several robustness checks and report results in columns (2) to (8) of Table 2.2. Our first check concerns the concentration of the bank's loan portfolio. Given that firms sharing similar technologies may operate in similar industry segments, we control for the segment similarity of the borrower and the bank's prior borrowers in column (2). A higher segment similarity indicates a larger overlap of the business lines of the borrowing firm and the bank's existing borrowers. It captures the bank's industry specialization to some extent, but primarily implies potentially higher industry concentration for the bank's loan portfolio. We empirically find that such higher portfolio concentration is positively associated with loan spreads. The negative relationship between borrower technology similarity and loan spreads, however, remains statistically significant at the 1% level after controlling for segment similarity.

Second, firms sharing similar technologies may lead to greater product market competition, which is expected to cause larger loan spreads (see, e.g. Campello & Gao, 2017; Croci et al., 2021; Hasan et al., 2020, 2021).[16] To account for such a possibility, we control for borrower product market rivalry using three different measures from Hoberg and Phillips (2016) and report results in columns (3) to (5) in Table 2.2. We find that, as expected, a higher product market HHI indicating less competition is negatively related to borrower loan spreads, and a larger product market similarity or fluidity is positively related to loan spreads. In all cases, borrower technology similarity remains negatively and significantly associated with loan spreads with similar-sized coefficient estimates as the baseline.

[13] We use the two-digit SIC codes to identify borrower industry and our results are robust to the use of alternative industry classifications such as four-digit SIC codes.

[14] Alternatively, our results are robust to clustering standard errors by bank or by borrower industry.

[15] The borrower technology similarity has a sample standard deviation of 0.06 and an estimated coefficient of -0.38 in our baseline model. Since the sample mean value of the natural logarithm of loan spread is 5.066, the reduction in loan spread is $e^{5.066} - e^{5.066-0.38\times0.06} \approx 4$ bps.

[16] However, we note that a higher technology similarity between two firms does not necessarily imply stronger direct competition. More importantly, even if technology similarity results in increased market competition, we should expect that the borrower firm faces higher bank loan costs. Therefore, it can only lead to bias against us finding a negative association between technology similarity and loan spreads.

Third, we consider the effect of borrower technology value on loan spreads. A borrower with higher technology value may receive favorable loan spreads regardless of its technology similarity with bank's prior borrowers. If technology similarity is then correlated with the firm's technology value,[17] our baseline model would suffer from omitted variable bias. As such, in columns (6) to (8) of Table 2.2, we control for the borrower's patent value and patent stock, as well as segment similarity and the three product market competition proxies, respectively. We find that in these most conservative specifications, our baseline result on the negative association between borrower technology similarity and loan spreads remains qualitatively unchanged. The robust effect of technology similarity in reducing loan costs suggests that technology similarity contains information beyond bank loan portfolio concentration, borrower firm competition, and technology value.[18]

2.4.2 Bank industry specialization

Because firms from different industries can share similar technologies, banks' technology knowledge accumulated from past lending transcends industry boundaries. However, we cannot rule out completely an overlap of technology expertise with industry specialization. In the baseline results, we partially address the concern by controlling for borrower segment similarity. A higher segment similarity indicates a larger overlap of the business lines of the borrower and the bank's existing borrowers, so that the bank may better leverage its industry specialization. We find that our results are robust to controlling for segment similarity.

To rule out the possibility that bank industry specialization drives our results, we repeat all of the above regressions but include the lender times industry times year fixed effects, which allow for time-varying bank industry specialization. Table 2.3 presents the results. We continue to find a negative and statistically significant relationship between borrower technology similarity and loan spreads across all model specifications. Moreover, the sizes of the estimated coefficients of technology similarity are even larger than in the baseline results. Therefore, bank specialization does not drive our results and bank technology expertise contains information beyond industry specialization. Given that including lender times industry times year fixed effects reduces our sample size, we focus on our baseline model specification in the following analyses, but we note that all of our results are robust to controlling for bank industry specialization.

[Insert Table 2.3 about here]

[17] For example, this correlation may happen when a bank has a strong preference for borrowers with high-value (or low-value) technologies in certain technology classes. We control for lender fixed effects throughout, which to some extent mitigates the concern.

[18] In Table 2.A2 in the Appendix, we repeat all of the above regressions but replace the dependent variable all-in-drawn spreads with the natural logarithm of total loan costs from Berg et al. (2016). The use of total loan costs includes the various fees specific to each loan facility but reduces our sample size due to data availability. Nevertheless, we find again a robust negative association between borrower technology similarity and loan costs, statistically significant at the 1% level across all model specifications.

In addition, we consider alternative industry classifications to further corroborate our results, including the Fama-French 48 industry classification and Hoberg and Phillips (2016) 10-K text-based 100 industries classification. Table 2.A3 in the Appendix shows that our results are robust to alternative industry classifications with and without controlling for bank industry specialization via lender times industry times year fixed effects. As such, we are confident that our results are robust to industry definition and that bank industry specialization does not drive our results.

2.4.3 Alternative construct of technology similarity

So far, our measure of bank technology knowledge of the borrower at loan origination is constructed by averaging the pairwise similarities of the borrower with each of the bank's prior borrowers. This method leads to a concern that we may accidentally introduce an attenuation bias. For example, a bank can have a strong technology knowledge of both biotechnology and semiconductor since it has existing borrowers that are respectively pure-play firms in biotechnology and semiconductor. Therefore, the bank should be well positioned in screening a new borrower that works in the intersection of biotechnology and semiconductor. However, the averaging of pairwise similarities could indicate that there is only limited overlap with existing borrowers, thus underestimating the bank's knowledge of the new borrower's technology.

To address this concern, we construct an alternative technology similarity measure. We first aggregate all of bank's prior borrowers as if they were one and build a portfolio of all patents of this "single" past borrower. We then compute the pairwise technology similarity between the new borrower and this technology portfolio aggregating all recent borrowers' patents.

This alternative measure allows the bank's technology knowledge learned from prior borrowers to be complementary and additive. As in the baseline, we use a 5-year window to construct the aggregate patent portfolio. Table 2.A4 in the Appendix shows that our baseline results remain qualitatively unchanged. We note that the estimated coefficients of the alternative technology similarity measure are smaller in size than in the baseline. This, however, does not suggest that our baseline method is free of attenuation bias given that the alternative measure is a non-linear transformation. In Table 2.A5 in the Appendix, we include lender times industry times year fixed effects and continue to find a negative and statistically significant relation between the alternative technology similarity measure and loan spreads across all model specifications.

On the other hand, both our baseline and the portfolio aggregation methods may still underestimate the bank's technology knowledge of the borrower, especially when the bank lends to many firms of different technologies in the past. In an extreme case, if the bank's past borrowers collectively own all technology classes, both methods will show a low level of technology similarity when the new borrower uses a small subset of technologies. To address this concern, we construct another alternative measure of borrower technology similarity that is the maximum pairwise similarity of the borrowing firm and the bank's past borrowers. We exclude the current borrowing firm from past borrowers because it creates a mechanical maximum pairwise similarity. Table 2.A6 in the Appendix shows that, controlling for time-varying bank industry specialization, the relationship between the alternative measure based on maximum pairwise similarity and loan spreads remain negative and statistically significant across all model specifications.

2.4.4 Firm innovation, bank expertise, and relationship lending

Further, variations in technological similarity can come from both banks having no expertise and firms having no technology. Specifically, a technology similarity measure of zero can be result of either 1) the borrowing firm has patents, but these patents are dissimilar to the experience set of the bank (no technology expertise), or 2) the borrowing firm has no patents (no firm innovation). Given that the median technology similarity is zero in our sample, it is important to disentangle these two channels. We investigate the issue by removing borrowers without patents. As reported in Table 2.A7 in the Appendix, we continue to find a negative relationship (albeit weaker) between borrower technology similarity and loan spreads. This result suggests that our results are not driven by the second channel.

In addition, to address the concern that technology similarity may be higher when the bank has fewer recent borrowers and when the focal borrower is a major prior borrower, we perform two sub-sample analyses in Table 2.A8 in the Appendix. In columns (1) to (4), we remove the loans originated by the banks whose numbers of recent borrowers are in the bottom quartile. In columns (5) to (8), we remove the loans where the borrower is a major prior borrower of the bank defined by its loan amount in the top quartile. In both cases, we continue to find a negative association between borrower technology similarity and loan spreads, statistically significant at the 1% level.

2.4.5 Patent value and citation

Next, the vast majority of patents granted may be relatively marginal, and only a small subset significantly impacts a firm's future performance or ability to repay loans. Our technology similarity measure, however, does not differentiate patents by their importance and assumes that the bank equally learns from all borrowers' patents. To address this concern, we re-construct the technology similarity measures using only the patents with above-median economic value (Kogan et al., 2017) within their technology class-year group. As shown in Table 2.A9 in the Appendix, we continue to find a negative and significant association between loan spreads and the refined borrower technology similarity measure. Moreover, the estimated coefficients of the refined measure across all specifications have larger sizes than the original measure in baseline model. Table 2.A10 in the Appendix reveals a similar pattern when we use patent citations to identify valuable patents.

The fact that the technology similarity effect strengthens when focusing on valuable and influential patents suggests that banks are developing genuine technological expertise, particularly when evaluating complex, important innovations - exactly where we expect bank technological specialization to be most valuable for reducing information asymmetry. These results demonstrate that our technology similarity measure is capturing economically meaningful relationships rather than spurious correlations.

2.4.6 Non-bank lenders

Another remaining concern is that the lead bank may involve non-bank lenders of specialized technology knowledge in the syndicate. As a result, the participation of non-bank lenders may correlate with the borrower's technology profile and with our measure of the bank's knowledge of the borrower's technology. To address this concern, we additionally control for the participation of non-bank lenders in our baseline regression. Specifically, we identify bank lenders as the commercial banks, classified by DealScan as "African bank", "Asia-Pacific bank", "Eastern European/Russian bank", "foreign bank", "Middle Eastern bank", "mortgage bank", "thrift/S&L", "US bank" or "Western European bank". Next, we make several corrections based on the lender's primary SIC code. We classify a lender with a SIC code 6211 ("Security brokers, dealers, and flotation companies"), which is labelled as commercial banks by DealScan, as a non-bank lender. For the lenders with missing institution type, they are classified as bank lenders if their SIC code starts with 60 (depository institutions). After these steps, for a few lenders classified as a bank but have a non-missing SIC code that does not start with 6, we classify them as non-bank lenders. Finally, all other lenders are classified as non-banks. For each loan, the participation of non-bank lenders is measured as the ratio of the number of non-bank lenders to the number of all lenders. As shown in Table 2.A11 in the Appendix, we find that the coefficient of non-bank lenders participation is positively associated with loan spreads and statistically significant at the 1% level across all specifications. More importantly, however, the negative and significant relation between loan spreads and borrower technology similarity remains largely unchanged after controlling for the participation of non-bank lenders in the loan syndicate.

2.4.7 Placebo test

Lastly, we conduct a placebo test on whether interest rates are lower on borrowers who have similar technology to future borrowers. If our conjecture is correct, then there should be no relationship between a borrowing firm's technology similarity with the bank's future borrowers and the spreads of current loans. Specifically, at each loan's origination (time t), we compute the technology similarity of the borrowing firm and the bank's future borrowers from five years to ten years after loan origination ($t+6$ to $t+10$) so that there is no time overlap of when counting patent classes of the current and future borrowers (we use a five-year window to capture a firm's technology profile). Table 2.A12 in the Appendix replicates the baseline model using the placebo technology similarity. As expected, we find no significant relationship between the placebo measure and loan spreads across all specifications. This placebo test lends further support to our findings that banks accumulate technology knowledge from past lending activities, which helps reducing loan costs for future borrowers sharing similar technologies.

2.4.8 Bank performance

Our empirical results document a robust effect of banks charging lower loan spreads for borrowing firms with a higher technology similarity with the banks' prior borrowers. This is likely due to cost savings in bank screening and monitoring from accumulated technology knowledge. We now examine whether such specialization in technology affects bank future performance. First, we construct a bank-year panel of the average borrower technology similarity of the loans lead arranged by the bank. Specifically, we aggregate loan-level borrower technology similarity to the bank-level using the loan-size-weighted average. This measure captures a bank's technology knowledge relevant in loan contracting in a year.[19] We then regress bank future ROA and distance to default on bank technology knowledge, controlling for bank size, total loans, non-deposit leverage, deposits and capital. We include year fixed effects and bank fixed effects, and cluster standard errors at the bank level.

[Insert Table 2.4 about here]

Table 2.4 shows that when banks lend more to borrowers sharing similar technologies with their prior borrowers, they have higher future ROA (for at least five years in the future) and larger distance to default (for up to four years in the future). These results suggest that bank specialization in technologies is positively associated with bank future performance, which also lends support to banks rationally charging lower loan spreads for borrowers sharing similar technologies with banks' prior borrowers. Nevertheless, we formally investigate the economic mechanisms in the next section.

[19] If a bank does not lead arrange any loan in a year, its technology knowledge for the year is missing. We do not forward fill missing values with the last available value because it would require the assumption of a depreciation rate and that our bank-year measure is technically not a stock measure.

2.5 Economic mechanisms

We now move on to study economic mechanisms underlying the negative effect of borrower technology similarity and loan costs. We start by showing that the technology similarity between a borrower and the bank's prior borrowers is informative about the borrower's creditworthiness. We then discuss the empirical challenges for identification and present our solutions based on structural matching model and difference-in-differences estimation.

2.5.1 The information content of technology similarity

In the screening process, why should a bank care about a borrower's technology similarity with the bank's prior borrowers? Extant studies have documented a vector of factors, beyond borrower fundamentals, from lending specialization, product market competition, supply chain relationship, innovation outputs to other soft information such as tax avoidance, stock price fragility and so on (e.g. Campello & Gao, 2017; Chava et al., 2017; Hasan et al., 2014, 2021). The literature also highlights the importance of firms' technology profiles on future performance (e.g. Kogan et al., 2017; Manso, 2011). We argue that the technology similarity facilitates the bank to acquire opaque information from the borrower at reduced costs given its accumulated knowledge of prior borrowers' technology.

We empirically study the information content of a borrower's technology similarity with the bank's prior borrowers by studying the explanatory power of such technology similarity on the borrower's fundamentals and credit risks. Specifically, we regress the absolute difference in the borrowing firm's and the bank's prior borrowers' creditworthiness measures on their technology similarity, controlling for their segment similarity and a range of absolute differences in other firms' characteristics. Heteroskedasticity-robust standard errors are clustered at the borrower firm level.[20]

Table 2.A16 in the Appendix shows that a higher technology similarity is negatively associated with the absolute difference in borrowers' Altman Z-score, Merton (1974) default probability, profitability and cash holdings, all significant at the 1% level. These results imply that borrowers with similar technology profiles exhibit similar levels of creditworthiness and their capacities to service debt are relatively equal. Moreover, in Table 2.A17 in the Appendix, we find that a higher technology similarity predicts smaller absolute differences in borrowers' Altman Z-score, Merton (1974) default probability, profitability and cash holdings in the next five years. The empirical evidence is consistent to there being valuable information content embedded in the borrowers' overlapping technology capabilities and that this is relevant information for assessing the borrower's credit risk given the bank's knowledge of the prior borrowers' creditworthiness.

[20] Alternatively, our results are robust to clustering standard errors by bank or by borrower industry.

2.5.2 Identification challenges and structural matching model

Given that technology similarity is informative about firm creditworthiness, a bank could potentially save on credit risk assessment costs when screening a borrower sharing similar technologies with the bank's prior borrowers. We find, in our baseline results, that banks pass on at least part of such savings to the borrower. However, an exact identification is challenging for two reasons. First, an exogenous shock to the observed borrower technology similarity is unlikely. Because our measure of technology similarity is based on the borrowing firm's technology profile and that of the bank's prior borrowers, an ideal shock can only affect the borrowing firm's technology profile – extant borrowers' technology profiles are historical and cannot be affected, which then implies that the lending bank cannot be changed. Therefore, a candidate shock is one that exogenously alters the borrowing firm's technology profile and does not cause the firm to switch bank. We as econometricians, however, cannot know whether a bank-borrower matching formation and termination is a result of switching or other factors such as the firm's capital requirement. On the other hand, a changed technology profile is likely associated with significant changes in the borrower's business strategies and other fundamental aspects. Hence, a shock to the borrowing firm's technology profile more or less has impacts on other firm characteristics, thereby affecting bank loan contracting. Simply put, it is challenging to employ traditional identification strategies. For example, to the best of our knowledge, prior studies like Lee et al. (2019), McLemore et al. (2022), and Bereskin et al. (2023) encounter similar identification challenges.

To address the identification challenges, we first employ a structural bank-borrower matching model similar to Fox (2017, 2018) and Schwert (2018) to show that technology similarity plays a positive role in bank's value maximization. The details of the model are presented in Section A.1 in the Appendix. The structural matching model allows us to identify drivers of bank-borrower matching assignments in the absence of unobservable non-matching assignments.[21] Specifically, both banks and firms in our model maximize their respective value. A loan contract contains transfer payments (e.g., interests and fees). Because we do not have data on the unobservable counterfactual loans, we derive an inequality condition for simultaneous value maximization without the transfer payment component. Assuming that observed actual bank-borrower matches (i.e., loans) yield higher value than unobservable counterfactual loans, we perform a semi-parametric estimation for the loan determinants at the bank-borrower level (Schwert, 2018). Our results suggest that the total economic surplus for banks and borrowers can be enhanced by matching banks to borrowers whose technology profiles are similar to that of the banks' prior borrowers. Lower loan spreads for borrowers sharing similar technology profiles with the banks' prior borrowers are economically optimal for both banks and borrowers.

[21] We can observe only the loans that have been originated, not the potential loans that could have been originated if borrowers chose different lenders.

2.5.3 Difference-in-differences results

As discussed previously, it is empirically challenging to find an exogenous shock to the technology similarity measure between a bank's new borrower and the bank's prior borrowers. However, if reduced information asymmetry and increased efficiency in screening and monitoring due to bank's accumulated technology knowledge are the mechanisms underlying the negative relationship between borrower technology similarity and loan spreads, we may alternatively exploit exogenous shocks to the stock of bank's technology knowledge using difference-in-differences (DiD) estimations, such as bank mergers and acquisitions (M&As). Bank M&As arguably increase acquirer banks' stock of accumulated technology knowledge, but do not affect their *extant* borrowers' creditworthiness and fundamentals. New borrowers of the acquirer bank (or the consolidated bank) after a M&A may confound our identification as their matching with the bank can be a result of endogenous selection. We therefore expect that after a bank M&A, extant borrowers of the acquirer bank will receive cheaper loans due to the exogenous positive shock on the bank's technology knowledge, compared to borrowers of other banks (without M&As).

Specifically, we start with a sample of 285 bank M&A deals from 1987 to 2019 from the updated DealScan lender link table by Schwert (2018). These M&As involve acquirer and/or target banks in the syndicated loan market and may provide acquirer banks better opportunities to improve efficiency via absorbing target banks' accumulated knowledge in assessing borrowers' technology profiles (Sapienza, 2002). We then filter out 12 bank-year M&A events as our shocks.[22] For each bank-year M&A event, we construct a cohort of treated banks and control banks with a five-year event window starting from two years before to two years after the event year. In each cohort, the treated bank is the bank that made one or more M&As in the event year, and control banks are those banks that have no M&A deals in the entire event window. If a bank has made a M&A deal, it is removed from all later cohorts as a control bank. Control banks in each cohort are the top-ten comparable banks selected using propensity score matching based on their bank size, non-deposit leverage, deposit ratio, ROA, total loans, and loan-to-deposit shortfall. Table 2.A14 in the Appendix verifies that our control banks and treated banks are comparable after the matching. Using the sample of loans made by the treated banks to their extant borrowers, we then estimate the following DiD model:

$$\begin{aligned} \ln(\mathit{Loan\ Spread}_{m,i,l,c}) =& \beta_1 \mathit{Post}_{t,c} \times \mathit{Treated}_{m,i,l,c} + \beta_2 \mathit{Post}_{t,c} + \beta_3 \mathit{Treated}_{m,i,l,c} \\ &+ \beta_4 \Gamma_{i,l,c} + \text{Fixed Effects} + \varepsilon_{m,i,l,c} \end{aligned} \tag{2.4}$$

[22]We identify 31 unique acquirers (or the surviving entities after mergers) in the M&A deals by GVKEY in the updated DealScan lender link table by Schwert (2018). If a bank makes multiple M&A deals in a year, we collapse them into one event. Of the resulting 42 acquirer-year events, we keep 26 events by removing acquisitions by the same acquirer within any two consecutive years to avoid overlapping event windows. We further remove 3 acquisitions before the start of our sample period. Lastly, we drop the bank-year M&A events where the acquirer banks have no observations in Compustat Bank or loans in DealScan for the years either before or after the event year, Table 2.A13 in the Appendix lists the 12 bank-year M&A events and associated M&A deals.

where $Treated_{m,i,l,c}$ is a dummy variable that equals to one, and zero otherwise, if loan l's lead bank m is the treated bank in cohort c. $Post_{t,c}$ is a dummy variable that equals one (zero) if year t in cohort c is after (before) the event year. $\Gamma_{i,l,t}$ represents loan and borrower characteristics as in the baseline. We include loan type and loan purpose fixed effects, borrower industry fixed effects and lender fixed effects. Heteoskedasticity-robust standard errors are double clustered at the bank and year levels.[23]

[Insert Table 2.5 about here]

Table 2.5 shows that the coefficient estimates of the interaction term, $Post \times Treated$, are negative and statistically significant at the 5% level at least across all specifications, controlling for segment similarity, product market rivalry, patent stock, and patent value. Moreover, Table 2.A15 in the Appendix reports the results of a dynamic DiD model, where we replace the *Post* dummy variable in Equation 2.4 with the indicator variable $D_{c,\tau}$ ($\tau = -1, 0, 1, 2$) that equals one if the loan is issued τ years after the event year in cohort c and zero otherwise. We confirm that the treatment effect occurs only in the year(s) after the event year, as shown in Figure 2.A2 in the Appendix. These results are consistent with our expectation that, as acquirer banks obtain the accumulated technology knowledge from target banks after a M&A, the increased stock of technology knowledge allows the banks to improve their screening and monitoring on borrowers' technology profiles, leading to reduced loan spreads.[24]

However, we acknowledge that bank M&As could occur for many reasons and affect more than the stock of acquirer banks' stock of technology knowledge. To partly mitigate this concern, columns (3) and (4) of Table 2.5 report the results using alternative cohorts of bank M&As where acquirer banks learned technology knowledge from target banks. We take steps to construct these cohorts. First, for each loan issued by an acquirer (treated) bank in the two years after the M&A, we compute the counterfactual technology similarity between the borrower and the target bank at the time of M&A. This counterfactual technology similarity captures the target bank's knowledge of the acquirer bank's future borrower. Further, if such counterfactual technology similarity (between the target bank and the borrower after M&A) is higher than factual technology similarity (between the acquirer bank and the borrower after M&A), we consider the acquirer bank as potentially having learned technology knowledge. Intuitively, the target bank is more familiar with the borrower's technologies than the acquirer bank. As such, the acquirer bank should have accumulated larger stock of technology knowledge of the borrower after M&A. Finally, we restrict to only cohorts of bank M&As where on average the acquirer banks have learned technology knowledge. We find that our DiD results remain largely unchanged. Interestingly, the mild reduction in the number of loans implies that most acquirer

[23] Our results are robust to clustering standard errors at the borrower level.

[24] Our results are also consistent with Erel (2011) that bank mergers decrease loan spreads when gains from cost savings outweigh the increase of bank market power.

banks potentially learned technology knowledge from target banks. Overall, our results suggest that the documented negative relation between borrower technology similarity and loan costs is likely causal.

2.6 Further results

Given that banks' technological expertise reduces loan spreads by mitigating information asymmetry and adverse selection, in this section, we conduct further tests to investigate the heterogeneous effect of borrower technology similarity under conditions that require varying levels of bank screening and monitoring, and that yield differing benefits to banks.

2.6.1 Borrower technological obsolescence

First, from a borrower's perspective, firms are technologically more opaque when their technologies are cutting-edge. Advanced technologies require greater screening efforts from banks because they can be harder to comprehend and may be of higher value. In this case, banks' accumulated knowledge from previous lending to firms sharing similar technologies plays a more important role in reducing information asymmetry and screening costs.

We calculate borrowers' technological obsolescence as in Ma (2021). Following Ma (2021), the technology base for a firm i in year t is predetermined as all other firms' patents cited by firm i up to year $t - \omega$. The technological obsolescence of firm i in year t is then calculated as the negative of the log difference in the external citations of the firm's technology base in year t and in year $t - \omega$. Our estimates are comparable to that in Ma (2021).[25] But to minimise the impact of measurement error, we calculate two proxies. The first is a dummy variable that takes the value of 1 if a borrower's technological obsolescence is below the annual median, and 0 otherwise. The second is the annual decile rank of borrowers' technological obsolescence. We then interact each of the two proxies with borrower technology similarity in the baseline model. Table 2.6 presents the results.

[Insert Table 2.6 about here]

Columns (1) and (2) of Table 2.6 show that the interaction term of borrower technology similarity and low technological obsolescence dummy is negative and statistically significant, whether or not we control for other known technology measures such as patent stock and patent value.

[25] In Ma (2021), a firm's technology base includes a mean (median) of 2,001 (219) patents. In our calculation, the mean (median) technology base is 2,223 (173). The technological obsolescence estimates have a mean of 12.9% ($\omega = 3$) and 19.39% ($\omega = 5$) in Ma (2021), and a mean of 11.3% ($\omega = 3$) and 14.9% ($\omega = 5$) in our estimates. The discrepancies may be due to different filters applied on the sample.

Moreover, we find that the coefficient estimates of technology similarity itself are no longer significant, which suggests that banks' technological expertise may only reduce loan spreads when the borrowers' technologies are not obsolete. In columns (3) and (4), we confirm that the more obsolete a borrower's technologies are, the lower the effect of technology similarity on loan spreads. These results support our Hypothesis 2.

2.6.2 Borrower product disclosure

Second, borrowers are more opaque when they have less product disclosure. Cao et al. (2018) validate that a firm's product disclosure contains proprietary information about the firm using interviews and behaviour surveys. Therefore, less product disclosure translates to greater firm opacity and information asymmetry, demanding higher bank screening efforts. Because banks' accumulated technological expertise should be more valuable in such cases, we expect the effect of borrower technology similarity to be stronger for borrowers that have lower product disclosure.

Specifically, we collect all product-related announcements from the Capital IQ Key Developments database.[26] As in Liu et al. (2024), we keep only press releases by firms via a newswire or firms' website, excluding news by third-party media. We then use textual analysis to identify product-related announcements into four categories: 1) new product launches, 2) product updates, 3) progress toward new products, and 4) others. The classification is based on a list of keywords for each category from Liu et al. (2024), similar to Cao et al. (2018).[27] A product-related announcement is classified into a category if it contains the highest keyword count for that category. We measure a firm's product disclosure as the total word count of all product-related press releases in the past five years after removing common stop words. In addition, we calculate an alternative measure based on the total word count of press releases related to the new product launch category, because it represents the highest degree of information asymmetry. In contrast, product updates and progresses may not have comparable technological opacity. We interact the natural logarithm of the product disclosure words with borrower technology similarity. Table 2.7 presents the results.

[Insert Table 2.7 about here]

Table 2.7 shows that the coefficient of borrower technology similarity remains negative and statistically significant across all specifications. However, columns (1) and (2) show that greater product disclosure mitigates the negative association between loan spreads and borrower technology similarity, whether or not we control for borrower product market competition and its

[26] Capital IQ Key Developments starts in 2002, hence the reduction in loan numbers in Table 2.7. Product-related announcements are identified using Key Development Type ID 41.

[27] The detailed list can be found in Liu et al. (2024). We use Python package "spacy" to expand the list to include all variants and synonyms using the NLTK's WordNet.

contribution to the industry segment concentration of the bank's portfolio. Using new product launch, columns (3) and (4) document a similar mitigating effect. Consistent with our expectation, these result show that banks' technological expertise is more valuable when borrower is more opaque with less product disclosure, lending further support to our Hypothesis 2.

2.6.3 Borrower riskiness

Apart from opacity, we expect borrower's ex ante riskiness to affect bank's willingness to pass on the cost savings from technological expertise accumulated from past lending. To examine such heterogeneity, we interact various borrower riskiness measures and borrower technology similarity. Table 2.8 reports the results.

[Insert Table 2.8 about here]

In column (1) of Table 2.8, we interact the borrower creditworthiness measured by the Altman Z-score and borrower technology similarity. We find that the negative and significant coefficient of the interaction terms drives out the statistical significance of borrower technology similarity, which continues to have a negative estimate. This result suggests that the documented negative effect of borrower technology similarity on loan costs mainly comes from safer borrowers.[28] In column (2), we investigate borrower leverage and find that the coefficient of its interaction with borrower technology similarity is positive and significant. This result shows that less-leveraged borrowers are granted lower loan spreads than more-leveraged ones when they share similar technology profiles with the bank's extant borrowers. In column (3), we interact borrower profitability and our technology similarity measure and find a negative and significant coefficient estimate of the interaction term, which drives out the statistical significance of technology similarity. This implies that banks are more willing to reduce loan spreads for profitable borrowers when they have a higher technology similarity with banks' prior borrowers. These results corroborate our Hypothesis 3.

2.6.4 Bank characteristics

From a bank's perspective, our Hypothesis 4 conjectures that accumulated technological knowledge should be more valuable for smaller, less-capitalized or low-liquidity banks. These banks are relatively more constrained and may thus value more the cost savings from accumulated knowledge from past lending. To test this, we include the relevant bank characteristics and their interaction with technology in the baseline model.

[28] We find similar results using alternative borrower creditworthiness measures such as the borrower default probability and whether the borrower has a credit rating.

[Insert Table 2.9 about here]

Table 2.9 shows the heterogeneous effects of borrower technology similarity on bank loan spreads from the bank's perspective. Column (1) shows that borrower technology similarity remains negatively and significantly associated with loan spreads after the inclusion of bank size and its interaction with borrower technology similarity. The positive and significant coefficient of the interaction term confirms that smaller banks cut loan spreads more than larger banks given a higher borrower technology similarity. In column (2), we include the bank's capital ratio and its interaction with borrower technology similarity. We find that the negative and significant association between borrower technology similarity and loan spreads remains qualitatively unchanged, and is stronger for banks with a lower capital ratio. In column (3), we include bank (il)liquidity measure, the loan-to-deposit shortfall (Acharya & Mora, 2015), and its interaction with borrower technology similarity. We find that the interaction between bank loan-to-deposit shortfall and borrower technology similarity is negative and statistically significant, which implies that less liquid banks reduce loan spreads for borrowers with similar technology profiles compared to prior borrowers. In summary, we show that it is the smaller, less-capitalized, or less-liquid banks that tend to charge lower loan spreads for borrowers sharing a higher technology similarity with their prior borrowers, supporting our Hypothesis 4.

2.6.5 Loan sales and bank business model

Lastly, banks do not always hold to maturity loans issued and frequently sell their loan shares after origination. Blickle et al. (2022) show that only 3-4% of credit lines are sold, whereas 40-50% of term loans are sold. Within term loans, Term B loans are significantly more likely to be sold than Term A loans. For loans originated to distribute, there is arguably reduced screening and monitoring incentives. As a result, a bank's accumulated technological expertise may not be as important in loan pricing. We hence expect the effect of borrower technology similarity to be muted for Term B loans.

To investigate this conjecture, we assume that banks accumulate and utilize technological expertise primarily through facilities they intend to hold and monitor, i.e., Term A loans and credit lines (Revolver/Line $\geq$ 1 Yr., Revolver/Line $<$ 1 Yr., 364-Day Facility). We then conduct two tests using two loan samples. The first includes all loans while the second focuses on Term B loans only, serving as a falsification test. Specifically, in the falsification test, technology similarity is measured between the borrower of Term B loan and the bank's prior borrowers of Term A loans and credit lines. As shown in Table 2.10, we find strong results in the full sample similar to the baseline. However, as expected, we find no significant relationship between technology similarity and loan spreads in the falsification test using only Term B loans.

[Insert Table 2.10 about here]

The absence of technological expertise's impact on the price of Term B loans confirms that banks accumulate and utilize technology expertise primarily through facilities they intend to hold and monitor. With reduced screening and monitoring incentives, bank's technological expertise no longer reduce loan spreads because of reduced potential cost savings. These results further corroborate our proposed economic mechanisms.

2.7 Conclusion

In this study, we empirically examine the impact of borrower technology similarity on the cost of bank loans. We show that banks charge lower loan spreads for borrowers that share a similar technology profile with the banks' prior borrowers, likely due to the cost savings from the reduced due diligence needed in assessing the borrowers' technologies and implications for their credit risk profiles. Such effect is robust to alternative constructs of borrower technology similarity, controlling for the industry segment similarity between the borrower and prior borrowers which affects the bank loan portfolio's industry concentration, and controlling for the intensity of product market competition faced by the borrower. Furthermore, the borrower's technology profile itself as measured by patent value and stock does not absorb the effect of its similarity with the bank's prior borrowers on loan costs, even after controlling for relationship lending. We further rule out the possibility that bank industry specialization drives our results.

Despite the identification challenges, we show that borrower technology similarity is informative about firm creditworthiness and debt service capability. We estimate a structural bank-borrower matching model to show that borrower technology similarity is an important determinant in bank lending decisions, which plays a positive role in the simultaneous value maximization of both banks and borrowers. The total economic surplus for banks and borrowers can be enhanced by matching banks to borrowers with a similar technology profile to the banks' prior borrowers. We then use difference-in-differences estimations to show that after exogenous positive shocks to banks' accumulated technology knowledge, banks reduce loan spreads to their borrowers. Specifically, we find that smaller, less-capitalized, or less-liquid banks give larger discounts in loan pricing for borrowers with a higher technology similarity to the banks' prior borrowers. For these borrowers, banks are more likely to charge lower loan spreads when they are ex ante more opaque with lower technological obsolescence and less product disclosures. Furthermore, banks reward the safer, more profitable and less indebted borrowers. We find that these borrowers are systematically granted cheaper loans by bank lenders in the syndicated loan market.

Figure 2.1: **Borrower Technology Similarity Over Time**

Figure 2.1 shows the annual average of bank-level borrower technology similarity from 1990 to 2020 by small and large banks based on annual median size. The detailed definition of borrower technology similarity is explained in Section 2.3. Intuitively, a higher value indicates that banks are more likely to lend to borrowers sharing similar technologies with their prior borrowers.

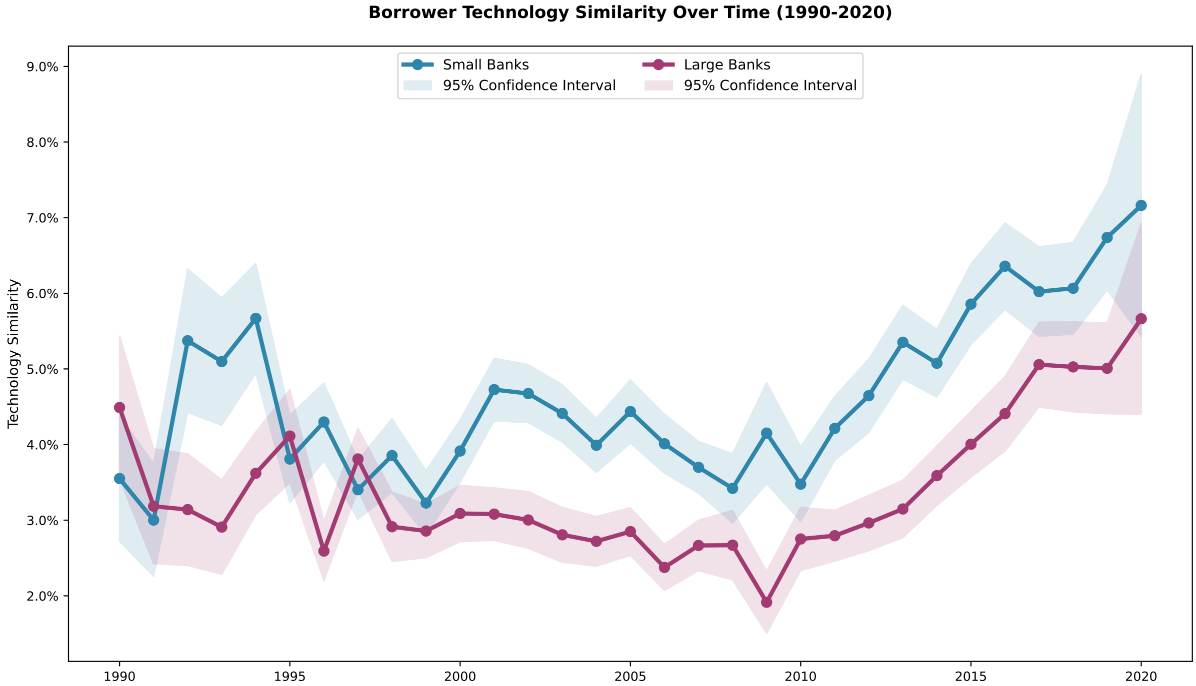
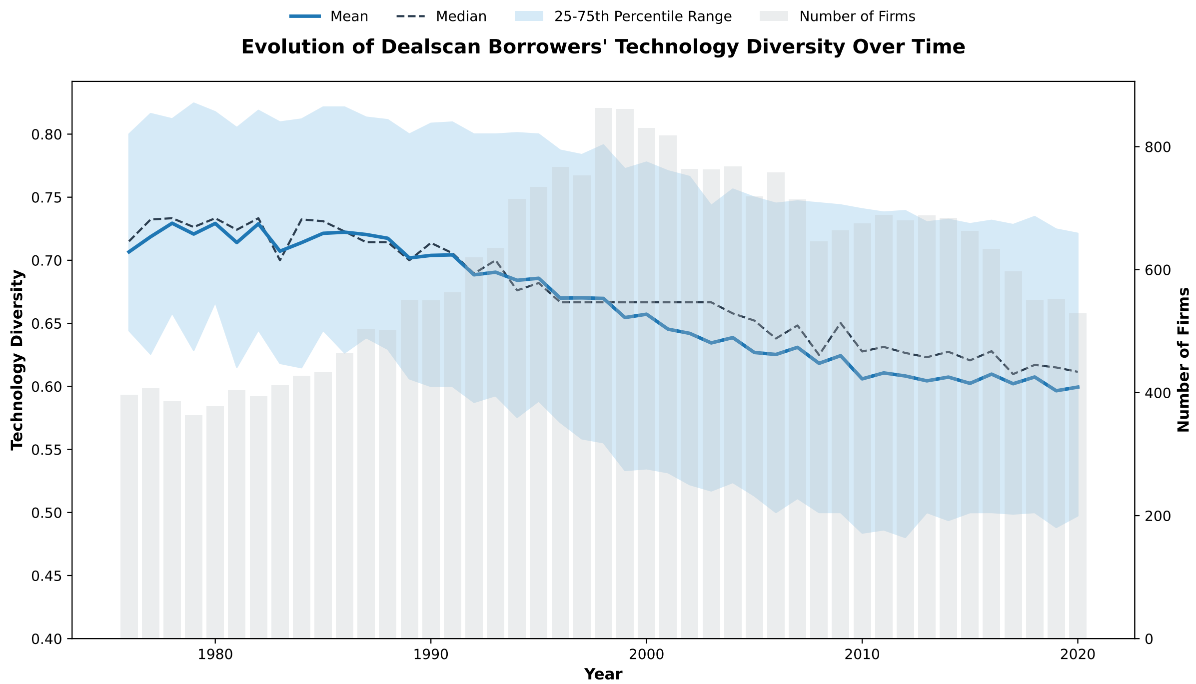


Figure 2.2: **Borrowing Firms' Technology Diversity Over Time**

Figure 2.2 plots the evolution of borrowers' technology diversity from 1990 to 2020. Following Hsu et al. (2018), firm technology diversity is measured using a Herfindahl-based index of the distribution of a firm's patents across CPC technology classes over rolling 5-year windows. The solid blue line shows the mean technology diversity, the dashed line shows the median, and the shaded area represents the 25th-75th percentile range. The bars indicate the number of firms with non-zero technology diversity scores in each year. A higher diversity score indicates a firm's patents are more evenly distributed across different technology classes, while a lower score suggests greater technological specialization.



Note: Based on the sample with non-zero Technology Diversity scores

Table 2.1: **Summary Statistics**
Table 2.1 presents the summary statistics of our loan sample sourced from DealSCan, which consists of 36,166 loans to U.S. borrowers (excluding utility and financial firms) from January 1990 to December 2020. Definitions of the variables are provided in Table 2.A1 in the Appendix. All continuous variables are winsorized by year at the 1$^{\text{st}}$ and 99$^{\text{th}}$ percentiles.

	Observations	Mean	Standard Deviation	10^{th} Percentile	Median	90^{th} Percentile
Bank-borrower characteristics						
Technology Similarity	36,166	0.039	0.060	0.000	0.000	0.132
Segment Similarity	36,166	0.063	0.056	0.008	0.052	0.125
Prior Relationship	36,166	0.355	0.379	0.000	0.239	1.000
Borrower In Bank Top Industries	36,166	0.190	0.392	0	0	1
Lending Relationship Time	36,166	1.940	2.838	0.000	1.000	6.000
Loan characteristics						
Loan Spread (bps)	36,166	205.310	141.397	50.000	175.000	375.000
ln(Loan Spread)	36,166	5.066	0.782	3.912	5.165	5.927
Loan Size ($ millions)	36,166	425.155	739.765	14.201	166.000	1045.000
Loan Maturity (months)	36,166	48.193	22.538	12.000	60.000	72.000
Loan Secured	36,166	0.533	0.499	0	1	1
Borrower characteristics						
Borrower Product Market HHI	36,166	0.283	0.264	0.049	0.184	0.689
Borrower Product Market Similarity	36,166	3.133	3.991	1.019	1.578	7.050
Borrower Product Market Fluidity	36,166	6.624	3.426	2.888	5.961	11.254
Borrower Patent Stock	36,166	0.566	4.398	0.000	0.000	0.492
Borrower Citation Stock	36,166	1.865	2.800	0.000	0.000	6.372
Borrower Patent Value	36,166	5.607	26.663	0.000	0.000	10.959
Borrower Size	36,166	7.074	1.942	4.498	7.101	9.642
Borrower Total Assets ($ billions)	36,166	6.566	19.569	0.088	1.213	15.399
Borrower Leverage	36,166	0.314	0.212	0.037	0.297	0.584
Borrower Z-score	36,166	0.016	0.014	0.001	0.015	0.032
Borrower Profitability	36,166	0.128	0.089	0.048	0.125	0.227
Borrower Market-to-Book	36,166	0.028	0.053	0.007	0.021	0.059
Borrower Cash	36,166	0.074	0.090	0.005	0.040	0.189
Borrower Has Credit Rating	36,166	0.519	0.500	0	1	1
Borrower Default Probability	28,120	0.027	0.083	0.000	0.000	0.066
Bank characteristics						
Bank Size	30,399	13.154	1.429	11.202	13.401	14.674
Bank Capital	30,399	0.076	0.022	0.044	0.079	0.102
Bank Deposits	30,399	0.576	0.122	0.424	0.576	0.714
Bank Total Loans	30,399	0.461	0.138	0.294	0.473	0.636
Bank Loan-to-Deposit Shortfall	30,399	-0.115	0.092	-0.209	-0.116	-0.005
Bank ROA	30,399	0.009	0.005	0.002	0.010	0.014

Table 2.2: **Borrower Technology Similarity and Loan Spread**

Table 2.2 reports the results of regressing the natural logarithm of all-in-drawn loan spreads on borrower technology similarity. Specifically, column (1) reports the baseline result. Column (2) controls for the borrower's segment similarity with bank's prior borrowers. Columns (3) through (5) control for the competition faced by the borrower using three of Hoberg and Phillips (2016) product market competition measures, respectively. Columns (6) to (8) additionally control for the borrower's patent value and patent stock, as well as the segment similarity. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity	-0.380***	-0.412***	-0.362***	-0.373***	-0.350***	-0.309***	-0.318***	-0.291***
	(0.093)	(0.094)	(0.092)	(0.093)	(0.092)	(0.094)	(0.095)	(0.094)
Segment Similarity		0.224**				0.218**	0.219**	0.194**
		(0.094)				(0.094)	(0.094)	(0.093)
Borrower Product Market HHI			-0.113***			-0.112***		
			(0.020)			(0.020)		
Borrower Product Market Similarity				0.004***			0.004***	
				(0.001)			(0.001)	
Borrower Product Market Fluidity					0.014***			0.014***
					(0.002)			(0.002)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Borrower Patent Stock						-0.001	-0.001	-0.001
						(0.001)	(0.001)	(0.001)
Prior Relationship	-0.019*	-0.022**	-0.020*	-0.020*	-0.020*	-0.024**	-0.025**	-0.024**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Borrower Size	-0.085***	-0.085***	-0.088***	-0.086***	-0.089***	-0.083***	-0.081***	-0.084***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Borrower Leverage	0.566***	0.566***	0.574***	0.569***	0.571***	0.564***	0.559***	0.560***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Borrower Z-score	-2.007***	-1.984***	-1.819***	-1.838***	-1.369***	-1.888***	-1.896***	-1.436***
	(0.442)	(0.443)	(0.441)	(0.445)	(0.447)	(0.439)	(0.442)	(0.444)
Borrower Profitability	-1.136***	-1.135***	-1.139***	-1.138***	-1.144***	-1.113***	-1.111***	-1.116***
	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.062)	(0.062)	(0.062)
Borrower Market-to-Book	-0.323***	-0.323***	-0.315***	-0.331***	-0.338***	-0.275***	-0.291***	-0.297***
	(0.073)	(0.073)	(0.073)	(0.073)	(0.074)	(0.070)	(0.071)	(0.072)
Borrower Cash	0.090*	0.091*	0.061	0.075	0.037	0.065	0.078	0.040
	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.048)
Borrower Has Credit Rating	0.039***	0.039***	0.039***	0.039***	0.040***	0.035***	0.035***	0.036***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.011)
ln(Loan Size)	-0.097***	-0.096***	-0.097***	-0.097***	-0.097***	-0.097***	-0.097***	-0.097***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.016**	0.016**	0.017**	0.017**	0.017**	0.014**	0.014*	0.014*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Loan Secured	0.385***	0.385***	0.384***	0.385***	0.380***	0.383***	0.384***	0.380***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	36,166	36,166	36,166	36,166	36,166	36,166	36,166	36,166
Adjusted R^2	0.653	0.653	0.654	0.653	0.655	0.656	0.655	0.657

Table 2.3: **Borrower Technology Similarity and Loan Spread Controlling for Time-varying Bank Industry Specialization**

Table 2.3 reports the results of estimating all the model specifications as in Table 2.2, additionally controlling for time-varying bank industry specialization via lender times industry times year fixed effects. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity	-0.444***	-0.483***	-0.426***	-0.438***	-0.408***	-0.371***	-0.380***	-0.341***
	(0.125)	(0.128)	(0.122)	(0.124)	(0.123)	(0.127)	(0.130)	(0.128)
Segment Similarity		0.296				0.335*	0.335*	0.272
		(0.199)				(0.200)	(0.200)	(0.197)
Borrower Product Market HHI			-0.150***			-0.151***		
			(0.025)			(0.025)		
Borrower Product Market Similarity				0.004***			0.005***	
				(0.002)			(0.002)	
Borrower Product Market Fluidity					0.016***			0.016***
					(0.002)			(0.002)
Borrower Patent Stock						-0.001	-0.001	-0.001
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Prior Relationship	-0.046***	-0.046***	-0.047***	-0.047***	-0.046***	-0.049***	-0.049***	-0.048***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Borrower Size	-0.097***	-0.097***	-0.101***	-0.098***	-0.102***	-0.095***	-0.091***	-0.095***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Borrower Leverage	0.665***	0.665***	0.673***	0.667***	0.665***	0.660***	0.654***	0.651***
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
Borrower Z-score	-2.859***	-2.850***	-2.605***	-2.661***	-2.164***	-2.723***	-2.761***	-2.276***
	(0.537)	(0.538)	(0.536)	(0.541)	(0.544)	(0.534)	(0.539)	(0.542)
Borrower Profitability	-1.191***	-1.191***	-1.193***	-1.193***	-1.199***	-1.161***	-1.161***	-1.166***
	(0.082)	(0.082)	(0.082)	(0.082)	(0.082)	(0.081)	(0.082)	(0.082)
Borrower Market-to-Book	-0.381***	-0.383***	-0.376***	-0.391***	-0.391***	-0.341***	-0.357***	-0.355***
	(0.087)	(0.087)	(0.086)	(0.088)	(0.089)	(0.083)	(0.085)	(0.086)
Borrower Cash	0.160**	0.160**	0.120*	0.143**	0.100	0.123**	0.145**	0.102
	(0.064)	(0.064)	(0.063)	(0.064)	(0.063)	(0.063)	(0.063)	(0.063)
Borrower Has Credit Rating	0.038***	0.038***	0.037***	0.038***	0.040***	0.033**	0.034**	0.035***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
ln(Loan Size)	-0.081***	-0.081***	-0.081***	-0.081***	-0.082***	-0.082***	-0.082***	-0.082***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.034***	0.034***	0.034***	0.034***	0.034***	0.032***	0.033***	0.032***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Loan Secured	0.352***	0.352***	0.349***	0.351***	0.347***	0.350***	0.352***	0.347***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender × Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,578	31,578	31,578	31,578	31,578	31,578	31,578	31,578
Adjusted R^2	0.729	0.729	0.730	0.729	0.731	0.731	0.731	0.732

Table 2.4: **Bank Technology Knowledge and Future Performance**

Table 2.4 examines the predictive power of a bank's accumulated technology knowledge for its future performance. Specifically, we regress the bank ROA and distance to default on bank technology knowledge, controlling for bank size, non-deposit leverage, deposits, and capital. The dependent variable is measured at year $t+h$ and all independent variables are measured at year t. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the lender level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
	h=1	h=2	h=3	h=4	h=5
Bank ROA (t+h)					
Bank Technology Knowledge	0.007***	0.009***	0.008***	0.006*	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Bank Size	-0.001	-0.002**	-0.001	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Bank Total Loans	-0.005*	-0.006**	-0.004	-0.004	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Bank Non-Deposit Leverage	-0.001	-0.004	-0.006	-0.006	-0.008**
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Bank Deposits	0.004	0.004	0.003	0.002	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Bank Capital	0.012	-0.010	-0.008	-0.006	0.015
	(0.022)	(0.020)	(0.018)	(0.019)	(0.015)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	1,774	1,690	1,608	1,526	1,444
Adjusted R^2	0.406	0.430	0.438	0.451	0.459
Bank Distance to Default (t+h)					
Bank Technology Knowledge	4.977***	6.750***	5.476**	6.752***	4.201
	(1.684)	(1.905)	(2.382)	(2.394)	(2.682)
Bank Size	-0.090	-0.430	-0.729*	-0.902*	-1.103**
	(0.334)	(0.360)	(0.407)	(0.454)	(0.528)
Bank Total Loans	-2.559**	-3.014**	-2.743	-1.995	0.416
	(1.211)	(1.370)	(1.662)	(1.996)	(2.410)
Bank Non-Deposit Leverage	-7.574*	-4.252	-4.086	-2.068	-4.818
	(4.517)	(3.616)	(4.165)	(5.626)	(7.455)
Bank Deposits	13.338***	13.250***	11.181***	9.499**	2.451
	(4.193)	(3.461)	(3.552)	(4.495)	(6.384)
Bank Capital	31.300**	28.793**	17.282	14.509	7.864
	(14.373)	(12.860)	(12.288)	(12.457)	(12.945)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	1,205	1,154	1,103	1,047	990
Adjusted R^2	0.803	0.799	0.792	0.778	0.769

Table 2.5: **Difference-in-Differences Estimation: Bank M&A**

Table 2.5 shows the results of the difference-in-differences (DiD) estimation using bank M&As as exogenous shocks to the stock of acquirer banks' technology knowledge. *Treated* and *Post* dummies variables are defined in Section 2.5.3. We include the same set of controls as in the baseline. We control for loan type fixed effects, loan purpose fixed effects, borrower industry fixed effects, lender fixed effects and year fixed effects across all specifications. Definitions of other variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the lender and year levels. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
	All M&As		Technology Learned	
Post × Treated	-0.152**	-0.154**	-0.156**	-0.157**
	(0.063)	(0.063)	(0.066)	(0.066)
Post	-0.006	-0.004	-0.008	-0.005
	(0.018)	(0.017)	(0.017)	(0.016)
Treated	-0.022	-0.012	-0.023	-0.014
	(0.050)	(0.047)	(0.050)	(0.045)
Borrower Patent Stock		-0.010		-0.007
		(0.012)		(0.013)
Borrower Patent Value		-0.001		-0.001
		(0.001)		(0.001)
Segment Similarity	-0.117	-0.052	-0.119	-0.054
	(0.445)	(0.435)	(0.465)	(0.460)
Borrower Product Market HHI	-0.277***	-0.272***	-0.306***	-0.302***
	(0.087)	(0.086)	(0.094)	(0.094)
Prior Relationship	-0.026	-0.022	-0.016	-0.013
	(0.047)	(0.047)	(0.052)	(0.054)
Borrower Size	-0.171***	-0.162***	-0.176***	-0.168***
	(0.015)	(0.015)	(0.017)	(0.017)
Borrower Leverage	0.770***	0.748***	0.798***	0.778***
	(0.131)	(0.127)	(0.131)	(0.126)
Borrower Z-score	-5.737***	-5.859***	-5.767***	-5.902***
	(1.744)	(1.738)	(1.899)	(1.849)
Borrower Profitability	-1.443***	-1.397***	-1.413***	-1.366***
	(0.174)	(0.159)	(0.198)	(0.176)
Borrower Market-to-Book	-0.974**	-0.934**	-1.087***	-1.041***
	(0.372)	(0.386)	(0.350)	(0.366)
Borrower Cash	0.714***	0.711***	0.748***	0.755***
	(0.211)	(0.207)	(0.201)	(0.203)
Borrower Has Credit Rating	0.025	0.023	0.022	0.020
	(0.022)	(0.023)	(0.023)	(0.025)
ln(Loan Size)	-0.086***	-0.086***	-0.088***	-0.088***
	(0.022)	(0.022)	(0.024)	(0.024)
ln(Loan Maturity)	0.031**	0.029**	0.033*	0.031*
	(0.015)	(0.014)	(0.016)	(0.016)
Loan Secured	0.464***	0.469***	0.458***	0.462***
	(0.037)	(0.037)	(0.037)	(0.037)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes
Borrower Industry Fixed Effects	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	4,700	4,700	4,267	4,267
Adjusted R^2	0.759	0.760	0.761	0.762

Table 2.6: **Borrower Technology Similarity and Technological Obsolescence**

Table 2.6 examines the heterogeneous effects of borrower technology similarity on loan spread conditional on borrowers' technological obsolescence as in Ma (2021). Specifically, columns (1) and (2) report the results based on the low Technological Obsolescence dummy that equals to 1 if the borrowers' Technological Obsolescence is below the annual median. Columns (3) and (4) alternatively use the annual decile rank of borrowers' technological obsolescence. ω is the window length used in computing technological obsolescence as in Ma (2021). In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
Technology Similarity	0.138	0.153	-0.760***	-0.731***
	(0.192)	(0.191)	(0.272)	(0.273)
Technology Similarity × Low Technology Obsolescence ($\omega = 5$)	-0.527**	-0.522**		
	(0.221)	(0.222)		
Technology Similarity × Technology Obsolescence Rank ($\omega = 5$)			0.116***	0.114***
			(0.041)	(0.041)
Low Technology Obsolescence ($\omega = 5$)	0.037	0.036		
	(0.024)	(0.024)		
Technology Obsolescence Rank ($\omega = 5$)			-0.006	-0.005
			(0.004)	(0.004)
Borrower Patent Stock		-0.001		-0.001
		(0.001)		(0.001)
Borrower Patent Value		-0.001***		-0.001***
		(0.000)		(0.000)
Segment Similarity	0.329	0.328	0.334	0.333
	(0.248)	(0.248)	(0.249)	(0.249)
Borrower Product Market HHI	-0.086**	-0.088**	-0.086**	-0.089**
	(0.035)	(0.035)	(0.035)	(0.035)
Prior Relationship	-0.071***	-0.072***	-0.070***	-0.071***
	(0.023)	(0.023)	(0.023)	(0.023)
Borrower Size	-0.072***	-0.062***	-0.072***	-0.062***
	(0.011)	(0.011)	(0.011)	(0.011)
Borrower Leverage	0.694***	0.675***	0.690***	0.671***
	(0.063)	(0.063)	(0.063)	(0.063)
Borrower Z-score	-2.578**	-2.800***	-2.633***	-2.854***
	(1.013)	(1.013)	(1.017)	(1.017)
Borrower Profitability	-1.760***	-1.692***	-1.756***	-1.688***
	(0.160)	(0.159)	(0.160)	(0.159)
Borrower Market-to-Book	-0.505***	-0.445***	-0.500***	-0.440***
	(0.131)	(0.126)	(0.131)	(0.126)
Borrower Cash	0.263**	0.274**	0.258**	0.270**
	(0.117)	(0.115)	(0.117)	(0.115)
Borrower Has Credit Rating	-0.002	-0.009	-0.001	-0.009
	(0.023)	(0.023)	(0.023)	(0.023)
ln(Loan Size)	-0.110***	-0.110***	-0.110***	-0.110***
	(0.010)	(0.010)	(0.010)	(0.010)
ln(Loan Maturity)	0.071***	0.068***	0.071***	0.068***
	(0.013)	(0.013)	(0.013)	(0.013)
Loan Secured	0.389***	0.392***	0.389***	0.392***
	(0.021)	(0.021)	(0.021)	(0.021)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes
Observations	10,311	10,311	10,311	10,311
Adjusted R^2	0.723	0.724	0.723	0.724

Table 2.7: **Borrower Technology Similarity and Product Disclosure**

Table 2.7 examines the heterogeneous effects of borrower technology similarity on loan spread conditional on borrowers' product disclosure as in Cao et al. (2018). Specifically, columns (1) and (2) report the results based the natural logarithm of one plus the total number of words in all press releases related to product development issued by the borrower in the last five years. Alternatively, columns (3) and (4) use the product disclosure words related to the new product announcement. We use natural language processing and the list of key words in Liu et al. (2024) to classify new product announcement. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
Technology Similarity	-0.612***	-0.608***	-0.569***	-0.578***
	(0.201)	(0.201)	(0.177)	(0.177)
Technology Similarity × ln(Product Disclosure)	0.063**	0.055*		
	(0.029)	(0.029)		
Technology Similarity × ln(New Product Disclosure)			0.065**	0.059**
			(0.028)	(0.028)
ln(Product Disclosure)	-0.005**	-0.006**		
	(0.002)	(0.002)		
ln(New Product Disclosure)			-0.005**	-0.006**
			(0.002)	(0.002)
Segment Similarity		0.259*		0.257*
		(0.142)		(0.142)
Borrower Product Market HHI		-0.079***		-0.079***
		(0.025)		(0.024)
Borrower Patent Stock	-0.002**	-0.002**	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Borrower Patent Value	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Prior Relationship	-0.068***	-0.069***	-0.067***	-0.069***
	(0.016)	(0.016)	(0.016)	(0.016)
Borrower Size	-0.056***	-0.057***	-0.056***	-0.057***
	(0.006)	(0.006)	(0.006)	(0.006)
Borrower Leverage	0.486***	0.488***	0.487***	0.489***
	(0.033)	(0.033)	(0.033)	(0.033)
Borrower Z-score	-4.028***	-3.893***	-4.002***	-3.859***
	(0.550)	(0.550)	(0.547)	(0.547)
Borrower Profitability	-1.023***	-1.018***	-1.024***	-1.019***
	(0.091)	(0.091)	(0.091)	(0.091)
Borrower Market-to-Book	-0.197***	-0.190***	-0.197***	-0.190***
	(0.073)	(0.073)	(0.074)	(0.073)
Borrower Cash	0.150**	0.135**	0.148**	0.132**
	(0.063)	(0.062)	(0.062)	(0.062)
Borrower Has Credit Rating	0.034**	0.033**	0.034**	0.033**
	(0.013)	(0.013)	(0.013)	(0.013)
ln(Loan Size)	-0.099***	-0.099***	-0.099***	-0.099***
	(0.006)	(0.006)	(0.006)	(0.006)
ln(Loan Maturity)	0.083***	0.083***	0.083***	0.083***
	(0.012)	(0.012)	(0.012)	(0.012)
Loan Secured	0.329***	0.328***	0.329***	0.327***
	(0.012)	(0.012)	(0.012)	(0.012)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes
Observations	17,794	17,794	17,794	17,794
Adjusted R^2	0.661	0.662	0.661	0.662

Table 2.8: **Borrower Technology Similarity and Borrower Riskiness**

Table 2.8 examines the heterogeneous effects of borrower technology similarity on loan spread for different borrowers. Specifically, we estimate the model as in column (6) of Table 2.2 and additionally include the interaction of technology similarity and three borrower characteristics: Altman's Z-score, leverage, and profitability, respectively. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
Technology Similarity	-0.157	-0.534***	0.188
	(0.129)	(0.156)	(0.158)
Technology Similarity × Borrower Z-score	-17.190***		
	(5.990)		
Technology Similarity × Borrower Leverage		0.905**	
		(0.413)	
Technology Similarity × Borrower Profitability			-4.382***
			(0.982)
Borrower Z-score	-8.491***		
	(0.431)		
Borrower Leverage		0.580***	
		(0.028)	
Borrower Profitability			-1.110***
			(0.058)
Segment Similarity	0.239**	0.273***	0.266***
	(0.095)	(0.095)	(0.095)
Borrower Product Market HHI	-0.090***	-0.131***	-0.101***
	(0.020)	(0.020)	(0.020)
Borrower Patent Stock	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Borrower Patent Value	-0.002***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
Prior Relationship	-0.022**	-0.037***	-0.022**
	(0.011)	(0.011)	(0.011)
Borrower Size	-0.077***	-0.075***	-0.082***
	(0.006)	(0.005)	(0.006)
Borrower Market-to-Book	-0.429***	-0.479***	-0.325***
	(0.074)	(0.081)	(0.073)
Borrower Cash	-0.171***	0.037	-0.257***
	(0.048)	(0.051)	(0.048)
Borrower Has Credit Rating	0.058***	0.031***	0.081***
	(0.012)	(0.012)	(0.012)
ln(Loan Size)	-0.101***	-0.110***	-0.097***
	(0.005)	(0.004)	(0.005)
ln(Loan Maturity)	0.011	0.002	0.015**
	(0.007)	(0.007)	(0.007)
Loan Secured	0.422***	0.417***	0.425***
	(0.010)	(0.010)	(0.010)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes
Observations	36,166	36,166	36,166
Adjusted R^2	0.638	0.639	0.638

Table 2.9: **Heterogeneous Effects of Borrower Technology Similarity: Banks**

Table 2.9 examines the heterogeneous effects of borrower technology similarity on loan spread for different banks. Specifically, we estimate the model as in column (6) of Table 2.2 and additionally include bank characteristics and their interaction with technology similarity. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
Technology Similarity	-2.615***	-0.913***	-0.544***
	(0.613)	(0.319)	(0.153)
Technology Similarity × Bank Size	0.178***		
	(0.047)		
Technology Similarity × Bank Capital		8.010**	
		(3.743)	
Technology Similarity × Bank Loan-to-Deposit Shortfall			-1.706**
			(0.785)
Bank Size	-0.082***		
	(0.018)		
Bank Capital		-1.780***	
		(0.489)	
Bank Loan-to-Deposit Shortfall			0.058
			(0.080)
Segment Similarity	0.284***	0.311***	0.302***
	(0.099)	(0.100)	(0.099)
Borrower Product Market HHI	-0.117***	-0.118***	-0.118***
	(0.021)	(0.021)	(0.021)
Borrower Patent Stock	-0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)
Borrower Patent Value	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
Prior Relationship	-0.015	-0.016	-0.015
	(0.011)	(0.011)	(0.011)
Borrower Size	-0.091***	-0.091***	-0.091***
	(0.006)	(0.006)	(0.006)
Borrower Leverage	0.608***	0.606***	0.607***
	(0.028)	(0.028)	(0.028)
Borrower Z-score	-1.822***	-1.884***	-1.857***
	(0.482)	(0.481)	(0.483)
Borrower Profitability	-1.083***	-1.086***	-1.081***
	(0.065)	(0.066)	(0.066)
Borrower Market-to-Book	-0.272***	-0.269***	-0.268***
	(0.073)	(0.073)	(0.074)
Borrower Cash	0.054	0.060	0.059
	(0.051)	(0.052)	(0.052)
Borrower Has Credit Rating	0.041***	0.041***	0.041***
	(0.012)	(0.012)	(0.012)
ln(Loan Size)	-0.092***	-0.093***	-0.093***
	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.003	0.002	0.002
	(0.008)	(0.008)	(0.008)
Loan Secured	0.367***	0.369***	0.368***
	(0.010)	(0.010)	(0.010)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes
Observations	30,365	30,365	30,365
Adjusted R^2	0.639	0.639	0.638

Table 2.10: **Borrower Technology Similarity Based on Loan Types**

Table 2.10 reports the results of regressing the natural logarithm of all-in-drawn loan spreads on borrower technology similarity. Specifically, this technology similarity is measured between a borrower and the bank's prior borrowers of Term A loans and revolving credit facilities. Panel A presents results for the full loan sample, while Panel B focuses only on Term B loans to provide a falsification test. In both panels, column (1) reports the baseline result. Column (2) controls for the borrower's segment similarity with bank's prior borrowers, which is also constructed using the same relationship loan types. Columns (3) through (5) control for the competition faced by the borrower using three of Hoberg and Phillips (2016) product market competition measures, respectively. Columns (6) to (8) additionally control for the borrower's patent value and patent stock, as well as the segment similarity. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: All Loans

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity (Term A Loan and Credit Lines)	-0.418***	-0.448***	-0.399***	-0.411***	-0.393***	-0.354***	-0.364***	-0.342***
	(0.091)	(0.092)	(0.090)	(0.091)	(0.090)	(0.091)	(0.092)	(0.091)
Segment Similarity (Term A Loan and Credit Lines)		0.324**				0.263**	0.260**	0.230*
		(0.126)				(0.126)	(0.125)	(0.125)
Borrower Product Market HHI			-0.124***			-0.122***		
			(0.020)			(0.020)		
Borrower Product Market Similarity				0.004***			0.004***	
				(0.001)			(0.001)	
Borrower Product Market Fluidity					0.014***			0.014***
					(0.002)			(0.002)
Borrower Patent Stock						-0.000	-0.000	-0.000
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Prior Relationship	-0.022**	-0.026**	-0.022**	-0.022**	-0.022**	-0.028***	-0.028***	-0.028**
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Borrower-level Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,472	33,472	33,472	33,472	33,472	33,472	33,472	33,472
Adjusted R^2	0.665	0.665	0.666	0.665	0.667	0.667	0.667	0.668

Panel B: Falsification Test (Term B Loans)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity (Term A Loan and Credit Lines)	-0.206	-0.212	-0.206	-0.210	-0.206	-0.275	-0.272	-0.272
	(0.178)	(0.175)	(0.178)	(0.179)	(0.178)	(0.174)	(0.174)	(0.174)
Segment Similarity (Term A Loan and Credit Lines)		0.073				0.122	0.088	0.106
		(0.317)				(0.317)	(0.319)	(0.320)
Borrower Product Market HHI			0.021			0.023		
			(0.042)			(0.041)		
Borrower Product Market Similarity				0.002			0.002	
				(0.003)			(0.003)	
Borrower Product Market Fluidity					0.000			0.000
					(0.004)			(0.004)
Borrower Patent Stock						0.005***	0.005***	0.005***
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						0.000	0.000	0.000
						(0.001)	(0.001)	(0.001)
Prior Relationship	-0.098***	-0.099***	-0.098***	-0.099***	-0.098***	-0.096***	-0.096***	-0.096***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Borrower-level Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,333	2,333	2,333	2,333	2,333	2,333	2,333	2,333
Adjusted R^2	0.463	0.463	0.463	0.463	0.463	0.464	0.464	0.464

Table 2.A1: **Variable Definition**

Variable	Definition	Source
Bank-borrower level variables		
Technology Similarity	The cosine similarity of the technology profiles between the current borrower and banks' lending portfolios over the past five years	USPTO
Segment Similarity	The cosine similarity of the product market segments between the current borrower and banks' prior lending portfolios	Compustat Segment
Borrower In Bank Top Industries	A dummy variable equals to one if the borrower is within the bank's top-five 2-digit SIC lending industries by total loan volume each year	DealScan
Prior Relationship	Bharath et al. (2011) relationship lending measure: the total amount of loan by the lead bank to the current borrower in the last five years divided by the total amount of loans by the borrower in the last five years	DealScan
Geographic Distance	The distance in kilometre (km) between the borrower and the bank based on their headquarters' ZIP codes.	Compustat
Loan level variables		
Loan Spread	The all-in-drawn loan spread measured in basis points	DealScan
Loan Size	Total amount of a loan facility in millions of US dollars	DealScan
Maturity	Total number of months to maturity of a loan facility	DealScan
Loan Secured	A dummy variable equals to one if the loan facility is secured	DealScan
Total Bank Loan Cost	Total bank loan cost constructed by Berg et al. (2016) including all fees charged by lenders	DealScan & Berg et al. (2016)
Borrower level variables		
Borrower Product Market HHI	The Hoberg and Phillips (2016) 10-K Text-based Network (TNIC) Industry Herfindahl-Hirschman Index	Hoberg-Phillips Data Library
Borrower Product Market Similarity	The Hoberg and Phillips (2016) 10-K Text-based Network (TNIC) Industry total similarity of each firm to the product market, calculated by firm-by-firm pairwise cosine similarity	Hoberg-Phillips Data Library
Borrower Product Market Fluidity	The Hoberg et al. (2014) 10-K based product market fluidity measuring how intensively the product market around a firm is changing in each year	Hoberg-Phillips Data Library
Borrower Patent Stock	The borrower patent stock created by capitalizing the number of granted patents in the last five years with 20% depreciation rate as in Chava et al. (2017)	USPTO
Borrower Patent Value	The borrower average patent value computed as the total Kogan et al. (2017) patent value at the firm level scaled by the number of patents granted	USPTO & Kogan et al. (2017)
Borrower Size	The natural logarithm of borrower total assets (AT)	Compustat
Borrower Leverage	The borrower financial leverage measured as the ratio of total debt (sum of long-term debt (DLTT) and debt in current liabilities (DLC)) to total assets (AT)	Compustat
Borrower Z-score	The borrower modified Altman's Z-score = (1.2×working capital (WCAP) + 1.4×retained earnings (RE) + 3.3×pretax-income (PI) + 0.999×total sales (SALE))/total assets (AT). We follow Hasan et al. (2014) and ignore the ratio of market value of equity to book value of total debt, since we control for a similar term borrower market-to-book ratio in our regressions	Compustat
Borrower Profitability	The borrower earnings before interest, taxes, depreciation, and amortization (EBITDA) scaled by total assets (AT)	Compustat
Borrower Market-to-Book	The borrower market value of equity scaled by the book value of equity ((PRCC_F×CSHO)/CEQ)	Compustat
Borrower Cash	The borrower cash and marketable securities (CHE) scaled by borrowers' total assets (AT)	Compustat
Borrower Has Credit Rating	A dummy variable equals to 1 if the borrower has the public credit rating	Compustat

Table 2.A1: **Continued**

Variable	Definition	Source
Borrower Bank-Dependent	The Schwert (2018) borrower bank-dependent indicator: a dummy variable equals to 1 if the borrower has no public credit rating	Compustat
Borrower Default Probability	The probability of default estimated using the Bharath and Shumway (2008) naive distance-to-default measure	Compustat & CRSP
Low Technological Obsolescence	A dummy variable equals 1 if a borrower's technological obsolescence as in Ma (2021) is below the annual median.	USPTO
Technological Obsolescence Rank	The annual decile rank of a borrower's technological obsolescence as in Ma (2021).	USPTO
Bank-level variables		
Bank Size	The natural logarithm of bank's total asset (AT)	Compustat Bank
Bank Capital	Bank common equity (CEQ) normalized by bank's total assets(AT)	Compustat Bank
Bank Total Loans	Bank total loans (LNTAL) normalized by bank's total assets(AT)	Compustat Bank
Bank ROA	Bank net income (loss) (NI) normalized by bank's total assets(AT)	Compustat Bank
Bank Non-Deposit Leverage	The Gropp and Heider (2010) bank non-deposit leverage: the ratio of bank debt, excluding deposit, to bank assets = debt in current liabilities (DLCQ) + long-term debt (DLTTQ) / total assets (ATQ)	Compustat Bank
Bank Loan-to-Deposit Shortfall	The Acharya and Mora (2015) loan-to-deposit shortfall: [total loans (LNTAL) - deposits (DPTC)]/total assets (AT)	Compustat Bank

Table 2.A1: **Borrower Technology Similarity Using Alternative Time Windows**

Table 2.A1 reports the robustness check of the baseline results that measures borrower technology similarity using banks' past 1-year, 3-year, 7-year and all history lending portfolios. The dependent variable is the natural logarithm of loan spreads. Additionally, we construct the segment similarity using the corresponding alternative time window. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	1Y Window		3Y Window		7Y Window		All History	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity	-0.158**	-0.130**	-0.236***	-0.198***	-0.332***	-0.274***	-0.438***	-0.366***
	(0.062)	(0.060)	(0.077)	(0.075)	(0.090)	(0.088)	(0.104)	(0.101)
Segment Similarity	0.103		0.162**		0.272***		0.306***	
	(0.064)		(0.079)		(0.092)		(0.098)	
Borrower Product Market HHI		-0.114***		-0.113***		-0.113***		-0.111***
		(0.020)		(0.020)		(0.020)		(0.020)
Borrower Patent Stock	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Borrower Patent Value	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Prior Relationship	-0.023**	-0.023**	-0.024**	-0.022**	-0.024**	-0.022**	-0.024**	-0.022**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Borrower Size	-0.081***	-0.084***	-0.081***	-0.084***	-0.080***	-0.083***	-0.079***	-0.082***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Borrower Leverage	0.558***	0.566***	0.557***	0.565***	0.555***	0.564***	0.554***	0.563***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Borrower Z-score	-2.049***	-1.875***	-2.066***	-1.898***	-2.073***	-1.908***	-2.089***	-1.918***
	(0.440)	(0.438)	(0.440)	(0.438)	(0.440)	(0.437)	(0.439)	(0.437)
Borrower Profitability	-1.110***	-1.113***	-1.110***	-1.113***	-1.110***	-1.113***	-1.109***	-1.113***
	(0.063)	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)
Borrower Market-to-Book	-0.286***	-0.276***	-0.284***	-0.275***	-0.282***	-0.275***	-0.284***	-0.277***
	(0.071)	(0.070)	(0.071)	(0.070)	(0.071)	(0.070)	(0.071)	(0.070)
Borrower Cash	0.091*	0.060	0.092*	0.063	0.094*	0.064	0.095*	0.065
	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.048)
Borrower Has Credit Rating	0.035***	0.035***	0.035***	0.035***	0.035***	0.035***	0.036***	0.035***
	(0.012)	(0.011)	(0.012)	(0.011)	(0.012)	(0.011)	(0.012)	(0.011)
ln(Loan Size)	-0.097***	-0.098***	-0.097***	-0.097***	-0.097***	-0.097***	-0.097***	-0.097***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.014*	0.014**	0.014*	0.014**	0.014*	0.014**	0.014*	0.014**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Loan Secured	0.386***	0.384***	0.385***	0.384***	0.385***	0.383***	0.384***	0.383***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	36,166	36,166	36,166	36,166	36,166	36,166	36,166	36,166
Adjusted R^2	0.654	0.655	0.654	0.655	0.655	0.655	0.655	0.656

Table 2.A2: **Borrower Technology Similarity and Total Loan Cost**

Table 2.A2 reports the results of regressing the natural logarithm of total loan costs from Berg et al. (2016) on borrower technology similarity. Specifically, column (1) controls for the borrower's segment similarity with bank's prior borrowers. Columns (2) through (4) control for competition faced by the borrower using three of Hoberg and Phillips (2016) product market competition measures, respectively. Columns (5) to (7) additionally control for the borrower's patent stock. In all specifications, we control for the bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity	-0.433***	-0.454***	-0.406***	-0.413***	-0.387***	-0.385***	-0.394***	-0.359***
	(0.118)	(0.121)	(0.117)	(0.118)	(0.118)	(0.121)	(0.123)	(0.122)
Segment Similarity		0.128				0.120	0.125	0.099
		(0.123)				(0.121)	(0.122)	(0.121)
Borrower Product Market HHI			-0.123***			-0.123***		
			(0.023)			(0.023)		
Borrower Product Market Similarity				0.007***			0.007***	
				(0.002)			(0.002)	
Borrower Product Market Fluidity					0.015***			0.015***
					(0.002)			(0.002)
Borrower Patent Stock						-0.000	-0.000	-0.001
						(0.003)	(0.003)	(0.003)
Borrower Patent Value						-0.001**	-0.001**	-0.001***
						(0.000)	(0.000)	(0.000)
Prior Relationship	-0.019*	-0.021*	-0.020*	-0.020*	-0.020*	-0.023**	-0.023*	-0.022*
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Borrower Size	-0.096***	-0.096***	-0.100***	-0.098***	-0.101***	-0.097***	-0.095***	-0.098***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Borrower Leverage	0.678***	0.678***	0.689***	0.683***	0.684***	0.684***	0.677***	0.678***
	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
Borrower Z-score	-3.929***	-3.921***	-3.761***	-3.679***	-3.212***	-3.793***	-3.711***	-3.239***
	(0.562)	(0.562)	(0.562)	(0.561)	(0.566)	(0.561)	(0.560)	(0.566)
Borrower Profitability	-1.171***	-1.171***	-1.173***	-1.170***	-1.186***	-1.156***	-1.154***	-1.169***
	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)
Borrower Market-to-Book	-0.578***	-0.579***	-0.558***	-0.596***	-0.607***	-0.522***	-0.561***	-0.570***
	(0.125)	(0.125)	(0.124)	(0.124)	(0.124)	(0.123)	(0.123)	(0.123)
Borrower Cash	0.005	0.004	-0.030	-0.020	-0.056	-0.027	-0.017	-0.053
	(0.067)	(0.067)	(0.066)	(0.067)	(0.066)	(0.066)	(0.066)	(0.066)
Borrower Has Credit Rating	-0.022	-0.022	-0.022	-0.022	-0.022	-0.023*	-0.023*	-0.024*
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
ln(Loan Size)	-0.042***	-0.042***	-0.043***	-0.043***	-0.042***	-0.043***	-0.043***	-0.042***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	-0.208***	-0.208***	-0.207***	-0.207***	-0.207***	-0.209***	-0.209***	-0.209***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
Loan Secured	0.509***	0.509***	0.506***	0.507***	0.502***	0.506***	0.507***	0.502***
	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)	(0.012)	(0.013)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,469	17,469	17,469	17,469	17,469	17,469	17,469	17,469
Adjusted R^2	0.804	0.804	0.805	0.805	0.806	0.805	0.805	0.806

Table 2.A3: **Alternative Industry Definition**

Table 2.A3 reports the results of regressing the natural logarithm of all-in-drawn loan spreads on the borrower technology similarity measure using alternative industry definitions. Specifically, Panel A use the Fama-French 48 industry classification. Columns (1) to (4) report the results controlling for industry–year fixed effects. Columns (5) to (8) control for borrower industry–lender–year fixed effects. Panel B employs Hoberg and Phillips (2016) 10-K text-based fixed industries with 100 classifications. In all specifications, we control for borrower characteristics, loan-level characteristics, bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type and loan purpose. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Fama-French 48 Industry (FF48)								
Technology Similarity	-0.294***	-0.242***	-0.258***	-0.223**	-0.347***	-0.300**	-0.319***	-0.297**
	(0.094)	(0.092)	(0.093)	(0.092)	(0.125)	(0.121)	(0.124)	(0.122)
Segment Similarity	0.166*				0.143			
	(0.087)				(0.148)			
Borrower Product Market HHI		-0.105***				-0.116***		
		(0.019)				(0.024)		
Borrower Product Market Similarity			0.005***				0.004**	
			(0.001)				(0.002)	
Borrower Product Market Fluidity				0.015***				0.013***
				(0.002)				(0.002)
Borrower-level and Loan-level Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FF48 Industry $\times$ Year Fixed Effects	Yes	Yes	Yes	Yes				
Lender Fixed Effects	Yes	Yes	Yes	Yes				
Lender $\times$ FF48 Industry $\times$ Year Fixed Effects					Yes	Yes	Yes	Yes
Observations	36,049	36,049	36,049	36,049	32,117	32,117	32,117	32,117
Adjusted R^2	0.655	0.655	0.655	0.657	0.730	0.731	0.731	0.732
Panel B: Text-based Fixed Industry 100 Classifications (FIC100)								
Technology Similarity	-0.330***	-0.281***	-0.293***	-0.270***	-0.393***	-0.354***	-0.375***	-0.349***
	(0.094)	(0.092)	(0.093)	(0.092)	(0.134)	(0.132)	(0.133)	(0.131)
Segment Similarity	0.196**				0.051			
	(0.090)				(0.162)			
Borrower Product Market HHI		-0.095***				-0.123***		
		(0.020)				(0.027)		
Borrower Product Market Similarity			0.004**				0.003*	
			(0.001)				(0.002)	
Borrower Product Market Fluidity				0.015***				0.015***
				(0.002)				(0.002)
Borrower-level and Loan-level Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FIC100 Industry $\times$ Year Fixed Effects	Yes	Yes	Yes	Yes				
Lender Fixed Effects	Yes	Yes	Yes	Yes				
Lender $\times$ FIC100 Industry $\times$ Year Fixed Effects					Yes	Yes	Yes	Yes
Observations	35,935	35,935	35,935	35,935	30,592	30,592	30,592	30,592
Adjusted R^2	0.662	0.662	0.662	0.664	0.747	0.747	0.747	0.748

Table 2.A4: **Alternative Borrower Technology Similarity Measure**

Table 2.A4 reports the results of regressing the natural logarithm of all-in-drawn loan spreads on the alternative technology similarity measure calculated as the pairwise similarity between the borrower and the aggregate patent portfolio of the bank's recent borrowers within the 5-year window. Specifically, column (1) reports the baseline result. Column (2) controls for the borrower's segment similarity with bank's prior borrowers. Columns (3) through (5) control for the competition faced by the borrower using three of Hoberg and Phillips (2016) product market competition measures, respectively. Columns (6) to (8) additionally control for the borrower's patent value and patent stock, as well as the segment similarity. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity (Portfolio)	-0.135***	-0.143***	-0.129***	-0.132***	-0.124***	-0.102***	-0.104***	-0.095***
	(0.034)	(0.034)	(0.033)	(0.034)	(0.033)	(0.034)	(0.034)	(0.034)
Segment Similarity (Portfolio)		0.109***				0.094**	0.094**	0.078**
		(0.041)				(0.040)	(0.040)	(0.040)
Borrower Product Market HHI			-0.113***			-0.110***		
			(0.020)			(0.020)		
Borrower Product Market Similarity				0.004***			0.004***	
				(0.001)			(0.001)	
Borrower Product Market Fluidity					0.014***			0.014***
					(0.002)			(0.002)
Borrower Patent Stock						-0.001	-0.001	-0.001
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Prior Relationship	-0.022**	-0.027**	-0.022**	-0.022**	-0.022**	-0.029***	-0.029***	-0.028***
	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Borrower Size	-0.085***	-0.086***	-0.088***	-0.086***	-0.089***	-0.084***	-0.081***	-0.084***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Borrower Leverage	0.566***	0.567***	0.574***	0.569***	0.571***	0.565***	0.560***	0.561***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Borrower Z-score	-2.005***	-1.940***	-1.817***	-1.839***	-1.369***	-1.855***	-1.876***	-1.415***
	(0.442)	(0.442)	(0.440)	(0.445)	(0.446)	(0.438)	(0.442)	(0.444)
Borrower Profitability	-1.136***	-1.138***	-1.139***	-1.137***	-1.143***	-1.115***	-1.113***	-1.118***
	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.062)	(0.062)	(0.062)
Borrower Market-to-Book	-0.324***	-0.323***	-0.315***	-0.332***	-0.339***	-0.275***	-0.291***	-0.298***
	(0.073)	(0.073)	(0.073)	(0.073)	(0.074)	(0.070)	(0.071)	(0.072)
Borrower Cash	0.095*	0.091*	0.065	0.080	0.041	0.065	0.079	0.041
	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.048)	(0.049)	(0.048)
Borrower Has Credit Rating	0.039***	0.039***	0.039***	0.039***	0.040***	0.035***	0.035***	0.036***
	(0.012)	(0.012)	(0.011)	(0.012)	(0.011)	(0.011)	(0.012)	(0.011)
ln(Loan Size)	-0.096***	-0.096***	-0.097***	-0.097***	-0.097***	-0.097***	-0.097***	-0.097***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.016**	0.016**	0.016**	0.016**	0.016**	0.014*	0.014*	0.014*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Loan Secured	0.385***	0.384***	0.383***	0.384***	0.380***	0.382***	0.383***	0.379***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	36,166	36,166	36,166	36,166	36,166	36,166	36,166	36,166
Adjusted R^2	0.653	0.653	0.654	0.653	0.655	0.656	0.655	0.657

Table 2.A5: **Alternative Borrower Technology Similarity Measure Controlling for Time-varying Bank Industry Specialization**

Table 2.A5 reports the results of regressing the natural logarithm of all-in-drawn loan spreads on the alternative technology similarity measure calculated as the pairwise similarity between the borrower and the aggregate patent portfolio of the bank's recent borrowers within the 5-year window. Specifically, column (1) reports the baseline result. Column (2) controls for the borrower's segment similarity with bank's prior borrowers. Columns (3) through (5) control for the competition faced by the borrower using three of Hoberg and Phillips (2016) product market competition measures, respectively. Columns (6) to (8) additionally control for the borrower's patent value and patent stock, as well as the segment similarity. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry–lender–year. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity (Portfolio)	-0.150***	-0.154***	-0.143***	-0.146***	-0.137***	-0.112***	-0.114***	-0.103**
	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.043)	(0.043)
Segment Similarity (Portfolio)		0.102				0.088	0.095	0.062
		(0.066)				(0.066)	(0.066)	(0.066)
Borrower Product Market HHI			-0.151***			-0.149***		
			(0.025)			(0.025)		
Borrower Product Market Similarity				0.004***			0.004***	
				(0.002)			(0.002)	
Borrower Product Market Fluidity					0.016***			0.016***
					(0.002)			(0.002)
Borrower Patent Stock						-0.001	-0.001	-0.001
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Prior Relationship	-0.046***	-0.049***	-0.047***	-0.047***	-0.046***	-0.051***	-0.051***	-0.050***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Borrower Size	-0.093***	-0.093***	-0.097***	-0.093***	-0.097***	-0.091***	-0.088***	-0.091***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Borrower Leverage	0.677***	0.678***	0.685***	0.679***	0.678***	0.672***	0.666***	0.664***
	(0.033)	(0.033)	(0.033)	(0.033)	(0.032)	(0.032)	(0.033)	(0.032)
Borrower Z-score	-2.915***	-2.856***	-2.659***	-2.719***	-2.226***	-2.736***	-2.783***	-2.309***
	(0.537)	(0.537)	(0.536)	(0.541)	(0.544)	(0.533)	(0.537)	(0.541)
Borrower Profitability	-1.184***	-1.188***	-1.187***	-1.187***	-1.192***	-1.159***	-1.159***	-1.163***
	(0.083)	(0.082)	(0.082)	(0.083)	(0.083)	(0.081)	(0.082)	(0.081)
Borrower Market-to-Book	-0.382***	-0.383***	-0.376***	-0.391***	-0.392***	-0.340***	-0.355***	-0.354***
	(0.087)	(0.087)	(0.086)	(0.088)	(0.089)	(0.084)	(0.085)	(0.086)
Borrower Cash	0.163**	0.158**	0.123*	0.147**	0.103	0.121*	0.144**	0.101
	(0.064)	(0.064)	(0.063)	(0.064)	(0.063)	(0.063)	(0.063)	(0.063)
ln(Loan Size)	-0.080***	-0.080***	-0.081***	-0.081***	-0.081***	-0.081***	-0.081***	-0.081***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.034***	0.035***	0.034***	0.035***	0.035***	0.033***	0.033***	0.033***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Loan Secured	0.354***	0.354***	0.351***	0.353***	0.349***	0.351***	0.353***	0.349***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender × Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,578	31,578	31,578	31,578	31,578	31,578	31,578	31,578
Adjusted R^2	0.729	0.729	0.730	0.729	0.731	0.731	0.730	0.732

Table 2.A6: **Maximum Pairwise Borrower Technology Similarity Controlling for Time-varying Bank Industry Specialization**

Table 2.A6 reports the results of regressing the natural logarithm of all-in-drawn loan spreads on the alternative technology similarity measure calculated as the largest pairwise similarity between the borrower and the bank's recent borrowers within the 5-year window, additionally controlling for time-varying bank industry specialization via lender times industry times year fixed effects. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity (Maximum)	-0.055***	-0.059***	-0.054***	-0.054***	-0.053***	-0.040**	-0.039**	-0.037**
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Segment Similarity (Maximum)		0.018				0.016	0.018	0.014
		(0.016)				(0.016)	(0.016)	(0.016)
Borrower Product Market HHI			-0.152***			-0.151***		
			(0.025)			(0.026)		
Borrower Product Market Similarity				0.004***			0.005***	
				(0.002)			(0.001)	
Borrower Product Market Fluidity					0.016***			0.016***
					(0.002)			(0.002)
Borrower Patent Stock						-0.001	-0.001	-0.001
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Borrower Size	-0.092***	-0.092***	-0.096***	-0.093***	-0.096***	-0.091***	-0.087***	-0.090***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Borrower Leverage	0.679***	0.678***	0.687***	0.681***	0.679***	0.671***	0.665***	0.664***
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
Borrower Z-score	-2.975***	-2.969***	-2.722***	-2.780***	-2.279***	-2.824***	-2.862***	-2.372***
	(0.540)	(0.540)	(0.539)	(0.544)	(0.547)	(0.536)	(0.541)	(0.544)
Borrower Profitability	-1.181***	-1.181***	-1.184***	-1.184***	-1.190***	-1.155***	-1.155***	-1.160***
	(0.083)	(0.083)	(0.082)	(0.083)	(0.083)	(0.082)	(0.082)	(0.082)
Borrower Market-to-Book	-0.382***	-0.383***	-0.376***	-0.392***	-0.391***	-0.339***	-0.355***	-0.353***
	(0.087)	(0.087)	(0.086)	(0.088)	(0.089)	(0.084)	(0.085)	(0.086)
Borrower Cash	0.178***	0.179***	0.138**	0.162**	0.118*	0.140**	0.162**	0.118*
	(0.064)	(0.064)	(0.064)	(0.064)	(0.063)	(0.063)	(0.063)	(0.063)
ln(Loan Size)	-0.081***	-0.081***	-0.082***	-0.082***	-0.082***	-0.082***	-0.082***	-0.082***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.035***	0.035***	0.035***	0.036***	0.035***	0.033***	0.034***	0.033***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Loan Secured	0.356***	0.356***	0.353***	0.355***	0.351***	0.353***	0.355***	0.351***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Lender Fixed Effects × Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,578	31,578	31,578	31,578	31,578	31,578	31,578	31,578
Adjusted R^2	0.728	0.728	0.730	0.729	0.730	0.731	0.730	0.731

Table 2.A7: **Sub-sample Analysis: Removing Firms without Patents**

Table 2.A7 reports the sub-sample analysis results where we restrict to the loans to borrowers with patents. Specifically, column (1) reports the baseline result. Column (2) controls for the borrower's segment similarity with bank's prior borrowers. Columns (3) through (5) control for the competition faced by the borrower using three of Hoberg and Phillips (2016) product market competition measures, respectively. Columns (6) to (8) additionally control for the borrower's patent value and patent stock, as well as the segment similarity. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity	-0.234**	-0.269**	-0.215**	-0.228**	-0.202*	-0.201*	-0.210*	-0.180*
	(0.107)	(0.109)	(0.106)	(0.107)	(0.106)	(0.108)	(0.109)	(0.109)
Segment Similarity		0.224*				0.224*	0.214*	0.193
		(0.125)				(0.125)	(0.124)	(0.123)
Borrower Product Market HHI			-0.134***			-0.135***		
			(0.025)			(0.025)		
Borrower Product Market Similarity				0.006***			0.006***	
				(0.002)			(0.002)	
Borrower Product Market Fluidity					0.014***			0.015***
					(0.003)			(0.002)
Borrower Patent Stock						-0.000	-0.000	-0.000
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Prior Relationship	-0.027*	-0.029*	-0.026*	-0.026*	-0.025*	-0.030**	-0.030**	-0.028*
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Borrower Size	-0.082***	-0.082***	-0.087***	-0.083***	-0.086***	-0.081***	-0.077***	-0.080***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
Borrower Leverage	0.645***	0.644***	0.655***	0.652***	0.655***	0.639***	0.636***	0.638***
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Borrower Z-score	-1.919***	-1.911***	-1.624***	-1.622***	-1.120*	-1.761***	-1.732***	-1.239**
	(0.593)	(0.594)	(0.590)	(0.600)	(0.609)	(0.587)	(0.595)	(0.605)
Borrower Profitability	-1.150***	-1.148***	-1.152***	-1.155***	-1.169***	-1.117***	-1.120***	-1.134***
	(0.092)	(0.093)	(0.092)	(0.093)	(0.092)	(0.090)	(0.092)	(0.091)
Borrower Market-to-Book	-0.520***	-0.519***	-0.514***	-0.533***	-0.541***	-0.450***	-0.468***	-0.475***
	(0.110)	(0.110)	(0.109)	(0.110)	(0.111)	(0.104)	(0.106)	(0.107)
Borrower Cash	0.083	0.083	0.033	0.056	0.015	0.038	0.058	0.019
	(0.068)	(0.068)	(0.067)	(0.068)	(0.067)	(0.066)	(0.068)	(0.067)
Borrower Has Credit Rating	0.022	0.022	0.023	0.023	0.023	0.018	0.018	0.019
	(0.017)	(0.017)	(0.016)	(0.017)	(0.017)	(0.016)	(0.017)	(0.016)
ln(Loan Size)	-0.104***	-0.104***	-0.105***	-0.104***	-0.104***	-0.105***	-0.104***	-0.104***
	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
ln(Loan Maturity)	0.033***	0.033***	0.033***	0.033***	0.033***	0.030***	0.030***	0.030***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Loan Secured	0.421***	0.421***	0.417***	0.420***	0.415***	0.417***	0.419***	0.415***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20,881	20,881	20,881	20,881	20,881	20,881	20,881	20,881
Adjusted R^2	0.682	0.682	0.683	0.683	0.684	0.685	0.684	0.685

Table 2.A8: **Sub-sample Analysis: Removing Banks with Few Borrowers and Loans to Major Customers**

Table 2.A8 reports sub-sample analysis results. Specifically, in columns (1) to (4), we remove the loans originated by the banks whose numbers of recent borrowers are in the bottom annual quartile. In columns (5) to (8), we remove the loans where the borrower is a major prior borrower of the bank defined by its loan amount in the top annual quartile. In all specifications, we control for borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Remove banks with few recent borrowers				Remove loans to major prior borrowers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity	-0.348***	-0.317***	-0.343***	-0.309***	-0.386***	-0.334***	-0.386***	-0.331***
	(0.098)	(0.096)	(0.098)	(0.096)	(0.106)	(0.103)	(0.106)	(0.103)
Segment Similarity	0.132		0.158		0.249**		0.249**	
	(0.106)		(0.106)		(0.117)		(0.117)	
Borrower Product Market HHI		-0.104***		-0.104***		-0.096***		-0.096***
		(0.021)		(0.021)		(0.022)		(0.022)
Prior Relationship			-0.037***	-0.036***	-0.036***		-0.036***	-0.035***
			(0.011)	(0.011)	(0.013)		(0.013)	(0.013)
Borrower Size	-0.078***	-0.081***	-0.079***	-0.082***	-0.073***	-0.075***	-0.073***	-0.076***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Borrower Leverage	0.552***	0.560***	0.554***	0.561***	0.546***	0.552***	0.546***	0.553***
	(0.028)	(0.028)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Borrower Z-score	-2.285***	-2.121***	-2.239***	-2.079***	-2.434***	-2.324***	-2.434***	-2.289***
	(0.453)	(0.451)	(0.453)	(0.451)	(0.466)	(0.466)	(0.466)	(0.465)
Borrower Profitability	-1.133***	-1.137***	-1.129***	-1.132***	-1.007***	-1.013***	-1.007***	-1.009***
	(0.067)	(0.066)	(0.067)	(0.066)	(0.068)	(0.068)	(0.068)	(0.068)
Borrower Market-to-Book	-0.318***	-0.311***	-0.320***	-0.313***	-0.336***	-0.327***	-0.336***	-0.329***
	(0.073)	(0.072)	(0.073)	(0.072)	(0.078)	(0.078)	(0.078)	(0.078)
Borrower Cash	0.088*	0.061	0.078	0.051	0.067	0.047	0.067	0.040
	(0.051)	(0.051)	(0.051)	(0.051)	(0.052)	(0.052)	(0.052)	(0.052)
Borrower Has Credit Rating	0.036***	0.036***	0.036***	0.036***	0.038***	0.038***	0.038***	0.038***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
ln(Loan Size)	-0.097***	-0.098***	-0.096***	-0.097***	-0.098***	-0.099***	-0.098***	-0.098***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.036***	0.036***	0.035***	0.036***	0.032***	0.033***	0.032***	0.032***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Loan Secured	0.381***	0.379***	0.380***	0.378***	0.367***	0.366***	0.367***	0.366***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,555	32,555	32,555	32,555	24,316	24,316	24,316	24,316
Adjusted R^2	0.658	0.658	0.658	0.659	0.634	0.634	0.634	0.634

Table 2.A9: **Borrower Technology Similarity Based on a High-Value Patent**

Table 2.A9 reports the results of regressing the natural logarithm of all-in-drawn loan spreads on borrower technology similarity. To focus on economically significant technological innovations, we construct the technology similarity measure using only patents with above-median economic value (based on Kogan et al. (2017) patent value measure) within their technology class-year group. This refined approach helps capture similarities in valuable technological expertise rather than general patent portfolios. Specifically, column (1) reports the baseline result. Column (2) controls for the borrower's segment similarity with bank's prior borrowers. Columns (3) through (5) control for the competition faced by the borrower using three of Hoberg and Phillips (2016) product market competition measures, respectively. Columns (6) to (8) additionally control for the borrower's patent value and patent stock, as well as the segment similarity. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity (High Value Patents)	-0.410***	-0.434***	-0.408***	-0.406***	-0.398***	-0.334***	-0.330***	-0.319***
	(0.113)	(0.112)	(0.112)	(0.112)	(0.111)	(0.109)	(0.110)	(0.108)
Segment Similarity (High Value Patents)		0.424***				0.364***	0.361***	0.321**
		(0.133)				(0.133)	(0.132)	(0.132)
Borrower Product Market HHI			-0.115***			-0.113***		
			(0.020)			(0.019)		
Borrower Product Market Similarity				0.004***			0.004***	
				(0.001)			(0.001)	
Borrower Product Market Fluidity					0.014***			0.014***
					(0.002)			(0.002)
Borrower Patent Stock						-0.001	-0.001	-0.001
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Prior Relationship	-0.019*	-0.025**	-0.020*	-0.020*	-0.020*	-0.027***	-0.027***	-0.026**
	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Borrower Size	-0.080***	-0.081***	-0.083***	-0.081***	-0.084***	-0.079***	-0.076***	-0.079***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Borrower Leverage	0.579***	0.578***	0.587***	0.582***	0.584***	0.574***	0.569***	0.571***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Borrower Z-score	-1.998***	-1.956***	-1.811***	-1.829***	-1.363***	-1.879***	-1.893***	-1.434***
	(0.444)	(0.444)	(0.442)	(0.446)	(0.448)	(0.439)	(0.442)	(0.444)
Borrower Profitability	-1.127***	-1.129***	-1.130***	-1.129***	-1.134***	-1.107***	-1.105***	-1.110***
	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.062)	(0.062)	(0.062)
Borrower Market-to-Book	-0.321***	-0.321***	-0.312***	-0.329***	-0.335***	-0.273***	-0.290***	-0.296***
	(0.073)	(0.073)	(0.073)	(0.073)	(0.074)	(0.070)	(0.071)	(0.072)
Borrower Cash	0.091*	0.088*	0.061	0.076	0.038	0.063	0.077	0.040
	(0.049)	(0.049)	(0.049)	(0.050)	(0.049)	(0.049)	(0.049)	(0.049)
ln(Loan Size)	-0.096***	-0.095***	-0.096***	-0.096***	-0.096***	-0.096***	-0.096***	-0.096***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.016**	0.017**	0.017**	0.017**	0.017**	0.014**	0.014*	0.014*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Loan Secured	0.387***	0.387***	0.385***	0.386***	0.382***	0.385***	0.386***	0.381***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	36,166	36,166	36,166	36,166	36,166	36,166	36,166	36,166
Adjusted R^2	0.653	0.653	0.654	0.653	0.655	0.655	0.655	0.656

Table 2.A10: **Borrower Technology Similarity Based on High Citation Patents**

Table 2.A10 reports the results of regressing the natural logarithm of all-in-drawn loan spreads on borrower technology similarity. To focus on influential technological innovations, we construct the technology similarity measure using only patents with above-median citations within their technology class-year group. This refined approach helps capture similarities in impactful technological expertise rather than general patent portfolios. Specifically, column (1) reports the baseline result. Column (2) controls for the borrower's segment similarity with bank's prior borrowers. Columns (3) through (5) control for the competition faced by the borrower using three of Hoberg and Phillips (2016) product market competition measures, respectively. Columns (6) to (8) additionally control for the borrower's patent value and patent stock, as well as the segment similarity. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity (High Citation Patents)	-0.388***	-0.418***	-0.380***	-0.384***	-0.363***	-0.327***	-0.329***	-0.304***
	(0.086)	(0.087)	(0.086)	(0.086)	(0.085)	(0.085)	(0.086)	(0.085)
Segment Similarity (High Citation Patents)		0.450***				0.386***	0.384***	0.341**
		(0.134)				(0.134)	(0.133)	(0.132)
Borrower Product Market HHI			-0.114***			-0.112***		
			(0.020)			(0.020)		
Borrower Product Market Similarity				0.004***			0.004***	
				(0.001)			(0.001)	
Borrower Product Market Fluidity					0.014***			0.014***
					(0.002)			(0.002)
Borrower Patent Stock						-0.000	-0.000	-0.001
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Prior Relationship	-0.019*	-0.025**	-0.019*	-0.019*	-0.019*	-0.027**	-0.027**	-0.026**
	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Borrower Size	-0.081***	-0.082***	-0.084***	-0.082***	-0.085***	-0.079***	-0.077***	-0.080***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Borrower Leverage	0.580***	0.578***	0.587***	0.583***	0.585***	0.575***	0.570***	0.572***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Borrower Z-score	-2.053***	-2.014***	-1.865***	-1.883***	-1.418***	-1.925***	-1.940***	-1.482***
	(0.444)	(0.443)	(0.442)	(0.446)	(0.448)	(0.439)	(0.442)	(0.444)
Borrower Profitability	-1.126***	-1.128***	-1.129***	-1.128***	-1.134***	-1.106***	-1.104***	-1.110***
	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.062)	(0.062)	(0.062)
Borrower Market-to-Book	-0.327***	-0.327***	-0.318***	-0.335***	-0.341***	-0.278***	-0.294***	-0.300***
	(0.073)	(0.073)	(0.073)	(0.073)	(0.074)	(0.070)	(0.071)	(0.072)
Borrower Cash	0.093*	0.090*	0.063	0.078	0.040	0.065	0.079	0.041
	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.048)
ln(Loan Size)	-0.096***	-0.096***	-0.097***	-0.096***	-0.096***	-0.097***	-0.096***	-0.097***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.017**	0.017**	0.018**	0.017**	0.017**	0.015**	0.015**	0.015**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Loan Secured	0.387***	0.386***	0.385***	0.386***	0.382***	0.384***	0.385***	0.381***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	36,166	36,166	36,166	36,166	36,166	36,166	36,166	36,166
Adjusted R^2	0.653	0.653	0.654	0.653	0.655	0.655	0.655	0.656

Table 2.A11: **Borrower Technology Similarity and Loan Spread Controlling for Non-bank Lenders**

Table 2.A11 reports the results of regressing the natural logarithm of all-in-drawn loan spreads on borrower technology similarity controlling for the proportion of non-bank lenders on each loans. Specifically, column (1) reports the baseline result. Column (2) controls for the borrower's segment similarity with bank's prior borrowers. Columns (3) through (5) control for the competition faced by the borrower using three of Hoberg and Phillips (2016) product market competition measures, respectively. Columns (6) to (8) additionally control for the borrower's patent value and patent stock, as well as the segment similarity. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity	-0.392***	-0.425***	-0.374***	-0.385***	-0.362***	-0.322***	-0.331***	-0.304***
	(0.093)	(0.094)	(0.092)	(0.093)	(0.092)	(0.094)	(0.095)	(0.094)
Segment Similarity		0.229**				0.223**	0.225**	0.200**
		(0.093)				(0.093)	(0.093)	(0.091)
Borrower Product Market HHI			-0.113***			-0.113***		
			(0.020)			(0.020)		
Borrower Product Market Similarity				0.004***			0.004***	
				(0.001)			(0.001)	
Borrower Product Market Fluidity					0.014***			0.014***
					(0.002)			(0.002)
Borrower Patent Stock						-0.001	-0.001	-0.001
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Prior Relationship	-0.010	-0.012	-0.010	-0.010	-0.010	-0.015	-0.015	-0.015
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Borrower Size	-0.089***	-0.089***	-0.092***	-0.090***	-0.093***	-0.087***	-0.084***	-0.087***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)
Borrower Leverage	0.557***	0.557***	0.565***	0.560***	0.562***	0.555***	0.550***	0.551***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Borrower Z-score	-1.994***	-1.970***	-1.805***	-1.818***	-1.358***	-1.875***	-1.877***	-1.425***
	(0.442)	(0.443)	(0.440)	(0.445)	(0.446)	(0.438)	(0.442)	(0.444)
Borrower Profitability	-1.128***	-1.127***	-1.131***	-1.129***	-1.135***	-1.104***	-1.102***	-1.107***
	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.062)	(0.062)	(0.062)
Borrower Market-to-Book	-0.325***	-0.325***	-0.316***	-0.333***	-0.340***	-0.276***	-0.293***	-0.299***
	(0.073)	(0.073)	(0.072)	(0.073)	(0.074)	(0.070)	(0.071)	(0.071)
Borrower Cash	0.093*	0.094*	0.064	0.078	0.040	0.068	0.081*	0.044
	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.048)	(0.049)	(0.048)
Borrower Has Credit Rating	0.040***	0.040***	0.040***	0.040***	0.041***	0.037***	0.037***	0.037***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.011)
ln(Loan Size)	-0.099***	-0.099***	-0.100***	-0.100***	-0.100***	-0.100***	-0.100***	-0.100***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Loan Maturity)	0.015**	0.015**	0.016**	0.015**	0.015**	0.013*	0.013*	0.012*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Loan Secured	0.384***	0.384***	0.382***	0.383***	0.379***	0.382***	0.383***	0.379***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Non-bank Lenders	0.021***	0.021***	0.021***	0.021***	0.021***	0.021***	0.021***	0.021***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	36,166	36,166	36,166	36,166	36,166	36,166	36,166	36,166
Adjusted R^2	0.654	0.654	0.655	0.654	0.656	0.657	0.656	0.658

Table 2.A12: **Placebo Test: Technology Similarity with Future Borrowers**

Table 2.A12 reports the results of regressing the natural logarithm of all-in-drawn loan spreads on the technology similarity of current borrower with future borrowers from five years to ten years after loan origination. Specifically, column (1) reports the baseline result. Column (2) controls for the borrower's segment similarity with bank's prior borrowers. Columns (3) through (5) control for the competition faced by the borrower using three of Hoberg and Phillips (2016) product market competition measures, respectively. Columns (6) to (8) additionally control for the borrower's patent value and patent stock, as well as the segment similarity. In all specifications, we control for bank-borrower prior lending relationship, borrower characteristics, loan characteristics, and fixed effects for loan type, loan purpose, borrower industry – year, and lender. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technology Similarity (Future Borrowers)	-0.161	-0.162	-0.163	-0.159	-0.140	-0.165	-0.161	-0.140
	(0.103)	(0.104)	(0.103)	(0.103)	(0.102)	(0.103)	(0.103)	(0.102)
Segment Similarity (Future Borrowers)		0.038				0.021	0.029	0.013
		(0.232)				(0.229)	(0.230)	(0.232)
Borrower Product Market HHI			-0.107***			-0.110***		
			(0.028)			(0.028)		
Borrower Product Market Similarity				0.003			0.004*	
				(0.002)			(0.002)	
Borrower Product Market Fluidity					0.011***			0.011***
					(0.003)			(0.003)
Borrower Patent Stock						-0.000	-0.000	-0.000
						(0.001)	(0.001)	(0.001)
Borrower Patent Value						-0.001***	-0.001***	-0.001***
						(0.000)	(0.000)	(0.000)
Prior Relationship	-0.049***	-0.049***	-0.049***	-0.049***	-0.048***	-0.049***	-0.050***	-0.048***
	(0.018)	(0.018)	(0.017)	(0.018)	(0.018)	(0.017)	(0.017)	(0.017)
Borrower Size	-0.075***	-0.075***	-0.079***	-0.076***	-0.078***	-0.071***	-0.067***	-0.070***
	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Borrower Leverage	0.651***	0.651***	0.661***	0.656***	0.662***	0.640***	0.635***	0.641***
	(0.046)	(0.046)	(0.046)	(0.045)	(0.045)	(0.046)	(0.045)	(0.045)
Borrower Z-score	-1.484**	-1.482**	-1.188*	-1.317**	-0.889	-1.355**	-1.441**	-1.027
	(0.663)	(0.662)	(0.665)	(0.671)	(0.688)	(0.659)	(0.663)	(0.681)
Borrower Profitability	-1.336***	-1.336***	-1.344***	-1.340***	-1.352***	-1.298***	-1.294***	-1.306***
	(0.104)	(0.104)	(0.104)	(0.104)	(0.104)	(0.102)	(0.102)	(0.102)
Borrower Market-to-Book	-0.385***	-0.385***	-0.379***	-0.390***	-0.396***	-0.303***	-0.314***	-0.319***
	(0.110)	(0.110)	(0.109)	(0.110)	(0.111)	(0.103)	(0.104)	(0.105)
Borrower Cash	0.036	0.036	-0.004	0.020	-0.014	0.008	0.030	-0.002
	(0.078)	(0.078)	(0.077)	(0.078)	(0.077)	(0.075)	(0.077)	(0.076)
ln(Loan Size)	-0.110***	-0.110***	-0.111***	-0.110***	-0.110***	-0.111***	-0.110***	-0.110***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
ln(Loan Maturity)	0.048***	0.048***	0.048***	0.048***	0.048***	0.045***	0.045***	0.045***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Loan Secured	0.409***	0.409***	0.406***	0.408***	0.405***	0.404***	0.407***	0.404***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry × Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,108	17,108	17,108	17,108	17,108	17,108	17,108	17,108
Adjusted R^2	0.688	0.688	0.689	0.688	0.689	0.691	0.690	0.691

Table 2.A13: **List of Bank Mergers and Acquisitions (M&A) Events**

Table 2.A13 lists the M&A events used for the difference-in-differences estimations in Table 2.5.

Event Year	Acquirer Bank	Target Bank
1996	JPMorgan Chase Bank	Chase Manhattan Corp
1996	JPMorgan Chase Bank	Chemical Securities Asia
1996	JPMorgan Chase Bank	Chase Manhattan Bank of Canada
1996	JPMorgan Chase Bank	Chase Manhattan Plc
1996	JPMorgan Chase Bank	Chase Manhattan Asia Ltd
1996	JPMorgan Chase Bank	Chase Manhattan Australia
1996	JPMorgan Chase Bank	Chemical Bank
1996	JPMorgan Chase Bank	Chase Securities
1996	JPMorgan Chase Bank	Chase Manhattan Bank
1996	JPMorgan Chase Bank	Chase Manhattan Australia
1996	JPMorgan Chase Bank	Chase Securities
1996	JPMorgan Chase Bank	Chemical Bank of Canada
1996	JPMorgan Chase Bank	Chemical Bank Australia
1996	JPMorgan Chase Bank	Chase Manhattan Investment Bank Ltd
1996	JPMorgan Chase Bank	Chemical Bank New Jersey NA
1996	JPMorgan Chase Bank	Chase Manhattan Bank of Canada
1996	JPMorgan Chase Bank	Chase Securities Australia
1996	JPMorgan Chase Bank	Chase Manhattan Bank
1997	Bankers Trust Co	BT Alex Brown Inc
1998	Bank of the West	First Hawaiian Bank
1998	Norwest Bank	Foothill Capital Corp
1998	Norwest Bank	Wells Fargo - Texas
1998	Norwest Bank	Wells Fargo Bank Texas NA
1998	Norwest Bank	Foothill Group
1998	Norwest Bank	Wells Fargo Bank
1999	Fleet Bank	Bank Boston Trust Co
1999	Fleet Bank	Michigan National Bank
1999	Fleet Bank	BankBoston NA
1999	Fleet Bank	Bank Boston
1999	Fleet Bank	Shawmut Bank Connecticut
1999	Fleet Bank	BankBoston Corp
1999	Fleet Bank	BankBoston Retail Finance Inc
1999	Fleet Bank	Shawmut Capital Corp
1999	Fleet Bank	BankBoston Capital
1999	Fleet Bank	Bank Boston Singapore
2004	National City Bank	Provident Bank
2005	Zions Bank	Amegy Bank NA
2006	Regions	AmSouth Bank
2008	PNC Bank	National City Bank
2008	PNC Bank	National City Business Credit
2008	Barclays	Lehman Commercial Paper Inc
2015	Royal Bank of Canada	City National Bank
2016	KeyBank	First Niagara Financial Group Inc
2016	KeyBank	First Niagara Bank

Table 2.A14: **Comparison of Treated and Control Banks**

Table 2.A14 shows the t-tests to examine whether the treated and control banks are comparable after propensity score matching on bank size, non-deposit leverage, deposits ratio, ROA, total loans, and loan-to-deposits shortfall. Columns (1) and (2) report the mean value of characteristics for treated and control banks respectively. Column (3) reports the difference. Columns (4) and (5) report the t-statistic and p-value. Column (6) reports the number of observations. Definitions of the variables are provided in Table 2.A1 in the Appendix.

	Mean			t-test		Observations
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Treated	Control	Difference (1)-(2)	t	$p > \|t\|$	Total
Bank Size	11.654	11.339	0.315	0.510	0.615	120
Bank Non-Deposit Leverage	0.214	0.181	0.059	0.930	0.361	120
Bank Deposits	0.609	0.668	-0.059	-1.090	0.287	120
Bank ROA	0.011	0.011	0.000	0.000	0.784	120
Bank Total Loans	0.536	0.562	-0.026	-0.410	0.685	120
Bank Loan-to-Deposit Shortfall	-0.073	-0.107	0.034	0.760	0.458	120

Table 2.A15: **Dynamic Difference-in-Differences Estimations: Bank M&A**

Table 2.A15 shows the results of the dynamic difference-in-differences estimation using bank M&As as exogenous positive shocks to the stock of acquirer banks' technology knowledge and hence value of technology similarity in loan pricing. The specifications follow Table 2.5, but we replace the single *Post* dummy variable with a series of indicators $\{D_j\}$, where D_j takes the value of one if the loan is issued in the j-th year after the event year, and zero otherwise. Definitions of other variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the lender and year levels. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
	All M&As		Technology Learned	
D_{-1} × Treated	0.048	0.062	0.047	0.060
	(0.076)	(0.076)	(0.077)	(0.076)
D_0 × Treated	0.029	0.031	0.044	0.044
	(0.082)	(0.084)	(0.081)	(0.083)
D_1 × Treated	-0.163**	-0.173**	-0.172**	-0.183**
	(0.065)	(0.067)	(0.070)	(0.072)
D_2 × Treated	-0.027	-0.032	-0.026	-0.032
	(0.078)	(0.076)	(0.080)	(0.078)
D_{-1}	-0.021	-0.021	-0.022	-0.023
	(0.024)	(0.024)	(0.026)	(0.026)
D_0	-0.022	-0.023	-0.027	-0.026
	(0.021)	(0.021)	(0.026)	(0.026)
D_1	-0.015	-0.014	-0.016	-0.012
	(0.019)	(0.019)	(0.021)	(0.020)
D_2	-0.028	-0.026	-0.026	-0.024
	(0.017)	(0.017)	(0.018)	(0.017)
Treated	-0.060	-0.059	-0.065	-0.063
	(0.056)	(0.054)	(0.059)	(0.057)
Borrower Patent Stock		-0.015		-0.014
		(0.011)		(0.012)
Borrower Patent Value		-0.001		-0.001*
		(0.001)		(0.001)
Segment Similarity	-0.091	-0.020	-0.113	-0.040
	(0.440)	(0.423)	(0.484)	(0.464)
Borrower Product Market HHI	-0.289***	-0.282***	-0.306***	-0.299***
	(0.069)	(0.067)	(0.071)	(0.070)
Prior Relationship	-0.057	-0.052	-0.049	-0.046
	(0.057)	(0.056)	(0.059)	(0.057)
Loan Level Controls	Yes	Yes	Yes	Yes
Borrower Level Controls	Yes	Yes	Yes	Yes
Loan Type and Purpose Fixed Effects	Yes	Yes	Yes	Yes
Borrower Industry Fixed Effects	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	6,074	6,074	5,524	5,524
Adjusted R^2	0.751	0.752	0.751	0.752

Table 2.A16: **Information Content of Technology Similarity**

Table 2.A16 examines the explanatory power of a borrower's technology similarity with its bank's prior borrowers for the difference in their creditworthiness. Specifically, we regress the absolute difference of borrower's and bank's prior borrowers' average creditworthiness measures on their technology similarity, controlling for their segment similarity and absolute differences across an array of firm characteristics. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable (Absolute Difference):	Z-score	Default Probability	Profitability	Cash Holding
	(1)	(2)	(3)	(4)
Technology Similarity	-0.005***	-0.037***	-0.056***	-0.038***
	(0.002)	(0.013)	(0.010)	(0.011)
Segment Similarity	-0.004	0.008	-0.040***	-0.037***
	(0.002)	(0.015)	(0.013)	(0.012)
Absolute difference in size	0.001***	-0.002**	0.004***	0.002***
	(0.000)	(0.001)	(0.001)	(0.001)
Absolute difference in leverage	0.017***	0.122***	0.062***	0.048***
	(0.001)	(0.009)	(0.006)	(0.005)
Absolute difference in market-to-book ratio	0.009***	-0.022	0.204***	0.050***
	(0.002)	(0.015)	(0.020)	(0.015)
Absolute difference in sales growth	0.000***	-0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Absolute difference in tangibility	0.003***	0.009	-0.010**	0.012**
	(0.001)	(0.007)	(0.005)	(0.005)
Absolute difference in patent stock	-0.037	0.293	-0.110	-0.055
	(0.025)	(0.219)	(0.093)	(0.194)
Absolute difference in patent value	0.000**	-0.000**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Industry and Year Fixed Effects	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes
Observations	36,166	28,074	36,166	36,166
Adjusted R^2	0.238	0.258	0.209	0.130

Table 2.A17: **Information Content of Technology Similarity for Future Borrower Creditworthiness**

Table 2.A17 examines the predictive power of a borrower's technology similarity with its bank's prior borrowers for the difference in their future creditworthiness. Specifically, we regress the future absolute difference of borrower's and bank's prior borrowers' average forward creditworthiness measures on their technology similarity, controlling for their segment similarity and absolute differences across an array of firm characteristics. The dependent variable is measured at year $t+h$ and all independent variables are measured at year t. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are heteroskedasticity-robust standard errors clustered at the borrower level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Absolute Difference ($h=1$):	Z-score	Default Probability	Profitability	Cash Holding
	(1)	(2)	(3)	(4)
Technology Similarity	-0.002	-0.046***	-0.054***	-0.002
	(0.003)	(0.012)	(0.013)	(0.012)
Absolute difference controls	Yes	Yes	Yes	Yes
Industry and Year Fixed Effects	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes
Observations	33,636	26,040	33,652	33,579
Adjusted R^2	0.160	0.268	0.168	0.129

Absolute Difference ($h=2$):	Z-score	Default Probability	Profitability	Cash Holding
	(1)	(2)	(3)	(4)
Technology Similarity	-0.006*	-0.068***	-0.072***	-0.013
	(0.004)	(0.014)	(0.015)	(0.012)
Absolute difference controls	Yes	Yes	Yes	Yes
Industry and Year Fixed Effects	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes
Observations	30,750	23,455	30,797	30,660
Adjusted R^2	0.028	0.264	0.070	0.140

Absolute Difference ($h=3$):	Z-score	Default Probability	Profitability	Cash Holding
	(1)	(2)	(3)	(4)
Technology Similarity	-0.006**	-0.069***	-0.057***	-0.024*
	(0.003)	(0.014)	(0.015)	(0.012)
Absolute difference controls	Yes	Yes	Yes	Yes
Industry and Year Fixed Effects	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes
Observations	27,793	21,081	27,853	27,684
Adjusted R^2	0.142	0.281	0.178	0.126

Absolute Difference ($h=4$):	Z-score	Default Probability	Profitability	Cash Holding
	(1)	(2)	(3)	(4)
Technology Similarity	-0.008**	-0.039**	-0.059***	-0.027**
	(0.003)	(0.018)	(0.016)	(0.013)
Absolute difference controls	Yes	Yes	Yes	Yes
Industry and Year Fixed Effects	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes
Observations	25,108	18,758	25,174	24,984
Adjusted R^2	0.221	0.274	0.203	0.121

Absolute Difference ($h=5$):	Z-score	Default Probability	Profitability	Cash Holding
	(1)	(2)	(3)	(4)
Technology Similarity	-0.007*	-0.065***	-0.078***	-0.033**
	(0.003)	(0.018)	(0.017)	(0.014)
Absolute difference controls	Yes	Yes	Yes	Yes
Industry and Year Fixed Effects	Yes	Yes	Yes	Yes
Lender Fixed Effects	Yes	Yes	Yes	Yes
Observations	22,616	16,922	22,690	22,475
Adjusted R^2	0.115	0.283	0.147	0.119

Figure 2.A1: **Pairwise Technology Similarity**

Figure 2.A1 illustrates the pairwise technology similarity calculation for a borrower firm i, at loan origination time t, for the two prior borrowers j and k of the bank in the five years leading to t. The degree of technology similarity is computed based on the patent portfolios of borrower j and k as at their respective borrowing time, instead of time t, and the portfolio of firm i as at time t. This specification builds on the assumption that the lending bank learns about a borrower's patent portfolio at loan origination.

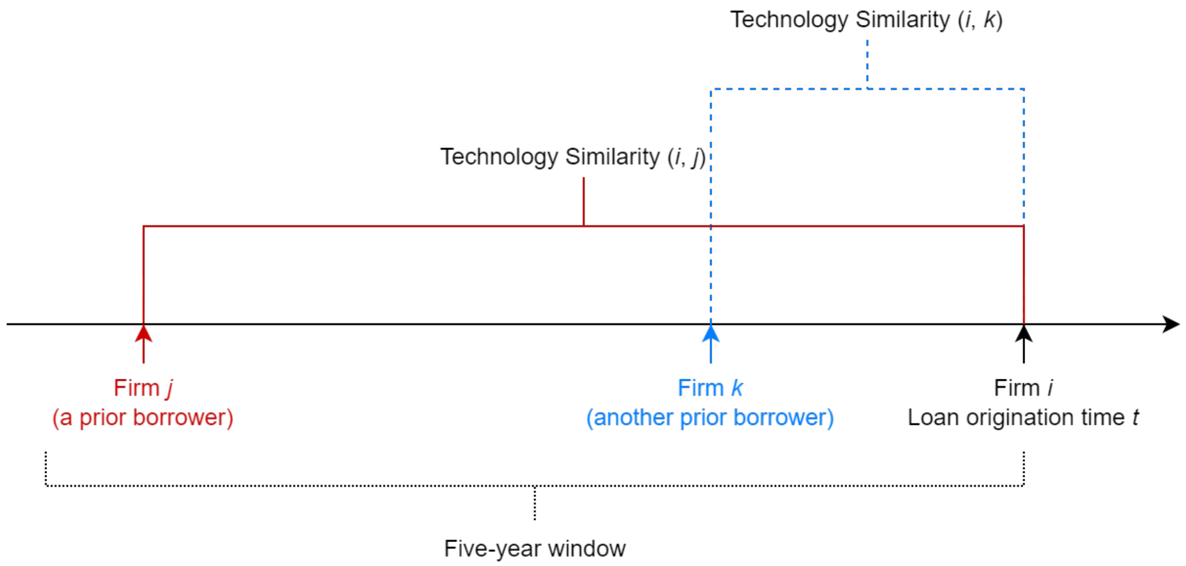
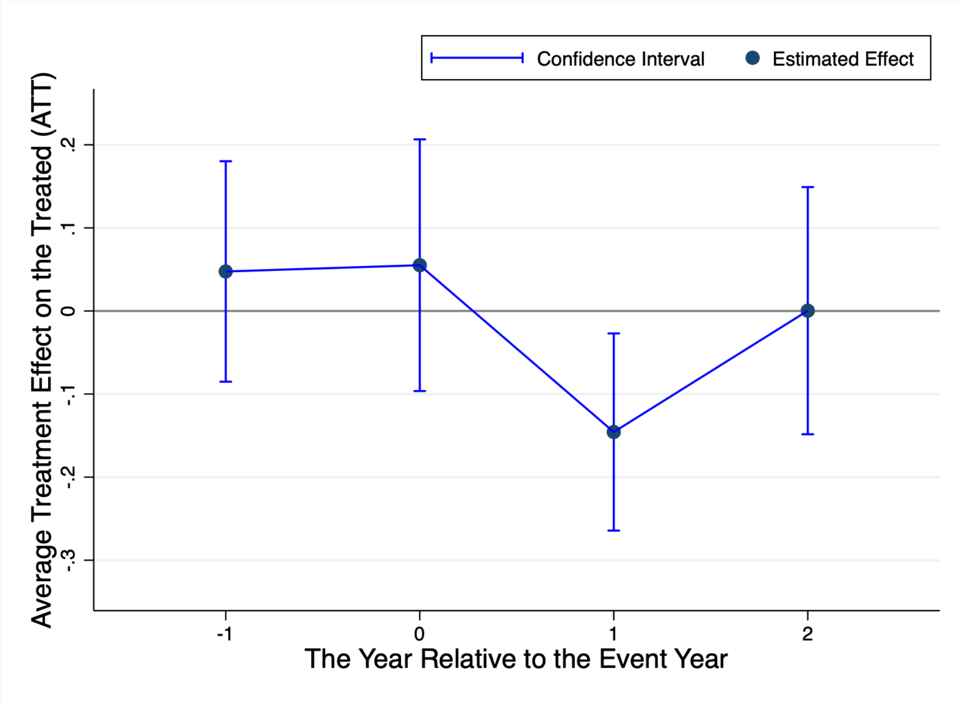


Figure 2.A2: **Treatment Effect Around the Bank M&A Event Year**

Figure 2.A2 shows the coefficient estimates of the interaction of the time dummies and the treated dummy in the dynamic difference-in-differences regression in Table 2.5. The figure also shows the 90% confidence interval of the coefficient estimates.



Chapter 3

Connected and Contagious: Unveiling Hidden Bank Vulnerabilities through AI-driven Co-Lending Networks

"Complex links among financial market participants and institutions are a hallmark of the modern global financial system. Across geographic and market boundaries, agents within the financial system engage in a diverse array of transactions and relationships that connect them to other participants." (Yellen, 2013)

3.1 Introduction

This study develops an innovative deep learning model, the Co-Lending Graph Neural Network (CoLGNN), to analyze risk spillovers among financial institutions within the syndicated lending market. In today's rapidly evolving global economy, interconnectedness among banks has become increasingly recognized as a crucial determinant of systemic risk. Extensive literature highlights the critical role that network structures play in propagating financial distress and amplifying risks across institutions (e.g. Acemoglu et al., 2015; Allen & Gale, 2000; Cabrales et al., 2017; Elliott et al., 2014). Lessons from the global financial crisis (GFC) clearly illustrate how localized shocks can quickly cascade through the financial system (Yellen, 2013). More recent events, including the Dexia bank bailout (2011), the Cyprus banking crisis (2013), the collapse of Banco Espírito Santo (2014), and the bankruptcy of Silicon Valley Bank (2023), further emphasize this vulnerability. Although regulations such as Basel capital requirements are intended to mitigate individual banks' risks (Acharya et al., 2017), the systemic nature of risk transmission demands more nuanced measures that accurately capture the broader network of interbank exposures.

One critical yet relatively underexplored area of risk propagation involves the cross-sectional interdependencies within syndicated loan markets. With a global scale exceeding $4.7 trillion

USD in 2022, syndicated loans represent a significant segment of corporate financing. This market's inherently asymmetric structure creates unique channels through which risks propagate, especially through monitoring-intensive revolving credit facilities that maintain persistent relationships between lead arrangers and participant banks. While term loans may be frequently sold in secondary markets, revolving facilities require continual monitoring, making them especially potent vectors for risk transmission (Blickle et al., 2020).

Additionally, existing bank-risk metrics often depend on publicly available stock returns, leveraging the efficiency of market-based information (e.g., Fama, 1970, 1991; Merton, 1974; Nagel & Purnanandam, 2020; Tobias & Brunnermeier, 2016). However, this reliance poses a critical limitation: private banks, a significant and integral component of the banking sector's interconnected structure, are excluded due to the absence of market data. Omitting private banks thus overlooks crucial aspects of risk transmission mechanisms, potentially impairing regulatory and institutional responses to systemic risk.

To bridge these gaps, we construct a directed co-lending network representing the asymmetric relationships inherent in syndicated lending, and develop a customized deep learning CoLGNN model to analyze risk spillover within this network. Our novel CoLGNN framework leverages deep learning techniques specifically designed for complex, non-Euclidean network structures.[1] Unlike traditional network analyses that rely solely on node centrality measures, our approach integrates comprehensive bank-level data, loan characteristics, and the intricate topological structure of lending relationships. Employing the "message-passing paradigm" (Chami et al., 2022), our framework captures nuanced, directional risk transmission among banks within the syndicated loan market.

In the syndicated lending market, banks manage credit exposure by jointly extending large loans—a practice known as loan syndication. These lending syndicates depend critically on the lead arranger, who originates and monitors the loan (e.g., Gopalan et al., 2011; Ivashina, 2009; Pichler & Wilhelm, 2001). A defining feature of this structure is the information asymmetry between lead arrangers and participant banks, as the latter typically lack oversight in both borrower screening and post-origination monitoring (Ivashina, 2009; Ivashina & Scharfstein, 2010b). Although lead arrangers often offload substantial portions of their loan exposures in the secondary market (Blickle et al., 2020), certain facilities—particularly revolving credit lines—embed ongoing monitoring responsibilities due to their contingent drawdown nature. These monitoring-intensive facilities thus become key channels for risk propagation. Motivated by this asymmetry, we model co-lending ties as directed edges from lead arrangers to participant banks. Since banks may serve as arrangers in some syndicates and participants in others, we

[1] Non-Euclidean data lack fixed spatial coordinates, unlike standard panel data. Social, supply chain, and financial networks are prime examples due to their complex connectivity structures (Bronstein et al., 2017).

define nodes at the bank holding company level and encode co-lending relationships as directed edges, thereby capturing the directional flow of risk from lead to participant institutions.

However, the complex topological structure of financial networks presents significant methodological challenges. Traditional empirical approaches, such as centrality metrics, offer only a partial perspective on multidimensional relationships and risk transmission because they primarily emphasize the relative importance of individual nodes. Our study departs from these conventional methods by applying a deep learning model, CoLGNN, to capture bank-level risk spillovers within the co-lending network. As a state-of-the-art framework tailored for non-Euclidean domains,[2] graph neural networks (GNNs) have emerged as the leading modeling tools for prediction tasks on network-structured data across a variety of domains, including social networks, supply chains, and computer vision (Wu et al., 2020; Xu et al., 2019).[3]

GNNs address the limitations of traditional empirical tools by jointly modeling high-dimensional node features and complex network structures. In our context, the CoLGNN model integrates granular bank-level characteristics, the topological configuration of co-lending relationships, and the detailed attributes of syndicated loans. It leverages the message-passing paradigm to capture how risk propagates through the network. To operationalize this, we develop a graph diffusion convolution operator that systematically aggregates information from both neighboring nodes and associated edge features. These diffusion convolution modules are specifically designed to detect directional spillovers within the co-lending network by incorporating bank attributes, network topology, and loan-level characteristics into the deep learning process.

Empirically, we apply the CoLGNN model to a comprehensive sample of U.S. syndicated loans from Refinitiv DealScan, combined with bank holding company data from the FR Y-9C database, to generate a co-lending network risk score (CLN score) for each bank—public and private—at the year-quarter level. At each quarter-end, we construct rolling-window co-lending networks based on all syndicated loans originated over the prior five years. Each network is represented as an attributed graph, where node features capture bank characteristics and edge features reflect loan-level attributes.

Using a semi-supervised deep learning framework, we train the CoLGNN model to estimate CLN scores across the network. The model is trained on a labeled subset of public banks for which risk labels—based on stock price performance—are observable. Specifically, motivated by the efficient market hypothesis, which posits that stock prices reflect all publicly available

[2]Non-Euclidean data lack properties such as global parameterization, a common coordinate system, vector space structure, or shift invariance (Bronstein et al., 2017). In contrast, Euclidean data, such as panel datasets, are organized in fixed dimensions. Networks like supply chains and social systems are inherently non-Euclidean due to their complex and irregular connectivity. Notably, if one computes centrality scores over time and arranges them as a panel, the resulting structure becomes Euclidean.

[3]GNN-based models consistently outperform traditional methods on graph-related benchmark tasks in deep learning. See: https://paperswithcode.com/task/node-classification.

information, we classify banks in the top decile of raw quarterly stock returns as "safe" and those in the bottom decile as "risky."[4]

The trained CoLGNN model estimates a CLN score for each bank in each year-quarter by jointly leveraging node features (bank characteristics), edge features (loan attributes), and the topological structure of the co-lending network through its graph convolution modules. To enhance model interpretability, we further apply the integrated gradients method to quantify the contribution of each input feature to the resulting CLN score, thereby identifying key drivers of network-based bank risk.

We validate the effectiveness of the CLN score in capturing risk spillovers—rather than merely reflecting common exposure to distressed corporate borrowers—using two quasi-natural experiments. First, we exploit S&P credit rating downgrade events to examine whether idiosyncratic shocks to focal banks propagate to their syndicate partners. Using a stacked-cohort difference-in-differences design, we find that banks directly connected to downgraded institutions experience subsequent and significantly larger increases in their CLN scores compared to matched control banks without such exposure, consistent with risk contagion. Second, we analyze the collapse of Lehman Brothers—a shock originating outside its corporate lending activities—as a clean setting to isolate network-driven transmission from borrower-side contagion. The financial distress that led to Lehman's collapse stemmed primarily from concentrated investments in commercial real estate and a heavy reliance on unstable short-term funding, rather than from deterioration in its corporate loan portfolio. Because the shock did not originate from credit risk in shared corporate borrowers, any observed effect on its syndicate lending partners must arise through the co-lending network. Consistent with directional spillovers, we find that banks with prior co-lending ties to Lehman exhibit a 28–29% increase in their CLN scores relative to the sample mean following its failure.

After validating the effectiveness of the CLN score in capturing risk spillovers, we next examine its predictive power for future bank risk using a comprehensive panel dataset spanning 1996 to 2020. Our analysis shows that the CLN score significantly forecasts future loan loss provisions up to eight quarters ahead. Notably, this predictive ability extends to both unlabeled public and private banks, highlighting the CLN score's value as an early-warning indicator—particularly for banks that lack market-based risk measures. Incorporating the CLN score into predictive models materially improves forecast accuracy, as evidenced by higher adjusted R-squared values and lower root-mean-square errors, whereas the traditional Z-score contributes minimal incremental explanatory power. The results remain robust after controlling for a wide set of bank characteristics, lending specializations, year-quarter fixed effects, and even bank fixed effects.

[4] We use raw stock returns rather than factor-adjusted or idiosyncratic returns because both systematic and idiosyncratic risks are integral to assessing bank stability. Raw returns provide a comprehensive, forward-looking proxy that captures both market-wide and firm-specific information. This aligns with the efficient market hypothesis (Fama, 1970). Moreover, Campbell et al. (2008) demonstrate that both components of stock returns offer valuable insights into the risk profile of financial institutions.

Moreover, we show that the CLN score significantly outperforms standard network centrality measures in predicting future bank risk.

Recognizing that lead arrangers often sell substantial portions of their initial term loan allocations in the secondary market, we construct a modified co-lending network that focuses exclusively on revolving credit facilities—loan contracts in which lead arrangers retain ongoing monitoring responsibilities. We train a separate CoLGNN model using this revolver-based network to more precisely identify the channels through which risk propagates. We find that the CLN score derived solely from revolving facilities exhibits stronger risk predictive ability than the baseline CLN score. As a placebo test, we construct a co-lending network using Term Loan B and above, which refers to the tranches of syndicated loans—including Term Loans B, C, and beyond—that are frequently held by non-bank investors and traded in the secondary market. Consistent with our expectation, the CLN score based on Term Loan B and above shows predominantly insignificant predictive power. This contrast provides empirical support for the role of monitoring incentives in driving risk spillovers, highlighting that risk contagion is more likely to occur through revolving credit facilities where lead arrangers maintain active involvement, rather than through term loans that are commonly sold in the secondary market.

Furthermore, because the CLN score captures risk spillovers from neighboring banks within the co-lending network, we hypothesize that more vulnerable banks—specifically those with smaller size, weaker earnings performance, higher return volatility, and lower capital adequacy—are more sensitive to these spillovers. Our results confirm this conjecture. In addition, the CLN score demonstrates predictive power for a range of risk measures beyond loan loss provisions, including non-performing loans, default probability (Merton, 1974), modified bank default probability (Nagel & Purnanandam, 2020), and stock return idiosyncratic volatility. These findings collectively affirm that the CLN score, generated by the deep learning–based CoLGNN framework, offers valuable insights into future bank risk and network-driven vulnerability.

Consistent with its risk forecasting ability, we also find that a higher CLN score predicts lower future bank profitability. Importantly, this predictive power remains robust after controlling for bank stock returns and is especially pronounced among unlabeled public banks with stock performance in the interquartile range. In this subsample, the CLN score significantly outperforms stock returns in forecasting both risk and profitability outcomes.[5] Finally, we document persistent predictive performance of the CLN score across both halves of the sample period, with notably stronger effects during the global financial crisis, when systemic risk and contagion intensified.

[5]If the CLN score merely reflected banks' exposure to common shocks shared with the labeled "risky" banks, controlling for stock returns should have largely eliminated its predictive power. However, our results indicate otherwise. Specifically, within the subsample of unlabeled public banks, the CLN score significantly outperforms quarterly stock returns in predicting future non-performing loans and return on assets (ROA), supporting its interpretation as a measure of network-driven spillover risk.

Our study contributes to the literature in several important ways. First, we pioneer the application of graph neural networks (GNNs) to the universe of syndicated loans by developing a deep learning Co-Lending Graph Neural Network (CoLGNN) model. This design is motivated by the literature on information asymmetry and risk sharing in syndicated lending (e.g., Gopalan et al., 2011; Ivashina, 2009; Ivashina & Scharfstein, 2010a). While prior studies have shown that loan outcomes and bank risks are influenced by syndicate structures (e.g., Lim et al., 2014) and lending relationships (e.g., Bharath et al., 2011), we contribute by constructing a directed network representation that captures the asymmetry of co-lending relationships and directional risk spillovers. Our approach enriches the syndicated lending literature by leveraging deep learning to quantify the role of risk transmission in shaping future bank risk.

Second, we contribute to the broader literature on financial networks. A growing body of work emphasizes the importance of interbank interconnectedness in amplifying systemic risk (e.g., Acemoglu et al., 2015; Anderson et al., 2019; Battiston et al., 2012; Elliott et al., 2014; Golub & Jackson, 2012). We extend this line of inquiry to the syndicated lending market—an economically significant and underexplored context—by modeling the topological structure of co-lending relationships. Our approach allows us to develop an early-warning risk metric that exploits the network structure to detect risk spillovers among banks.

Third, our study advances the literature on bank risk measurement. Existing research primarily relies on market-based information, such as stock returns and equity volatility, to infer banks' default risk (e.g., Merton, 1974; Nagel & Purnanandam, 2020) and systemic risk (e.g., Tobias & Brunnermeier, 2016). We contribute a novel, semi-supervised deep learning framework that allows for the estimation of bank risk metrics even in the absence of market-based signals, thereby extending risk monitoring to private and thinly traded banks. This framework captures risk spillovers embedded in the co-lending network and provides a more comprehensive view of bank fragility.

Finally, we contribute to the emerging literature on the use of artificial intelligence (AI) in finance. As AI tools increasingly enable the analysis of high-dimensional, non-Euclidean data, recent studies have introduced new empirical methods (Gu et al., 2020), investment strategies (e.g., Cong et al., 2021), and alternative data sources (e.g., Cao et al., 2023; Jiang et al., 2023; Li et al., 2021). Our study adds to this growing field by introducing a cutting-edge deep learning framework tailored to financial networks. The proposed CoLGNN model demonstrates how AI can be applied to enhance risk management and early-warning systems in the banking sector.

The rest of the paper proceeds as follows. Section 3.2 develops the main hypotheses. Section 3.3 describes the construction of the co-lending network and introduces the deep learning CoLGNN model. Section 3.5 validates the risk transmission mechanism using two quasi-natural experiments. Section 3.4 presents the bank-level CLN score, outlines the data sources, and details key variable constructions. Section 3.7 concludes. The Appendix provides detailed variable definitions and additional empirical results.

3.2 Theoretical Motivation and Hypotheses

The global financial system is deeply interconnected, enabling financial shocks to propagate swiftly across institutions and markets. This interconnectedness is particularly salient in tightly coupled financial networks, where localized disturbances can cascade through both direct and indirect linkages, amplifying systemic vulnerabilities (e.g., Acemoglu et al., 2015; Anderson et al., 2019; Elliott et al., 2014). The structure and topology of such networks play a critical role in shaping the speed, intensity, and reach of contagion (Glasserman & Young, 2015; Haldane & May, 2011). These dynamics highlight the urgent need for risk assessment frameworks that account for the complexity of inter-institutional relationships and the potential for cross-institutional spillovers (Battiston et al., 2012; Brunnermeier & Oehmke, 2013).

This study focuses on the syndicated loan market—a setting that naturally embeds banks within a web of financial interdependencies. In a typical syndicate, multiple banks jointly extend credit to a single corporate borrower. Lead arrangers originate the loan, conduct due diligence, negotiate terms, and are primarily responsible for post-origination monitoring. Participant banks typically rely on the lead bank's reputation and informational advantage when deciding to join the syndicate (e.g., Gopalan et al., 2011; Holmstrom & Milgrom, 1987; Ivashina & Scharfstein, 2010a; Pichler & Wilhelm, 2001). While this structure promotes credit access and risk-sharing, it also creates a dense and often opaque network of exposure, where a disruption at one institution may have downstream consequences for others.

We argue that these syndicate-based interdependencies serve as channels for risk transmission. When a lead bank experiences distress—such as that triggered by borrower default, funding pressures, or internal fragility—the resulting disruption can spill over to participant banks through both financial and informational pathways. Importantly, such spillovers may arise even in the absence of a shared borrower default. For example, if a lead bank becomes financially impaired, it may scale back or abandon its monitoring responsibilities, weakening the syndicate's ability to detect borrower deterioration, enforce covenants, or coordinate timely interventions. This failure can leave participant banks exposed to latent credit risks that would otherwise have been mitigated by effective oversight. As a result, a bank's exposure to fragile syndicate partners can materially elevate its own risk, heightening its vulnerability to shocks propagating through the network.

To quantify and forecast such network-induced risk, we develop a novel deep learning framework based on graph neural networks (GNNs). Our model constructs a Co-Lending Network (CLN) risk measure by integrating information on co-lending ties, bank fundamentals, and loan-level characteristics. We label a subset of public banks as "safe" or "risky" based on their stock returns—a forward-looking, market-based proxy of bank health grounded in the principles of market efficiency and informativeness (Campbell et al., 2008; Fama, 1970, 1991; Fama & French, 2015; Merton, 1974; Nagel & Purnanandam, 2020)—and apply semi-supervised learning to propagate these labels through the network.

The CLN score generated by the GNN is not merely a function of a bank's own credit portfolio or observable risk characteristics. Instead, it reflects the bank's embeddedness within the broader co-lending network and its exposure to latent risks—those arising from the fragility, poor monitoring, or misjudgments of its syndicate partners. Banks with high CLN scores are more likely to be connected to distressed or fragile segments of the network or to occupy structurally vulnerable positions that heighten their exposure to partner-induced spillovers. These banks may face spillovers from impaired lead arrangers or deteriorating syndicates, manifesting in elevated future credit losses and bank risk. Accordingly, we propose the following hypothesis:

Hypothesis 1. Higher co-lending network risk of a bank predicts greater future bank risk.

Traditional market-based risk measures apply only to publicly traded banks, leaving the risk profiles of private banks largely unobserved. Our deep learning GNN framework overcomes this limitation by leveraging observed characteristics and co-lending ties to labeled public banks to infer the risks of private institutions. Through its message-passing mechanism, the GNN integrates network structure with node-level features to estimate latent risk exposures for unlabeled nodes. By doing so, our deep learning model effectively provides a quasi market-based risk measure for private banks, capturing their forward-looking network-embedded risks that are otherwise unobservable. Accordingly, we posit the following:

Hypothesis 2. The co-lending network risk measure exhibits strong predictive power for the future risks of private banks.

We further expect the predictive power of our co-lending network risk measure to vary systematically across banks with different financial and structural characteristics. Specifically, we hypothesize stronger effects for banks that are more vulnerable to risk spillovers transmitted through the co-lending network, due to limited internal buffers, reduced operational flexibility, or greater reliance on external relationships.

First, smaller banks typically have more concentrated loan portfolios and fewer resources for risk management, increasing their susceptibility to network-based contagion effects (Giannetti & Saidi, 2019). Second, banks experiencing negative earnings shocks or elevated return volatility often operate under greater financial pressure and reduced market confidence, limiting their ability to absorb spillovers. Third, banks with lower capital ratios possess diminished capacity to withstand losses, amplifying the consequences of externally transmitted shocks. Finally, institutions with greater operational complexity may face internal coordination challenges that hinder timely responses to risk transmission. We summarize these expectations as follows:

Hypothesis 3. The predictive power of the co-lending network risk measure is stronger for banks that are smaller in size, experience negative earnings shocks, exhibit higher return volatility, maintain lower capital ratios, or operate with greater organizational complexity.

3.3 Modeling Bank Risk via Graph Neural Network

3.3.1 Co-Lending Network Design in the Syndicate Lending Market

As discussed earlier, the syndicated lending market exhibits a distinctive structure wherein lead arrangers play a central role in screening borrowers, initiating deals, and monitoring loans. This hierarchical arrangement generates a web of financial interdependencies that motivates the construction of a directed co-lending network to capture potential risk spillovers. In this framework, banks are represented as nodes connected by co-lending relationships (edges), with a directed edge denoting the flow of information and risk from a lead arranger to a participant bank involved in a syndicated loan (Benmelech et al., 2012; Ivashina, 2009).

Directed edges in the co-lending network capture the transmission of information and risk, with their orientation indicating the direction of flow from lead banks to participant banks within the same loan syndicate. Each directed connection thus represents not only a co-lending relationship but also a potential channel for risk spillover. The network aggregates co-lending activities across all syndicated loans within a given time window, producing a dynamic sequence of co-lending networks that can be analyzed using rolling intervals (e.g., five-year windows). Formally, we define the co-lending network as an attributed graph as follows:

Co-Lending Network Each co-lending network is represented as an attributed graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, \dots, v_n\}$ is the set of banks and $\mathcal{E} = \{e_{i,j} : (v_i, v_j) \in \mathcal{V} \times \mathcal{V}\}$ is the set of edges. Each directed edge $e_{i,j}$ points from bank v_i to bank v_j. The set of neighboring banks of v_i is denoted as $\mathcal{N}(v_i) = \{v_j : (v_j, v_i) \in \mathcal{E}\}$. $\mathbf{X} \in \mathbb{R}^{n \times f}$ and $\mathbf{E} \in \mathbb{R}^{m \times s}$ represent the bank and loan characteristics matrices, where $n = |\mathcal{V}|$ and $m = |\mathcal{E}|$ denote the numbers of banks and co-lending relationships, and f and s are the dimensions of the bank and loan feature spaces, respectively. The adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ of the co-lending network captures directed co-lending relationships, where $a_{ij} = 1$ if there exists a directed edge from bank i to bank j, and $a_{ij} = 0$ otherwise.

3.3.2 Co-Lending Graph Neural Network (CoLGNN) Framework

To model risk transmission within the co-lending network, we introduce the **Co**-**L**ending **G**raph **N**eural **N**etwork framework (CoLGNN). Unlike traditional econometric methods, which often overlook the complex topological features inherent in financial networks, CoLGNN explicitly incorporates both the structural properties of the network and the attributes of individual banks to capture how risk propagates across interconnected institutions.

Framework Overview

The CoLGNN framework builds on recent advances in Graph Neural Networks (GNNs), which have become powerful tools for analyzing graph-structured data. GNNs perform learnable transformations across all components of a graph—nodes, edges, and global attributes—enabling the model to extract patterns directly from the underlying network structure. In our setting, this capability allows us to effectively trace how risk signals propagate from one bank to another through their co-lending relationships.

Figure 3.1 illustrates the core mechanism through which our model captures risk transmission. The process unfolds in several stages. In *Stage 1*, we classify banks based on their stock market performance, labeling a subset of public banks as either "safe" or "risky," while leaving the remaining public banks and all private banks as unlabeled. In *Stage 2*, we construct the co-lending network by mapping directed links from lead arrangers to participant banks. In *Stage 3*, we implement a graph diffusion convolution model that incorporates both bank-level characteristics (node features) and loan-level attributes (edge features).[6] Finally, in *Stage 4*, the model produces bank-level CLN scores that estimate each institution's risk profile.

[Insert Figure 3.1 about here]

A key innovation of the CoLGNN framework lies in its directional diffusion convolution modules, which capture the inherently asymmetric nature of risk transmission in syndicated lending. These modules are specifically designed to model the propagation of risk from lead arrangers to participant banks, rather than assuming symmetric information flow among all syndicate members. Figure 3.2 illustrates how the diffusion convolution modules process information across the network. CoLGNN differentiates between incoming and outgoing connections, thereby accounting for the distinct roles banks play within a syndicate. This architecture enables the model

[6] A graph convolution module is a specialized neural network technique designed for graph-structured data. It operates via *message passing*, an iterative process in which each node aggregates information from its neighbors, updates its representation based on these aggregated signals, and transmits the updated information in subsequent iterations. This mechanism enables information to diffuse throughout the network, allowing nodes to learn from both direct and indirect connections. The full mathematical specification of the message passing framework is provided in Appendix B.1.

to learn which types of relationships and bank characteristics are most predictive of risk outcomes by training initially on public banks with observable stock market performance. Through semi-supervised learning, the inferred patterns are then extended to the entire network, including private banks that lack observable market-based risk indicators. The full mathematical specification is provided in Appendix B.1.

[Insert Figure 3.2 about here]

A Simple Visulization of Graph Neural Network

To provide deeper insight into how the CoLGNN framework captures risk spillovers within the co-lending network, we present a simplified illustration of a single GNN layer in operation, as shown in Figure 3.3. This example uses a seven-bank network to demonstrate the core mechanism of the model, offering an intuitive understanding of the directional message-passing process prior to introducing the full multi-layer architecture.

[Insert Figure 3.3 about here]

The process begins with constructing the adjacency matrix, in which directed edges encode potential pathways of risk transmission from lead arrangers to participant banks. To facilitate stable and efficient information propagation, we compute the random walk Laplacian matrix $\hat{A}$ (also referred to as the random walk normalized adjacency matrix or diffusion probability matrix), which normalizes the adjacency structure and smooths the flow of information across the network. This step helps regulate the influence of connected versus unconnected nodes, enhancing the model's ability to capture the underlying network topology.

Next, edge features are represented using an edge feature matrix, where each entry corresponds to a set of attributes associated with the edge between two banks.[7] The model learns a low-dimensional embedding of these features, allowing it to focus on the most salient information relevant to risk propagation. These learned representations form the edge embedding matrix $\tilde{E}$, which is subsequently integrated with the normalized adjacency matrix. The resulting matrix, $\tilde{A} = \tilde{E} + \hat{A}$, captures both the structural and relational information within the graph, providing a richer and more informative representation of the co-lending network.

In the feature aggregation step, the GNN layer uses the combined matrix $\tilde{A}$ to aggregate node features X, producing a new feature matrix $Z = \tilde{A}X$. This operation enables each bank to incorporate information from its neighbors, effectively capturing the potential for risk spillovers within the network. Following aggregation, the model applies adaptive signal channel mixing

[7] For illustration purposes, we use a simplified example with three-dimensional edge features, while in our baseline model, high-dimensional edge features are used.

through a learnable weight matrix W, resulting in $H = ZW$. This step can be understood as learning which dimensions of information are most important for the risk assessment task.

Finally, the first dimension of the output matrix H is passed through a sigmoid activation function to generate a continuous risk score for each bank, denoted as $\hat{y} = \sigma(H_1)$. These scores range from 0 (least risky) to 1 (most risky). While the full CoLGNN model incorporates multiple layers and additional specialized components, this simplified example illustrates the core mechanism by which the GNN architecture transforms co-lending network information into meaningful risk assessments.

3.4 Measurement and Sample

3.4.1 Measuring Bank Co-Lending Network Risk

To implement the CoLGNN framework and construct the bank-level co-lending network risk score (CLN score), we apply our model to a comprehensive sample of bank characteristics from Form FR Y-9C and bank loans from Dealscan. Specifically, at each quarter end, we construct an attributed graph G_t, where nodes represent bank holding companies (BHC, or bank hereafter) and edges indicate co-lending relationships. This process yields a dynamic series of co-lending networks, $\mathcal{G} = \{G_1, ..., G_T\}$, spanning from 1996Q1 to 2020Q4, using a five-year rolling window.[8] As indicated in Section 3.3.1, each directed edge points from a lead bank to another bank within the syndicate structure. For multiple co-lending relationships pointing from one bank to another bank within the window, we aggregate them into a single edge. For each co-lending network G_t, we use the BHC characteristics at $t-1$ as the node feature and the lending activities and loan characteristics from $t-1$ to $t-20$ as the edge features.

In our semi-supervised learning framework, we leverage quarterly stock returns to assign "safe" or "risky" labels to banks based on their recent stock market performance. Stock returns are an ideal metric for this purpose, as stock prices are forward-looking and incorporate all publicly available information about a firm's financial health, market outlook, and risk exposure under the assumption of market efficiency. Following established practice in Berger and Bouwman (2013) and Fahlenbrach et al. (2018), we test multiple classification approaches, including tercile-based, quartile-based, and quintile-based thresholds, where public banks in the top proportion of stock returns are categorized as "safe" (label $Y_{i,t} = 0$), while those in the bottom proportion are classified as "risky" (label $Y_{i,t} = 1$).[9] The remaining public banks, along with all private banks, are treated as unlabeled, facilitating out-of-sample analysis.

[8] The results are robust to using a three-year rolling window alternative.

[9] Our findings are robust to alternative classification thresholds. In untabulated results, we confirm that using different threshold specifications (20%, 25%, 40%) yields qualitatively similar results, with consistent statistical significance and economic implications.

The training of our model is executed over 200 epochs, employing an Adam optimizer with appropriate learning rate and weight decay settings.[10] Our CoLGNN architecture includes two diffusion graph convolution modules, with each node passing through two layers of diffusion graph convolution. The highest performing model is selected based on the combined accuracy across training and validation sets, ensuring the model generalizes effectively to unseen data.

We frame the prediction task as node classification, using the probability of the "risky" class as the co-lending network risk score. Using the information from the labeled group, we estimate the CLN score for all banks without labels and update the scores for labeled banks to reflect changes in risk caused by the effects of the network. The CLN score for each bank ranges from 0 (safest) to 1 (riskiest), representing the probability that the bank belongs to the "risky" class based on both its characteristics and network position.

At each year-quarter t, we construct a co-lending network G_t, resulting in an estimated $CLN\ score_{i,t}$ for each bank i at year-quarter t.[11] The complete technical details of the CLN score calculation are provided in Appendix B.1. We apply the same set of hyperparameters to all networks, and the complete hyperparameter specifications are in Section B.1.4 of the Appendix B.1.

3.4.2 Sample Construction

Our cross-sectional analysis of bank-level co-lending network risk employs a sample of U.S. bank holding companies from 1991Q1 to 2020Q4. We collect syndicated loan data from Refinitiv LoanConnector DealScan, stock returns from Center for Research in Security Prices (CRSP), and bank characteristics from Form FR Y-9C. We combine bank characteristics and stock return data using the CRSP-FRB link table supplied by the Federal Reserve Bank of New York. We match data on syndicated loans from DealScan with bank characteristics based on the parent company of the lenders, using hand-matched bank name concordance files combined at the BHC level. Following (Ivashina, 2009), we identify the lead bank in the DealScan database if the bank's 'Primary Role' is listed as one of 'Arranges', 'Co-arranger', 'Co-lead arranger', 'Lead

[10] Adam is a popular extension of stochastic gradient descent (SGD) that adaptively adjusts the learning rate for each parameter. Weight decay refers to a regularization term that mitigates overfitting by penalizing large parameter values.

[11] We measure the bank-level CLN score using the output from the last layer of CoLGNN. For bank i in a given co-lending network G_t, the CLN score is estimated as:

$$CLN\ score_i = \sigma\left(\sum_{k=0}^{K-1}\left(\sum_i \tilde{\mathbf{a}}^k_{out,i}\,\tilde{\mathbf{h}}^{(L)}_{out,i}\mathbf{W}_{k,out} + \sum_i \tilde{\mathbf{a}}^k_{in,i}\,\tilde{\mathbf{h}}^{(L)}_{in,i}\mathbf{W}_{k,in}\right)\right)$$

where $\sigma(\cdot)$ represents the sigmoid activation function. The terms $\tilde{\mathbf{a}}^k_{out,i}$ and $\tilde{\mathbf{a}}^k_{in,i}$ are the out/in direction transition probabilities for bank i, while $\tilde{\mathbf{h}}^{(L)}_{out,i}$ and $\tilde{\mathbf{h}}^{(L)}_{in,i}$ represent the edge-augmented node features from the final layer L of the network. These edge-augmented features incorporate both bank characteristics (node attributes) and loan relationship characteristics (edge attributes), enabling the model to capture the risk transmission through syndicated loan relationships. This calculation is performed for each bank within each network, yielding a vector of risk scores across all banks. By applying this procedure to each quarterly co-lending network G_t, we obtain the time series of CLN scores, denoted as $CLN\ score_{i,t}$ for bank i at year-quarter t.

arranger', 'Mandated Lead arranger', 'Mandated arranger', 'Lead manager' or if the lender's name is listed in the 'Lead Arranger' column.

[Insert Table 3.1 about here]

Table 3.1 reports the summary statistics of loan-level and bank-level characteristics. Definitions of variables and data sources are provided in Table 3.A1 in the Appendix. All continuous variables are winsorized at the 1st and 99th percentiles for each year-quarter. Our key variable, the bank co-lending network risk score (*CLN score*), ranges from 0 to 1 by construction, with a value of 0 indicating the safest bank and 1 indicating the riskiest bank based on co-lending network risk. Table 3.A2 in the Appendix lists the top seven banks with the highest and lowest *CLN* scores. Our primary dependent variable, bank loan loss provisions, is measured as the ratio of loan loss provisions (BHCK4230) to total assets (BHCK2170), expressed in percentage points. The mean (median) loan loss provisions is 0.362 (0.163). The average (median) natural logarithm of the total asset size of banks in our sample is 16.164 (15.935). The average (median) return on assets is 0.540 (0.525). The average (median) loan size (total loans to total assets) is 0.628 (0.669). These bank characteristics align with prior literature (e.g. Dou et al., 2018; Ellul & Yerramilli, 2013). For example, Dou et al. (2018) report average (median) loan loss provisions of 0.335 (0.177). Ellul and Yerramilli (2013) report an average (median) bank size of 16.631 (16.66) and loan size of 0.626 (0.672).

3.4.3 Feature Importance

To enhance the interpretability of the CoLGNN model, we implement the integrated gradients method to quantify the contribution of each input feature to the predicted CLN risk score for banks.[12] Figure 3.4 presents the most influential loan and bank characteristics based on their attribution scores. Among loan features, the revolver indicator—which identifies revolving credit facilities—emerges as the most important predictor. This finding aligns with the monitoring-intensive nature of such facilities, where continuous lead arranger oversight and due diligence likely intensify risk transmission among syndicate members. It suggests that risk spillovers primarily occur through facilities requiring active engagement rather than through more easily traded instruments. We further validate this finding in Section 3.6.3.

[12]We adopt the integrated gradients method proposed by Sundararajan et al. (2017) to assess feature importance. This technique quantifies how much each input feature contributes to the CoLGNN's output (the CLN score), relative to a baseline scenario in which the feature is absent. Specifically, the method integrates the gradient of the model's output with respect to each input feature along a straight-line path from a reference point (typically zero, representing absence of signal) to the actual input values. Unlike simple gradient-based approaches that reflect only local sensitivity, integrated gradients capture the cumulative effect of each feature across its full range, making them well-suited for interpreting complex, non-linear models with interactive effects. For each feature, we compute the absolute value of its attribution score in each quarter and then average these values across all quarterly co-lending networks. This approach allows us to identify features that consistently exert strong influence on bank risk predictions over time.

Other loan characteristics such as security status, syndicate structure, interest spread, maturity, and covenant strength also rank highly in influencing the CLN score. On the bank side, financial fundamentals dominate. Total loan exposure is the top predictor, followed by profitability metrics and capital adequacy indicators. To address concerns about the potential overfitting or noise arising from a high-dimensional loan feature space, we conduct a robustness check using a reduced set of standard loan variables commonly employed in the syndicated lending literature. As shown in Figure 3.A2 in the Appendix, the resulting importance rankings remain qualitatively similar, suggesting that our findings are not driven by the inclusion of auxiliary features.

[Insert Figure 3.4 about here]

3.5 Validation

In this section, we validate that the *CLN score* indeed captures risk spillovers transmitted through co-lending networks, rather than merely reflecting shared exposure to borrower default or endogenous lender selection. To establish the underlying mechanism, we exploit two quasi-natural experiments.

3.5.1 Validating Risk Spillover via Credit Rating Downgrades

First, we examine how credit rating downgrades of specific banks affect the *CLN scores* of connected banks relative to comparable unconnected banks using a difference-in-differences framework. A credit rating downgrade is a pivotal event that impairs a bank's access to funding, particularly in wholesale markets and public bond issuances (Adelino & Ferreira, 2016). Such downgrades typically result in reduced market access, increased collateral requirements, and higher funding costs.

We investigate whether and how negative shocks originating from a focal bank's downgrade propagate through the co-lending network and influence the risk profiles of its connected peers. This transmission occurs via two main mechanisms. First, downgraded banks may scale back their screening and monitoring efforts to conserve capital and manage elevated funding costs. This reduction in due diligence shifts greater financial risk onto syndicate partners, weakening the overall quality of loan oversight. Second, the downgrade increases perceived counterparty risk, prompting co-lenders to reassess their exposure to the downgraded bank. This reassessment can lead to tighter credit conditions and greater risk aversion across the network, amplifying spillover effects on connected institutions.

Following Adelino and Ferreira (2016), we rely on Standard & Poor's (S&P) credit rating history rather than alternative sources. This choice reflects S&P's tendency to act more proactively and to lead other agencies in issuing rating changes (Kaminsky & Schmukler, 2002).[13]

We use these rating downgrades to trace how risk propagates through co-lending networks and to assess whether the *CLN score* reflects true spillover effects—rather than shared exposure to default-prone borrowers. A credit downgrade reflects worsening financial health at the focal bank and constitutes an exogenous shock for connected banks. Due to information asymmetry between credit rating agencies and syndicate members, these connected banks may not be aware of the specific conditions leading to the downgrade. Consequently, the downgrade triggers a reassessment of network risk—driven not by borrower fundamentals, but by interbank ties. This effect is compounded by potential regulatory scrutiny and market reevaluation of the focal bank's role and reliability. Additionally, the downgraded bank may reduce oversight of existing loans it previously lead-arranged due to internal strain, thereby creating unforeseen moral hazard concerns for its co-lending partners.

If the *CLN score* effectively captures such spillover risks, then the scores of banks connected to a downgraded institution should rise significantly after the event. To test this hypothesis, we employ a stacked cohort difference-in-differences (DiD) estimation approach. Specifically, we treat each credit rating downgrade event c for a focal bank at time t as a distinct event cohort and construct a matched subsample surrounding each event.

In our directed co-lending network (as detailed in Section 3.3.1), a directed edge flows from the lead bank to other participant banks within a syndicated loan. Within each event cohort, we designate as treated banks those with direct inward links from the focal (downgraded) bank. Unconnected banks—those not linked to the focal bank within the same loan syndicate—serve as the control group.

A potential concern is that credit rating downgrades may coincide with macroeconomic downturns or industry-wide shocks, which could violate the parallel trends assumption by systematically exposing treated banks to elevated risk even before the downgrade. To mitigate this concern and bolster the exogeneity of the downgrade shocks, we exclude observations from the Global Financial Crisis (GFC) and from quarters marked by concentrated downgrade activity.[14]

For each cohort, we use an event window for each credit downgrade event, spanning from two quarters before to three quarters after the event (excluding the focal downgrade-event quarter). We then stack all the credit downgrade event cohorts together and estimate the

[13] S&P rating announcements tend to have a stronger and less anticipated impact on stock prices, further supporting our reliance on this source (Adelino & Ferreira, 2016; Reisen & Von Maltzan, 1999). Consistent with the credit rating literature (e.g., Badoer et al., 2019; Xia, 2014), we focus on long-term issuer ratings.

[14] Specifically, we exclude downgrade events occurring during the GFC, defined as 2007Q2 to 2009Q1, as well as quarters with more than ten downgrade events, to limit confounding effects stemming from widespread economic distress.

following standard DiD regression specification:

$$CLN\ score_{i,c,t} = \beta_1 Connected\ Banks_{i,c} \times Post\ Credit\ Downgrade_{c,t} + \gamma_i + \theta_t + \varepsilon_{i,c,t} \tag{3.1}$$

where $CLN\ score_{i,c,t}$ is the risk score for bank i in cohort c at time t. The interaction term $Connected\ Banks_{i,c} \times Post\ Credit\ Downgrade_{c,t}$ equals 1 for treated banks post-downgrade, capturing the differential effect. Fixed effects $\gamma_{i,c}$ (cohort-bank) and $\theta_{c,t}$ (cohort-year-quarter) absorb the main effects of treatment and time, respectively, controlling for bank-specific and temporal heterogeneity within cohorts. We report standard errors clustered at the bank level, and, in robustness checks, also double-clustered at the bank and cohort levels.

[Insert Table 3.2 about here]

Table 3.2 presents the baseline DiD results. The interaction term is positive and statistically significant at the 5% level, indicating that banks directly connected to a downgraded institution experience a significant increase in their CLN scores relative to unconnected banks following the downgrade. This finding remains robust to the inclusion of bank-level control variables and the use of standard errors double-clustered at the bank and cohort levels.

To further examine the timing and persistence of these spillover effects, we incorporate the downgrade-event quarter into the analysis and estimate a dynamic difference-in-differences specification. This approach allows us to trace the evolution of CLN scores around the event and to rule out spurious effects arising from pre-existing trends. Using the downgrade quarter as the reference period, we find no evidence of differential pre-trends in CLN scores between treated and control banks, as shown in Part (a) of Figure 3.5. The figure shows that following the downgrade, treated banks exhibit a pronounced and statistically significant increase in their CLN scores, particularly two quarters after the event. These dynamic effects highlight the role of the co-lending network in propagating financial risk: a downgrade affecting one bank triggers a subsequent and persistent rise in the perceived risk of its connected counterparts.

Taken together, the results provide compelling evidence that the CLN score captures network-driven risk spillovers, rather than reflecting mere common exposure to borrower default or broader macroeconomic shocks.

3.5.2 Validating Risk Spillover via Lehman Brothers' Collapse

Second, we examine the collapse of Lehman Brothers—a shock that originated outside its corporate lending business—as a quasi-natural experiment to isolate network-driven spillover effects from common exposures to distressed corporate borrowers. A central empirical challenge in studying risk spillovers in banking networks is disentangling true contagion from correlated vulnerabilities to shared economic conditions. Following Chodorow-Reich (2014), we exploit Lehman's September 2008 failure as an exogenous disruption to the co-lending network. Lehman's financial distress was driven primarily by concentrated investments in commercial real estate and dependence on unstable short-term funding, rather than by deterioration in its corporate loan portfolio.[15] This clear separation between the causes of Lehman's collapse and its syndicated lending activities offers an ideal setting to test whether risk propagates through co-lending relationships rather than simply reflecting shared exposures to troubled corporate borrowers.

To isolate the effect of Lehman Brothers' failure on network-driven risk spillovers, we impose several key sample restrictions. First, we exclude major recipients of the Troubled Asset Relief Program (TARP)—including Citigroup, Bank of America, JPMorgan Chase, and other large financial institutions that received substantial government support. This step is essential, as direct government intervention may obscure the natural network-based transmission dynamics we seek to identify. Second, we remove institutions with significant exposure to Bear Stearns to ensure our estimates reflect the distinct effects of Lehman's collapse rather than residual impacts from Bear Stearns' earlier failure in March 2008. Third, we restrict the analysis to a window covering two quarters before and after Lehman's September 2008 bankruptcy. This window strikes a balance between temporal proximity to the shock and allowing enough time to detect risk propagation through the network, while minimizing contamination from concurrent macroeconomic events. Together, these restrictions strengthen the internal validity of our identification strategy by approximating a cleaner quasi-natural experiment. We estimate the following DiD regression specification:

$$CLN\ score_{i,t} = \beta_1 Lehman\ Co\text{-}Lender_i \times Post\ Lehman\ Failure_t + \gamma_i + \theta_t + \varepsilon_{i,t} \tag{3.2}$$

where $CLN\ score_{i,t}$ is the network risk score for bank i at time t. The indicator variable $Lehman\ Co\text{-}Lender_i$ equals 1 for banks with co-lending relationships with Lehman Brothers prior to its collapse. The indicator variable $Post\ Lehman\ Failure_t$ equals 1 for the quarters following September 2008. Fixed effects γ_i (bank) and θ_t (year-quarter) absorb time-invariant bank characteristics and common temporal shocks, respectively.

[15] As detailed in the bankruptcy examiner's report (Valukas, 2010), Lehman's collapse stemmed from its 2006 strategic expansion into proprietary commercial real estate investments and a 2007 "countercyclical growth strategy" that increased real estate exposure at a time when other institutions were scaling back.

[Insert Table 3.3 about here]

Panel A of Table 3.3 presents results from the standard DiD specification. The coefficient estimate on the interaction term is positive and statistically significant, indicating that banks with co-lending relationships with Lehman experienced a 28–29% increase in their CLN scores relative to the sample mean following Lehman's failure.[16] These results provide strong evidence that risk transmission operates through co-lending network connections, rather than being solely driven by shared exposure to distressed borrowers. As shown in Part (b) of Figure 3.5, the dynamic DiD estimation further supports this interpretation: it reveals no significant differences in pre-trends and shows that the treatment effect begins to emerge in the first quarter following Lehman's collapse.

Panel B of Table 3.3 reports results from the synthetic difference-in-differences (SDID) method of Arkhangelsky et al. (2021), which constructs counterfactual outcomes using matched control groups. We adopt SDID to strengthen causal inference and relax the parallel trends assumption by ensuring better balance in both covariates and pre-treatment trends between treated and control banks. In our primary specification (column 1), we construct synthetic controls based on lending-related characteristics—loan size, loan growth, and loan loss allowance. Since Lehman's collapse was driven by factors unrelated to its corporate lending activities, matching on these characteristics ensures that treated and control banks had comparable lending profiles prior to the shock. The SDID estimate yields an average treatment effect on the treated (ATT) of 0.140, statistically significant at the 5% level. In column 2, we confirm the robustness of this effect by incorporating the full set of bank-level control variables. These results imply a 26–30% increase in risk scores for Lehman's co-lenders relative to their synthetic counterparts.[17] Notably, both specifications provide strong validation that the CLN score captures genuine risk transmission through co-lending network channels, rather than reflecting common exposure to distressed corporate borrowers.

3.6 The Risk Predictability of Bank CLN Score

3.6.1 Predicting future bank-specific risks

Having validated that the *CLN score* captures risk spillovers transmitted through co-lending networks, we next investigate its predictive power for future bank risk. Specifically, we employ loan loss provisions scaled by total assets as the dependent variable in our baseline analysis.

[16] The percentage increases are calculated by dividing the coefficient estimates from Table 3.3 Panel A (0.128 and 0.134) by the sample mean CLN score of 0.465 from Table 3.1, resulting relative increases of 27.5% (0.128/0.465) and 28.8% (0.134/0.465), respectively.

[17] The relative increases of 30.1% and 26.7% are calculated by dividing the ATT estimates (0.140 and 0.124) by the sample mean CLN score of 0.465.

Loan loss provisions reflect the reserves that banks set aside for potential loan defaults and serve as a forward-looking measure of credit risk. These provisions not only impact earnings through expense recognition but also signal banks' expectations regarding future loan performance. To empirically test whether the *CLN score* predicts future risk, we estimate the following h-quarter-ahead predictive regression:

$$LLP_{i,t+h} = \beta_1 CLN\ score_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + \varepsilon_{i,t+h} \tag{3.3}$$

where $LLP_{i,t+h}$ is the loan loss provisions for bank i at time $t+h$, $CLN\ score_{i,t}$ refers to the bank-level co-lending network risk score generated by our CoLGNN model of bank i at time t, $X_{i,t}$ is the vector of bank characteristics at time t and we control for year-quarter fixed effects to capture the unobservable heterogeneity at each year-quarter level.[18] For bank-level characteristics, we control for bank size, equity capital ratio, deposits, return on assets, loan portfolio size, growth rate of loan portfolio, loan loss allowance, and liquidity ratio. To maintain consistency between the samples of public and private banks, we exclude market-based control variables due to the limited availability of such data for private banks.[19] Standard errors are clustered at the bank level. Table 3.4 reports the results.

[Insert Table 3.4 about here]

We present the results of predictive regressions ranging from 1-quarter-ahead to 8-quarter-ahead horizons. The results show that banks with higher co-lending network risk scores (*CLN score*) tend to have higher loan loss provisions in the future. Specifically, in Panel A of Table 3.4, which includes all sample banks, the coefficient estimates of $CLN\ score_{i,t}$ are consistently positive across all horizons and statistically significant at the 1% level for the first 7 quarters. For instance, column (1) shows that a one-standard-deviation increase in $CLN\ score_{i,t}$ corresponds to a 0.042-percentage-point rise (i.e., 0.458*0.093) in scaled quarterly loan loss provisions in the subsequent quarter. Economically, considering an average natural logarithm of bank size of 16.164 (in thousands), this translates to an expected annual loan loss of about $18 million. These findings strongly support Hypothesis 1 that higher co-lending network risk of a bank predicts greater future bank risks.

Panels B and C of Table 3.4, focusing on unlabeled and private banks respectively, validate the $CLN\ score_{i,t}$'s out-of-sample prediction efficacy.[20] The coefficient estimates of $CLN\ score_{i,t}$ remain positive and mostly significant at the 1% level. Consistent with Hypothesis 2, the results

[18] We do not control for bank fixed effects since our baseline results focus the cross-sectional risk predictive power of bank co-lending network risk score. However, our results remain robust after including the bank fixed effects, as shown in Table 3.A4 in the Appendix.

[19] In untabulated results, we additionally include the market-to-book equity ratio and quarterly buy-and-hold stock returns for public banks. Our results remain qualitatively unchanged.

[20] Panel B includes the remaining 50% unlabeled public banks and all private banks. Panel C includes only private banks.

in Panel C of Table 3.4 show that *CLN* $score_{i,t}$ exhibits a robust predictive ability for future risks of private banks. This underscores how our semi-supervised CoLGNN model effectively captures risk transmission by leveraging network topology and neighboring banks' risk profiles, thereby providing a robust risk assessment framework for private banks that traditionally lack market-based risk metrics. For example, a one-standard device increase in *CLN* $score_{i,t}$ is associated with a 0.042-percentage-point increase (i.e., 0.458×0.092) in the quarterly loan loss provisions scaled one year ahead (column (4) in Panel C), which is around 12% of the sample mean loan loss provisions of 0.362. Economically, it indicates a sizeable expected loan loss of about \$19 million per year.

We conduct several robustness checks. First, recognizing that co-lending network formation and banks' risk profiles may be influenced by their reputation and lending specialization, we include controls for these factors. Specifically, a bank's reputation is measured by its market share of lead-arranged syndicated loans, while lending specialization is assessed based on the focus of its lending activities within the syndicated loan market and its concentration in specific industries. Table 3.A3 in the Appendix shows that even after accounting for these factors, the coefficient estimates of *CLN* $score_{i,t}$ remain both statistically and economically significant.

Second, we conduct horse-racing tests to compare the in-sample and out-of-sample predictive performance of the CLN score against the bank Z-score, following the empirical approach of Barwick et al. (2023). The accounting-based bank Z-score is one of the few risk metrics available for banks regardless of stock market information availability and has been widely used to evaluate bank risk in recent literature (e.g. Houston et al., 2010; Laeven & Levine, 2009).[21]

Specifically, we perform tests on bank risk predictions on each time horizon using three samples: a) labeled banks only, b) unlabeled banks, and c) private banks. We consider four model specifications: 1) bank-level controls only, 2) bank Z-score and control variables, 3) bank CLN score and control variables, and 4) all variables combined. For simplicity, Table 3.5 reports results for 2, 4, and 6 quarters ahead horizons.[22] The results show that the CLN score consistently improves the adjusted in-sample R^2 across all time horizons (for example, Panel A shows a increase from 0.560 to 0.568), while the inclusion of the bank Z-score yields no discernible improvement.

Using the coefficient estimates from Panels A, D and G, we evaluate out-of-sample predictions for unlabeled banks and private banks. The results show that the CLN score consistently reduces prediction errors across all horizons. For instance, at the 2-quarter horizon, the inclusion of the CLN score reduces the RMSE from 0.599 (controls only) to 0.569 (CLN score and controls) for unlabeled banks and from 0.665 (controls only) to 0.629 (CLN score and controls) for private

[21] Following Laeven and Levine (2009), we construct the bank Z-score as Bank Z-score = $\frac{(ROA+CAR)}{\sigma(ROA)}$, where ROA is the return on assets, CAR is the ratio of equity to assets, and $\sigma(ROA)$ is the standard deviation of ROA.

[22] In untabulated results, we also examine the predictive regressions of bank Z-score and CLN score across all horizons from 1 to 8 quarters. The results remain qualitatively consistent throughout.

banks. This pattern persists at longer horizons, with the 6-quarter horizon still showing RMSE decreases from 0.669 to 0.655 for unlabeled banks and from 0.712 to 0.699 for private banks. In contrast, the bank Z-score either increases RMSE or offers negligible improvements, indicating its limited value for predictive accuracy. These findings validate the CLN score's superior ability to forecast future bank risks, particularly for private banks where traditional market-based risk measures are unavailable. The CLN score's predictive strength derives from its integration of accounting information with the topological dynamics of banks' co-lending networks and the risk profiles of neighboring banks.

[**Insert Table 3.5 about here**]

Lastly, we include bank-fixed effects to account for unobserved, time-invariant heterogeneity across banks. Table 3.A4 in the Appendix confirms that the predictive power of $CLN\ score_{i,t}$ remains robust and statistically highly significant even after controlling for bank-fixed effects. Meanwhile, the economic magnitudes of the coefficient estimates are comparable to those in the baseline results presented in Table 3.4.

3.6.2 Comparison with traditional network centrality

This section evaluates our proposed co-lending network (CLN) risk score against traditional network centrality measures in predicting future bank risks. Metrics such as eigenvector centrality, closeness centrality, and degree centrality are widely utilized in financial network analysis (e.g., Battiston et al., 2012; El-Khatib et al., 2015; Rossi et al., 2018) and primarily measure a node's structural importance within a network. However, these metrics fall short in capturing dynamic risk spillovers from neighboring banks, a critical determinant of future bank risks. For instance, a highly central bank in the co-lending network may be more stable due to diversification or more vulnerable due to heightened exposure, depending on the characteristics of its connections. Likewise, a peripheral bank might seem insulated from systemic shocks but could be critically dependent on a few key relationships. Traditional centrality measures overlook the risk profiles of neighboring nodes, which are essential in understanding risk propagation within co-lending networks. By contrast, the CLN score integrates both the network's structural topology and the risk characteristics of connected banks, offering a more comprehensive and nuanced approach to assessing future bank risks.

We compare the predictive power of the CLN score against various centrality measures, including both in-degree and out-degree versions of these measures. Traditional centrality measures are typically designed for undirected networks. To account for the centrality of a bank from both the lead arranger and participant perspectives, we estimate both in-degree and out-degree versions of network centrality. Appendix B.2 provides detailed discussions and definitions of all network centrality measures.

Figure 3.6 displays the coefficient estimates of the CLN score alongside these centrality measures in predicting bank loan loss provisions four quarters ahead.[23] The figure shows that the coefficient estimates of the CLN score remain statistically significant across all specifications and samples. In contrast, traditional network centrality measures fail to predict future bank risks. This limitation likely arises from their inability to differentiate the nature of connections—whether a bank is predominantly linked to risky or safe counterparties—or to account for the specific attributes of lending relationships.

[Insert Figure 3.6 about here]

3.6.3 Revolver-based CLN score

As discussed earlier, lead arrangers often sell substantial portions of their initial loan allocations, undermining the assumption of sustained exposure typically associated with term loans. To address this concern, we construct the *CLN score* based on co-lending networks formed exclusively through revolving credit facilities—an instrument less likely to be traded in secondary markets due to its continuous drawdown structure and the need for intensive, ongoing monitoring. This design ensures that the co-lending ties captured by the *CLN score* more accurately reflect persistent exposure and monitoring responsibilities, thereby providing a more credible channel for risk transmission.

To formalize this approach, we build a modified co-lending network and re-estimate the CoL-GNN model using only revolving credit facilities. The literature documents important structural differences between revolving facilities and term loans (Blickle et al., 2020). Because revolvers are rarely sold in secondary markets, they create stronger monitoring incentives: future drawdowns are contingent on borrower performance and creditworthiness. Consequently, lead lenders must remain actively engaged with borrowers over the life of the facility, reinforcing long-term exposure and amplifying the potential for risk spillovers through the co-lending network.

Our implementation follows the same semi-supervised learning framework as in Section 3.4.1, but restricts both the network structure and edge features to those derived from revolving credit lines. As a robustness check, we conduct a placebo test by constructing an alternative *CLN score* using only Term Loan B and above—tranches more likely to be held by institutional investors such as CLOs and asset managers and sold in the secondary market after loan initiation (Lim et al., 2014). These tranches are less likely to reflect enduring lender-borrower relationships, providing a useful benchmark for evaluating the strength and credibility of our measure.

[23] The results remain consistent across different predictive horizons.

Table 3.6 presents the findings, with particular emphasis on private banks. These institutions offer an ideal setting for analyzing risk transmission mechanisms, as they operate under less regulatory scrutiny, depend more heavily on relationship-based lending, and, crucially, lack publicly observable market-based risk measures.

[Insert Table 3.6 about here]

Panel A shows that the coefficient estimates for the *CLN score (Revolver)* are positive and statistically significant across all forecast horizons, consistently exhibiting larger economic effects than those reported for our baseline *CLN score* in Panel C of Table 3.4. This enhanced predictive power supports the view that risk transmission in co-lending networks is primarily driven by ex-post monitoring relationships. In contrast, the placebo test in Panel B of Table 3.6, which uses Term Loan B and above, yields predominantly insignificant and economically modest coefficient estimates, with magnitudes substantially smaller than those associated with revolvers.

The superior performance of the revolver-based *CLN score* highlights that risk transmission operates primarily through monitoring-intensive relationships, where lead arrangers retain ongoing oversight responsibilities and economic interests. These enduring ties—anchored in administrative obligations, informational advantages, and reputational concerns—serve as the principal channels through which financial risk propagates across the co-lending network.

3.6.4 Cross-Sectional Variation in the Risk Predictive Power

We now investigate the cross-sectional variation in the risk predictive power of the CLN score. As proposed in Hypothesis 3, we expect that the predictive power of a bank's CLN score is stronger for banks that are inherently more vulnerable within the co-lending network. These include banks characterized by smaller size, poorer performance, higher return volatility, lower capital adequacy, and greater operational complexity. Such institutions are less equipped to absorb negative shocks, as their limited resilience makes risk management processes more challenging, particularly in the face of risk propagation within the network. Consequently, the risk profiles of these banks are likely to be more sensitive to the risk dynamics of their neighboring banks in the co-lending network.

Table 3.7 presents the results on the cross-sectional variation in the risk predictive power of the CLN score, focusing on private banks where market-based risk signals are lacking and our method offers a contribution. Untabulated results for public banks are qualitatively similar. Each panel investigates a distinct vulnerability characteristic that may heighten a bank's sensitivity to network-induced risk transmission.

[Insert Table 3.7 about here]

First, smaller banks with lower market share are less equipped to internalize negative spillovers (Giannetti & Saidi, 2019). Specifically, these banks often lack sufficient capital buffers and the sophisticated risk management capabilities necessary to mitigate the effects of adverse shocks originating from their co-lending partners. The coefficient estimates of the interaction terms between $CLN\ score_{i,t}$ and $Small\ bank_{i,t}$ are positive and mostly statistically significant in Panel A of Table 3.7. For example, at the five-quarter horizon (h=5), smaller banks are 2.41 times more sensitive to co-lending network risk than their larger counterparts.[24]

Second, banks experiencing negative performance shocks face immediate operational pressures that can compromise their due diligence and monitoring capabilities (Gopalan et al., 2011). Specifically, we calculate unexpected earnings as the year-over-year change in quarterly earnings per share (EPS) and standardize these unexpected earnings (SUE) by dividing them by their standard deviation.[25] We then introduce an indicator, $Negative\ earnings\ shock_{i,t}$, which equals 1 for bank i at time t if the SUE is negative, indicating a negative performance shock. Additionally, we estimate accounting-based earnings volatility using ROA volatility, a measure applicable to all banks irrespective of stock price availability. ROA volatility is calculated as the standard deviation of quarterly ROA over the preceding five-year period for each bank. To identify banks experiencing significant earnings fluctuations in a given quarter, we introduce an indicator $High\ ROA\ volatility_{i,t}$. The results in Panels B and C of Table 3.7 suggest that the bank CLN score has greater predictive power for banks with poorer recent performance and volatile returns. At two quarters ahead, while the baseline effect for banks without earnings shocks is 0.032, the total effect for banks with negative earnings shocks is 0.032+0.136=0.168, representing a 5.3-fold increase in sensitivity to co-lending network risk.

Third, we examine whether banks with greater structural complexity exhibit heightened vulnerability to risk spillovers. We employ a similar accounting "Disaggregation" measure proposed by Chen et al. (2015), by using the ratio of non-missing BHCK items to total BHCK items reported in the FR Y-9C filings.[26] The rationale is that a higher proportion of filled items in FR Y-9C filings indicates a broader scope of banking operations, encompassing not only standard balance-sheet entries but also detailed off-balance-sheet activities such as unused loan commitments and credit derivatives. As banks engage in a wider range of financial products, their business operations become inherently more complex, rendering effective risk management more difficult to implement. The FR Y-9C, with its extensive list of 2,374 distinct line items with the BHCK prefix as of 2020, provides a robust framework for capturing such operational

[24] The baseline effect of the CLN score for larger banks is 0.049, while for smaller banks it is 0.049+0.069=0.118.

[25] Unexpected earnings (UE) for bank i at time t are defined as $UE_{i,t} = \text{EPS}_{i,t} - \text{EPS}_{i,t-4}$, and standardized unexpected earnings (SUE) are calculated as $SUE_{i,t} = \frac{UE_{i,t}}{\sigma(UE_{i,t})}$.

[26] Chen et al. (2015) use the ratio of non-missing items to total items in Compustat. We argue that the proportion of non-missing items in FR Y-9C filings serves as a proxy for the complexity of a bank's business operations.

complexity.[27] Each non-missing entry in these filings signifies active engagement in a specific banking activity or service, making the ratio of non-missing to total items an effective proxy for a bank's business complexity. This measure reflects the diversity and intricacy of banking operations, which are closely tied to the challenges and nuances of managing risks in complex financial environments.

More complex banks face coordination challenges in aligning risk management practices across multiple business domains, potentially amplifying their susceptibility to network-induced risks. Panel D of Table 3.7 reveals that the interaction terms are positive and statistically significant across multiple prediction horizons. For example, at the two-quarter horizon, the CLN score coefficient for highly complex banks (0.053+0.072=0.125) is 2.4 times larger than for less complex banks. These findings suggest that organizationally complex banks are indeed more vulnerable to risk transmission through co-lending networks. Additionally, we investigate the predictive power of the CLN score conditional on bank opacity. We employ discretionary loan loss provisions as a proxy for bank opacity following Jiang et al. (2016).[28] Table 3.A5 in the Appendix reveals that the coefficient estimates for the interaction terms between $Bank\ Opacity_{i,t}$ and $CLN\ score_{i,t}$ are mostly positive and statistically significant across all three samples. These findings suggest that the predictive power of bank CLN score is amplified for more opaque banks.

Finally, Panel E of Table 3.7 examines how capital adequacy affects the predictive power of bank CLN score. Banks with lower capital adequacy are more susceptible to financial distress because they have weaker capital buffers to absorb losses. Reduced loss-absorbing capacity can increase banks' sensitivity to risk spillovers inside the co-lending network. We identify banks with low capital adequacy using an indicator variable that captures institutions in the bottom quartile of risk-based capital ratios. The interaction terms between $Low\ capital\ adequacy_{i,t}$ and $CLN\ score_{i,t}$ are strongly significant at shorter horizons (quarters 1-3), suggesting that capital-constrained banks face immediate vulnerability to risk spillovers. The significance wanes at longer horizons, potentially reflecting regulatory pressure or market discipline that encourages capital-constrained banks to improve their positions over time, thus reducing their susceptibility to network risks in the longer term.

[27] By comparison, Compustat lists only 974 items as of 2024, highlighting the FR Y-9C's broader scope and its ability to reflect the multifaceted nature of banking operations.

[28] The bank opacity is determined by the natural logarithm of the absolute value of residuals from a regression of loan loss provisions on changes in non-performing assets, among other factors, incorporating state-quarter fixed effects. The regression specification is: $LLP_{i,t,j} = \alpha_1 \Delta NPA_{i,j,t+1} + \alpha_2 \Delta NPA_{i,j,t} + \alpha_3 \Delta NPA_{i,j,t-1} + \alpha_4 Size_{i,j,t-1} + \alpha_5 \Delta Loan_{i,j,t} + \delta_{i,t} + \epsilon_{i,j,t}$ where $LLP_{i,j,t}$ represents loan loss provisions scaled by lagged total loans for bank i in state j at quarter t, $\Delta NPA_{i,j,t}$ denotes the change in non-performing assets for bank i in state j from quarter $t-1$ to t, scaled by lagged total loans. $Size_{i,j,t-1}$ is the natural logarithm of the bank's total assets at $t-1$, and $\Delta Loan_{i,j,t}$ captures the change in total loans from $t-1$ to t. $\delta_{j,t}$ represents state-quarter fixed effects, capturing regional and temporal variations. The regression model includes both lead and lag of $\Delta NPA_{i,j,t}$ to reflect the banks' use of forward-looking and historical non-performing asset data in provisioning for loan losses.

3.6.5 Further Results

Predicting other bank risk metrics

In this section, we show that the risk predictive ability of the bank CLN score extends to other bank risk metrics beyond loan loss provisions. Specifically, we use non-performing loans (NPLs) as an alternative dependent variable. Unlike loan loss provisions, which may be influenced by managerial discretion, NPLs represent loans where borrowers have ceased making interest or principal payments. As such, NPLs provide a direct measure of the bank's credit risk exposure, reflecting actual loan performance. Table 3.8 reports the results of predicting future NPLs.

[Insert Table 3.8 about here]

In Panel A, the coefficient estimates of $\mathit{CLN\ score}_{i,t}$ are positive and highly significant at the 1% level across all prediction horizons. In terms of economic significance, for instance, a one-standard-deviation increase in $\mathit{CLN\ score}_{i,t}$ is, on average, related to a 0.099-percentage-point rise (i.e., 0.458*0.216) in the ratio of NPLs to total assets, representing approximately 9.7% of the sample mean NPL ratio of 1.019. This translates into an estimated annual increase in non-performing loans of approximately $41 million. Panels B and C further confirm the out-of-sample predictive strength of the CLN score for unlabeled and private banks, with results remaining qualitatively consistent.

Additionally, we validate the risk predictive power of bank-level CLN score on well-known and widely-used public bank risk metrics, including the default probability (Merton, 1974), the modified default probabilities (Nagel & Purnanandam, 2020),[29], and the natural logarithm of stock return idiosyncratic volatility (ln(IVOL)). Due to the reliance on stock market information, the sample is restricted to public banks. We employ specifications similar to Equation (3.3), replacing the dependent variable with these public bank risk metrics. Table 3.9 reports the results.

[Insert Table 3.9 about here]

The results show that the coefficient estimates for bank CLN score are positive and highly statistically significant across all three market-based bank risk measures and various predictive horizons. These findings provide strong evidence of the CLN score's effectiveness in predicting future bank risks.

[29] The default probability is estimated via a KMV iterative approach. The modified default probability is estimated using the data and code provided by Stefan Nagel, available at https://voices.uchicago.edu/stefannagel/code-and-data/.

Predicting future bank profitability

We further examine the predictive power of bank CLN score on future bank profitability, anticipating that the risk spillover effects within the co-lending network may influence banks' profitability. To test this, we estimate predictive regression models similar to Equation (3.3), replacing the bank risk measures with the profitability metric, return on assets (ROA). Table 3.10 reports the results.

[Insert Table 3.10 about here]

The results indicate that the coefficient estimates for bank $CLN\ score_{i,t}$ are negative and statistically significant at the 1% level across all specifications in Panel A and most specifications in Panels B and C. These findings demonstrate that a higher bank-level $CLN\ score_{i,t}$ predicts lower bank profitability over the next eight quarters. Economically, column (1) of Panel A reveals that a one-standard-deviation increase in $CLN\ score_{i,t}$ is associated with a 0.121-percentage-point decrease (i.e., -0.265 $\times$ 0.458) in ROA in the next quarter, representing approximately 22% of the sample mean ROA of 0.54%.

The predictability after controlling for bank stock performance

We further examine the predictive power of bank CLN score after controlling for bank stock performance, measured by quarterly stock returns, within the subsample of public banks. In Panel A of Table 3.A6 in the Appendix, we find that controlling for quarterly bank stock returns does not reduce the predictive strength of $CLN\ score_{i,t}$. Its coefficient estimates remain significant across all dependent variables—loan loss provisions, non-performing loans, and ROA.

Importantly, if the predictive power of the CLN score were due to banks with high CLN scores experiencing shared negative shocks (e.g., common major borrower defaults), then controlling for the stock returns of these banks should significantly diminish the CLN score's predictive power. However, our findings suggest otherwise. Specifically, bank CLN score continues to significantly predict future loan loss provisions, non-performing loans, and ROA after controlling for bank stock returns, consistent with its role in capturing risk spillovers through the co-lending network.

Notably, Panel B of Table 3.A6 further shows that bank CLN score significantly outperforms quarterly bank stock returns in predicting future non-performing loans and ROA for the subsample of unlabeled public banks. Our CoLGNN provide more informative signals than middle-ranked stock returns. We find similar patterns for market-based risk measures in Table 3.A7, where the CLN score maintains significant predictive power for Merton default probability, modified default probability, and idiosyncratic volatility even after controlling for contemporaneous stock returns. These findings collectively highlight the utility of the CLN score in capturing bank risks beyond conventional stock performance metrics.

Time-varying predictive power of bank CLN score

We further examine the predictive power of bank CLN score across different subperiods within the full sample period. Table 3.A8 in the Appendix presents the results of this analysis, where the sample is equally split into two halves to validate the predictive ability of bank-level CLN score. We find that the $CLN\ score_{i,t}$ consistently demonstrates robust and statistically significant predictive power across all subsamples and different dependent variables—loan loss provisions, non-performing loans, and ROA.

Additionally, Table 3.A9 in the Appendix focuses on the predictive power of bank-level CLN score during the Global Financial Crisis (GFC). We introduce a GFC indicator, set to 1 for year-quarters between mid-2007 and the end of 2008, and 0 otherwise. The interaction term between $CLN\ score_{i,t}$ and the GFC indicator is both statistically and economically significant, indicating that the predictive power of bank CLN score is particularly pronounced during periods of heightened financial instability. During the GFC, the interconnected nature of financial institutions amplifies the transmission of risk throughout the banking sector. As banks face simultaneous liquidity constraints and credit defaults, risk propagation through co-lending relationships becomes more severe, underscoring the importance of network-based risk measures. These results highlight the effectiveness of the CLN score in capturing network spillover risks during crisis periods, when traditional risk metrics often fail to capture the full extent of interconnected risks.

3.7 Conclusion

This paper develops the Co-Lending Graph Neural Network (CoLGNN), an AI-driven framework designed to model risk transmission in syndicated lending markets. By integrating the topological structure of the co-lending network with rich lender and loan features, our deep learning approach generates a bank-level CLN risk score that robustly predicts future bank distress and profitability. The model significantly enhances early-warning risk prediction, particularly for private and smaller banks that are often overlooked by traditional market-based risk metrics.

We first validate the CLN score's ability to capture network-based risk spillovers using two complementary quasi-natural experiments. Credit rating downgrades serve as targeted, bank-level external shocks, while the collapse of Lehman Brothers provides a large, non-lending shock. In both settings, we find that banks with inward co-lending ties to affected institutions experience subsequent and significant increases in their CLN scores. These findings confirm that the CLN score reflects genuine risk spillovers through the co-lending network rather than shared exposure to distressed corporate borrowers.

Having validated the CLN score's ability in capturing network risk spillovers, we next document the strong predictive power of the CLN score for future bank risk up to two years ahead. Its forecasting accuracy is especially pronounced among more vulnerable institutions—those that are smaller, less profitable, more volatile, less capitalized, and more operationally complex. This risk predictive ability is particularly valuable for private banks, where traditional market-based risk signals are often unavailable.

The findings of the study provide implications for financial institutions and regulators. They highlight the need to adopt a holistic approach to risk management that extends beyond firm-level assessments to encompass the broader financial ecosystem. For regulators, the results advocate for supervisory frameworks that explicitly account for the interconnected nature of banking networks. By doing so, financial systems can better absorb shocks originating from individual institutions and mitigate systemic vulnerabilities.

The CoLGNN framework represents a significant advancement in financial risk modeling. As an AI-powered and scalable tool, it enables regulators and market participants to detect hidden interdependencies and quantify network-based risk spillovers with greater precision. Integrating network-aware metrics such as the CLN score into stress testing, early-warning systems, and macroprudential policy can enhance the resilience of modern banking systems. Future research may build on this approach to examine dynamic contagion mechanisms, cross-border spillovers, and the design of targeted regulatory interventions in complex financial networks.

Figure 3.1: **Estimating Bank Risk via Co-Lending Networks and CoLGNN**

Figure 3.1 shows the flow chart of our estimation of bank co-lending network risk via Co-Lending Networks and CoLGNN. Specifically, the figure illustrates the co-lending network at time t. At stage 1, we classify banks with top-performed stock market performance to the "safe" group and classify banks with bottom-performed stock market performance to the "risky" group. The remaining public banks and private banks are treated as unlabeled samples. At stage 2, we construct the co-lending network by utilizing the all syndicated loans originated in the last five years (from $t-20$ to $t-1$). Each co-lending network is a directed network with edges points from each lead arrange to other banks within the syndicate. At stage 3, we estimate bank CLN score CoLGNN model as described in section B.1.2. At stage 4, we perform empirical experiment showing the Bank CLN score effectively captures risk spillovers in the co-lending network, serving as an early-warning indicator of each bank's future risk and performance.

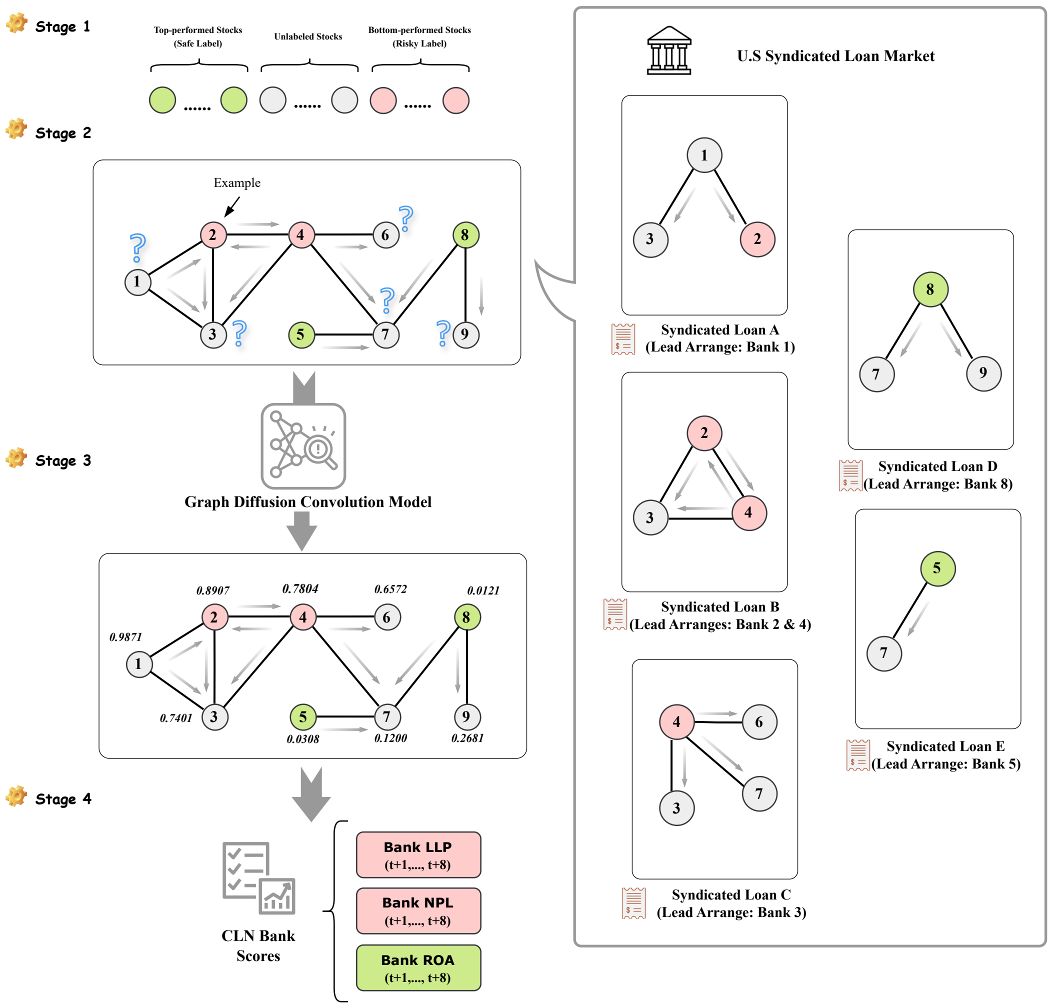


Figure 3.2: **Graph Diffusion Module**

Figure 3.2 visually illustrate the CoLGNN framework which includes a graph convolution with edge embedding. At stage 3.1, we prepare the high dimensional node features and edge features using bank characteristics from FR Y9-C and the loan characteristics from DealScan, respectively. At stage 3.2, we aggregate the all features at node level. To capture the risk spillover in the co-lending network, we utilize the message-passing paradigm design in the graph neural network by creating in-flow aggregation and out-flow aggregation. At stage 3.3, we calculate the node-wise representation for each layer of CoLGNN and use the softmax activation function to introduce non-linearity into the output of a neuron. At stage 3.4, the parameter matrix of CoLGNN is estimated by minimizing the binary cross-entropy (BCE) loss function. At stage 3.5, we consider the estimated probabilities to the "risky" class as the final co-lending network risk score for bank i at time t.

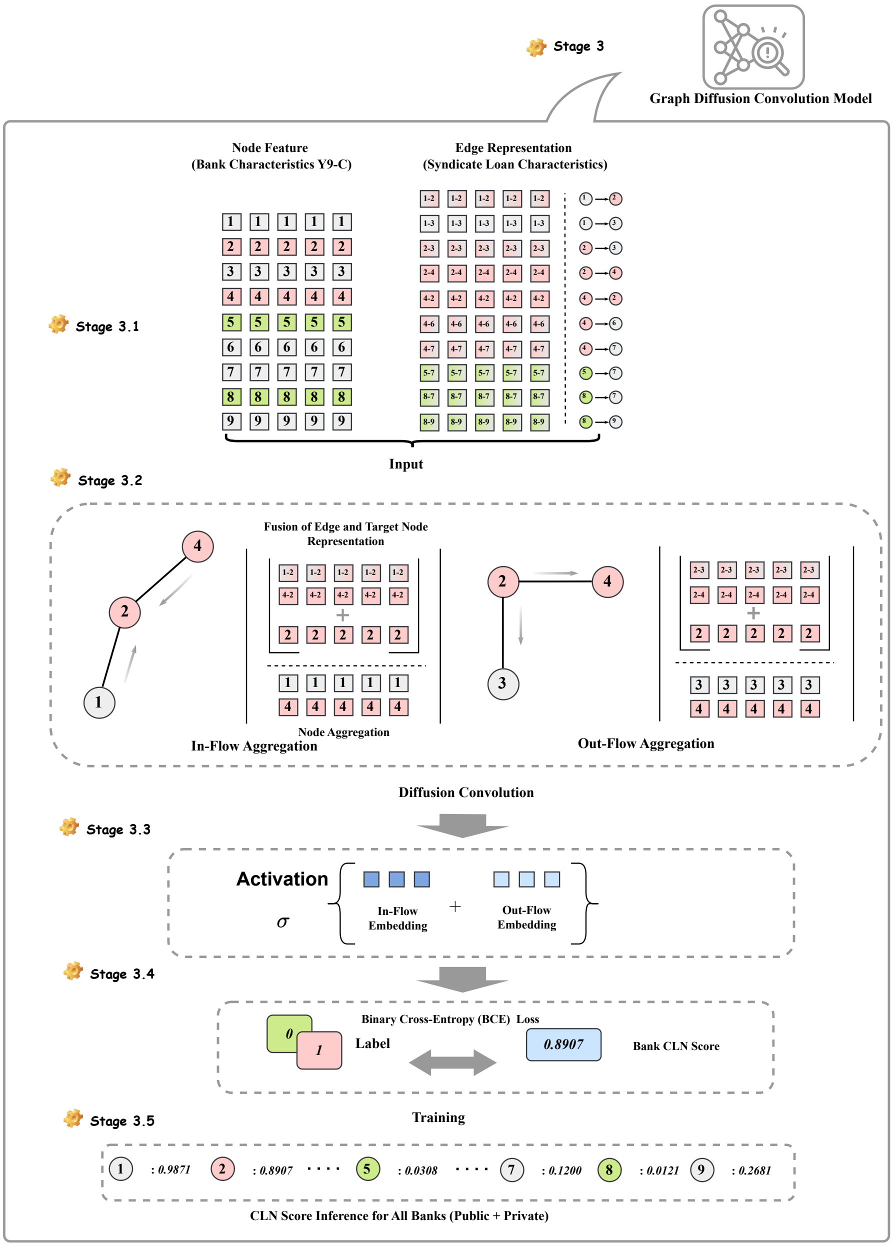


Figure 3.3: **CoLGNN One-Layer Graph Neural Network Visualization Example**

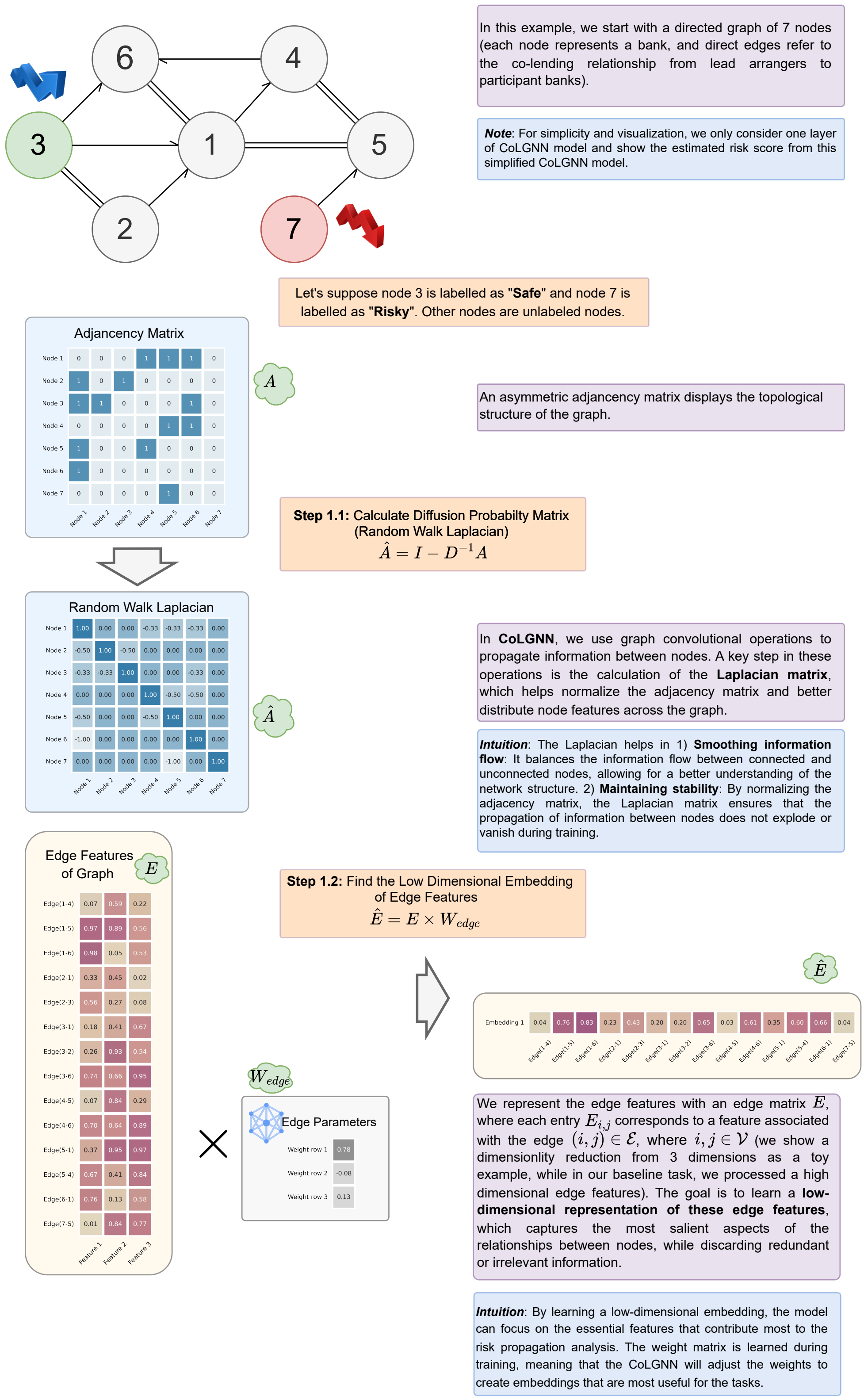


Figure 3.3: **Continued**

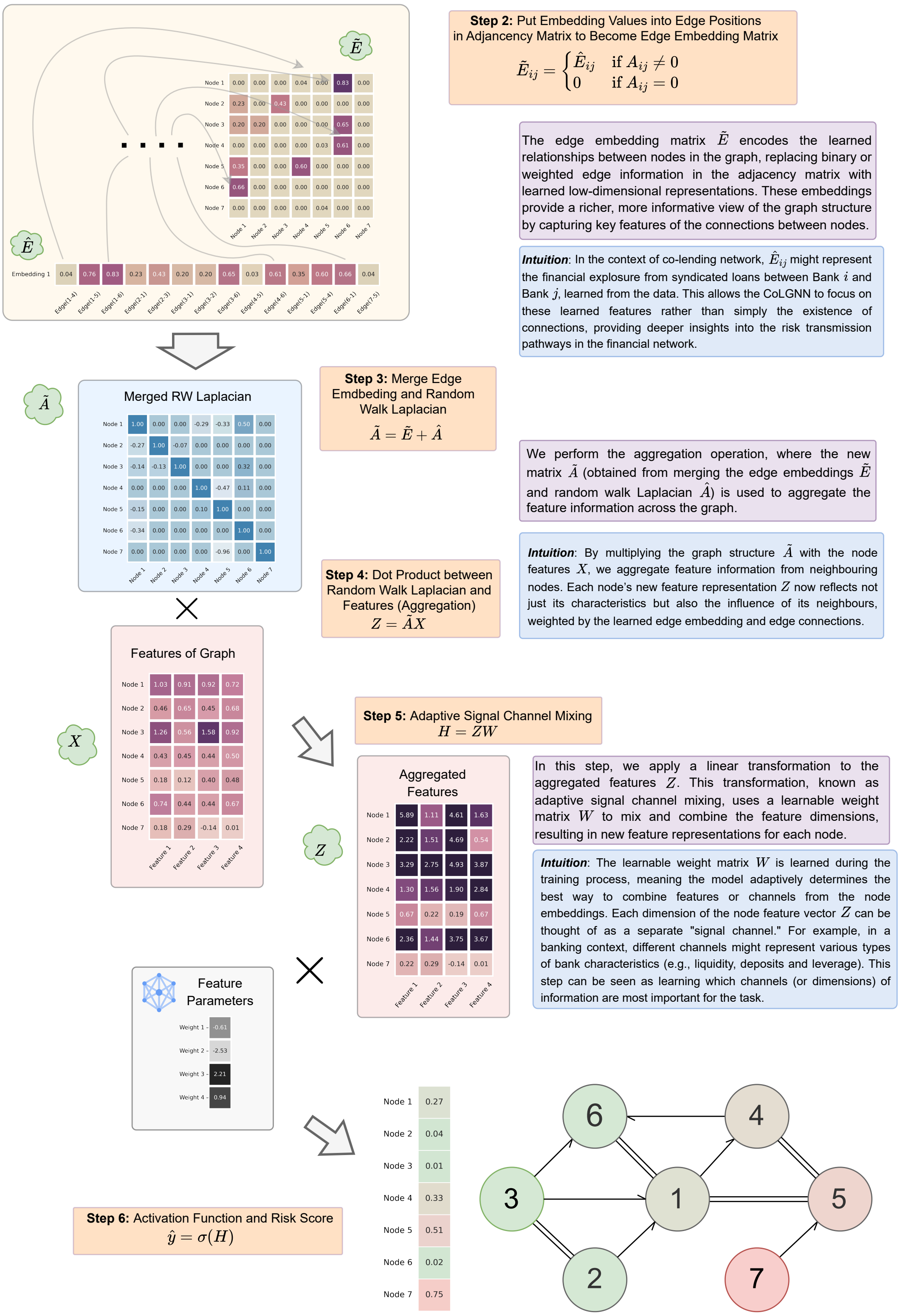
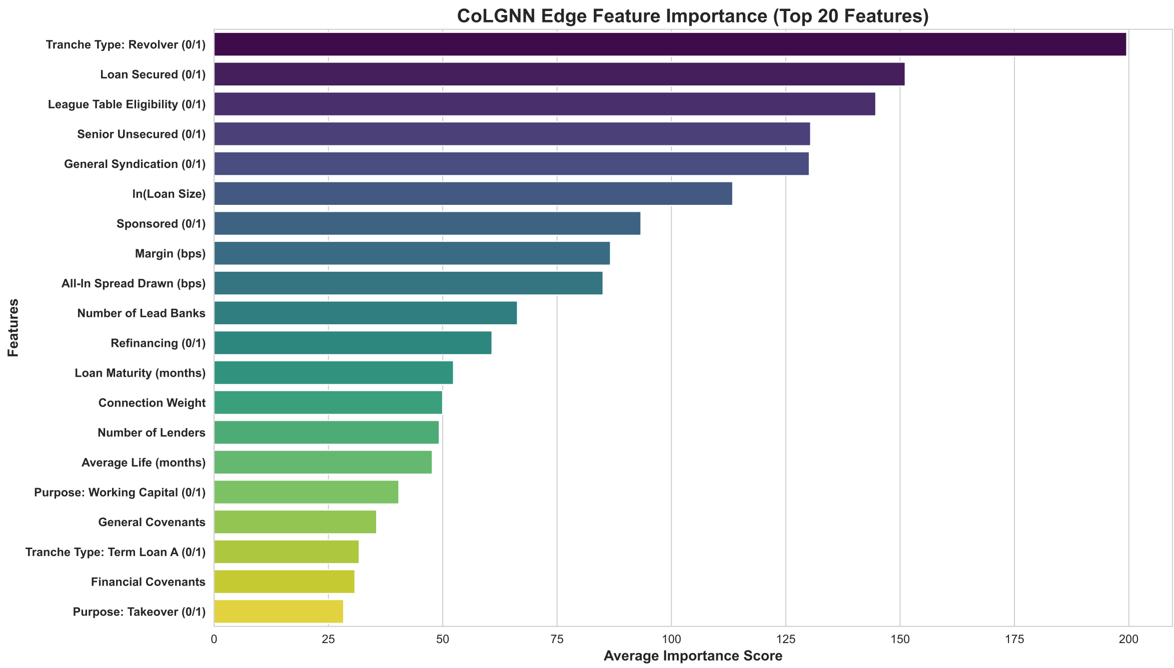
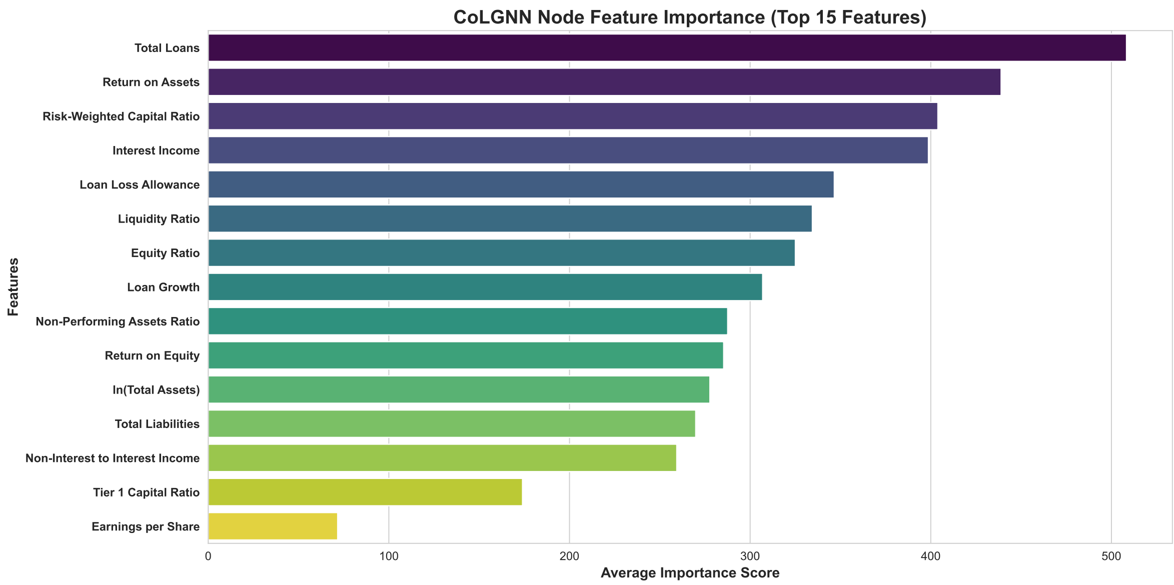


Figure 3.4: **CoLGNN Feature Importance**

Figure 3.4 displays the importance scores of the top-ranked (a) edge features (loan characteristics) and (b) node features (bank characteristics) based on Integrated Gradients attributions, averaged across all quarterly co-lending networks in our sample. The importance scores quantify each feature's contribution to the CLN risk score prediction using the Integrated Gradients method (Sundararajan et al., 2017), which measures attribution by integrating the partial derivatives of model output with respect to each input feature along a straight-line path from a reference point to the actual input values. For each feature, we compute the absolute values of these attributions and average them across all networks.



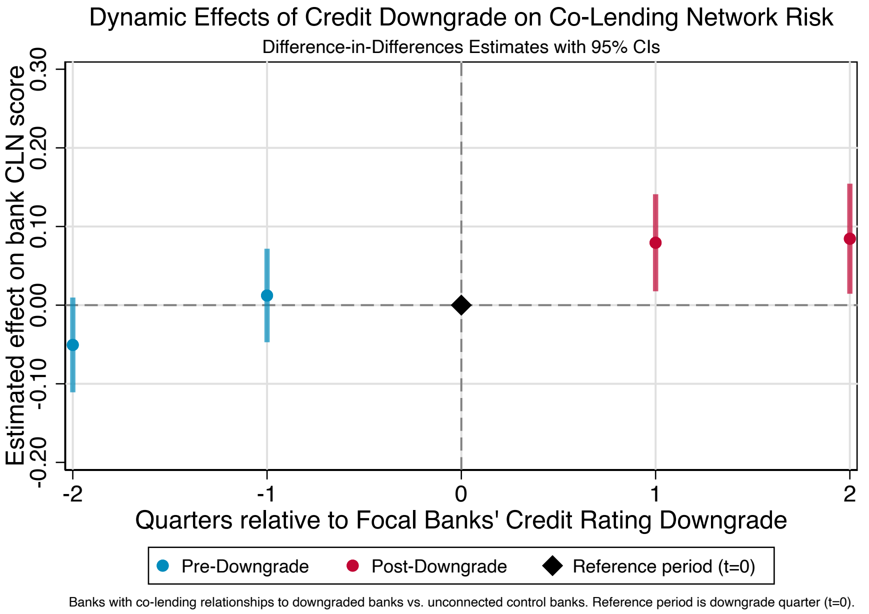
(a) CoLGNN Edge Feature Importance



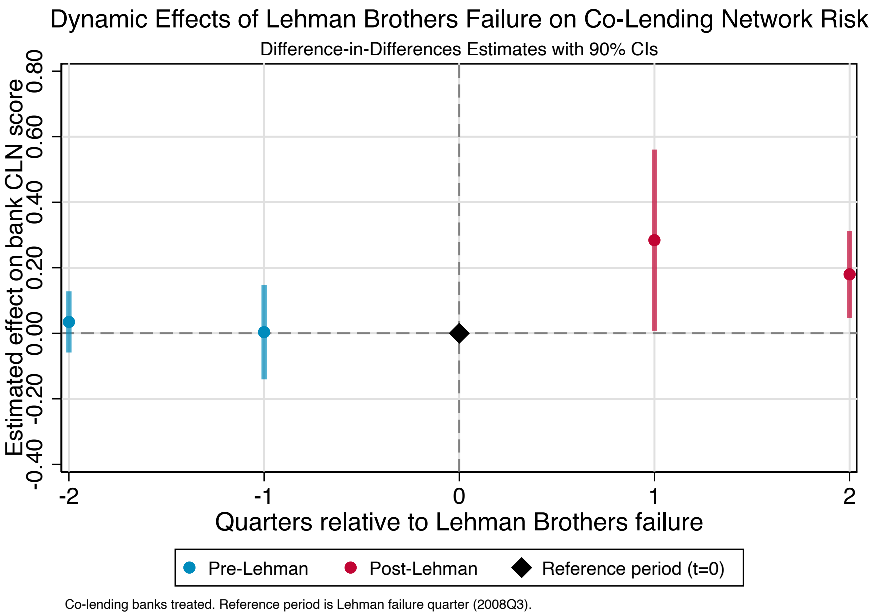
(b) CoLGNN Node Feature Importance

Figure 3.5: **Dynamic Effects of Risk Transmission in Bank Co-Lending Networks**

Figure 3.5 presents event study plots showing the dynamic effects of risk transmission through co-lending networks. Figure (a) shows the dynamic DiD effects on bank CLN scores following the credit downgrades. Banks with co-lending relationships to focal downgraded banks are treated banks, and other unconnected banks are control banks. Figure (b) shows the dynamic estimated effects on bank CLN scores following the Lehman Brothers bankruptcy in 2008Q3. Banks with co-lending relationships to Lehman Brothers are treated banks, and others are control banks. Each plot displays difference-in-differences estimates with 90% confidence intervals. The reference period (t=0) is indicated by the diamond marker.



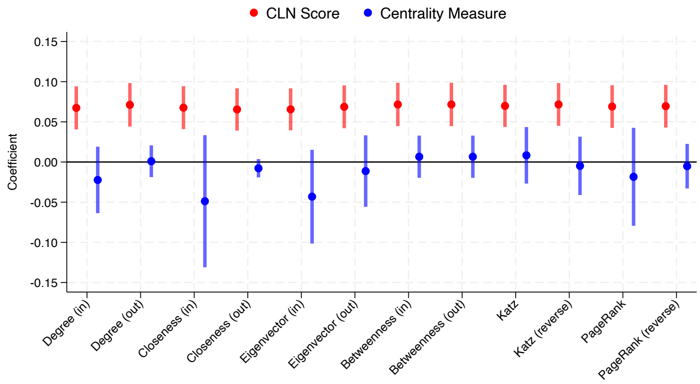
(a) Credit Rating Downgrade Transmission Effect



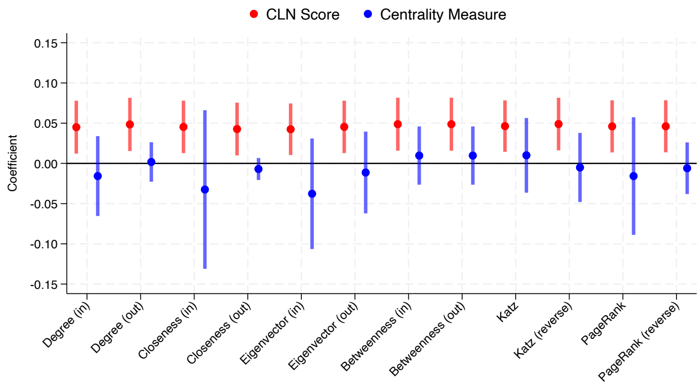
(b) Non-lending shock: Lehman Brothers Failure Transmission Effect

Figure 3.6: **CLN Score and Centrality Measure Coefficient Estimates**

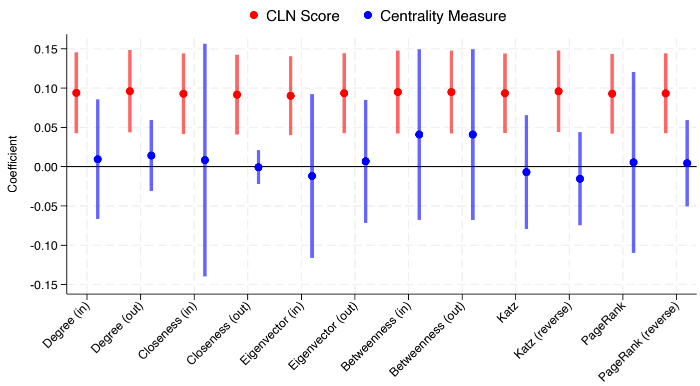
Figure 3.6 compares the coefficient estimates of our CLN score (red dots) alongside various traditional network centrality measures (blue dots) in predicting bank loan loss provisions four quarters ahead. Each centrality measure was included in a separate regression alongside the CLN score, with all regressions including the same control variables. While the CLN score consistently shows statistical significance across all samples, none of the traditional network centrality measures demonstrate significant predictive power. The centrality measures are defined in Appendix B.2.



(a) All banks



(b) Unlabeled banks



(c) Private banks

Table 3.1: **Summary Statistics**

Table 3.1 presents the summary statistics of our study. The loan-level samples from January 1990 to December 2020. Definitions of the variables are provided in Table 2.A1 in the Appendix. All continuous variables are winsorized by year at the 1^{st} and 99^{th} percentiles.

	Observations	Mean	10^{th} Percentile	Median	90^{th} Percentile	Standard Deviation
Bank-level Samples						
CLN score	11662	0.465	0.000	0.313	1.000	0.458
Loan loss provision	11662	0.362	0.003	0.163	0.865	0.688
Size	11662	16.164	13.854	15.935	18.892	1.984
Equity capital	11662	0.104	0.069	0.098	0.139	0.042
ROA	11662	0.540	0.109	0.525	1.193	0.827
Loan size	11662	0.628	0.415	0.669	0.784	0.161
Loan growth	11662	0.035	-0.023	0.015	0.068	1.165
Loan loss allowance	11662	0.946	0.431	0.869	1.550	0.521
Liquidity	11662	0.199	0.080	0.179	0.350	0.112
Deposits	11662	0.549	0.293	0.579	0.743	0.180
Lead share	11662	0.004	0.000	0.000	0.007	0.014
Specialization in syndicated loan	11662	0.007	0.000	0.000	0.001	0.311
Specialization in industry	11662	0.226	0.000	0.000	1.000	0.341
Bank Z-score	11662	0.323	0.150	0.297	0.505	0.187
Bank nonperforming loans	10586	1.019	0.140	0.567	2.160	1.581
Standardized unexpected earnings	11294	0.039	-0.000	0.000	0.000	2.810
Bank complexity	11662	0.044	-0.172	0.062	0.242	0.154
Bank risk capital	8309	0.143	0.108	0.132	0.181	0.066
Modified default probability	6252	0.209	0.089	0.174	0.369	0.134
Default probability	6252	0.131	0.018	0.065	0.316	0.174
IVOL	8190	1.594	0.762	1.245	2.649	1.224
Loan-level Samples						
ln(Loan Size)	66596	4.998	3.122	5.017	6.909	1.451
Loan Spread (bps)	66596	227.286	50.000	200.000	425.000	154.598
Loan Maturity (months)	66596	53.088	12.000	60.000	84.000	30.476
Lead Banks	66596	1.965	1.000	1.000	4.000	1.646
Lenders	66596	9.197	2.000	6.000	20.000	8.923
General covenants	66596	1.222	0.000	0.000	6.000	2.924
Financial covenants	66596	0.413	0.000	0.000	2.000	0.965
Loan secured	66596	0.400	0	0	1	0.490

Table 3.2: **Difference-in-Differences Estimation: Credit Downgrade Spillover**

Table 3.2 shows the results of the difference-in-differences estimation using the focal banks' credit downgrade as exogenous shocks to the *CLN* score of neighboring banks in the co-lending network. The treatment events are long-term credit downgrade at the entity level from S&P credit ratings for banks in our co-lending network. In each event cohort, treatment groups are banks that have a inward direction with the focal bank (from the focal bank to treatment banks) and control groups are banks that are not connected with the focal banks. *Treat* equals to 1 for treatment banks and 0 for control banks. In each cohort, we use a two quarters before and two quarters after ($t-2$, $t+2$) event window within each cohort. *Post* indicator equals to 0 (1) for all quarters before (after) the credit event in each cohort, and time dummies d_j equal to 1 for the year that is j quarter(s) after the treatment. Bank controls are lagged-one-period. Definitions of other variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are are two-tailed t-statistics. Standard errors are clustered at the bank level or double-clustered at both the cohort and bank levels, depending on the specification. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
Post Credit Downgrade × Connected Banks	0.109***	0.109**	0.102***	0.102**
	(2.700)	(2.354)	(2.633)	(2.279)
Size			0.241**	0.241**
			(2.567)	(2.490)
Equity capital			1.769**	1.769**
			(2.498)	(2.463)
Loan size			-0.055	-0.055
			(-0.140)	(-0.129)
Loan growth			-0.010***	-0.010***
			(-7.363)	(-8.410)
Deposits			0.339*	0.339
			(1.710)	(1.649)
ROA			-0.077***	-0.077**
			(-3.791)	(-2.394)
Loan loss allowance			-0.076**	-0.076
			(-2.040)	(-1.499)
Liquidity			-0.043	-0.043
			(-0.114)	(-0.108)
Cohort-YrQtr Fixed Effects	Yes	Yes	Yes	Yes
Cohort-Bank Fixed Effects	Yes	Yes	Yes	Yes
Clustering	Bank	Bank & Cohort	Bank	Bank & Cohort
Observations	5,939	5,939	5,939	5,939
Adjusted R^2	0.277	0.277	0.286	0.285

Table 3.3: **Difference-in-Differences Estimation: Lehman Brothers Spillover**

Table 3.3 shows the results of the difference-in-differences estimation using the Lehman Brothers failure as an exogenous shock to the *CLN* score of neighboring banks in the co-lending network. The financial distress that led to Lehman's collapse stemmed primarily from its real estate investments and unstable funding, not from deterioration in its corporate loan portfolio, helping isolate network transmission effects from common exposure to troubled borrowers. The treatment group consists of banks with co-lending relationships with Lehman Brothers prior to its collapse. The control group includes banks without such connections. We use a two-quarter window before and after the Lehman failure (t-2, t+2), excluding major recipients of the Troubled Asset Relief Program and banks with connections to Bear Stearns. Panel A presents the standard DiD results with the interaction term capturing the effect on treated banks after Lehman's collapse. Panel B shows results from Arkhangelsky et al. (2021) synthetic difference-in-differences, where we construct counterfactual outcomes using matched control groups based on lending characteristics and full bank controls. Bank controls are lagged-one-period. Definitions of all variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are standard errors clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Standard DiD

	(1)	(2)
Lehman Brothers' Co-Lender × Post Lehman Brothers Failure	0.128*	0.134**
	(0.071)	(0.067)
Size		0.213
		(0.463)
Equity capital		-3.065**
		(1.470)
ROA		0.004
		(0.015)
Loan size		0.304
		(0.825)
Loan growth		-0.444
		(0.521)
Loan loss allowance		0.070
		(0.064)
Liquidity		0.430
		(0.594)
Bank Fixed Effects	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes
Observations	472	472
Adjusted R^2	0.459	0.466

Panel B: Synthetic DiD

	(1)	(2)
Average Treatment Effect on the Treated	0.140**	0.124*
	(0.067)	(0.072)
Controls	Lending Characteristics	Full
Observations	435	435

Table 3.4: **Bank Co-Lending Network Risk and Bank Loan Loss Provisions**

Table 3.4 presents the h-quarter-ahead prediction results of bank-level *CLN score* for bank loan loss provisions. The bank-level *CLN score* and control variables are measured at time t, and the dependent variable is measured at time $t+h$. Panel A reports the results using all observations. Panel B shows the results using unlabeled banks (banks who do not have assigned labels). Panel C shows the results using private banks only. Panel B and C have extactly same control variables as Panel A. For simplicity, we do not report the coefficients estimates of the control variables in the Panel B and C. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Loan loss provisions	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel A: All banks								
CLN score	0.091***	0.090***	0.090***	0.071***	0.079***	0.059***	0.054***	0.024
	(7.401)	(7.191)	(6.765)	(5.179)	(5.222)	(3.833)	(3.439)	(1.342)
Size	0.035***	0.036***	0.036***	0.039***	0.039***	0.040***	0.042***	0.045***
	(3.719)	(3.782)	(3.676)	(3.810)	(3.715)	(3.590)	(3.712)	(3.900)
Equity capital	-0.675	-0.531	-0.408	0.149	0.247	0.275	0.392	0.631
	(-0.697)	(-0.712)	(-0.606)	(0.279)	(0.497)	(0.549)	(0.773)	(1.242)
Deposits	-0.017	-0.028	-0.063	-0.081	-0.102	-0.132	-0.121	-0.115
	(-0.147)	(-0.247)	(-0.533)	(-0.646)	(-0.776)	(-0.990)	(-0.914)	(-0.837)
ROA	-0.131***	-0.092***	-0.069***	-0.106***	-0.058**	-0.052*	-0.057	-0.074*
	(-4.328)	(-3.326)	(-2.706)	(-2.818)	(-2.269)	(-1.714)	(-1.496)	(-1.708)
Loan size	-0.494***	-0.402***	-0.301**	-0.163	-0.079	-0.014	0.038	0.113
	(-3.558)	(-2.957)	(-2.119)	(-1.097)	(-0.525)	(-0.089)	(0.237)	(0.700)
Loan growth	0.001	0.001**	-0.035	-0.074	-0.018	-0.006	0.126*	0.232**
	(0.566)	(2.042)	(-0.507)	(-1.254)	(-0.309)	(-0.088)	(1.653)	(2.190)
Loan loss allowance	0.576***	0.530***	0.478***	0.400***	0.347***	0.315***	0.275***	0.234***
	(10.745)	(8.993)	(7.483)	(6.287)	(5.325)	(4.521)	(3.958)	(3.356)
Liquidity	-0.288*	-0.305**	-0.333**	-0.318**	-0.322**	-0.307**	-0.300**	-0.278*
	(-1.801)	(-2.003)	(-2.217)	(-2.143)	(-2.216)	(-2.087)	(-1.983)	(-1.805)
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11308	10966	10631	10303	9980	9666	9365	9073
Adjusted R^2	0.555	0.510	0.483	0.470	0.449	0.442	0.440	0.441
Panel B: Unlabeled banks								
CLN score	0.063***	0.058***	0.053***	0.048***	0.059***	0.042**	0.038**	0.012
	(4.047)	(3.425)	(3.009)	(2.861)	(3.013)	(2.297)	(2.154)	(0.565)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6907	6699	6495	6294	6098	5901	5702	5518
Adjusted R^2	0.510	0.476	0.455	0.448	0.424	0.429	0.429	0.435
Panel C: Private banks								
CLN score	0.069***	0.076***	0.089***	0.092***	0.111***	0.079***	0.072***	0.039
	(2.717)	(2.863)	(3.058)	(3.399)	(3.559)	(2.697)	(2.641)	(1.232)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3882	3742	3607	3476	3348	3223	3103	2988
Adjusted R^2	0.495	0.466	0.445	0.438	0.416	0.415	0.409	0.410

Table 3.5: **Incremental R^2 and Predictive Power vs Z-score**

Table 3.5 presents comparative prediction analyses of loan loss provisions. For simplicity, we report results for 2, 4, and 6 quarters ahead prediction horizons, though our analysis confirms consistent patterns across all forecasting horizons. We compare four specifications: (1) bank-level controls only, (2) controls plus the Z-score risk measure, (3) controls plus our CLN score, and (4) all variables combined. Panels A, D, and G report in-sample adjusted R^2 from the labeled public banks used for training. Panels B, E, and H report out-of-sample root-mean-square-error (RMSE) for unlabeled banks. Panels C, F, and I report out-of-sample RMSE specifically for private banks. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1) Controls only	(2) Z-score	(3) CLN score	(4) Both
Prediction Horizon: 2 Quarters Ahead				
Panel A: Labeled public banks				
CLN score			0.135***	0.136***
			(7.273)	(7.223)
Bank Z-score		0.003		-0.018
		(0.036)		(-0.214)
Controls and Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes
Observations	4267	4267	4267	4267
Adjusted R^2	0.560	0.560	0.568	0.568
Panel B: Unlabeled banks				
RMSE	0.599	0.600	0.569	0.566
Controls and Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes
Panel C: Private Banks)				
RMSE	0.665	0.665	0.629	0.626
Controls and Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes
Prediction Horizon: 4 Quarters Ahead				
Panel D: Labeled public banks				
CLN score			0.116***	0.116***
			(5.187)	(5.157)
Bank Z-score		0.015		-0.005
		(0.168)		(-0.062)
Controls and Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes
Observations	4009	4009	4009	4009
Adjusted R^2	0.515	0.515	0.521	0.521
Panel E: Unlabeled banks				
RMSE	0.663	0.666	0.648	0.646
Controls and Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes
Panel F: Private Banks				
RMSE	0.702	0.705	0.686	0.685
Controls and Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes
Prediction Horizon: 6 Quarters Ahead				
Panel G: Labeled public banks				
CLN score			0.091***	0.091***
			(4.169)	(4.148)
Bank Z-score		0.029		0.015
		(0.282)		(0.149)
Controls and Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes
Observations	3765	3765	3765	3765
Adjusted R^2	0.482	0.482	0.486	0.485
Panel H: Unlabeled banks				
RMSE	0.669	0.675	0.655	0.658
Controls and Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes
Panel I: Private Banks				
RMSE	0.712	0.717	0.699	0.701
Controls and Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes

Table 3.6: **Revolver-based CLN score**

Table 3.6 presents h-quarter-ahead prediction results for bank loan loss provisions using CLN scores constructed from different loan types, focusing specifically on private banks where our methodology offers unique insights unavailable through market-based metrics. To address concerns about loan sales in secondary markets (Blickle et al., 2020), we conduct two tests: Panel A shows results using only revolving credit facilities, which have documented higher retention rates and ongoing monitoring requirements (Sufi, 2007). Panel B presents a placebo test using Term Loan B and above, which are frequently sold by lead arrangers shortly after origination. Both panels include the full set of controls as in our baseline model. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Loan loss provisions	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel A: CLN score using revolving credit facilities								
CLN score (Revolver)	0.070**	0.074***	0.102***	0.085***	0.112***	0.126***	0.114***	0.047
	(2.472)	(2.786)	(3.363)	(3.070)	(3.439)	(4.007)	(3.150)	(1.523)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3423	3283	3152	3026	2907	2792	2683	2581
Adjusted R^2	0.506	0.475	0.453	0.440	0.415	0.413	0.407	0.407
Panel B: CLN score using term loan B and above								
CLN score (Term loan B above)	0.054	0.062**	0.049	0.029	0.067	0.040	0.014	-0.021
	(1.494)	(2.158)	(1.462)	(0.724)	(1.427)	(1.131)	(0.452)	(-0.587)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2003	1945	1888	1831	1776	1722	1671	1621
Adjusted R^2	0.569	0.540	0.514	0.486	0.464	0.464	0.448	0.441

Table 3.7: **The Cross-Sectional Variation in the Risk Predictive Power**

Table 3.7 examines the cross-sectional variation in the risk predictive power of the CLN score. Each panel tests for heterogeneous effects based on bank size (Panel A), earnings performance (Panel B), earnings volatility (Panel C), operational complexity (Panel D), and capital adequacy (Panel E). Small bank dummy equals 1 if bank size is below the median. Negative earnings shock equals 1 if the bank has negative standardized unexpected earnings. High ROA volatility equals 1 for banks in the top quartile of ROA standard deviation. High complexity equals 1 if bank regulatory filing complexity is above the median. Low capital adequacy equals 1 for banks in the bottom quartile of risk-based capital. For simplicity, we do not report the coefficient estimates of the control variables. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Loan loss provisions	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel A: Bank size								
CLN score	0.066***	0.059***	0.064***	0.051***	0.049***	0.037**	0.027	-0.002
	(4.422)	(4.263)	(3.769)	(3.122)	(3.213)	(2.360)	(1.470)	(-0.089)
Small bank dummy	-0.004	0.012	0.032	0.045	0.041	0.051	0.048	0.045
	(-0.112)	(0.328)	(0.814)	(1.143)	(0.988)	(1.237)	(1.087)	(1.046)
CLN score × Small bank dummy	0.055**	0.069**	0.061**	0.047	0.069**	0.054	0.066*	0.063*
	(2.198)	(2.464)	(1.986)	(1.572)	(2.279)	(1.647)	(1.916)	(1.718)
Observations	11308	10966	10631	10303	9980	9666	9365	9073
Adjusted R^2	0.555	0.511	0.484	0.471	0.451	0.443	0.442	0.442
Panel B: Negative earnings								
CLN score	0.033**	0.032**	0.045***	0.030**	0.032**	0.033*	0.031*	0.001
	(2.485)	(2.494)	(3.394)	(2.044)	(2.168)	(1.953)	(1.709)	(0.069)
Negative earnings shock	-0.003	0.004	0.024	0.015	0.022	0.044**	0.032	0.010
	(-0.169)	(0.211)	(1.275)	(0.725)	(1.122)	(2.128)	(1.490)	(0.442)
CLN score × Negative earnings shock	0.135***	0.136***	0.103***	0.094***	0.107***	0.051*	0.048	0.051
	(4.586)	(4.686)	(3.378)	(3.217)	(3.691)	(1.821)	(1.555)	(1.581)
Observations	11308	10966	10631	10303	9980	9666	9365	9073
Adjusted R^2	0.558	0.514	0.486	0.473	0.453	0.444	0.442	0.442
Panel C: Earnings volatility								
CLN score	0.081***	0.081***	0.085***	0.057***	0.067***	0.046***	0.045**	0.010
	(6.494)	(5.780)	(5.378)	(3.514)	(4.243)	(2.868)	(2.591)	(0.479)
Highly ROA volatility	0.067**	0.046	0.041	0.015	0.004	-0.027	-0.032	-0.042
	(1.978)	(1.336)	(1.242)	(0.486)	(0.111)	(-0.799)	(-0.956)	(-1.213)
CLN score × Highly ROA volatility	0.074**	0.073*	0.041	0.080*	0.071	0.079**	0.050	0.060*
	(2.004)	(1.693)	(0.934)	(1.714)	(1.543)	(2.008)	(1.375)	(1.690)
Observations	11007	10671	10337	10013	9695	9387	9094	8808
Adjusted R^2	0.561	0.517	0.490	0.475	0.453	0.448	0.447	0.444
Panel D: Bank complexity								
CLN score	0.051***	0.053***	0.062***	0.052***	0.049***	0.037**	0.021	-0.012
	(3.455)	(3.807)	(3.304)	(2.967)	(3.139)	(2.366)	(1.084)	(-0.541)
High complexity dummy	-0.022	-0.011	0.000	0.016	0.020	0.037	0.031	0.019
	(-0.692)	(-0.315)	(0.005)	(0.446)	(0.544)	(0.998)	(0.786)	(0.454)
CLN score × High complexity dummy	0.078***	0.072***	0.057*	0.039	0.061**	0.047	0.071**	0.076*
	(3.170)	(2.668)	(1.854)	(1.328)	(2.135)	(1.582)	(2.133)	(1.964)
Observations	11308	10966	10631	10303	9980	9666	9365	9073
Adjusted R^2	0.555	0.511	0.483	0.471	0.450	0.443	0.442	0.442
Panel E: Capital adequacy								
CLN score	0.066***	0.070***	0.077***	0.070***	0.077***	0.060***	0.058***	0.030
	(5.822)	(5.778)	(5.355)	(4.488)	(4.610)	(3.518)	(3.326)	(1.507)
Low capital adequacy	-0.029	-0.015	0.006	0.039	0.025	0.031	0.031	0.033
	(-1.026)	(-0.610)	(0.226)	(1.346)	(0.902)	(1.078)	(1.061)	(0.974)
CLN score × Low capital adequacy	0.099***	0.078**	0.054	0.001	0.006	-0.007	-0.016	-0.027
	(3.296)	(2.537)	(1.489)	(0.040)	(0.170)	(-0.210)	(-0.453)	(-0.787)
Observations	11308	10966	10631	10303	9980	9666	9365	9073
Adjusted R^2	0.555	0.511	0.483	0.471	0.450	0.442	0.440	0.441
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.8: **Bank Co-Lending Network Risk and Bank Non-performing Loans**

Table 3.8 presents the h-quarter-ahead prediction results of bank-level *CLN score* for bank non-performing loans. The bank-level *CLN score* and control variables are measured at time t, and the dependent variable is measured at time $t + h$. Panel A reports the results using all observations. Panel B shows the results using unlabelled banks (banks who do not have assigned labels). Panel C shows the results using private banks only. Panel B and C have the same control variables as Panel A. For simplicity, we do not report the coefficients estimates of the control variables in the Panel B and C. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Non-peforming Loans	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel A: All banks								
CLN score	0.216***	0.228***	0.221***	0.222***	0.229***	0.213***	0.211***	0.200***
	(3.695)	(3.991)	(3.883)	(4.078)	(3.986)	(3.776)	(3.511)	(3.300)
Size	-0.010	-0.011	-0.010	-0.012	-0.012	-0.010	-0.007	-0.003
	(-0.308)	(-0.311)	(-0.274)	(-0.315)	(-0.296)	(-0.244)	(-0.184)	(-0.078)
Equity capital	0.434	0.199	0.116	-0.298	-0.625	-0.727	-0.799	-0.820
	(0.205)	(0.093)	(0.055)	(-0.134)	(-0.264)	(-0.312)	(-0.352)	(-0.378)
Deposits	0.160	0.147	0.107	0.049	0.041	0.013	-0.001	0.006
	(0.395)	(0.351)	(0.249)	(0.111)	(0.088)	(0.027)	(-0.001)	(0.012)
ROA	-0.447***	-0.408***	-0.343***	-0.253***	-0.180**	-0.155*	-0.109	-0.076
	(-3.164)	(-3.168)	(-3.117)	(-2.745)	(-2.007)	(-1.873)	(-1.322)	(-1.011)
Loan size	-0.743	-0.698	-0.561	-0.434	-0.326	-0.171	-0.052	0.093
	(-1.534)	(-1.359)	(-1.036)	(-0.743)	(-0.552)	(-0.296)	(-0.091)	(0.167)
Loan growth	-0.484**	0.001	-0.295	-0.176	-0.132	0.039	0.254	0.317
	(-2.001)	(0.405)	(-1.571)	(-0.949)	(-0.661)	(0.185)	(1.049)	(1.251)
Loan loss allowance	1.297***	1.263***	1.186***	1.093***	0.990***	0.879***	0.788***	0.675***
	(5.398)	(5.050)	(4.715)	(4.200)	(3.986)	(3.828)	(3.636)	(3.335)
Liquidity	-1.297***	-1.387***	-1.443***	-1.523***	-1.581***	-1.598***	-1.614***	-1.612***
	(-3.351)	(-3.433)	(-3.424)	(-3.416)	(-3.356)	(-3.235)	(-3.125)	(-2.976)
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10279	10036	9797	9504	9270	9039	8812	8585
Adjusted R^2	0.431	0.395	0.358	0.322	0.299	0.283	0.272	0.263
Panel B: Out-of-sample (unlabeled banks)								
CLN score	0.184***	0.180***	0.164**	0.182***	0.201***	0.179**	0.169**	0.156**
	(2.666)	(2.684)	(2.480)	(2.754)	(2.815)	(2.547)	(2.224)	(2.001)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6340	6191	6040	5862	5714	5559	5400	5250
Adjusted R^2	0.422	0.378	0.341	0.300	0.271	0.257	0.249	0.237
Panel C: Out-of-sample (private banks)								
CLN score	0.267**	0.276***	0.265**	0.314***	0.360***	0.317***	0.291**	0.289**
	(2.529)	(2.670)	(2.564)	(2.995)	(3.081)	(2.692)	(2.357)	(2.231)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3628	3520	3415	3301	3198	3093	2991	2889
Adjusted R^2	0.450	0.401	0.356	0.306	0.266	0.244	0.224	0.205

Table 3.9: **Bank Co-Lending Network Risk and Public Bank Risk Metrics**

Table 3.9 presents the h-quarter-ahead prediction results of bank-level *CLN score* for market-based public bank risk measures. We use stock market-based risk measures including Merton (1974) default probability, Nagel and Purnanandam (2020) modified default probability, the natural logarithm of idiosyncratic volatility, and use only the sample of public banks. The bank-level *CLN score* and control variables are measured at time t, and the dependent variable is measured at time $t+h$. In all specifications, we include the same set of controls as in the baseline and control for year-quarter fixed effects. For simplicity, we do not report the coefficient estimates of the control variables. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Dependent: Merton default probability								
CLN score	0.029***	0.034***	0.040***	0.038***	0.034***	0.030***	0.027***	0.025***
	(5.980)	(6.859)	(7.643)	(7.150)	(6.122)	(5.406)	(4.835)	(4.320)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6077	5902	5728	5553	5377	5206	5042	4879
Adjusted R^2	0.729	0.724	0.711	0.699	0.688	0.675	0.664	0.660
Dependent: Modified default probability								
CLN score	0.037***	0.037***	0.039***	0.036***	0.034***	0.031***	0.029***	0.025***
	(8.165)	(7.837)	(7.938)	(7.589)	(6.739)	(6.025)	(5.406)	(4.591)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6077	5902	5728	5553	5377	5206	5042	4879
Adjusted R^2	0.688	0.678	0.657	0.639	0.622	0.602	0.585	0.575
Dependent: ln(IVOL)								
CLN score	0.053***	0.060***	0.044***	0.046***	0.039***	0.025**	0.021*	0.016
	(4.677)	(5.209)	(3.816)	(3.953)	(3.228)	(2.091)	(1.690)	(1.226)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7976	7764	7554	7347	7138	6935	6740	6547
Adjusted R^2	0.685	0.684	0.679	0.679	0.678	0.675	0.674	0.672

Table 3.10: **Bank Co-Lending Network Risk and Bank Profitability**

Table 3.10 presents the h-quarter-ahead prediction results of bank-level *CLN score* for bank return on assets (ROA). The bank-level *CLN score* and control variables are measured at time t, and the dependent variable is measured at time $t+h$. Panel A reports the results using all observations. Panel B shows the results using unlabelled banks (banks who do not have assigned labels). Panel C shows the results using private banks only. Panel B and C have extactly same control variables as Panel A. For simplicity, we do not report the coefficients estimates of the control variables in the Panel B and C. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Return on assets	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel A: All banks								
CLN score	-0.265***	-0.237***	-0.218***	-0.199***	-0.195***	-0.172***	-0.147***	-0.120***
	(-10.864)	(-10.331)	(-9.276)	(-8.148)	(-8.123)	(-6.434)	(-6.326)	(-4.669)
Size	0.040**	0.036*	0.034*	0.034*	0.032	0.029	0.027	0.028
	(2.218)	(1.954)	(1.834)	(1.769)	(1.626)	(1.474)	(1.338)	(1.312)
Equity capital	4.349***	3.659***	3.415***	3.247***	3.274**	3.160**	3.054*	2.958*
	(5.086)	(3.942)	(3.214)	(2.669)	(2.395)	(2.167)	(1.947)	(1.869)
Deposits	0.020	0.031	0.072	0.112	0.132	0.152	0.175	0.221
	(0.106)	(0.155)	(0.354)	(0.529)	(0.602)	(0.664)	(0.747)	(0.940)
Loan size	0.822***	0.706***	0.603***	0.480**	0.350	0.280	0.197	0.123
	(3.580)	(3.156)	(2.684)	(2.096)	(1.517)	(1.189)	(0.860)	(0.562)
Loan growth	-0.001*	-0.000	0.046	0.221**	0.103	-0.054	-0.075	-0.411**
	(-1.732)	(-0.833)	(0.422)	(2.370)	(0.911)	(-0.539)	(-0.664)	(-2.147)
Loan loss allowance	-0.408***	-0.344***	-0.299***	-0.236***	-0.171**	-0.149**	-0.117*	-0.102*
	(-4.348)	(-3.953)	(-3.661)	(-3.168)	(-2.504)	(-2.211)	(-1.913)	(-1.804)
Liquidity	0.627***	0.589***	0.568***	0.568***	0.553**	0.502**	0.445**	0.381*
	(3.140)	(2.850)	(2.695)	(2.657)	(2.561)	(2.277)	(1.983)	(1.701)
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11310	10968	10633	10305	9982	9667	9365	9073
Adjusted R^2	0.310	0.277	0.262	0.247	0.240	0.235	0.232	0.226
Panel B: Unlabeled banks								
CLN score	-0.210***	-0.184***	-0.157***	-0.140***	-0.143***	-0.126***	-0.113***	-0.079**
	(-7.679)	(-6.142)	(-5.039)	(-4.294)	(-4.317)	(-4.038)	(-3.765)	(-2.156)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6909	6701	6497	6296	6100	5902	5702	5518
Adjusted R^2	0.299	0.252	0.245	0.227	0.218	0.221	0.226	0.231
CLN score	-0.262***	-0.237***	-0.216***	-0.182***	-0.204***	-0.160***	-0.146***	-0.079
	(-6.555)	(-5.464)	(-4.443)	(-3.372)	(-4.029)	(-3.430)	(-3.059)	(-1.324)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3884	3744	3609	3478	3350	3224	3103	2988
Adjusted R^2	0.274	0.243	0.234	0.226	0.222	0.212	0.202	0.193

Table 3.A1: **Variable Definition**

Variable	Definition	Source
Bank level variables		
CLN score	Bank-level Co-Lending Network (CLN) risk score estimated by the CoLGNN model using FR Y-9C bank features and DealScan loan features. The detailed estimation process is described in Section B.1.2 and Section 3.4.1.	CoLGNN
Loan loss provisions	The ratio of loan loss provisions (BHCK4230) to total assets (BHCK2170) in percentage points.	FR Y-9C
Nonperforming loans	The ratio of nonperforming loans (we use the sum of BHCK5525 and BHCK5526 for the period before 2017Q4 and the sum of BHCK1403 and BHCK1407 after 2017) to total assets (BHCK2170) in percentage points.	FR Y-9C
ROA	Banks' net income (BHCK4340) divided by bank total assets (BHCK2170)	FR Y-9C
Size	The natural logarithm of bank total assets (BHCK2170)	FR Y-9C
Equity capital	Total equity capital (BHCK3210) divided by bank total assets (BHCK2170).	FR Y-9C
Deposits	Total interest-bearing deposits (BHCK3517) devided by bank total assets (BHCK2170).	FR Y-9C
Loan size	Total loans (BHCK2122) divided by bank total assets (BHCK2170).	FR Y-9C
Loan growth	The annual percentage growth rate of total loans (BHCK2122).	FR Y-9C
Loan loss allowance	The allowance for loan and lease losses (BHCK3123) divided by bank total assets (BHCK2170).	FR Y-9C
Liquidity	The sum of cash (BHCK0010) and short-term securities (BHCK1773) divided by bank total assets (BHCK2170).	FR Y-9C
Standardized earnings change	The year-over-year quarterly change of earnings per share (BHCK4340 divided by BHCK3459) normalized by the rolling-window standard deviation of the earning change.	FR Y-9C
Bank capital adequacy	The ratio of loan loss provisions (BHCK4230) to total assets (BHCK2170) in percentage points.	FR Y-9C
Lead share	The total dollar amount of syndicated loans lead-arranged by a bank holding company normalized by the total amount of all syndicated loans in the past 12 months.	DealScan
Specialization in syndicated loan	The total amount of the syndicated loans a bank participates over the past 12 months divided by the bank's total loans (BHCK2122).	DealScan & FR Y-9C
Specialization in industry	The borrower's industry concentration of a bank's syndicated loan portfolio over the past 12 months, measured by the Herfindahl–Hirschman (HHI) index based on the borrowers' 2-digit SIC codes and the total loan amounts.	DealScan
Merton default probability	The default probability estimated by Merton (1974) model via a KMV iterative approach.	Nagel and Purnanandam (2020)
Modified default probability	Bank modified default probability using the data and code from Nagel and Purnanandam (2020).	Nagel and Purnanandam (2020)
ln(IVOL)	The natural logarithm of idiosyncratic volatility (IVOL) of bank stock returns, calculated by the standard deviation of the residuals from the Fama-French three-factor model estimated for each year-quarter.	CRSP & Kenneth R. French Data Library

Table 3.A2: **Banks with High and Low Co-Lending Network Risk**

Table 3.A2 presents the top and bottom six banks in our sample, categorized based on their co-lending network risks, as measured by the *CLN* score. This table specifically includes banks with a presence spanning more than an economic cycle, defined as 7 years or 28 quarters. Notably, two RSSDIDs became inactive after the end of our sample period: People's United Finance, Inc.(RSSDID: 3650152) and Umpqua Holdings Corporation(RSSDID: 2747644). People's United Finance was acquired and fully integrated by the third quarter of 2022. Umpqua Holdings Corporation was acquired by Columbia Banking System, Inc., with the merger concluding on March 1, 2023.

RSSDID	Bank	Headquarter State	Average CLN score
Top Seven			
3650152	People's United Finance, Inc. (Inactive)	Connecticut	0.786
2333663	Berkshire Hills Bancorp, Inc.	Massachusetts	0.767
2747644	Umpqua Holdings Corporation (Inactive)	Oregon	0.755
2132932	New York Community Bancorp, Inc.	New York	0.717
1562859	Ally Financial Inc.	Michigan	0.672
1098303	Old National Bancorp	Indiana	0.669
1078846	Synovus Financial Corp.	Georgia	0.653
Bottom Seven			
2461016	Enterprise Bancorp, Inc.	Massachusetts	0.265
1107205	Amarillo National Bancorp, Inc.	Texas	0.259
3635319	Servisfirst Bancshares, Inc.	Alabama	0.228
1208906	Lakeland Financial Corporation	Indiana	0.221
1399073	Heartland Banccorp	Ohio	0.211
1059715	American National Corporation	Nebraska	0.202
1862036	Guaranty Bancshares, Inc.	Texas	0.199

Table 3.A3: **Bank Co-Lending Network Risk and Bank Loan Loss Provisions: Controlling for Lending Specialization**

Table 3.A3 presents the h-quarter-ahead prediction results of bank-level *CLN score* for bank loan loss provisions. The bank-level *CLN score* and control variables are measured at time t, and the dependent variable is measured at time $t + h$. Panel A reports the results using all observations. Panel B shows the results using unlabeled banks (banks who do not have assigned labels). Panel C shows the results using private banks only. Panel B and C have extactly same control variables as Panel A. For simplicity, we do not report the coefficients estimates of the control variables in the Panel B and C. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Loan loss provisions	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel A: All banks								
CLN score	0.092***	0.091***	0.093***	0.074***	0.082***	0.062***	0.059***	0.028
	(7.439)	(7.316)	(6.895)	(5.453)	(5.486)	(4.135)	(3.735)	(1.635)
Size	0.029***	0.030***	0.030***	0.032***	0.030***	0.030**	0.031**	0.033***
	(3.223)	(3.185)	(2.994)	(3.018)	(2.652)	(2.509)	(2.575)	(2.712)
Equity capital	0.211	0.176	0.103	0.438	0.336	0.370	0.496	0.736
	(0.420)	(0.386)	(0.213)	(0.866)	(0.686)	(0.745)	(0.978)	(1.440)
Deposits	-0.041	-0.050	-0.082	-0.097	-0.125	-0.155	-0.144	-0.138
	(-0.369)	(-0.445)	(-0.705)	(-0.771)	(-0.946)	(-1.155)	(-1.081)	(-0.998)
ROA	-0.139***	-0.099***	-0.075***	-0.108***	-0.058**	-0.052*	-0.057	-0.073*
	(-4.689)	(-3.610)	(-2.925)	(-2.855)	(-2.288)	(-1.711)	(-1.483)	(-1.692)
Loan size	-0.482***	-0.373***	-0.260*	-0.109	-0.024	0.052	0.113	0.194
	(-3.646)	(-2.840)	(-1.866)	(-0.743)	(-0.161)	(0.337)	(0.722)	(1.218)
Loan growth	0.001	0.002**	-0.052	-0.081	-0.023	-0.011	0.121	0.226**
	(0.896)	(2.340)	(-0.817)	(-1.417)	(-0.410)	(-0.178)	(1.623)	(2.175)
Loan loss allowance	0.568***	0.520***	0.468***	0.393***	0.343***	0.310***	0.269***	0.226***
	(10.612)	(8.854)	(7.357)	(6.170)	(5.259)	(4.445)	(3.865)	(3.252)
Liquidity	-0.276*	-0.270*	-0.284**	-0.256*	-0.253*	-0.224	-0.204	-0.175
	(-1.891)	(-1.889)	(-1.968)	(-1.765)	(-1.816)	(-1.589)	(-1.398)	(-1.176)
Lead share	1.156	1.432	1.667	1.986*	2.574**	2.774**	2.948**	3.119**
	(1.225)	(1.428)	(1.562)	(1.704)	(2.030)	(2.116)	(2.142)	(2.191)
Specialization in syndicated loan	-0.337***	-0.353***	-0.376***	-0.470***	-6.393**	-5.685*	-5.209	-4.877
	(-12.054)	(-13.202)	(-12.586)	(-10.215)	(-1.997)	(-1.840)	(-1.636)	(-1.496)
Specialization in industry	0.014	-0.012	-0.017	-0.012	-0.012	-0.022	-0.036*	-0.033
	(0.787)	(-0.687)	(-0.898)	(-0.690)	(-0.674)	(-1.112)	(-1.705)	(-1.632)
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11308	10966	10631	10303	9980	9666	9365	9073
Adjusted R^2	0.569	0.521	0.490	0.474	0.452	0.445	0.443	0.444
Panel B: Urivate banks								
CLN score	0.067***	0.061***	0.058***	0.052***	0.062***	0.046**	0.042**	0.016
	(4.275)	(3.634)	(3.232)	(3.172)	(3.257)	(2.580)	(2.396)	(0.795)
Controls and Lending Specializations	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6907	6699	6495	6294	6098	5901	5702	5518
Adjusted R^2	0.534	0.495	0.468	0.453	0.427	0.433	0.433	0.439
Panel C: Out-of-sample (private banks)								
CLN score	0.077***	0.081***	0.099***	0.100***	0.115***	0.083***	0.077***	0.044
	(3.091)	(3.167)	(3.413)	(3.909)	(3.880)	(2.988)	(2.944)	(1.510)
Controls and Lending Specializations	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3882	3742	3607	3476	3348	3223	3103	2988
Adjusted R^2	0.526	0.489	0.460	0.444	0.419	0.417	0.412	0.413

Table 3.A4: **Controlling for Bank Fixed Effects**

Table 3.A4 presents the h-quarter-ahead prediction results of bank-level *CLN score* with bank fixed effects. The bank-level *CLN score* and control variables are measured at time t, and the dependent variable is measured at time $t + h$. Panel A reports the results using all observations. Panel B shows the results using unlabeled banks (banks who do not have assigned labels). Panel C shows the results using private banks only. For simplicity, we do not report the coefficients estimates of the control variables in the Panel B and C. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1) h=1	(2) h=2	(3) h=3	(4) h=4	(5) h=5	(6) h=6	(7) h=7	(8) h=8
Panel A: All banks								
Dependent variable: Loan loss provisions								
CLN score	0.074***	0.071***	0.073***	0.055***	0.066***	0.050***	0.050***	0.020
	(7.673)	(7.053)	(6.369)	(4.568)	(4.968)	(4.203)	(3.978)	(1.400)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11303	10963	10628	10297	9973	9658	9360	9065
Adjusted R^2	0.655	0.615	0.597	0.587	0.575	0.573	0.576	0.577
Dependent variable: Non-performing loans								
CLN score	0.125***	0.138***	0.131***	0.132***	0.139***	0.138***	0.139***	0.130***
	(4.211)	(4.554)	(4.205)	(3.915)	(3.605)	(3.607)	(3.704)	(3.367)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10263	10022	9782	9496	9262	9028	8806	8576
Adjusted R^2	0.747	0.731	0.714	0.701	0.695	0.689	0.690	0.692
Dependent variable: Return on assets								
CLN score	-0.177***	-0.153***	-0.137***	-0.123***	-0.125***	-0.104***	-0.089***	-0.061***
	(-7.911)	(-8.158)	(-7.822)	(-6.571)	(-6.383)	(-4.355)	(-4.771)	(-3.307)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11305	10965	10630	10300	9975	9658	9360	9065
Adjusted R^2	0.440	0.418	0.412	0.409	0.414	0.419	0.426	0.448
Panel B: Out-of-sample (unlabeled banks)								
Dependent variable: Loan loss provisions								
CLN score	0.067***	0.061***	0.058***	0.052***	0.062***	0.046**	0.042**	0.016
	(4.275)	(3.634)	(3.232)	(3.172)	(3.257)	(2.580)	(2.396)	(0.795)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6907	6699	6495	6294	6098	5901	5702	5518
Adjusted R^2	0.534	0.495	0.468	0.453	0.427	0.433	0.433	0.439

Table 3.A4: **Continued**

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Dependent variable: Non-performing loans								
CLN score	0.098***	0.104***	0.093**	0.111***	0.135***	0.131***	0.128***	0.109**
	(2.698)	(2.802)	(2.474)	(2.746)	(2.904)	(2.821)	(2.695)	(2.233)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6321	6176	6024	5854	5702	5548	5392	5239
Adjusted R^2	0.747	0.733	0.722	0.710	0.704	0.699	0.695	0.692
Dependent variable: Return on assets								
CLN score	-0.132***	-0.111***	-0.085***	-0.082***	-0.095***	-0.079***	-0.072***	-0.034
	(-5.802)	(-4.508)	(-3.599)	(-3.117)	(-3.261)	(-3.019)	(-3.171)	(-1.437)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6896	6691	6490	6288	6091	5894	5695	5507
Adjusted R^2	0.468	0.426	0.430	0.412	0.416	0.434	0.461	0.540
Panel C: Out-of-sample (private banks)								
Dependent variable: Loan loss provisions								
CLN score	0.069***	0.073***	0.093***	0.089***	0.114***	0.087***	0.073**	0.030
	(3.076)	(3.033)	(3.168)	(3.310)	(3.596)	(3.040)	(2.520)	(0.979)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3351	3225	3105	2988	2873	2763	2663	2559
Adjusted R^2	0.638	0.611	0.598	0.583	0.568	0.572	0.568	0.561
Dependent variable: Non-performing loans								
CLN score	0.138**	0.162***	0.164***	0.195***	0.233***	0.217***	0.205***	0.188**
	(2.598)	(2.996)	(3.025)	(3.393)	(3.392)	(3.027)	(2.679)	(2.357)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3617	3509	3405	3294	3189	3085	2988	2883
Adjusted R^2	0.765	0.751	0.740	0.727	0.716	0.705	0.704	0.703
Dependent variable: Return on assets								
CLN score	-0.171***	-0.154***	-0.140***	-0.116***	-0.146***	-0.112***	-0.119***	-0.047
	(-5.271)	(-4.500)	(-4.182)	(-2.968)	(-3.936)	(-3.743)	(-3.648)	(-1.198)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3877	3739	3605	3475	3344	3216	3101	2982
Adjusted R^2	0.465	0.455	0.451	0.457	0.469	0.481	0.492	0.552

Table 3.A5: **Bank Co-Lending Network Risk, Bank opacity and Bank Loan Loss Provisions**

Table 3.A5 examines the heterogeneous effects of bank co-lending network (CLN) risk score on predicting h-quarter-ahead bank loan loss provisions for different banks. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Loan loss provisions	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel A: All banks								
CLN score	0.526***	0.343***	0.259***	0.380***	0.275***	0.087	0.002	0.079
	(5.035)	(3.529)	(2.630)	(3.639)	(3.099)	(0.873)	(0.025)	(0.744)
Bank opacity	0.025**	0.026**	0.022*	0.023	0.024	0.031**	0.035**	0.039**
	(2.144)	(2.136)	(1.761)	(1.627)	(1.555)	(1.992)	(2.342)	(2.262)
CLN score × Bank opacity	0.066***	0.038***	0.025*	0.047***	0.030**	0.005	-0.008	0.009
	(4.618)	(2.821)	(1.753)	(3.194)	(2.279)	(0.322)	(-0.626)	(0.586)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9086	8909	8735	8552	8375	8202	8039	7879
Adjusted R^2	0.578	0.528	0.497	0.485	0.460	0.453	0.450	0.455
Panel B: Unlabeled banks								
CLN score	0.439***	0.219*	0.125	0.302**	0.249**	0.093	-0.029	0.084
	(3.346)	(1.771)	(1.053)	(2.181)	(2.483)	(0.768)	(-0.267)	(0.585)
Bank opacity	0.039***	0.037**	0.031*	0.042**	0.039**	0.051**	0.056***	0.058***
	(2.636)	(2.362)	(1.906)	(2.498)	(2.077)	(2.487)	(3.013)	(2.693)
CLN score × Bank opacity	0.058***	0.025	0.011	0.039**	0.030**	0.009	-0.010	0.011
	(3.269)	(1.460)	(0.654)	(2.025)	(2.029)	(0.486)	(-0.612)	(0.563)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5608	5498	5392	5282	5169	5052	4937	4829
Adjusted R^2	0.536	0.496	0.470	0.465	0.434	0.439	0.439	0.450
Panel C: Private banks								
CLN score	0.663***	0.404**	0.358**	0.783***	0.578***	0.374**	0.220	0.375*
	(3.393)	(2.299)	(2.207)	(4.134)	(4.217)	(2.361)	(1.601)	(1.886)
Bank opacity	0.058***	0.053**	0.032	0.053**	0.061**	0.075***	0.066**	0.081***
	(2.917)	(2.418)	(1.437)	(2.464)	(2.487)	(2.732)	(2.600)	(2.684)
CLN score × Bank opacity	0.092***	0.050**	0.040*	0.106***	0.073***	0.047**	0.023	0.052*
	(3.451)	(2.032)	(1.683)	(4.027)	(3.592)	(2.033)	(1.103)	(1.916)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3213	3128	3048	2967	2889	2810	2734	2659
Adjusted R^2	0.528	0.490	0.460	0.466	0.436	0.435	0.422	0.433

Table 3.A6: **Controlling for bank stock performance**

Table 3.A6 examines the predictive power of bank-level CLN score for future bank outcomes when controlling for quarterly stock returns. Panel A reports results for all public banks, while Panel B shows results for unlabeled public banks (those not used in model training). For each panel, we present results for loan loss provisions, non-performing loans, and return on assets across different prediction horizons (h=1 to h=8 quarters ahead). All specifications include the same control variables as in Table 3.4, though coefficient estimates for these controls are not reported for brevity. Definitions of variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel A: All public banks								
Dependent variable: Loan loss provisions								
CLN score	0.092***	0.077***	0.084***	0.062***	0.066***	0.042**	0.043**	0.009
	(6.412)	(6.570)	(6.143)	(4.056)	(3.893)	(2.580)	(2.430)	(0.446)
Quarterly stock returns	-0.048	-0.398***	-0.343***	-0.202**	-0.110	-0.326***	-0.242***	-0.265***
	(-0.591)	(-4.198)	(-3.956)	(-2.371)	(-1.430)	(-3.922)	(-3.821)	(-3.217)
Observations	7972	7753	7535	7323	7113	6909	6714	6525
Adjusted R^2	0.604	0.555	0.527	0.510	0.493	0.489	0.491	0.495
Dependent variable: Non-performing loans								
CLN score	0.223***	0.225***	0.219***	0.183***	0.168***	0.187***	0.203***	0.196***
	(3.750)	(3.707)	(3.573)	(3.342)	(3.067)	(3.378)	(3.326)	(3.234)
Quarterly stock returns	-0.396**	-0.468**	-0.597***	-0.648***	-0.688***	-0.463*	-0.320*	-0.191
	(-2.502)	(-2.307)	(-3.733)	(-2.839)	(-2.721)	(-1.897)	(-1.749)	(-1.044)
Observations	7183	7032	6880	6691	6547	6408	6270	6130
Adjusted R^2	0.414	0.393	0.368	0.349	0.338	0.326	0.321	0.323
Dependent variable: Return on assets								
CLN score	-0.102***	-0.116***	-0.144***	-0.125***	-0.137***	-0.120***	-0.104***	-0.081***
	(-7.492)	(-6.474)	(-7.719)	(-4.868)	(-5.106)	(-3.854)	(-5.185)	(-4.085)
Quarterly stock returns	0.016	0.440***	0.383***	0.398***	0.318***	0.394***	0.173**	0.237**
	(0.219)	(2.727)	(2.828)	(3.467)	(3.455)	(4.231)	(2.589)	(2.555)
Observations	7972	7753	7535	7323	7113	6909	6714	6525
Adjusted R^2	0.595	0.411	0.330	0.317	0.292	0.288	0.290	0.289
Panel B: Unlabeled public banks								
Dependent variable: Loan loss provisions								
CLN score	0.049***	0.042***	0.039**	0.023	0.018	0.018	0.019	-0.006
	(2.955)	(2.620)	(2.104)	(1.169)	(0.915)	(0.956)	(0.930)	(-0.253)
Quarterly stock returns	0.191*	-0.010	-0.036	0.124	0.231*	-0.118	-0.207*	-0.179
	(1.681)	(-0.057)	(-0.271)	(0.806)	(1.847)	(-1.114)	(-1.751)	(-1.512)
Observations	3579	3494	3407	3322	3239	3152	3059	2977
Adjusted R^2	0.585	0.540	0.523	0.504	0.490	0.504	0.507	0.526
Dependent variable: Non-performing loans								
CLN score	0.160*	0.146*	0.139*	0.113	0.108	0.127*	0.148*	0.133
	(1.922)	(1.896)	(1.797)	(1.583)	(1.584)	(1.950)	(1.898)	(1.642)
Quarterly stock returns	0.203	-0.024	0.199	0.202	0.301	0.357	0.301	0.406*
	(0.823)	(-0.062)	(0.684)	(0.806)	(1.329)	(1.513)	(1.269)	(1.708)
Observations	3249	3193	3129	3054	2996	2933	2864	2802
Adjusted R^2	0.327	0.319	0.322	0.314	0.308	0.307	0.313	0.317
Dependent variable: Return on assets								
CLN score	-0.084***	-0.078***	-0.064***	-0.050*	-0.057	-0.063*	-0.059**	-0.054**
	(-3.495)	(-4.094)	(-2.774)	(-1.774)	(-1.549)	(-1.782)	(-2.021)	(-2.025)
Quarterly stock returns	-0.067	0.016	-0.028	0.045	-0.099	0.004	-0.055	0.055
	(-0.616)	(0.091)	(-0.209)	(0.283)	(-0.784)	(0.035)	(-0.491)	(0.546)
Observations	3579	3494	3407	3322	3239	3152	3059	2977
Adjusted R^2	0.585	0.373	0.330	0.292	0.271	0.293	0.319	0.394
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.A7: **Controlling for stock performance: market-based risk measures**

Table 3.A7 examines the predictive power of bank-level CLN score for market-based risk measures when controlling for quarterly stock returns. Panel A reports results for all public banks, while Panel B shows results for unlabeled public banks (those not used in model training). For each panel, we present results for Merton (1974) default probability, Nagel and Purnanandam (2020) modified default probability, and stock return idiosyncratic volatility (ln(IVOL)) across different prediction horizons (h=1 to h=8 quarters ahead). All specifications include the same control variables as in Table 3.4, though coefficient estimates for these controls are not reported for brevity. Definitions of variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1) h=1	(2) h=2	(3) h=3	(4) h=4	(5) h=5	(6) h=6	(7) h=7	(8) h=8
Panel A: All public banks								
Dependent variable: Merton default probability								
CLN score	0.028***	0.032***	0.036***	0.034***	0.029***	0.026***	0.024***	0.021***
	(6.476)	(7.281)	(8.061)	(7.154)	(5.845)	(5.251)	(4.744)	(4.130)
Quarterly stock returns	-0.017	-0.037*	-0.045**	-0.070***	-0.080***	-0.078***	-0.060***	-0.062***
	(-1.061)	(-1.895)	(-2.128)	(-3.515)	(-4.423)	(-3.844)	(-3.227)	(-3.327)
Observations	6075	5898	5722	5545	5364	5185	5016	4850
Adjusted R^2	0.730	0.725	0.713	0.703	0.694	0.680	0.667	0.661
Dependent variable: Modified default probability								
CLN score	0.033***	0.034***	0.035***	0.032***	0.030***	0.028***	0.026***	0.023***
	(8.282)	(8.177)	(8.271)	(7.495)	(6.608)	(6.009)	(5.380)	(4.492)
Quarterly stock returns	-0.065***	-0.059***	-0.061***	-0.076***	-0.070***	-0.061***	-0.050***	-0.050***
	(-5.071)	(-3.580)	(-3.505)	(-5.217)	(-4.590)	(-3.519)	(-3.137)	(-3.163)
Observations	6075	5898	5722	5545	5364	5185	5016	4850
Adjusted R^2	0.691	0.681	0.661	0.645	0.629	0.608	0.589	0.578
Dependent variable: ln(IVOL)								
CLN score	0.043***	0.054***	0.039***	0.045***	0.037***	0.022**	0.019*	0.016
	(4.136)	(5.166)	(3.703)	(4.104)	(3.327)	(2.003)	(1.707)	(1.303)
Quarterly stock returns	-0.223***	-0.155***	-0.143***	-0.084*	-0.103*	-0.151***	-0.111***	-0.071**
	(-5.471)	(-3.666)	(-3.906)	(-1.845)	(-1.909)	(-3.667)	(-3.126)	(-2.139)
Observations	7956	7727	7502	7283	7066	6855	6654	6458
Adjusted R^2	0.689	0.687	0.683	0.682	0.681	0.679	0.678	0.677
Panel B: Unlabeled public banks								
Dependent variable: Merton default probability								
CLN score	0.013***	0.011***	0.014***	0.012***	0.009*	0.007	0.005	0.004
	(3.201)	(2.804)	(3.704)	(2.924)	(1.826)	(1.209)	(0.923)	(0.737)
Quarterly stock returns	0.083***	0.073***	0.091***	0.080***	0.036	0.025	0.029	0.022
	(4.354)	(3.397)	(4.020)	(3.542)	(1.568)	(1.153)	(1.350)	(1.173)
Observations	2491	2438	2382	2326	2257	2186	2109	2040
Adjusted R^2	0.745	0.753	0.745	0.731	0.709	0.701	0.688	0.691
Dependent variable: Modified default probability								
CLN score	0.009**	0.008**	0.012***	0.008**	0.009*	0.009*	0.009	0.005
	(2.515)	(2.235)	(2.888)	(1.980)	(1.972)	(1.790)	(1.561)	(0.933)
Quarterly stock returns	0.043***	0.043**	0.049***	0.036**	0.017	0.016	0.013	0.016
	(2.873)	(2.535)	(2.965)	(2.165)	(1.000)	(0.979)	(0.832)	(1.130)
Observations	2491	2438	2382	2326	2257	2186	2109	2040
Adjusted R^2	0.679	0.691	0.678	0.663	0.637	0.617	0.603	0.596
Dependent variable: ln(IVOL)								
CLN score	0.031**	0.039***	0.019	0.036**	0.021	0.012	-0.007	-0.002
	(1.993)	(2.718)	(1.301)	(2.595)	(1.510)	(0.847)	(-0.533)	(-0.110)
Quarterly stock returns	-0.193***	-0.145**	-0.075	-0.015	-0.004	-0.018	-0.008	-0.023
Observations	3576	3487	3399	3311	3226	3138	3041	2955
Adjusted R^2	0.688	0.688	0.674	0.679	0.671	0.675	0.680	0.670
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.A8: **Subperiod Analysis**

Table 3.A8 presents the h-quarter-ahead prediction results of bank-level *CLN score* on different sample Periods. We equally split the sample to two halves. The bank-level *CLN score* and control variables are measured at time t, and the dependent variable is measured at time $t + h$. Panel A reports the results using all observations. Panel B shows the results using unlabeled banks (banks who do not have assigned labels). Panel C shows the results using private banks only. For simplicity, we do not report the coefficients estimates of the control variables in the Panel B and C. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel A: All banks								
First-half Subperiod								
Dependent variable: Loan loss provisions								
CLN score	0.096***	0.108***	0.146***	0.106***	0.113***	0.090***	0.118***	0.064*
	(6.723)	(6.962)	(7.084)	(4.944)	(4.968)	(3.617)	(4.203)	(1.951)
Observations	5416	5308	5204	5101	4994	4891	4796	4706
Adjusted R^2	0.468	0.475	0.510	0.484	0.464	0.458	0.456	0.448
Dependent variable: Non-performing loans								
CLN score	0.260***	0.295***	0.292***	0.305***	0.338***	0.325***	0.312***	0.297***
	(5.899)	(5.915)	(5.019)	(4.120)	(4.111)	(3.482)	(3.091)	(2.859)
Observations	4387	4378	4370	4302	4284	4264	4243	4218
Adjusted R^2	0.481	0.455	0.415	0.387	0.366	0.347	0.331	0.317
Dependent variable: Return on assets								
CLN score	-0.216***	-0.212***	-0.250***	-0.221***	-0.226***	-0.211***	-0.207***	-0.156***
	(-7.385)	(-8.243)	(-7.061)	(-6.259)	(-6.549)	(-5.123)	(-5.345)	(-3.455)
Observations	5418	5310	5206	5103	4996	4892	4796	4706
Adjusted R^2	0.302	0.273	0.258	0.242	0.231	0.225	0.218	0.209
Second-half Subperiod								
Dependent variable: Loan loss provisions								
CLN score	0.153***	0.113***	0.059***	0.072***	0.059***	0.036**	-0.003	-0.002
	(7.299)	(5.792)	(3.100)	(3.835)	(3.136)	(2.171)	(-0.183)	(-0.124)
Observations	5892	5658	5427	5202	4986	4775	4569	4367
Adjusted R^2	0.578	0.527	0.429	0.416	0.396	0.363	0.268	0.273
Dependent variable: Non-performing loans								
CLN score	0.363***	0.318***	0.268***	0.212***	0.179**	0.145**	0.134*	0.113*
	(4.023)	(3.692)	(3.168)	(2.761)	(2.507)	(2.176)	(1.877)	(1.679)
Observations	5892	5658	5427	5202	4986	4775	4569	4367
Adjusted R^2	0.354	0.326	0.302	0.271	0.244	0.218	0.196	0.173
Dependent variable: Return on assets								
CLN score	-0.298***	-0.246***	-0.172***	-0.162***	-0.152***	-0.119***	-0.075***	-0.076***
	(-9.986)	(-9.072)	(-7.061)	(-6.238)	(-6.569)	(-5.228)	(-3.898)	(-4.010)
Observations	5892	5658	5427	5202	4986	4775	4569	4367
Adjusted R^2	0.321	0.296	0.314	0.298	0.298	0.303	0.329	0.327
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.A8: **Continued**

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel B: Unlabeled banks								
First-half Subsamples								
Dependent variable: Loan loss provisions								
CLN score	0.064***	0.065***	0.086***	0.059**	0.070**	0.053*	0.081**	0.027
	(3.683)	(3.553)	(3.138)	(2.374)	(2.339)	(1.688)	(2.421)	(0.672)
Observations	3302	3231	3164	3099	3030	2963	2891	2829
Adjusted R^2	0.436	0.462	0.488	0.466	0.436	0.452	0.451	0.453
Dependent variable: Non-performing loans								
CLN score	0.201***	0.206***	0.180***	0.188**	0.230**	0.224*	0.216	0.174
	(3.443)	(3.227)	(2.616)	(2.184)	(2.288)	(1.850)	(1.594)	(1.290)
Observations	2735	2723	2709	2667	2646	2621	2589	2561
Adjusted R^2	0.439	0.410	0.378	0.358	0.340	0.324	0.310	0.292
Dependent variable: Return on assets								
CLN score	-0.146***	-0.138***	-0.164***	-0.150***	-0.167***	-0.151***	-0.158***	-0.102
	(-5.600)	(-3.673)	(-3.481)	(-3.256)	(-3.179)	(-2.870)	(-3.021)	(-1.598)
Observations	3304	3233	3166	3101	3032	2964	2891	2829
Adjusted R^2	0.338	0.269	0.252	0.225	0.209	0.216	0.219	0.222
Second-half Subsamples								
Dependent variable: Loan loss provisions								
CLN score	0.116***	0.085***	0.041*	0.063***	0.054**	0.032	-0.007	-0.003
	(4.226)	(3.309)	(1.780)	(2.720)	(2.424)	(1.608)	(-0.404)	(-0.160)
Observations	3605	3468	3331	3195	3068	2938	2811	2689
Adjusted R^2	0.534	0.480	0.399	0.398	0.383	0.342	0.271	0.282
Dependent variable: Non-performing loans								
CLN score	0.382***	0.328***	0.280**	0.252**	0.215**	0.181*	0.162*	0.155
	(3.062)	(2.773)	(2.427)	(2.346)	(2.155)	(1.928)	(1.669)	(1.554)
Observations	3605	3468	3331	3195	3068	2938	2811	2689
Adjusted R^2	0.326	0.302	0.285	0.255	0.223	0.193	0.172	0.146
Dependent variable: Return on assets								
CLN score	-0.256***	-0.208***	-0.131***	-0.110***	-0.106***	-0.090***	-0.062**	-0.056**
	(-6.868)	(-6.176)	(-4.714)	(-3.771)	(-4.120)	(-3.462)	(-2.381)	(-2.125)
Observations	3605	3468	3331	3195	3068	2938	2811	2689
Adjusted R^2	0.273	0.260	0.296	0.296	0.297	0.299	0.313	0.309
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.A8: **Continued**

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel C: Private banks								
First-half Subsamples								
Dependent variable: Loan loss provisions								
CLN score	0.096***	0.109***	0.163***	0.115***	0.129***	0.086*	0.130***	0.045
	(3.264)	(3.648)	(3.699)	(2.988)	(2.869)	(1.815)	(2.780)	(0.712)
Observations	1852	1798	1746	1696	1647	1598	1552	1510
Adjusted R^2	0.443	0.459	0.481	0.469	0.444	0.453	0.447	0.437
Dependent variable: Non-performing loans								
CLN score	0.313***	0.349***	0.347***	0.365***	0.462***	0.418**	0.388*	0.361
	(3.530)	(3.527)	(3.178)	(2.719)	(2.829)	(2.083)	(1.703)	(1.537)
Observations	1598	1576	1554	1521	1497	1468	1440	1411
Adjusted R^2	0.469	0.438	0.395	0.370	0.340	0.315	0.291	0.266
Dependent variable: Return on assets								
CLN score	-0.209***	-0.208***	-0.254***	-0.197**	-0.232***	-0.193**	-0.206**	-0.097
	(-4.983)	(-3.849)	(-3.443)	(-2.512)	(-2.933)	(-2.344)	(-2.429)	(-0.922)
Observations	1854	1800	1748	1698	1649	1599	1552	1510
Adjusted R^2	0.315	0.280	0.263	0.249	0.236	0.228	0.206	0.182
Second-half Subsamples								
Dependent variable: Loan loss provisions								
CLN score	0.108**	0.085**	0.039	0.096**	0.086**	0.050	-0.009	0.004
	(2.503)	(2.195)	(1.054)	(2.444)	(2.235)	(1.420)	(-0.280)	(0.102)
Observations	2030	1944	1861	1780	1701	1625	1551	1478
Adjusted R^2	0.512	0.465	0.396	0.385	0.370	0.344	0.295	0.301
Dependent variable: Non-performing loans								
CLN score	0.526***	0.468**	0.409**	0.393**	0.337**	0.292*	0.254*	0.247*
	(2.679)	(2.452)	(2.149)	(2.157)	(1.996)	(1.820)	(1.712)	(1.681)
Observations	2030	1944	1861	1780	1701	1625	1551	1478
Adjusted R^2	0.328	0.299	0.277	0.247	0.209	0.173	0.141	0.115
Dependent variable: Return on assets								
CLN score	-0.296***	-0.240***	-0.157***	-0.148***	-0.165***	-0.118***	-0.080**	-0.057
	(-5.712)	(-5.003)	(-3.528)	(-2.991)	(-3.775)	(-2.998)	(-2.212)	(-1.567)
Observations	2030	1944	1861	1780	1701	1625	1551	1478
Adjusted R^2	0.261	0.245	0.269	0.259	0.269	0.273	0.296	0.303
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.A9: **Bank Co-Lending Network Risk and GFC**

Table 3.A9 presents the h-quarter-ahead prediction results of bank-level *CLN score* on GFC periods. The GFC indicator variable equals to 1 (0) if the year-quarter is within the mid of 2007 and the end of 2008. Panel A reports the results using all observations. Panel B shows the results using unlabelled banks (banks who do not have assigned labels). Panel C shows the results using private banks only. Panel B and C have extactly same control variables as Panel A. For simplicity, we do not report the coefficients estimates of the control variables. Definitions of the variables are provided in Table 2.A1 in the Appendix. Numbers in parentheses are two-tailed t-statistics. Standard errors are clustered at the bank level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable (t+h):	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Panel A: All banks								
Dependent variable: Loan loss provisions								
CLN score	0.109***	0.097***	0.079***	0.047***	0.045***	0.035**	0.027*	0.001
	(7.462)	(6.852)	(5.848)	(3.750)	(3.251)	(2.245)	(1.907)	(0.037)
CLN score × GFC	0.252***	0.236***	0.438***	0.744***	0.707***	0.544***	0.606***	0.593***
	(3.235)	(3.292)	(4.035)	(5.528)	(5.929)	(6.294)	(5.372)	(4.191)
Observations	11308	10966	10631	10303	9980	9666	9365	9073
Adjusted R^2	0.540	0.505	0.484	0.477	0.460	0.448	0.447	0.446
Dependent variable: Non-performing loans								
CLN score	0.289***	0.262***	0.222***	0.186***	0.157***	0.144***	0.142**	0.131**
	(4.522)	(4.282)	(3.713)	(3.435)	(3.073)	(2.803)	(2.515)	(2.326)
CLN score × GFC	0.718***	1.012***	1.187***	1.360***	1.625***	1.510***	1.354***	1.226***
	(3.353)	(4.138)	(4.362)	(4.422)	(4.310)	(4.402)	(4.062)	(3.824)
Observations	10279	10036	9797	9504	9270	9039	8812	8585
Adjusted R^2	0.396	0.372	0.346	0.323	0.309	0.292	0.280	0.270
Dependent variable: Return on assets								
CLN score	-0.231***	-0.207***	-0.173***	-0.139***	-0.143***	-0.136***	-0.103***	-0.070***
	(-10.977)	(-9.644)	(-8.278)	(-6.465)	(-6.786)	(-5.514)	(-5.663)	(-3.484)
CLN score × GFC	-0.558***	-0.475***	-0.710***	-0.931***	-0.787***	-0.544***	-0.673***	-0.745***
	(-3.329)	(-4.843)	(-4.368)	(-3.876)	(-4.113)	(-4.407)	(-4.106)	(-4.236)
Observations	11310	10968	10633	10305	9982	9667	9365	9073
Adjusted R^2	0.316	0.281	0.270	0.262	0.251	0.239	0.239	0.235
Panel B: Unlabeled banks								
Dependent variable: Loan loss provisions								
CLN score	0.076***	0.067***	0.050***	0.034**	0.033**	0.024	0.016	-0.014
	(4.182)	(3.636)	(2.994)	(2.280)	(1.988)	(1.273)	(1.006)	(-0.792)
CLN score × GFC	0.200*	0.123*	0.237**	0.487***	0.508***	0.359***	0.411***	0.472**
	(1.963)	(1.701)	(1.999)	(2.943)	(3.000)	(3.422)	(2.678)	(2.552)
Observations	6907	6699	6495	6294	6098	5901	5702	5518
Adjusted R^2	0.498	0.471	0.454	0.448	0.429	0.431	0.433	0.439
Dependent variable: Non-performing loans								
CLN score	0.284***	0.245***	0.203***	0.179**	0.151**	0.130*	0.115	0.100
	(3.319)	(3.027)	(2.624)	(2.471)	(2.188)	(1.904)	(1.594)	(1.360)
CLN score × GFC	0.568*	0.750**	0.803**	0.960**	1.240**	1.234***	1.194**	1.022**
	(1.878)	(2.348)	(2.373)	(2.504)	(2.564)	(2.753)	(2.517)	(2.273)
Observations	6340	6191	6040	5862	5714	5559	5400	5250
Adjusted R^2	0.361	0.335	0.313	0.291	0.274	0.260	0.253	0.241
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.A9: **Continued**

Dependent variable (t+h):	(1) h=1	(2) h=2	(3) h=3	(4) h=4	(5) h=5	(6) h=6	(7) h=7	(8) h=8
Dependent variable: Return on assets								
CLN score	-0.190***	-0.167***	-0.125***	-0.096***	-0.104***	-0.103***	-0.080***	-0.036
	(-7.425)	(-6.200)	(-4.832)	(-3.718)	(-4.024)	(-3.722)	(-3.487)	(-1.270)
CLN score × GFC	-0.314***	-0.269*	-0.499**	-0.683**	-0.600***	-0.360**	-0.498**	-0.642***
	(-3.548)	(-1.713)	(-2.384)	(-2.465)	(-2.648)	(-2.587)	(-2.370)	(-2.849)
Observations	6909	6701	6497	6296	6100	5902	5702	5518
Adjusted R^2	0.301	0.253	0.249	0.234	0.223	0.223	0.229	0.238
Panel C: Private banks								
Dependent variable: Loan loss provisions								
CLN score	0.077**	0.078***	0.065**	0.050**	0.052*	0.036	0.013	-0.022
	(2.568)	(2.715)	(2.365)	(1.980)	(1.937)	(1.185)	(0.528)	(-0.787)
CLN score × GFC	0.331**	0.253**	0.531***	0.899***	0.899***	0.593***	0.834***	0.830***
	(2.059)	(2.310)	(3.179)	(3.764)	(3.973)	(4.120)	(4.867)	(3.009)
Observations	3882	3742	3607	3476	3348	3223	3103	2988
Adjusted R^2	0.486	0.463	0.449	0.450	0.433	0.422	0.423	0.423
Dependent variable: Non-performing loans								
CLN score	0.387***	0.351***	0.298**	0.285**	0.259**	0.227**	0.198*	0.178
	(2.994)	(2.835)	(2.480)	(2.492)	(2.360)	(2.066)	(1.842)	(1.637)
CLN score × GFC	0.980*	1.340**	1.559***	1.746***	2.237***	2.070***	1.943**	1.939**
	(1.956)	(2.459)	(2.640)	(2.719)	(2.816)	(2.755)	(2.388)	(2.455)
Observations	3628	3520	3415	3301	3198	3093	2991	2889
Adjusted R^2	0.376	0.349	0.322	0.298	0.277	0.252	0.232	0.216
Dependent variable: Return on assets								
CLN score	-0.228***	-0.203***	-0.147***	-0.098***	-0.136***	-0.121***	-0.091***	-0.019
	(-6.126)	(-5.268)	(-3.786)	(-2.640)	(-3.921)	(-3.255)	(-2.718)	(-0.447)
CLN score × GFC	-0.551***	-0.534***	-1.084***	-1.277***	-1.020***	-0.599***	-0.816**	-0.898**
	(-3.661)	(-2.612)	(-3.650)	(-2.984)	(-2.873)	(-2.644)	(-2.517)	(-2.393)
Observations	3884	3744	3609	3478	3350	3224	3103	2988
Adjusted R^2	0.278	0.246	0.249	0.248	0.236	0.216	0.210	0.203
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Figure 3.A1: **CoLGNN Validation and Test Accuracy**

Figure 3.A1 presents shows the test accuracy and validation accuracy of our CoLGNN model.

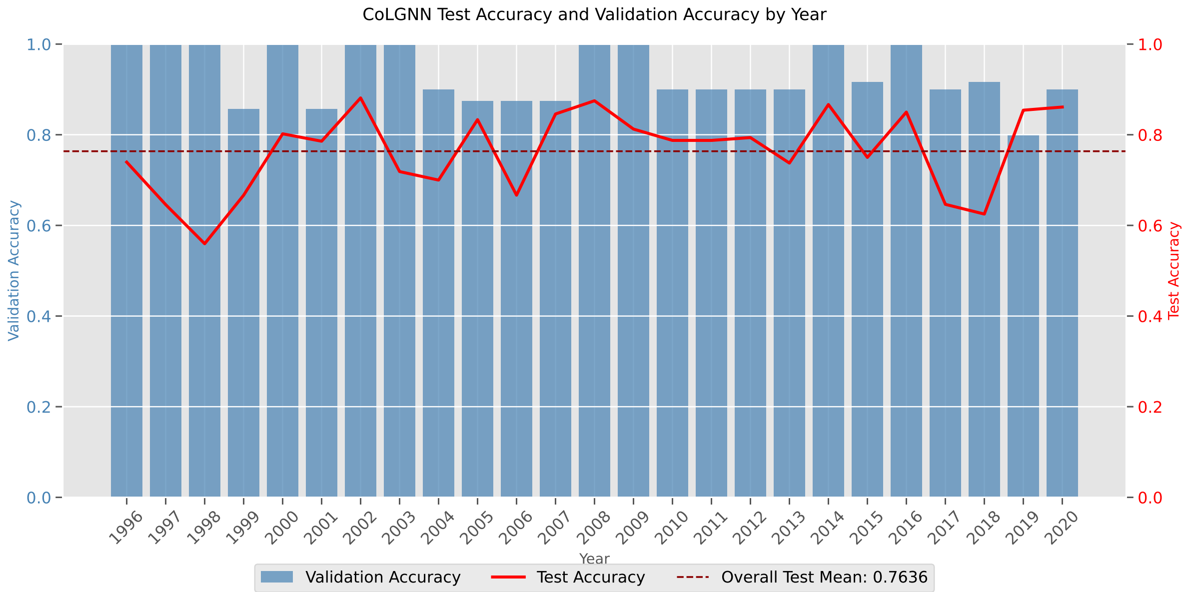
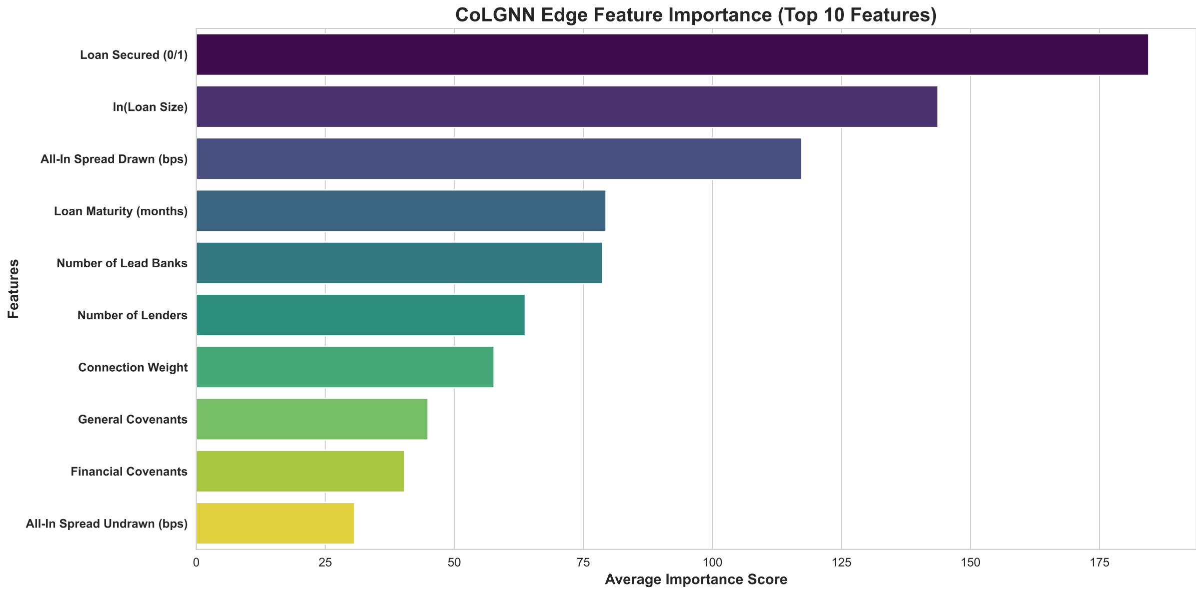
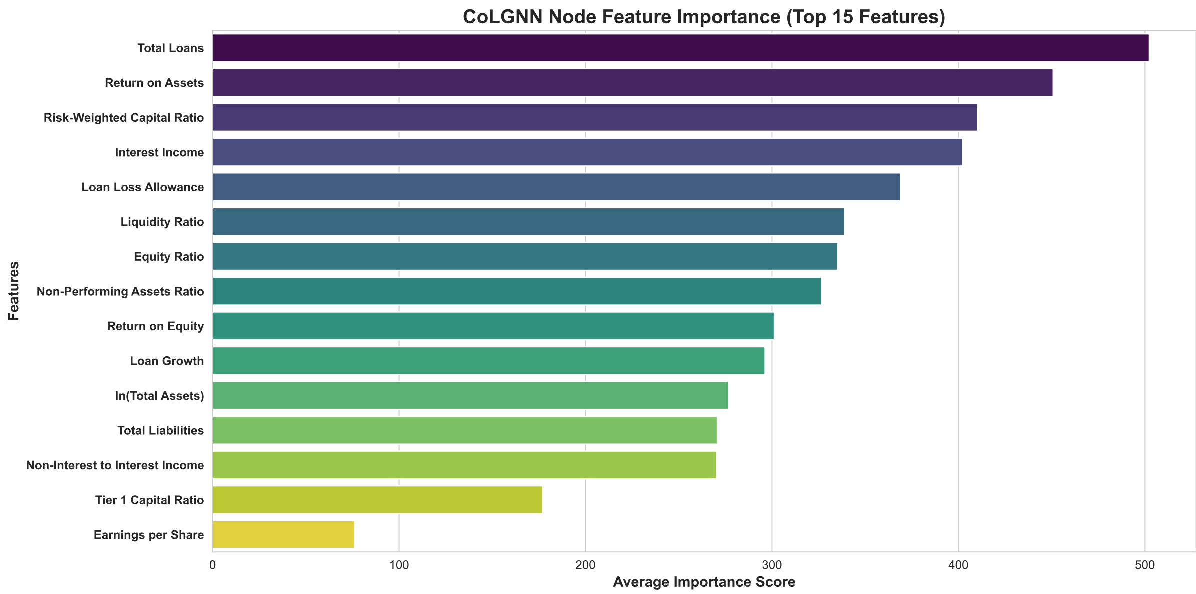


Figure 3.A2: **CoLGNN with Standard Bank Loan Control Variables**

Figure 3.A2 displays the importance scores of the top-ranked (a) edge features (loan characteristics) and (b) node features (bank characteristics) based on Integrated Gradients attributions for CoLGNN training using only standard control variables from the syndicated lending literature. This supplementary analysis focuses exclusively on the loan features that have been consistently used in prior banking literature: loan size, maturity, spread, covenant intensity, and other key elements. The importance scores are calculated and averaged similarly to Figure 3.4 using integrated gradient.



(a) CoLGNN Edge Feature Importance



(b) CoLGNN Node Feature Importance

Chapter 4

How Does Antitrust Enforcement Affect Corporations? Evidence from Government Contractors

"If we will not endure a King as a political power, we should not endure a King over the production, transportation, and sale of the necessaries of life."—Senator John Sherman (Sherman, 1890)

4.1 Introduction

Despite over a century of antitrust enforcement and its central role in maintaining competitive markets, empirical evidence on how these legal actions affect corporate performance remains surprisingly limited. Recently reignited interest in antitrust enforcement among policymakers, researchers, and media reflects growing concerns about market concentration (Baker, 2019; Bessen, 2016; Furman, 2016; Grullon et al., 2019; Shapiro, 2019), but studies have not reached consensus on whether antitrust actions benefit or harm affected markets and their participants.[1] This gap has significant implications for financial markets, as antitrust enforcement actions potentially trigger substantial wealth redistribution among stakeholders, affecting competitive dynamics, corporate valuations, and investment incentives in ways underexplored by market participants and policymakers.

Antitrust enforcement represents the legal and administrative mechanisms through which government authorities correct market failures arising from anti-competitive conduct. These enforcement actions aim to restore market efficiency by detecting and deterring practices that create deadweight losses, such as monopolies, market allocation, bid rigging, and other forms of

[1]While one strand of the literature argues that corporate market power can negatively impact employees, consumers, and productivity (Autor et al., 2020; Barkai, 2020; Gutiérrez & Philippon, 2017), another suggests that firms operating in less competitive markets may benefit from significant economies of scale. This allows a smaller number of large competitors to offer superior quality at lower prices, encourage innovation, and ultimately drive productivity growth (Bessen, 2020; Ganapati, 2018; Kang, 2025; Syverson, 2004; Van Reenen, 2018).

collusion that can harm markets by reducing competition (Tirole, 1988). Therefore, from a theoretical point of view, successful antitrust lawsuits are expected to reduce market entry barriers, resulting in an influx of new market participants, reduced prices, and increased total output (Babina et al., 2023). However, research also suggests that incumbent firms in affected product markets, whether defendants or non-defendants, experience a decline in profit margins—in some cases simply because enforcement actions remove a major market participant-which may have negative implications for their performance and financial stability (Aguzzoni et al., 2013; Besley et al., 2021; Cestone et al., 2021). With the additional concern that antitrust enforcement might impose significant costs—litigation expenses, regulatory uncertainty, disruption of network efficiencies—that could offset the intended benefits. This concern is often reflected in adverse financial market reactions to antitrust lawsuits (Bittlingmayer, 1993; Bittlingmayer & Hazlett, 2000; Wier, 1983).

We use antitrust lawsuits related to government procurement activities as an ideal economic laboratory. This setting is particularly compelling for several reasons. First, government procurement represents a substantial portion of economic activity, typically accounting for 10% to 15% of a country's GDP. The guaranteed government demand shuts down demand-side confounders, allowing us to isolate competitive supply-side dynamics. Second, procurement systems are highly susceptible to supplier collusion and corruption (Fazio & Zaldokas, 2024; Gallo et al., 1994; Goldman & Zeume, 2023), and a significant proportion of antitrust lawsuits—approximately 35%, according to our calculations—are related to government procurement activities.[2] Third, the formal debarment process creates a clean identification of market exclusion that would be unobservable in private markets. Last but not least, high-quality data on procurement contracts enables precise identification of these markets, their participants, and their structure.

We compile detailed information on all Department of Justice (DOJ) antitrust lawsuits using two complementary sources: Wolters Kluwer's VitalLaw platform and the official DOJ Antitrust Case Filings database. Our methodology leverages the capabilities of large language models with human verification to ensure comprehensive coverage and accuracy. By doing so, we identify lawsuits related to government procurement activities and extract data on the associated product markets and defendants. This approach also allows us to gather extensive information on the types of misconduct and other characteristics of antitrust lawsuits. Furthermore, we match this dataset with comprehensive information on government procurement contracts and

[2]Evidence of the importance for the government of ensuring fair competition in procurement markets is the creation in 2019 of the Procurement Collusion Strike Force (PCSF), an inter-agency partnership led by the U.S. Department of Justice's Antitrust Division. The stated goal of the agency is to protect taxpayer dollars from being misused due to collusion between companies seeking government contracts. To date, this agency has trained over 38,000 agents and procurement officials, opened more than 140 investigations, and recovered more than 65 million in fines and restitution.

detailed establishment-level data on government contractors, including their employment, sales, financial health metrics, and organizational characteristics.

We focus our analysis on the period from 2001 to 2021. We observe significant variation in antitrust lawsuits both over time and across different government procurement product markets, which provides the basis for our empirical analysis. Using granular establishment-level data on government contractors and procurement contracts, we apply a difference-in-differences research design (Babina et al., 2023; Kang, 2025; Sproul, 1993), and we observe an increase in government grants awarded to non-defendant government contractors' establishments operating in exposed product markets, relative to those in unaffected product markets. This effect is economically meaningful; our estimation results imply a 9% increase in government grants awarded to an establishment in our sample.

We also examine the impact of antitrust lawsuits on traditional measures of business outcomes. Specifically, we find that non-defendant establishments experience a substantial increase in the use of inputs, proxied by the number of employees, and in output, proxied by total sales. The evidence suggests that establishments expand their operations following antitrust enforcement actions, hiring additional workers and generating higher revenue streams. Importantly, we do not find evidence that this expansion adversely affects their financial health, as the increase in input use is proportional to the increase in output.

This stream of current and anticipated revenues positively affects the market valuation of non-defendant, publicly traded government contractors in our sample. This effect is economically significant and corresponds to an increase in shareholder wealth of $8.45 million for the average firm in this subsample. We rationalize this finding by considering the efficient market hypothesis (Fama, 1970), which suggests that stock prices reflect all available information accessible to investors. In anticipation of increased cash flows-documented in our analysis through the rise in government contracts awarded-investors raise their valuations of non-defendant firms.

A potential concern with our analysis is that we investigate lawsuits brought to court by the DOJ Antitrust Division, meaning that the treatment is not random. If illegal conduct emerges in product markets with unique trends that make them different from other markets, our approach does not allow us to establish a causal relationship. However, we use a day-by-day event study analysis to support the hypothesis that lawsuits are unexpected, as the market did not fully anticipate the drop in cash flow. We also corroborate the validity of our difference-in-differences analysis by investigating the dynamic effects of our outcome variables of interest and applying a robust inference test (Rambachan & Roth, 2023), providing empirical evidence that the parallel trend assumption is likely to hold. In addition, we show that our main findings are robust to broader product market and industry dynamics by controlling for aggregate industry trends and including linear and quadratic time trends at the product market level. Our results also remain robust when we refine the selection of the control group using a propensity score matching.

Furthermore, we propose a placebo test and find no evidence that establishments not engaged in government procurement activities but located in the same county and operating within the same product market are affected by antitrust lawsuits.

Additional results further support the hypothesis that the exclusion of colluding firms from procurement product markets is the key to explaining our main finding. We show that our results are driven by antitrust lawsuits in which corruption, fraud, and bribery have been identified as tools to illicitly gain market power. A distinctive feature of our setting is that, especially for this type of misconduct, defendant firms in procurement markets are subject to bans from bidding on or receiving government contracts (see the discussion in Karpoff et al. (1999)). This result suggests that antitrust lawsuits become particularly effective and beneficial for incumbent businesses when colluding firms are rationed out of the market.

To provide more direct evidence on the exclusion of colluding businesses from product markets, we identify a sample of establishments belonging to defendant firms and demonstrate that they experience a significant and substantial decline in government contract awards. These results further suggest that antitrust lawsuits caused a reallocation of sales across businesses and enhanced access to exposed product markets; simultaneously, our results clearly indicate that colluding establishments lose a substantial share of their market. We also provide evidence that antitrust lawsuits particularly benefit incumbent establishments in large procurement markets, where the market share reallocation is expected to be more significant.

At the product-market level, we document that antitrust lawsuits drive an influx of new participants into the affected markets. This increase in market participation is also reflected in a measurable decline in market concentration at the product market level. However, despite increased market participation, we do not find any evidence that antitrust lawsuits affect government acquisition costs. If anything, we document a small increase in the renegotiation of government contracts, which can drive government contract costs substantially higher (Bajari & Tadelis, 2001). This finding can potentially be explained by the winner's curse, where companies aggressively underbid in more competitive environments, only to realize ex post that they cannot fulfill the contract at the awarded price (Hong & Shum, 2002; Thaler, 1988). Furthermore, as discussed by Carril and Duggan (2020), the government's ability to leverage its monopsony power constrains suppliers' ability to set higher prices, even when they possess market power.

In the final section of the paper, we examine the establishments that benefit the most from antitrust enforcement and the characteristics of establishments participating in affected product markets. Antitrust lawsuits are generally expected to lower entry barriers, benefiting especially small firms and new entrants. However, antitrust lawsuits also generate increased competition, which requires establishments to make substantial investments to gain a larger share of reshaped product markets and thrive in this competitive environment.

Our findings reveal that there are no significant differences in the characteristics of the establishments taking part in the exposed product markets after an antitrust lawsuit. If any, our results suggest that they tend to be, on average, larger. At the same time, we observe that small and financially constrained businesses hold a relatively smaller market share in the reshaped, more competitive landscape. These results do not align with the Schumpeterian theory, where market dominance is expected to be disrupted by new, innovative entrants. Instead, the observed reshaping of market participants suggests a reallocation of dominance among established firms. In this sense, our findings align more closely with recent research by Faccio and McConnell (2020), emphasizing that replacing dominant firms often involves other existing large firms rather than new entrants.

4.2 Related literature and contribution

Our paper contributes to the literature investigating how antitrust enforcement affects corporate performance and industry dynamics.

Within this literature, some papers investigate the effects of antitrust enforcement on the equity prices of publicly traded firms under investigation, as well as other firms in the same industry, immediately following the release of indictment news. These studies find negative stock market reactions (Bittlingmayer, 1993; Bittlingmayer & Hazlett, 2000; Bizjak & Coles, 1995), and their findings have been widely cited to support the argument that antitrust lawsuits can create inefficiency and uncertainty, potentially harming corporations, consumers, and the economy as a whole.

Another strand of literature combines reduced-form and structural analysis to estimate the effects of antitrust sanctions on welfare, focusing on a single industry and documenting significant negative effects of collusion on customers' well-being (Barkley, 2023; Igami & Sugaya, 2022). Few studies, but likely more relevant to our paper, examine the impact of antitrust lawsuits across multiple industries and events, reporting mixed results. For example, Babina et al. (2023) examine DOJ antitrust lawsuits and find that antitrust enforcement actions in the U.S. permanently increase employment by 5.4% and business formation by 4.1% in non-tradable industries. In a similar vein, Reed et al. (2022) use data from Mexican cartel investigations to show that sanctions improve industry performance. Also investigating DOJ antitrust lawsuits, Kang (2025) examines how price-fixing cartels affect defendants' innovation, showing that collusion significantly enhances patent filings, R&D investment, and innovation breadth. These effects, primarily driven by financial gains and managerial expectations, vary across industries, and dissipate after collusion ends. Cestone et al. (2021), using the cases opened by the European Commission between 1991 and 2019, find that cartel investigations temporarily reduce profits for all firms in the affected industry and increase profits for their customers. In response to the negative shock to their profitability, firms engage in intense restructuring: they undertake mass layoffs and reduce employment; to lesser extents, they increase leverage, cut investment, and sell assets.

Our analysis complements these investigations, which have yielded contrasting conclusions on the impact of antitrust lawsuits on economic activities, by focusing on antitrust lawsuits related to government procurement activities in the U.S. and their corporate impact. This is important because, as we show in the paper, a large fraction of antitrust lawsuits are related to government procurement. Considering that government spending is a significant component of the country's GDP, the government has a strong interest in ensuring fair competition and the efficient allocation of resources in these markets. In addition, from an empirical point of view, this setting presents the advantage that these markets and their participants can be identified very clearly. Moreover, these antitrust lawsuits exhibit distinctive characteristics that deepen our understanding of how and when such legal actions positively impact businesses.

To the best of our knowledge, we are the first to identify the positive effects of antitrust lawsuits at the granular corporate level and to show the consequences of antitrust lawsuits for defendant firms. Furthermore, our research provides novel insights into the heterogeneous effects of antitrust enforcement, which depend on the size of the exposed product market, the ability of businesses to capitalize on these opportunities, and the effective exclusion of colluding firms from the market. Unfortunately, we do not find evidence that customers and taxpayers benefit from antitrust lawsuits, as we do not document any change in government acquisition costs. Furthermore, the benefits of these legal actions are not evenly distributed, as larger, well-established firms benefit more than smaller, financially constrained businesses.

Other papers investigate the impact of antitrust enforcement and competition policies on corporate and industry performance by leveraging regulation changes across different countries. Dasgupta and Žaldokas (2019) use country-level variation across amnesty programs to measure the effects of equilibrium changes in antitrust policy on investment and financing decisions. They find that firms step up investment and increase equity issuance as the equilibrium switches from collusion to oligopolistic competition. As a result, debt ratios fall. Buccirossi et al. (2013) use a country-level index of competition policy to show that countries with more robust competition policies have more significant productivity growth. Similarly, Faccio and Zingales (2022) provides evidence that rules promoting competition are associated with a lower concentration and lower prices. They find no evidence that pro-competition rules are associated with worse quality, lower investments, less employment, or lower wages. Besley et al. (2021) use a country-level antitrust enforcement index to show that firms operate with lower profit margins in countries with solid antitrust policies.

Our work also contributes to the literature investigating how corporate violations affect corporate performance and the performance of their peers. Within this literature, misconduct has been shown to impose significant negative consequences on firms. Such violations are met with legal and regulatory penalties (Karpoff et al., 2005), reputational damage (Armour et al., 2017; Murphy et al., 2009), and stakeholder conflicts (Barrage et al., 2020; Deng et al., 2014; Johnson et al., 2014; Ng & Rezaee, 2015). Contrary to our findings, however, corporate misconduct has been generally shown to have adverse spillover effects on other firms operating within the same sectors. Evidence has shown that peer firms, especially those that are potential targets for future regulations, experience an increase in litigation risk and reputational damage after an accusation (Donelson et al., 2022; Gande & Lewis, 2009; Gleason et al., 2008). On the other hand, in line with the competitive dynamics hypothesis arguing that peer misconduct could be beneficial for corporations because it creates opportunities for competitors to capture market share from weakened peers (Naumovska & Lavie, 2021), our results indicate that peer firms benefit from the misconduct of defendant firms and the breaking of anti-competitive barriers.

Finally, since many antitrust lawsuits involve corruption and bribery activities, our study contributes to the literature on how corruption and anti-corruption measures affect firms and economic activity. Shleifer and Vishny (1993) show that weak institutions increase corruption and that its illicit, secretive nature makes it more distortionary and costly than taxation, especially in less developed countries. Using cross-country data, Mauro (1995) finds that corruption reduces investment and economic growth. Christensen et al. (2024) document that stronger mid-2000s anti-corruption enforcement in Africa boosts local development while reducing perceived corruption and mitigating the political resource curse. Goldman and Zeume (2023) show that anti-corruption enforcement benefits unpunished firms in non-OECD industries, particularly those linked to foreign, productive, and less corruption-exposed groups, by increasing revenue, asset productivity, and profitability while lowering local corruption perceptions. Colonnelli et al. (2022) demonstrate that randomized anti-corruption audits in Brazil encourage firms to grow by adapting their investment strategies—increasing capital investment and borrowing—even as they lose government contracts, with effects varying by the level of corrupt involvement.

4.3 Antitrust lawsuits

4.3.1 Institutional background

The U.S. institutional framework for antitrust enforcement evolved to address distinct categories of market failure through three complementary mechanisms. The Sherman Act of 1890 established the DOJ's authority to prosecute anti-competitive conduct criminally, while the Federal Trade Commission Act of 1914 created the Federal Trade Commission (FTC) to pursue administrative remedies. Additionally, the Clayton Act of 1914 introduced private rights of action, enabling individuals and businesses to seek treble damages for harm caused by antitrust violations.

This tripartite structure aligns with economic theory on optimal deterrence. Criminal prosecution by the DOJ focuses on explicit collusion, where harm is evident and deterrence through penalties is most efficient. Administrative proceedings by the FTC address emerging competitive threats that require detailed economic analysis to assess their effects. Finally, private enforcement establishes market-based incentives for both detection and deterrence.

This institutional specialization has evolved to match enforcement technologies with the economic characteristics of different anti-competitive behaviors (Sawyer, 2019). Early enforcement under the Sherman Act primarily targeted obvious monopolization and price-fixing agreements. Modern enforcement addresses more sophisticated anti-competitive schemes, particularly in procurement markets where complex bid-rigging arrangements have replaced obvious forms of collusion. Recent innovations, such as the 2019 creation of the Procurement Collusion Strike Force, reflect the continued adaptation of enforcement mechanisms to emerging forms of anti-competitive conduct.

4.3.2 Collecting information on antitrust lawsuits

We draw information on antitrust lawsuits from two complementary sources: Wolters Kluwer's VitalLaw platform (which incorporated the Commerce Clearing House Trade Regulation Reporter, previously used by Babina et al. (2023) and Kang (2025)) and the official DOJ Antitrust Case Filings database. This dual-source approach enables extensive cross-validation and ensures complete coverage of enforcement actions.

VitalLaw serves as our primary source, providing detailed case summaries authored by legal experts specializing in antitrust law. Each summary contains standardized information about (1) case identifiers, including docket numbers and filing dates; (2) detailed descriptions of alleged violations; (3) geographic and temporal scope of violations; (4) market and industry characteristics; and (5) case outcomes including penalties and remedies. As argued and demonstrated by Babina et al. (2023), these summaries serve as an authoritative source. Unlike the DOJ's official website, which primarily contains recent cases, this database maintains comprehensive coverage across all periods. In addition, the summaries provide standardized information in a consistent format, facilitating systematic data collection.

While these alternative sources contain valuable information, extracting key metrics from individual antitrust lawsuits can be extraordinarily complex. These documents often use specialized terminology and draw upon dense statutory and economic frameworks, including detailed discussions of relevant markets, industry practices, and case law precedents. Such language frequently requires a substantial degree of legal and economic expertise to interpret reliably. While prior studies have tried human assistants to identify and code these data, this manual process is prone to inconsistencies, subjective biases, and substantial labor demands, factors that can compromise both the accuracy and scalability of traditional approaches. In the next section, we introduce a novel approach that addresses these challenges through the use of advanced language modeling techniques.

4.3.3 Leveraging LLM for high-fidelity legal analysis

Our data-collecting approach produces a complete and high-fidelity dataset by combining the analytical capability of the state-of-the-art large language models (LLM) with thorough human verification. We deploy OpenAI's GPT-4 Omni (gpt-4o) via a systematic API implementation built on a structured JSON schema and supported by robust error-handling protocols. The process begins with a primary classification phase that identifies procurement-related cases within our corpus. Following this initial classification, we use LLM to extract detailed information about violation types, defendant identities, and product market identifiers. This extraction follows predefined classification schemas and passes through multiple validation checkpoints to ensure accuracy.

Our approach is motivated by LLM's advanced capability of detecting subtle linguistic cues in complex legal documents. LLM identifies specialized legal phrasing, contextual references, and descriptive language that signal anti-competitive behaviors such as collusion or monopolistic practices—nuances that traditionally require extensive manual review by legal experts. Recognizing the limitations inherent in LLM, such as "hallucinations", we implemented a comprehensive human review protocol where we manually examine all outputs, assuring that our final dataset consistently reflects the actual information content of antitrust enforcement documents. By integrating advanced LLM with human verification, our methodology provides granular insights into procurement-related antitrust dynamics that would be difficult to capture through conventional manual coding or basic keyword-search approaches.

4.3.4 Antitrust lawsuits related to government procurement

We collect information on 3,438 cases from 1971 and identify antitrust lawsuits associated with government procurement activities.[3] We found that a large proportion of antitrust lawsuits

[3] Figure 4.A1 presents two complementary text visualizations: a frequency-weighted word cloud displaying key terms across all government procurement-related antitrust cases, and a specialized visualization highlighting procurement-specific terminology identified via LLM. The first visualization excludes common legal terminology (e.g., "defendant," "plaintiff," "court"), procedural language, temporal references, numerical expressions, generic business terms, common verbs/adjectives, and geographic designations to focus on case-specific content. The

belong to this category: 26% in our sample, which includes the period 2001-2021, as reported in Figure 4.1, or 35.46% for the whole sample of antitrust lawsuits available. We also observe significant heterogeneity in the timing of antitrust lawsuits that we exploit in our empirical analysis.

[Insert Figure 4.1 about here]

Another important aspect of our analysis is the identification of the procurement market affected by the antitrust lawsuits. Following Reed et al. (2022), we define specific product markets using the corresponding 6-digit North American Industry Classification System (NAICS) code, the most granular level of aggregation available. Our data collection follows a strict hierarchy: when cases are available on the Department of Justice's Antitrust Case Filings website,[4] we directly use the NAICS-6 classification officially assigned by the DOJ. For cases not available on the DOJ website, we use large language models to extract this information from VitalLaw summaries, followed by manual verification to ensure maximum fidelity to official designations while providing comprehensive coverage across our sample.

We present the distribution of antitrust lawsuits across industries (2-digit NAICS code) in Figure 4.2. The majority of events are concentrated in construction, manufacturing, and other service markets, which are the largest procurement sectors. Figure 4.A2 provides a more detailed breakdown at the product market level.

[Insert Figure 4.2 about here]

Another notable aspect of our approach is its ability to extract detailed information about the various facets of antitrust lawsuits, including the type of misconduct. Figure 4.3 highlights the diverse range of misconduct types associated with antitrust lawsuits, emphasizing the complexity of these cases. Among the violations, bid rigging and price fixing emerge as particularly notable, indicating efforts to manipulate competitive conditions. Bribery, wire fraud, and government fraud are also commonly used as tools to control market power, indicating substantial concerns about corruption and fraudulent activities in government procurement. This diversity underscores the multifaceted challenges in addressing and preventing anticompetitive behavior. Figure 4.A3 shows the spatial patterns in government procurement-related antitrust enforcement. Court venues are concentrated in major judicial centers like Washington DC, New York, and California, while defendant headquarters show greater dispersion across industrial and commercial hubs.

second uses a multi-role LLM to extract procurement-specific terms from each case (e.g., "bid rigging," "contract awards," "federal acquisition") that distinguish procurement-related violations from general antitrust cases.

[4] https://www.justice.gov/atr/antitrust-case-filings-alpha

As explained by Karpoff et al. (1999), a distinctive feature of our government procurement setting is that defendant firms are effectively excluded from procurement markets when antitrust lawsuits involve corruption, fraud, and other fraudulent activities. More specifically, these firms face temporary bans from bidding on or receiving government contracts during the investigation phase, as well as long-term bans (up to three years) from engaging in government contracts, which may apply to the entire company or specific divisions.

[Insert Figure 4.3 about here]

We show in Figure 4.4 the share of different settlement types and appeal statuses in our final sample. The majority (92.04%) of cases were resolved through plea agreements, followed by consent decrees (6.57%). Dismissed cases represent a smaller proportion (1.38%). Regarding appeal cases, most cases (95.56%) were not appealed, while 4.44% were appealed. These proportions are similar to the number that we observe for the whole sample of DOJ antitrust lawsuits.

[Insert Figure 4.4 about here]

We also examine whether antitrust lawsuits address absolute or relative monopolistic practices. This classification is interesting to extract from our source, as the Chicago School approach to antitrust emphasizes the enforcement against absolute monopolistic practices while advocating reduced enforcement against relative monopolistic practices (Lancieri et al., 2022). This perspective emphasizes price theory, industrial organization, and maximization of efficiency or consumer welfare, presenting these objectives as economically sound. At the same time, relative monopolistic practices are considered less significant, pro-competitive, and potentially beneficial to consumers. However, as reported in Figure 4.5, only a tiny share of antitrust lawsuits are related to absolute monopolistic practices. These statistics are also consistent when considering the entire sample of DOJ antitrust lawsuits.[5]

[Insert Figure 4.5 about here]

4.4 Data and Summary Statistics

After collecting and classifying information on antitrust lawsuits, we gathered additional data on government contractors, both at the establishment and corporate levels. Furthermore, we collected information on government procurement contracts to construct our final database and advance our empirical analysis.

[5] In Online Appendix C, we include five examples of antitrust lawsuits related to government procurement activities, along with other relevant information we have collected.

We use information from the National Establishment Time Series (NETS) database. This database contains information on the universe of establishments in the U.S., belonging to both private and public firms.

Using this database, we identify establishments listed as government contractors. This includes all establishments that have held government contracts or grants at any point during our period of analysis (Barrot & Nanda, 2020). In addition, this source provides information on the sales and employees. We also use the *PAYDEX* score, a business credit score issued by Dun & Bradstreet, to evaluate the impact of antitrust lawsuits on establishments' financial health and riskiness. This measure objectively assesses financial health by reflecting their payment behavior, and it is extensively used by creditors, suppliers, and financial institutions to evaluate credit risk. Furthermore, this score predicts future payment behavior, overall credit risk, and business failure (Chava et al., 2023). Finally, the NETS database also provides information on the location and industry of the establishments.

After identifying government contractors using the NETS database, we have been able to classify 174 establishments as defendants based on the extracted information from our antitrust lawsuit data.[6] Additionally, we perform fuzzy matching to align the headquarters' names of these establishments with corporate balance sheet data from Compustat and stock price information from CRSP. Through this process, we identified 196 public companies that are government contractors operating in product markets exposed to antitrust lawsuits, along with 9 defendant companies.

We gather comprehensive information on government contracts from *USAspending*. Detailed data on these contracts have been available since the Federal Funding Accountability and Transparency Act (FFATA) was enacted in 2006. This law was designed to enhance transparency in government spending, and the database includes records dating back to 2001 (Brogaard et al., 2021).

It is important to note that a single award may consist of multiple transactions, as some awards involve several payments or modifications over time. We collapse this information at the establishment and year level to obtain the total dollar amount of government contracts awarded to an establishment.[7] Additionally, this database provides detailed information on the characteristics of government procurement contracts, the relative procurement processes, and the winners. This enables us to conduct an empirical analysis to examine changes in market competitiveness, participant dynamics, and efficiency at the contract and product market levels.

[6] This analysis is limited by the fact that, in criminal antitrust cases, the DOJ Antitrust Division often prosecutes individuals rather than companies. Additionally, when confidentiality concerns are paramount, the DOJ may list individuals' names instead of companies' in its lawsuits.

[7] We merge this information with the NETS database, as both databases use the same establishment identifier, the DUNS number. Notably, the establishment identifier used for reporting establishment-level information in the procurement database changed after 2021, limiting our ability to extend the analysis beyond this period. Furthermore, the most recent and final release of the NETS database includes information up to the year 2022.

In our main analysis, we focus on establishments classified as government contractors (135,348 establishments), as indicated in the NETS database. After merging the antitrust lawsuit data with establishment information and excluding establishments belonging to defendant firms, our final dataset consists of 1,678,543 establishment-year observations spanning the period from 2001 to 2021. We report the summary statistics for this sample in Table 4.1. Government contractors in our sample have an average of approximately 72 employees and report average annual revenues of $16 million. On average, they receive $820 thousand in government obligations. Their average Paydex score is 72, out of a maximum possible value of 100, and 13% of the establishments operate in product markets affected by antitrust lawsuits.

[**Insert Table 4.1 about here**]

4.5 Empirical analysis

4.5.1 The determinants of antitrust lawsuits

Figure 4.1 shows the yearly variation in antitrust lawsuit intensity over time. We observe significant heterogeneity, with an increase in antitrust lawsuit activity during the financial crisis and at the beginning of the Obama administration. After a sharp decline, we observe increased use of antitrust lawsuits in the more recent years during the first Trump administration.

To better understand the dynamics of antitrust lawsuit activities, we begin our empirical analysis by examining their determinants. To do so, we estimate the following Equation:

$$Antitrust\ Lawsuit\ Event_{j,t} = \alpha_j + \gamma_t + \beta Industry\ Characteristics_{j,t} + \epsilon_{j,t} \tag{4.1}$$

Antitrust Lawsuit Event is an indicator variable equal to 1 if an antitrust lawsuit takes place in a product market j at time t. *Industry Characteristics* are alternative industry characteristics that we investigate. More specifically, we consider product market size, concentration (using the HHI index), and the number of participants.[8] α_j and θ_t are, respectively, product market and year fixed effects, that we use in alternative specifications.

We report estimation results in Table 4.2. Contrary to our expectations, we find that antitrust lawsuits are more common in less concentrated product markets, as well as in markets with a more significant number of competitors and greater overall size. More specifically, a 1% increase in the HHI reduces the likelihood of being treated by 0.005 percentage points, a 1% increase in

[8] We obtain this information by collapsing information from *USAspending* at the product-market and year level. We present the average values of these variables in the table containing the estimation results. A detailed description of each variable is available in the Online Appendix.

the number of firms increases it by 0.003 percentage points, and a 1% increase in government spending increases it by 0.002 percentage points. These findings align with the idea that anti-competitive and illegal behaviors are more likely where the benefits of such actions are higher.[9] However, when we control for industry and year-fixed effects, the coefficients are not statistically significant anymore, suggesting that the exact timing of antitrust lawsuits is exogenous.

[Insert Table 4.2 about here]

4.5.2 Access to the product market

We aim to investigate how antitrust lawsuits related to government activities affect non-defendant government contractors operating in the product market impacted by the lawsuits. In particular, antitrust lawsuits are expected to reduce barriers to market entry in affected product markets and lead to a reallocation of sales across establishments.

To explore this hypothesis in greater depth, we collect information on government procurement contracts. We aggregate the dollar amount of each transaction at the establishment-year level and merge this information with the NETS database using the DUNS number identifier available in both databases. Following previous literature (Babina et al., 2023; Kang, 2025; Sproul, 1993), we employ a difference-in-differences approach to compare affected government contractors with establishments operating in different, unaffected product markets before and after an antitrust lawsuit. In doing so, we exclude the establishments of identified defendant firms from the sample. More specifically, we estimate Equation (4.2):

$$IHS\ Government\ Contracts_{i,t} = \alpha_i + \gamma_t + \delta_{c,t} + \beta Antitrust\ Lawsuit_{i,j,t} + \epsilon_{i,j,t} \tag{4.2}$$

The outcome variable is the inverse hyperbolic sine transformation (IHS) of total government awards in an establishment i at time t. This transformation is appropriate due to the presence of many zeros and some negative values. In addition, it allows a similar interpretation of the regression results as the logarithmic transformation (Bahar & Rapoport, 2018; Carroll et al., 2003). In this specification, *Antitrust Lawsuit* is an indicator that takes a value equal to one after an establishment i that is operating in a product market j at time t is exposed to an antitrust lawsuit for the first time.[10] α_i and γ_t are respectively establishment and year fixed

[9] While it is well known that illicit conducts are significantly more frequent in environments with weak institutions, the theoretical effect of firm competition on the benefits that firms derive from illegal conducts is ambiguous (Cheung et al., 2021; Svensson, 2005).

[10] We identify the product market in which an establishment operates by using the NAICS-6 code that is reported when the establishment first appears in the NETS database. Our results do not change if we use the time-varying NAICS code reported in the database.

effects. Furthermore, in an additional specification, we control for county times year fixed effects ($\delta_{c,t}$), which enable us to account for time-invariant geographic characteristics.

We report the results in Table 4.3. We find a positive effect of antitrust lawsuits on the dollar amount of government contracts of non-defendant establishments. This effect is also economically meaningful; considering the results reported in Column (2), the coefficient implies an increase of 10% in federal obligations after an antitrust lawsuit concerning the average transformed outcome variable. Given that the average non-transformed government obligation is 820.66 thousand dollars, this translates to an increase of approximately 215–250 thousand dollars in government obligations per establishment. This confirms that the estimated effect is both statistically and economically significant.

[Insert Table 4.3 about here]

4.5.3 Establishments' performance, employment, and financial health.

To get a broader picture of how antitrust lawsuits affect corporations, we aim to estimate how these events affect the performance of these establishments. To do so, we estimate Equation (4.2) and use the natural logarithm transformations of employment and sales as alternative outcome variables. In addition, we also use the Paydex score, which measures the financial health of establishments.

We report estimation results in Table 4.4. We find a positive impact on sales and employees. More specifically, considering the results reported in Column (1), we find an increase of 3.9% in employment after an antitrust lawsuit concerning the average transformed outcome variable. On the other hand, considering the results reported in Column (2), we find an increase of 0.6% in sales after an antitrust lawsuit concerning the average transformed outcome variable. These results provide evidence that antitrust lawsuits benefit both corporations and employees in the affected product markets. More specifically, employment increases by 5.8 additional employees per establishment, while sales increase by $1.75 million per establishment.[11] These results suggest that corporations not only benefit from increased access to specific, exposed product markets but also experience overall growth.

Within this framework, we also consider prior literature suggesting that incumbent firms in affected product markets experience a decline in profit margins, potentially undermining their

[11] Note that the higher magnitudes in terms of establishment sales, compared to revenues attributed solely from government contracts, can be explained by the fact that: i. a large proportion of sales in the NETS database tends to be estimated (Barnatchez et al., 2017); and ii. government contracts include only federal government contracts, and the data before 2006 is incomplete, leading to measurement errors that can potentially bias the coefficients toward zero. However, it is also possible that higher government contracts enable establishments to expand into other product markets as well.

financial stability (Aguzzoni et al., 2013; Besley et al., 2021; Cestone et al., 2021). Accordingly, in Column (3), we examine the impact of antitrust lawsuits on the financial health of the establishments. However, we find a negative but statistically insignificant effect on the Payday score. This is likely due to the simultaneous increase in input use and sales.

[**Insert Table 4.4 about here**]

4.5.4 Identification

The validity of our empirical approach relies on the assumption that the treatment and control groups would have experienced the same trend in the outcome variable in the absence of treatment. This implies that any pre-treatment differences between the groups remain constant over time.

Our empirical strategy assumes the timing of the antitrust lawsuits in a product market is unexpected. This is supported by several features of antitrust enforcement: the DOJ typically conducts early investigations covertly to prevent evidence destruction and gather third-party information, creating uncertainty about if and when a lawsuit will occur. Even if firms suspect or know of an investigation, the likelihood and timing of litigation remain unclear.

To empirically investigate the validity of our identification strategy, we test whether there is any fundamental difference between our treatment and control groups. Table 4.A2 shows the normalized differences for both industry and establishment characteristics. Panel A reports characteristics at the industry level, while Panel B reports characteristics at the establishment level. In the two panels, ND is used to indicate normalized differences.

As reported in Imbens and Wooldridge (2009), two variables are considered similar if the normalized differences fall within the threshold of ± 0.25. Based on this criterion, we conclude that establishments in exposed and non-exposed product markets are identical regarding observable characteristics. However, the two markets are less similar. Specifically, as already shown in our previous results reported in Table 4.2, affected product markets are more competitive. Nonetheless, these differences should be accounted for by including time-invariant fixed effects.

To more explicitly test the validity of the parallel trend assumption, we consider the dynamic effects of our outcome variables. To do so, we use the estimator proposed by Callaway and Sant'Anna (2021). This approach also addresses the potential endogeneity of the difference-in-differences estimator when the treatment variable is staggered (Baker et al., 2022; De Chaisemartin & d'Haultfoeuille, 2020). Unlike traditional difference-in-differences methods that rely on a fixed reference year, this estimator compares treatment effects dynamically over multiple periods by leveraging units with the same initial treatment level as controls, allowing for non-binary and non-absorbing treatment processes.

[Insert Figure 4.6 about here]

Figure 4.6 provides evidence that our identified effects are long-lasting. There is a clear increase in employment, sales, and the amount of federal obligations awarded after the event; on the other hand, we do not find any effect on the Paydex score, which is consistent with our baseline results. It is important also to notice that the coefficients before the event are close to zero. However, the F-test on the joint statistical significance of the coefficients before the event rejects the hypothesis that the coefficients are jointly equal to zero, except for the coefficients related to the Paydex score.

4.5.5 A more credible approach to test the parallel trend assumption

Roth (2022) argues that coefficients close to zero but statistically significant might suggest that, while the difference in trends is small, the model is detecting consistent, systematic deviations from parallel trends due to a large sample size or low variability. In particular, if the pre-treatment coefficients are statistically significant but their magnitude is close to zero, it could reflect a practical, negligible deviation from parallel trends, particularly if the effect size is economically or substantively insignificant. Therefore, the significance of the coefficients alone cannot definitively confirm or refute the validity of the assumption of parallel trends. Instead, it should be evaluated in combination with additional sensitivity analyses.

For this reason, we implement the sensitivity test proposed by Rambachan and Roth (2023) to evaluate the robustness of our results to potential violations of the parallel trends assumption. This approach provides formal bounds on treatment effects under progressively relaxed constraints on the counterfactual trend's behavior. More specifically, our analysis follows two complementary approaches. First, the "relative magnitudes" method examines the sensitivity of period-specific point estimates to violations of parallel trends, testing how robust individual treatment effects are across different time periods. Second, the "average effects" method examines weighted averages of treatment effects throughout all post-treatment periods, which provides a more holistic assessment of overall impact while potentially improving precision. Using these two alternative approaches, we calculate the 95 % confidence intervals for our main estimators under different assumptions of the value M, the upper limit for the change between two consecutive periods in the slope of the underlying linear trend. A value of M equal to 0 on the x-axis corresponds with allowing for linear violations of parallel trends. Larger values of M allow for more significant deviations from linearity.

As shown in Figure 4.A4, our treatment effect estimates for employment and sales remain statistically significant across increasingly relaxed constraints on trend violations in both approaches. The confidence intervals widen as M increases, but the estimated effects remain positive and significant even at M=0.5. For government awards, while the confidence intervals are relatively wider—particularly in the period-specific analysis—the point estimates consistently remain positive across specifications, and the average effect shows greater stability. The stronger robustness in the average effects analysis suggests that our estimates of aggregate impact are less sensitive to potential non-linearities than period-specific estimates.

The consistency of these results, even under cautious assumptions about trend breaches, offers strong proof that our conclusions represent real treatment effects of antitrust enforcement rather than byproducts of parallel trend violations.

4.5.6 Stock market reactions to antitrust lawsuits.

To further validate our empirical approach and test whether antitrust lawsuits are unexpected to investors, we evaluate stock market returns for non-defendant firms around the filing dates. To do so, we identify affected firms by antitrust lawsuits as those operating within 6-digit NAICS codes subject to antitrust enforcement (excluding defendants).

As reported in Figure 4.7, we observe positive stock market reactions for this sample of 196 firms. More specifically, the figure presents the buy-and-hold abnormal return (BHAR) for these firms surrounding the days of antitrust enforcement filings. Before the enforcement event, BHAR remains relatively stable, fluctuating around zero. However, following the event, we observe positive abnormal returns, reaching approximately 1% by day 10. This translates to an increase in shareholder wealth of $8.45 million for the average firm in our sample (market capitalization of $845 million), with effects ranging from $1.96 million at the 25th percentile to $34.4 million at the 75th percentile of firm size. For our sample of 196 non-defendant firms, this represents a total wealth creation of approximately $1.66 billion.

[Insert Figure 4.7 about here]

These positive stock market reactions can be explained considering the efficient market hypothesis (Fama, 1970). More specifically, stock prices are expected to reflect all available information. Antitrust enforcement disrupts anti-competitive practices, thereby leveling the playing field for non-defendant firms. As market power shifts, non-defendant firms may gain access to previously restricted markets, secure higher sales, and achieve better contract terms. Investors, anticipating these positive outcomes, respond with higher valuations of non-defendant firms.

More formally, we can write that stock prices reflect the present value of expected future cash flows, as reported in the following Equation:

$$P_t = \sum_{t=1}^{\infty} \frac{E(CF_t)}{(1+r)^t}, \tag{4.3}$$

where P_t is the stock price at time t, $E(CF_t)$ is the expected cash flow at time t, and r is the discount rate, incorporating the risk-free rate and a risk premium.

Our empirical analysis is consistent with this framework; we show that establishments of non-defendant firms have increased access to market share, and government contracts raise $E(CF_t)$. At the same time, we did not document any change in business credit risk, proxied by the Paydex score.

Being more precise, we can also express the change in expected cash flows ($\Delta E(CF_{t,\text{After Antitrust}})$) as:

$$\Delta E(CF_{t,\text{After Antitrust}}) = \gamma \cdot \text{MarketSize} \cdot \Delta S, \tag{4.4}$$

where γ is the efficiency parameter capturing competitive firms' ability to capitalize on redistributed market share, ΔS is the market share lost by monopolistic firms, and Market Size is the total size of the affected product market. If one of these three factors is equal to 0, antitrust lawsuits are unlikely to bring any benefits to non-defendant businesses.

This simple framework allows us to develop the following set of hypotheses to explain the documented changes in cash flows and the benefits from antitrust lawsuits for non-defendant businesses:

- Establishments of defendant firms are excluded and lose a substantial share of their product markets.
- The benefits of non-defendant firms depend on the total size of the affected product market.
- The ability of non-defendant firms to gain from antitrust lawsuits depends on their capacity to capitalize on the redistributed market share.

4.6 Robustness checks

In this section, we provide additional results to demonstrate the robustness of our findings and the validity of our identification strategy.

4.6.1 Alternative methodological choices

First, a potential concern with our baseline difference-in-differences approach is that product markets exposed to antitrust lawsuits may differ systematically from unexposed markets. To address this selection concern, we combine our difference-in-differences approach with a propensity score matching based on product market characteristics. Specifically, we match treated and untreated markets using one-to-one and one-to-five nearest-neighbor matching algorithms based on market size, concentration, and competitive structure. After matching, normalized differences for our alternative samples confirm that both industries and establishments are similar in terms of observable characteristics. Table 4.A3 shows the estimation results using these matched samples. The coefficients remain positive and statistically significant, though slightly smaller in magnitude than our baseline estimates, falling within one standard error of the original results. This confirms that our findings are not driven by selection based on observable market characteristics.

In our main specification, we cluster the standard errors at the product market level, which is the level at which the treatment is assigned (Abadie et al., 2023; Bertrand et al., 2004). In this section, we demonstrate that our results hold when we consider alternative approaches. Specifically, our results remain robust when we adjust the standard errors for correlation at the following levels: product market and year, establishment, establishment and year, county, and county and year. The results are presented in Table 4.A4.

In addition, we consider alternative ways to measure government contracts. In our main results, we use the inverse hyperbolic sine (IHS) transformation of total government awards as the outcome variable, as this approach is suitable for handling the presence of many zeros and some negative values. We show that our findings remain robust when we use the level of the outcome variable or when we consider the average award values over two or three years, as well as the relative IHS transformation. In fact, since government contracts often span multiple years, using the average award value provides a more stable measure that smooths out year-to-year fluctuations. We report the estimation results in Table 4.A5.

4.6.2 Addressing potential confounding factors

We further test whether other product market trends explain our findings. Specifically, we first explicitly control for time-variant product market characteristics.[12] Next, we include industry (2-digit NAICS code)-by-year fixed effects to control for broader industry trends. Finally, we account for linear and quadratic time trends in the product market. We report consistent results in Table 4.A6. These results further indicate that broader industry or product market trends do not explain our findings.

Antitrust enforcement in one product market might affect related markets through competitive spillovers or resource reallocation. To investigate this possibility, we test whether establishments

[12]We control for the share of fixed-price contracts, the share of contracts awarded through competitive auctions, the average number of contract renegotiations, and the HHI index at the product procurement market level.

in adjacent markets—defined as operating in the same broader 4-digit NAICS sector but not in the specific 6-digit NAICS code exposed to antitrust lawsuits—experience indirect effects. We create an indicator variable, *Antitrust Lawsuit Spillover*, equal to one for establishments in these adjacent markets following an antitrust lawsuit (and zero otherwise). As shown in Table 4.A7, we do not find any statistically significant spillover effects, suggesting that the benefits of antitrust enforcement are concentrated within directly affected product markets rather than spread more broadly. This finding strengthens our identification strategy by confirming the specificity of the treatment effect to exposed markets.

4.6.3 Placebo test and triple difference-in-differences

First, we use establishments that are not government contractors but are located in the same county and operate in the same product market as government contractors as a placebo group. Using this substantially larger sample of 185,463,304 establishment-year observations, we estimate our baseline specification with employment, sales, and financial health as outcomes. As reported in Table 4.A8, we do not find any effect of antitrust lawsuits on the performance of this group of establishments.

As an additional robustness check, we implement a triple difference-in-differences framework using non-government contractors as a control group. The identification assumption for this estimator is that no other events in the product market are generating differential trends between government and non-government contractors that could affect their relative outcomes (Gruber, 1994). If this assumption holds, we can isolate the causal effect of government spending from other confounding factors, thereby clarifying its true impact on corporate outcomes. Notably, this triple difference-in-differences approach enables us to include product-by-year fixed effects, which control for time-varying product market characteristics. More specifically, we estimate:

$$Y_{i,t} = \alpha_i + \delta_{c,t} + \gamma_{j,t} + \beta(GovtContractor_i \times Antitrust\ Lawsuit_{j,t}) + \epsilon_{i,t} \tag{4.5}$$

where $Y_{i,t}$ is the outcome for establishment i at time t. α_i captures establishment fixed effects, $\delta_{c,t}$ controls for county-by-year effects, and $\gamma_{j,t}$ represents product market-by-year fixed effects. The main effects of $GovtContractor_i$ and $Antitrust\ Lawsuit_{j,t}$ are absorbed by the establishment and product market-by-year fixed effects, respectively. Our coefficient of interest, β, isolates the differential effect of antitrust lawsuits on government contractors.

Table 4.A9 presents these results. The coefficients on the triple interaction are positive and highly significant across all outcomes. Employment increases by 18.1% and sales by 30.7% for government contractors relative to non-contractors following antitrust enforcement. In addition, and in contrast to our baseline findings, the positive interaction coefficient we uncover when

using the Paydex score as the outcome variable suggests that contractors experience a 1.598-point improvement (a 2.2% increase relative to the mean) following an antitrust lawsuit.

4.7 Mechanisms

Our main results suggest that antitrust lawsuits halt anti-competitive practices and facilitate the entry and redistribution of sales across establishments within the exposed product market. In this section, we provide more direct evidence of these mechanisms.

4.7.1 Type of misconduct and exclusion effects

As reported in Equation (4.4), we expect that businesses in exposed product markets benefit from antitrust lawsuits when colluding businesses lose a share of the colluded markets.

To better understand how antitrust lawsuits affect corporations and product markets, we exploit the heterogeneity in their characteristics. More specifically, we identify antitrust lawsuits that encompass bribery, government fraud, mail fraud, wire fraud, tax evasion, money laundering, and obstruction of justice. We then estimate Equation (4.2) separately for this group of antitrust lawsuits and for the other group that is not linked to any of these misconducts. We report the two coefficients of interest in Figure 4.A5.

We find that our results are driven by antitrust lawsuits in which market power has been attempted to be controlled through fraud and corruption. For this type of misconduct, the law mandates debarment and exclusion from government contracting (Karpoff et al., 1999). In contrast, other anti-competitive practices, while serious, may result in corrective actions or less severe penalties. This result suggests that antitrust lawsuits become effective and benefit non-defendant businesses when defendants are effectively excluded from the markets.

4.7.2 Product market size effects

As reported in Equation (4.4), the benefits of antitrust lawsuits are expected to be more significant when the exposed product market is larger. To test this hypothesis, we divide the sample based on the median value of the size of the procurement market, proxied by the total spending, and re-estimate Equation (4.2).

We report the results in Table 4.A10. According to our hypothesis, we find that the coefficient is positive and statistically significant only for the sample of establishments operating in larger, exposed product markets. Given the substantial costs of antitrust enforcement for the government, our findings suggest that focusing investigations on firms operating in large product markets may be a more effective strategy to stimulate the economy.[13]

4.7.3 Market share reallocation

To provide more direct evidence that colluding firms are excluded from the markets, we identified 174 unique establishments that are government contractors and part of defendant firms. We use the same control group and compare their government award dynamics with those of establishments operating in product markets that have never been affected by antitrust lawsuits.

We next estimate the following Equation:

$$IHS\ Government\ Contracts_{i,t} = \alpha_i + \theta_{t,c} + \beta Antitrust\ Lawsuit_{i,t} + \epsilon_{i,t} \quad (4.6)$$

In this setting, *Antitrust Lawsuit* is an indicator variable equal to 1 after an antitrust lawsuit that affects the establishments of an affected firm. On the other hand, α_i and $\theta_{t,c}$ are establishment and county and year fixed effects, respectively.

We find a negative effect on the dollar amount of government contracts for these establishments, as reported in Table 4.5. More specifically, we find a statistically significant decrease of 35.34% in federal obligations for defendant establishments after an antitrust lawsuit concerning the average outcome variable. Thus, after an antitrust lawsuit, defendant firms lose an average of \$460–\$470 thousand dollars in government contracts.[14] Overall, these results provide supporting evidence that the exclusion of defendants facilitates the redistribution of sales across other establishments within the product market.[15]

[Insert Table 4.5 about here]

We also report the results from Equation (4.6), using the natural logarithm of employment and sales, as well as the paydex score, as outcome variables. The results are presented in Table 4.A11. We find a negative and statistically significant effect on employment. Regarding sales and the paydex score, the coefficients are negative and economically significant, but not

[13] Baker (2002) estimate that costs of antitrust enforcement, both for the government and private firms, are estimated at approximately \$1–2 billion annually.

[14] Notably, Heese and Pérez-Cavazos (2019) show that regulatory scrutiny does not always lead to contract reductions; they show that federal agencies can substitute cost-plus contracts with fixed-price contracts when firms face fraud allegations, though reductions occur after settlements.

[15] We also report the dynamic effects in Figure 4.A6. Although the low number of observations does not allow us to properly estimate these dynamics, the penalties appear to last for three years, as stipulated by the law (Karpoff et al., 1999).

statistically significant at the conventional 10% level. This could suggest that while defendant firms are penalized in exposed product markets, they may have expanded in other markets.

In a similar vein, stock market reactions around the days of the event for 9 identified defendants who are also public corporations are in line with the idea that investors anticipate negative cash flows and more significant risks for these corporations. Figure 4.A7 presents the buy-and-hold abnormal return (BHAR) for 9 identified defendant firms operating in the treated product market surrounding the days of antitrust enforcement filings. Following the antitrust lawsuit filings, BHAR exhibits a notable negative trend, suggesting a positive abnormal return, reaching approximately 2% by day 10. For the typical defendant company (market capitalization of \$9.54 billion), this unfavorable market response results in a significant decline in shareholder worth of \$191 million. This reflects a total wealth loss of around \$1.72 billion for our sample of defendant companies, in line with investors expecting major expenses from antitrust enforcement activities.

4.7.4 Product market competition

To further investigate whether antitrust lawsuits halt anti-competitive practices, we conduct an analysis at the product market level. Specifically, we examine whether the number of participants in exposed product markets increases following an antitrust lawsuit. Increased participation is also expected to decrease market concentration. Moreover, we examine whether changes in competition are linked to variations in the average contract value, a proxy for government acquisition costs.

To do so, we estimate the following Equation 4.7.

$$Outcome_{j,t} = \alpha_s + \gamma_t + \beta Antitrust\ Lawsuit_{j,t} + \epsilon_{j,t} \tag{4.7}$$

In this setting, *Outcome* is the number of participants in a product market, a measure of market concentration, the HHI index, the average contract value, and the protest rate.[16] *Antitrust lawsuits* is an indicator variable equal to 1 after the first antitrust lawsuits in a product market, and 0 otherwise. α_s and γ_t are product market and year-fixed effects, respectively.

The results are presented in Table 4.6. We find that antitrust lawsuits result in an increased number of market participants and a reduction in market concentration. Specifically, the number of participants increases by 7% relative to the average transformed value. This suggests that an antitrust lawsuit leads to an increase of approximately 16 additional establishments per product market, representing a substantial rise in market participation. At the same time, we document

[16]We constructed these variables using data from *USASpending*, except for the protest rate, which is derived from bid protest data from Canayaz et al. (2025). The estimation results table presents the average values of these variables, and a detailed description of each variable is provided in the Online Appendix.

a significant 2% decrease in the natural logarithm of the HHI index relative to its average transformed value. This coefficient implies that an antitrust lawsuit reduces the HHI index from approximately 2,027 to 1,750, indicating a meaningful decline in market concentration.

These results are economically meaningful and provide evidence that antitrust lawsuits lower entry barriers, benefit new entrants, and reduce market concentration. However, interestingly, as reported in the third column, we do not find evidence that the average contract values change, suggesting that antitrust lawsuits are not affecting government procurement costs. We investigate and discuss more on this point further in the next session of the paper.

Finally, we use data on bid protests from Canayaz et al. (2025) to investigate whether antitrust lawsuits effectively impact market fairness—or at least how market fairness is perceived by participants. This information is available for the spanning period from 2005 to 2016. Using the number of bid protests adjusted for the number of unique firms participating in a product market as an outcome variable, we identify a negative relationship between antitrust enforcement and the protest rate. More specifically, the coefficient indicates a 0.121 percentage point decrease in the protest rate following antitrust lawsuits, implying that enforcement actions may improve market transparency and reduce the need for firms to challenge procurement decisions. This effect represents a substantial 30.1% reduction relative to the sample average protest rate of 0.402%. However, despite this economically significant effect, it is not statistically significant at the conventional 90% level.

[**Insert Table 4.6 about here**]

4.8 Market participants and their characteristics

Antitrust lawsuits are generally expected to lower entry barriers and benefit small firms and new entrants (Babina et al., 2023). However, as we demonstrate in this paper, increased competition requires greater resources, which small firms may lack. Additionally, it is important to consider that in concentrated markets, significant economies of scale may exist, allowing a smaller number of large competitors to deliver products with both higher quality at lower prices (Bessen, 2020; Ganapati, 2018; Kang, 2025; Syverson, 2004; Van Reenen, 2018).

In this section, we examine how antitrust lawsuits impact the characteristics of market participants and other characteristics of government procurement. To do so, we merge our establishment-level database with government contracts. After identifying the main contract based on the earliest transaction and removing contracts with a value of zero or negative, our final database consists of 22,458,175 contracts.

4.8.1 Characteristics of market winners

After providing evidence of increased market participation and decreased market concentration, we examine the characteristics of market participants in reshaped product markets. Using our contract-level database, we estimate the following Equation:

$$\text{Winner Characteristic}_{c,i,j,t} = \eta_s + \theta_{tc} + \beta \text{Antitrust Lawsuit}_{j,t} + \epsilon_{c,i,j,t} \tag{4.8}$$

Winner Characteristic indicates the characteristics of the winner of a contract c to an establishment i that win award the contract in a product market s at time t. *Antitrust Lawsuit* is an indicator variable that takes a value equal to after the first antitrust lawsuit in a specific product market. When there has been more than one antitrust lawsuit in a product market, we consider the first event. η_s and θ_{tc} are product market and county-year fixed effects. Finally, we weigh each contract by (the natural logarithm of its) respective contract value.

We present our results in Table 4.7. As reported in Columns (1) and (2), we document that contract-winning establishments in exposed product markets are, on average, larger. The two coefficients are, indeed, large, positive, and statistically significant. In Column (3), the coefficient for the financial indicator (Paydex) suggests a large 27.6% improvement in the financial strength of the average establishment taking part in this market, although this estimate is not statistically significant at the 10% conventional level.

In Column (4), the coefficient of 0.014 indicates a small, statistically insignificant increase in the likelihood of being a public corporation. Column (5) reports a 0.133 increase in the age of winning establishments, yet this effect is not statistically significant. Likewise, Column (6) shows a -0.015 coefficient, implying a 1.5 percentage point reduction in the probability of being classified as a small firm, and Column (7) shows a -0.002 coefficient for minority ownership, both of which are not statistically significant.

Overall, these results suggest that the characteristics of businesses participating in the product market do not change significantly after an antitrust lawsuit. If anything, we document that establishments in affected markets tend to be, on average, larger and less financially constrained. This outcome may be explained by the fact that the ability of non-defendant firms to benefit from antitrust lawsuits depends on their capacity to capitalize on the redistributed market share, as shown in Equation (4.4). Furthermore, as we demonstrate in our paper, firms must make significant investments to enter and secure a share of the product market, which may be particularly challenging for small and financially constrained firms. These results partially contradict Schumpeterian theory, which posits that the market power of dominant firms is eroded over time by new, innovative entrants. Instead, our findings align more closely with recent research by Faccio and McConnell (2020), which emphasizes that replacing dominant firms often involves other existing firms rather than new entrants.

[Insert Table 4.7 about here]

4.8.2 Government acquisition costs

Antitrust lawsuits are widely regarded as a tool to increase competition, leading to a decrease in prices that benefits consumers (Brot et al., 2024; Clark et al., 2018). However, the result in the last column of Table 4.6 suggests that the average value of procurement contracts within an exposed product market does not decrease after an antitrust lawsuit.

A limitation of our data is that we do not observe unit procurement costs directly. However, to more carefully evaluate this point, we follow previous literature and investigate whether antitrust lawsuits impact renegotiation activities and the use of cost-plus contracts, which are primary drivers of government overspending (Bajari & Tadelis, 2001; Crocker & Reynolds, 1993). Furthermore, we analyze whether it influences the likelihood of a contract being awarded through a competitive bidding process. We use this information as alternative outcome variables in Equation (4.8).

We report the results in Table 4.8. We find no evidence that antitrust lawsuits affect the likelihood that a contract is awarded through a competitive bidding process or whether the use of cost-plus contracts becomes less common. However, we do find an increase in the number of renegotiation activities of each contract. Since the mean number of renegotiations per contract is 1.303, so an increase of 0.095 represents approximately a 6% increase in the frequency of renegotiations. This finding can potentially be explained by the winner's curse, where firms aggressively underbid in highly competitive and uncertain procurement environments, only to realize ex-post that they cannot fulfill the contract at the awarded price (Hong & Shum, 2002; Thaler, 1988). As a result, these firms often seek renegotiation to adjust for unforeseen losses, likely leading to higher government contract costs. This aligns with the incentives framework in Bajari and Tadelis (2001), which highlights how ex-ante competitive pressure coupled with ex-post contractual flexibility fosters renegotiation dynamics.

[Insert Table 4.8 about here]

Overall, we find no evidence that antitrust lawsuits lead to lower prices. This result may be explained by the nature of procurement markets. As recently demonstrated by Carril and Duggan (2020), the government's ability to dictate contract terms and leverage its monopsony power constrains suppliers' ability to set higher prices, even when they possess market power. This suggests that while colluding firms may have used their market power to secure government contracts, the government's strong bargaining position limited their ability to raise prices. However, a potential limitation of this analysis is that we do not observe the quality of the goods and services provided, which may have been affected by increased competition.

4.9 Conclusion

A growing literature in economics and finance debates whether government intervention in product markets effectively enhances competition, and whether increased competition, in turn, leads to improved economic outcomes. While previous literature presents mixed evidence regarding the consequences of antitrust lawsuits on the economy, our analysis at the corporate level offers fresh insights into the effectiveness of these legal actions.

To advance the literature, we use government contractors and antitrust lawsuits related to government procurement activities as a laboratory experiment. Our empirical analysis reveals that antitrust lawsuits improved access to the affected product markets for non-defendant businesses. Additional findings provide evidence that these events positively affect the performance and labor outcomes of these corporations without any negative consequence for their financial health. Conversely, defendant firms face reduced government contracts and diminished market power. At the product market level, we also document an increase in participants and a decrease in market concentration. Interestingly, post-lawsuit markets tend to be populated by larger and older establishments. At the same time, the competitive environment challenges small, financially constrained businesses. Finally, we do not find any change in government procurement costs.

These results support the hypothesis that antitrust lawsuits could significantly benefit non-defendant corporations and their employees. Our findings are particularly relevant in a setting where firms are rationed out of the market due to legal restrictions on collusive firms. Our conclusions may not directly apply in contexts where customers can still purchase from collusive firms, which might lower prices following an investigation.

Figure 4.1: **Antitrust Lawsuits Related to Government Activities**

Figure 4.1 shows the number of antitrust lawsuits related to government procurement activities over time, specifically for the period 2001–2021. VitalLaw case summaries serve as our source of information. We use LLMs to identify government-related antitrust cases.

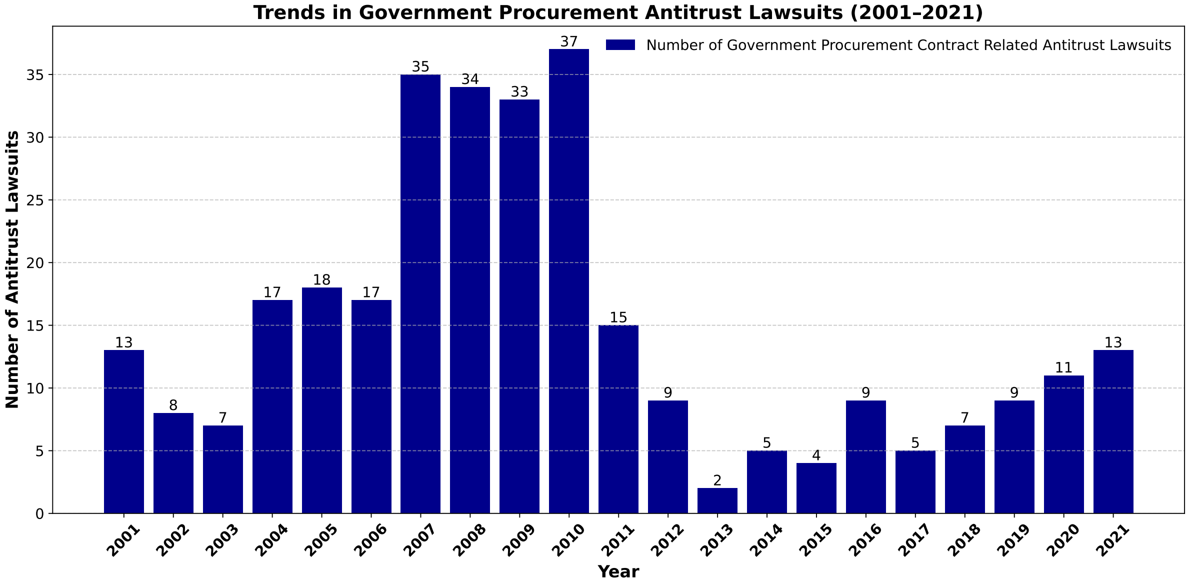


Figure 4.2: **Antitrust Lawsuits Related to Government Procurement Activities Across Sectors**

Figure 4.2 illustrates the frequency of antitrust lawsuits related to government procurement activities across product markets during the period 2001–2021. When information about the product market is unavailable from the Antitrust Case Filings database, we use LLMs to extract this information from VitalLaw case summaries.

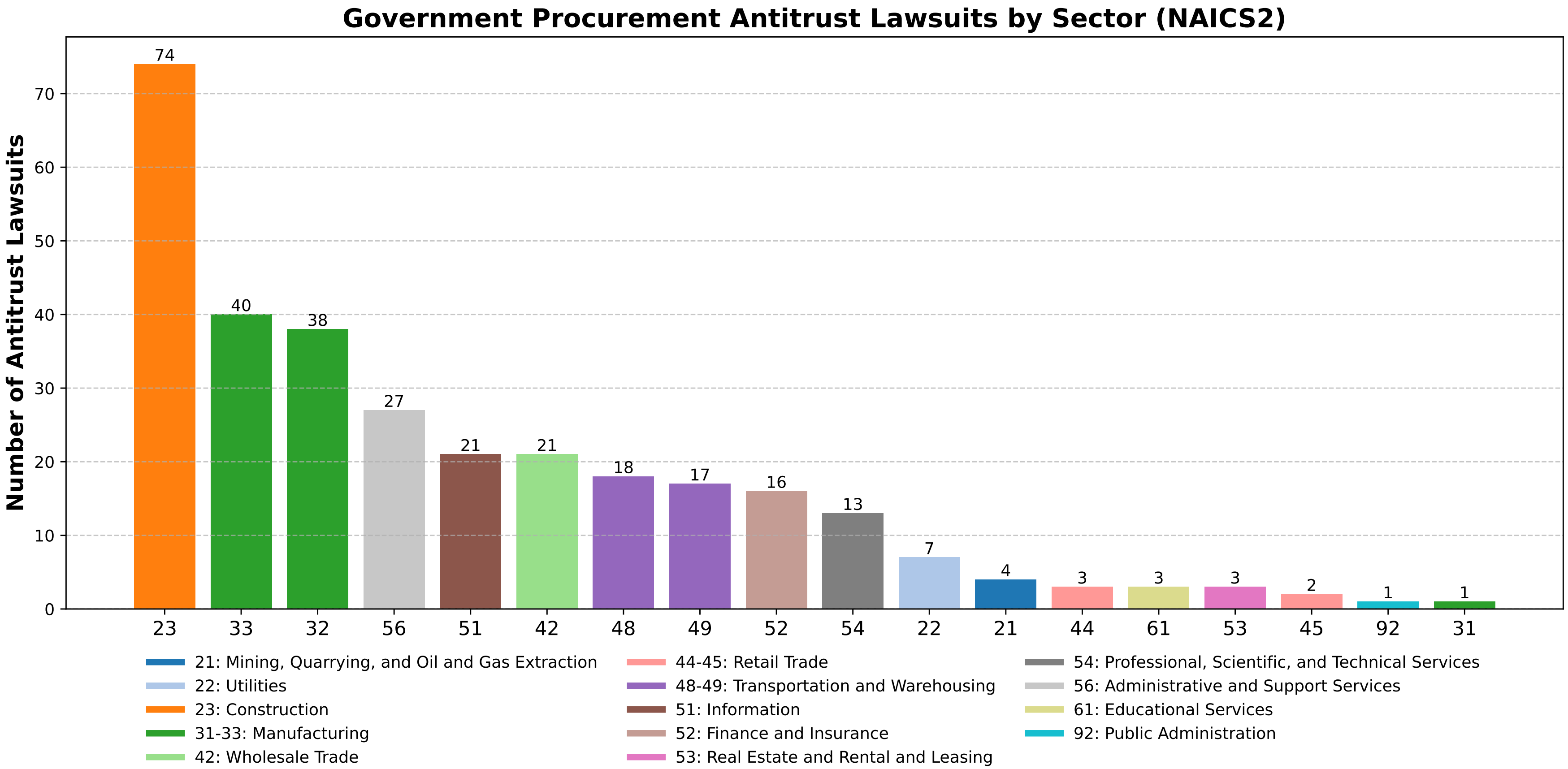


Figure 4.3: **Types of Antitrust Lawsuits Violation**

Figure 4.3 displays the frequency distribution of different violation types in antitrust lawsuits related to government procurement activities from 2001-2021. The horizontal bars show the count of each violation type, arranged in descending order of frequency. This analysis is based on DOJ antitrust case data from VitalLaw summaries, with violations identified and classified using LLM.

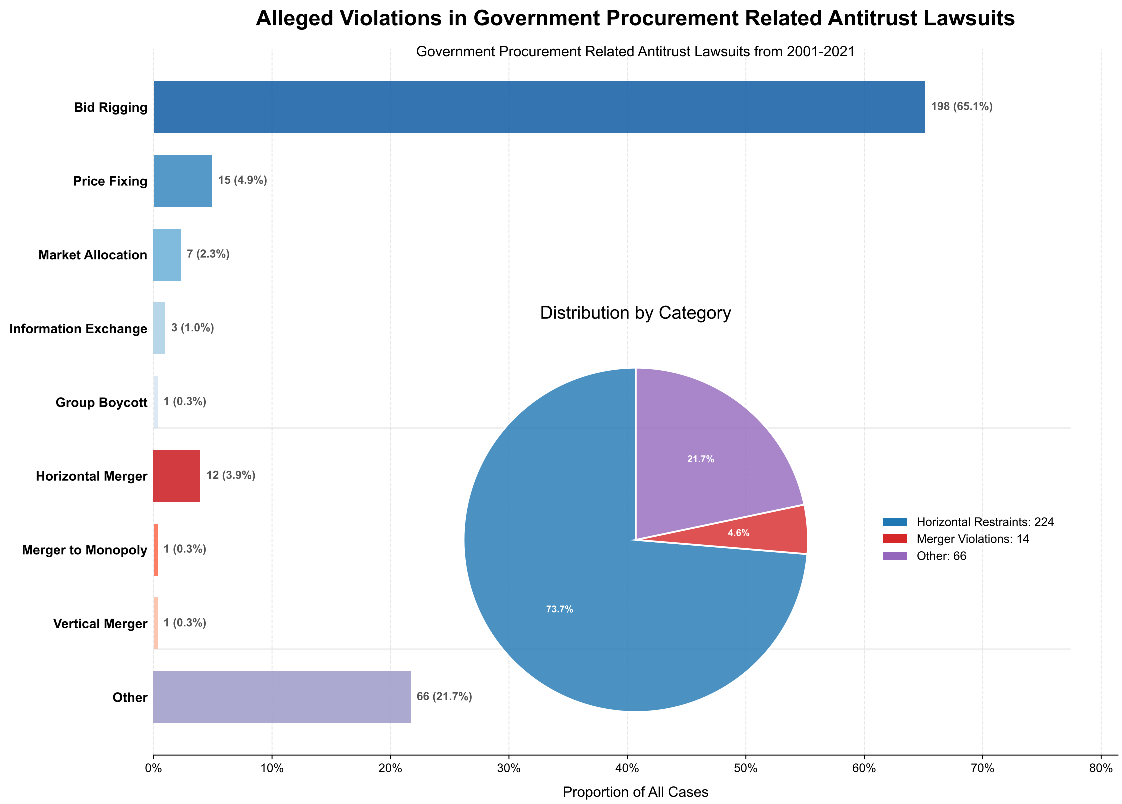
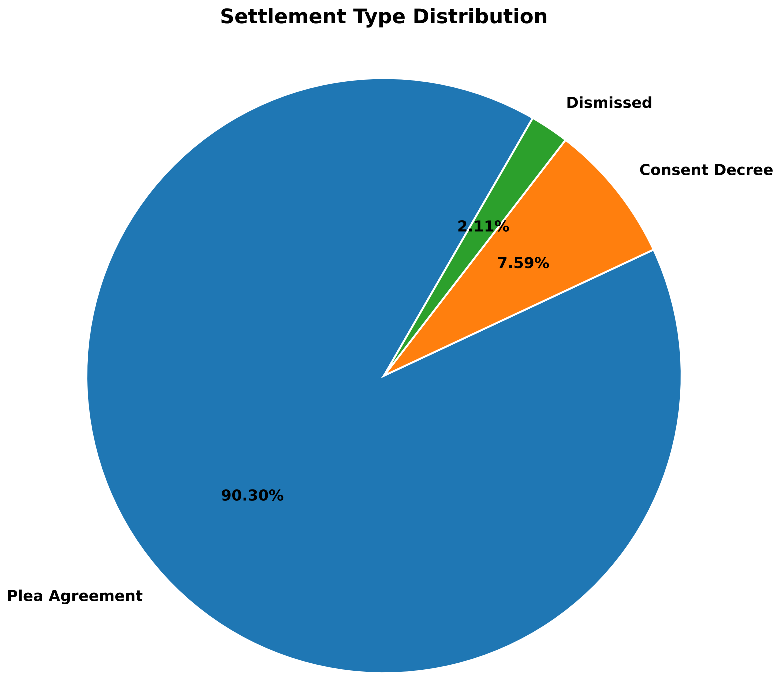
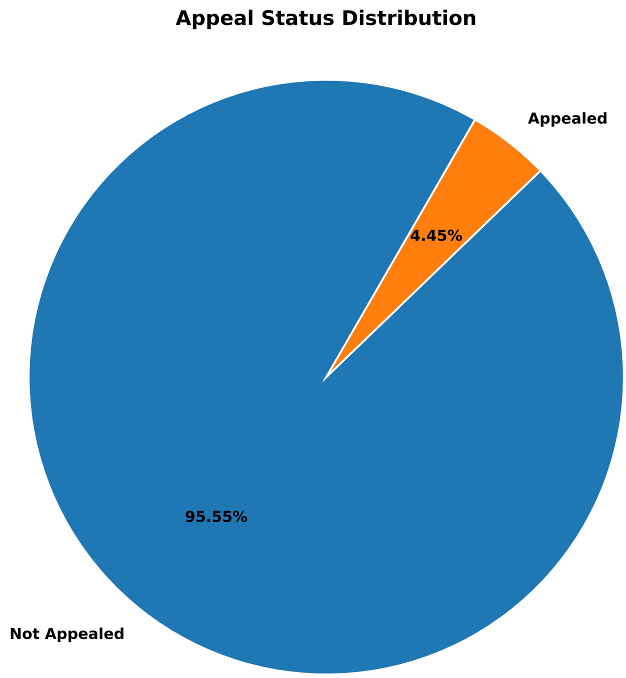


Figure 4.4: **Antitrust Lawsuits Legal Outcomes**

Figure 4.4 reports information on the relative share of different settlement types and appeal statuses in our final sample. VitalLaw case summaries serve as our source of information. We use LLMs to extract this information.



(a) Antitrust Lawsuits Settlement Type



(b) Antitrust Lawsuits Appeal Status

Figure 4.5: **Relative and Absolute Monopolistic Practices**

Figure 4.5 the share of different settlement types and appeal statuses in our final sample. VitalLaw case summaries serve as our source of information. We use LLMs to extract this information.

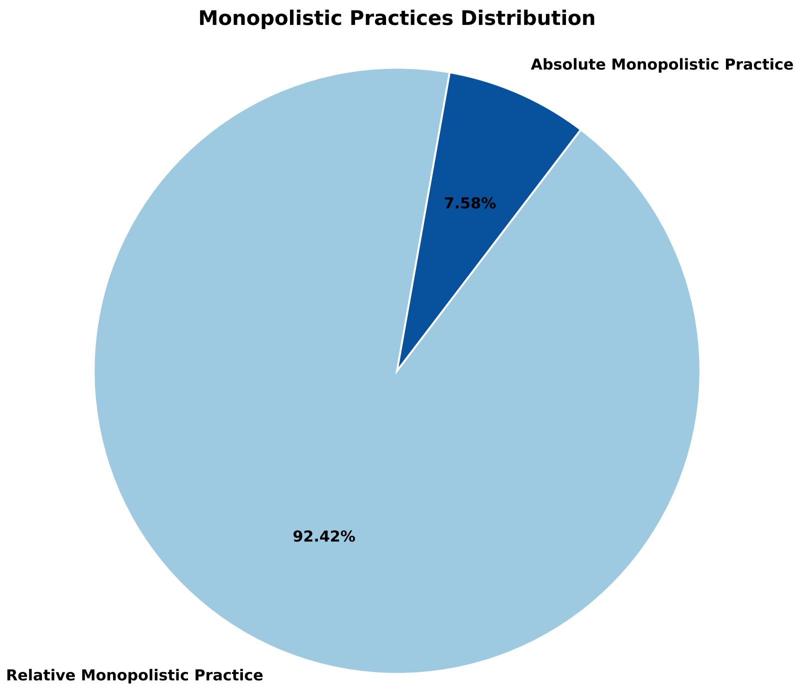
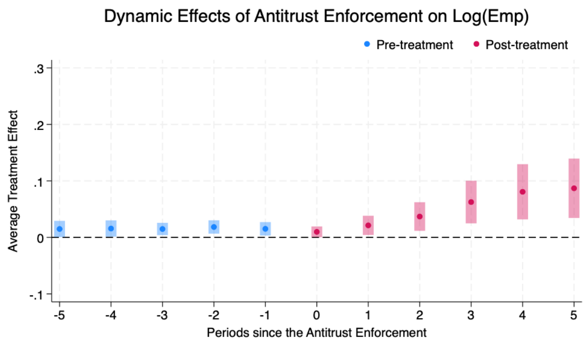
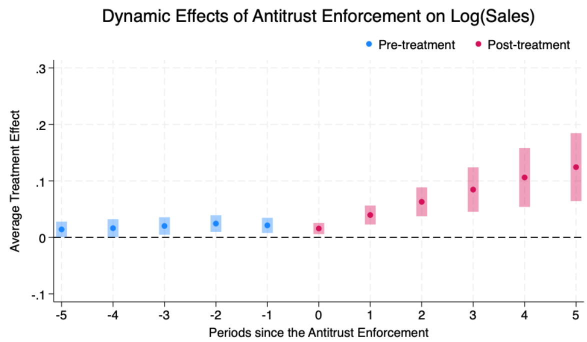


Figure 4.6: **Antitrust Lawsuits and Non-Defendants Outcomes: Dynamics Effects**

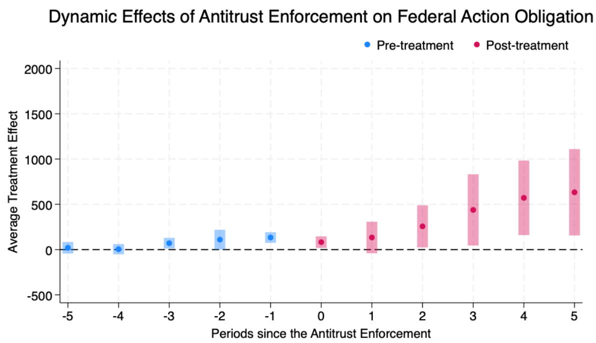
Figure 4.6 displays the dynamic effects of antitrust lawsuits on non-defendant establishments' outcomes to test the validity of our difference-in-differences identification strategy. Using the estimator proposed by Callaway and Sant'Anna (2021), we address potential endogeneity concerns in staggered treatment settings that traditional difference-in-differences methods may not fully account for (Baker et al., 2022). This approach compares treatment effects dynamically across multiple periods, leveraging units with identical initial treatment levels as controls, which allows for non-binary and non-absorbing treatment processes. We examine four key outcome variables: employment, sales, federal government contract awards, and financial health (Paydex score). The x-axis represents time periods relative to the filing date of antitrust lawsuits (event time), with negative values indicating pre-treatment periods and positive values indicating post-treatment periods. The y-axis shows the average treatment effect on the treated (ATT) for each outcome variable. The shaded areas represent 95% confidence intervals. The patterns in pre-event periods provide insights into the validity of the parallel trends assumption that underpins our causal interpretation, while post-event trajectories reveal the temporal dynamics of treatment effects following antitrust enforcement.



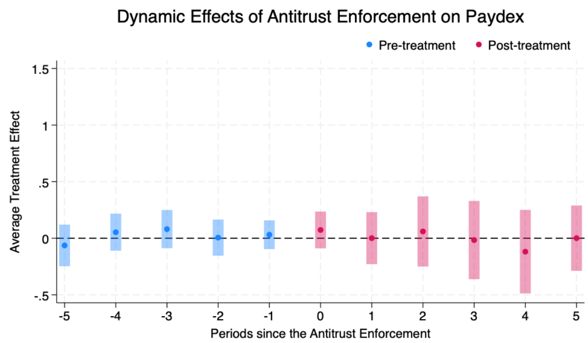
(a) Log(Employment)



(b) Log(Sales)



(c) Federal Government Awards



(d) Paydex

Figure 4.7: **Stock Market Reactions to Antitrust Lawsuits**

Figure 4.7 presents an event study analysis examining stock market reactions around antitrust lawsuit filing dates to test whether these events are unexpected to investors and provide market-based validation of our empirical approach. We analyze a sample of 196 publicly traded firms that operate within 6-digit NAICS codes affected by antitrust enforcement but were not defendants themselves. For each firm, we calculate buy-and-hold abnormal returns (BHAR) over a 21-day window surrounding the filing date (10 days before to 10 days after). The BHAR methodology accounts for the difference between a firm's actual return and its expected return based on a market model estimated during a pre-event period, capturing the abnormal performance attributable to the antitrust event. This approach aligns with the efficient market hypothesis (Fama, 1970), which suggests stock prices incorporate all available information and rapidly adjust to new information that might affect future cash flows. Under this framework, stock prices reflect the present value of expected future cash flows as shown in Equation (4.3), and market reactions reveal investors' expectations about how antitrust enforcement might affect competitive dynamics and non-defendant firms' future profitability. The x-axis represents trading days relative to the filing date (day 0), while the y-axis shows cumulative BHAR in percentage terms. The solid line represents the average BHAR across all firms, with dashed lines indicating 90% confidence intervals.

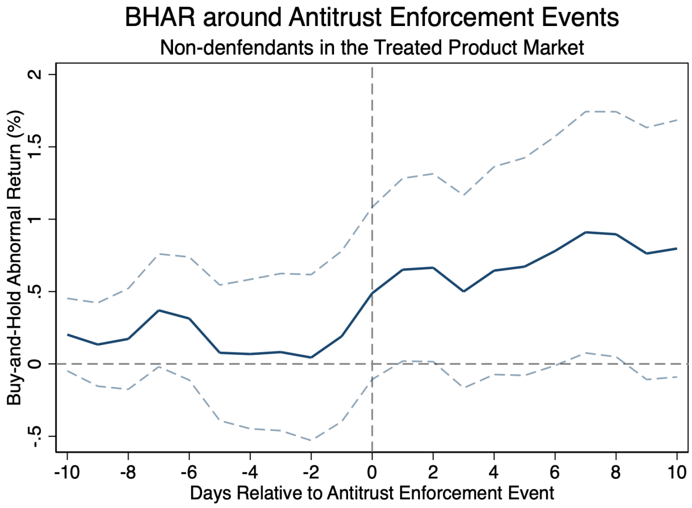


Table 4.1: **Summary Statistics: Establishment Level Database**

Table 4.1 provides information on the main sample we use in our empirical analysis. It reports the summary statistics of the variables in our establishment-level database. The period of analysis goes from 2001 to 2021. Definitions of the variables are provided in Table 4.A1 in the Appendix. All continuous variables are winsorized by year at the 1st and 99th percentiles.

Variables	Count	Mean	SD	p25	p75
Employment	1,678,543	70.888	372.100	3.000	40.000
Log(Employment)	1,678,543	2.477	1.714	1.099	3.689
Sales ($thousand)	1,678,543	15,812.024	135,943.546	262.500	6,000.020
Log(Sales)	1,678,543	7.215	2.111	5.570	8.700
Government Obligation ($thousand)	1,678,543	866.221	4,000.002	0.000	67.325
IHS Government Obligation	1,678,543	2.422	3.277	0.000	4.903
Paydex	1,197,437	72.572	8.931	69.500	79.500
Antitrust Lawsuit	1,678,543	0.136	0.343	0.000	0.000

Table 4.2: **The Determinants of Antitrust Lawsuits**

Table 4.2 shows regression results from Equation (4.1). We use an indicator variable equal to one if the product market has been exposed to an antitrust lawsuit as an outcome variable. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Treated	Treated	Treated	Treated	Treated	Treated
Log(HHI)	-0.006***	-0.001				
	(0.001)	(0.001)				
Log(Number of Firms)			0.004***	-0.001		
			(0.001)	(0.001)		
Log(Government Spending)					0.002***	-0.001
					(0.000)	(0.000)
Product Market FE	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes
Observations	23,636	23,618	23,636	23,618	23,636	23,618
R-squared	0.005	0.172	0.006	0.172	0.007	0.172

Table 4.3: **Antitrust Lawsuits and Access to the Product Market**

Table 4.3 shows regression results from Equation (4.2). We use the inverse hyperbolic sine transformation (IHS) of total government awards as an outcome variable. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)
Variables	IHS Government Contracts	IHS Government Contracts
Antitrust lawsuit	0.260**	0.227**
	(0.109)	(0.092)
Establishment FE	Yes	Yes
Year FE	Yes	No
County-Year FE	No	Yes
Observations	1,678,543	1,678,543
R-squared	0.547	0.561

Table 4.4: **Antitrust Lawsuits and Non-Defendants Outcomes**

Table 4.4 shows regression results from Equation (4.2), applying the same model to three different outcome variables: Log(Employment), Log(Sales), and Paydex. These variables measure establishment size, performance, and financial health, respectively. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)	(3)
Variables	Log(Employment)	Log(Sales)	Paydex
Antitrust Lawsuit	0.075***	0.103***	-0.090
	(0.024)	(0.024)	(0.142)
Establishment FE	Yes	Yes	Yes
County-Year FE	Yes	Yes	Yes
Observations	1,678,543	1,678,543	1,189,104
R-squared	0.935	0.943	0.525
Average Outcome	2.477	7.215	72.57

Table 4.5: **Market Share Loss for Defendant Establishments**

Table 4.5 examines whether defendant establishments experience market exclusion following antitrust enforcement. To provide direct evidence of market exclusion mechanisms, we identify 174 unique establishments that are government contractors and part of defendant firms. We compare their government contract dynamics with those of establishments operating in product markets that have never been affected by antitrust lawsuits using Equation (4.6). The Defendant Establishment variable captures the effect of antitrust lawsuits on establishments belonging to firms found to have violated antitrust laws. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)
Variables	IHS Government Contracts	IHS Government Contracts
Defendant Establishment	-0.949***	-0.964***
	(0.280)	(0.303)
Establishment FE	Yes	Yes
Year FE	Yes	No
County-Year FE	No	Yes
Observations	1,339,546	1,339,546
R-squared	0.535	0.551
Average Outcome	2.301	2.301

Table 4.6: **Industry-Level Effects of Antitrust Enforcement**

Table 4.6 presents product market-level analysis examining how antitrust lawsuits affect market structure, concentration, government acquisition costs, and perceived fairness. We estimate Equation (4.7) using four different outcome variables: the number of market participants (logarithm of unique firms), market concentration (logarithm of HHI index), government procurement costs (average contract value), and perceived market fairness (protest rate). This analysis complements our establishment-level findings by providing evidence on aggregate market dynamics following antitrust enforcement. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)	(3)	(4)
Variables	Log(# of Unique Firms)	Log(HHI)	Average Contract Value	Protest Rate (%)
Antitrust lawsuit	0.303***	-0.148**	-2.934	-0.121
	(0.074)	(0.070)	(26.755)	(0.155)
Product market FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	23,610	23,610	23,610	14,078
R-squared	0.874	0.642	0.532	0.432
Average Outcome	3.816	7.616	159.2	0.402

Table 4.7: **Characteristics of Contract Winners Following Antitrust Enforcement**

Table 4.7 examines how the characteristics of firms winning government contracts change following antitrust enforcement. We estimate Equation (4.8) using various establishment characteristics as dependent variables, including size (employment and sales), financial health (Paydex), ownership structure (public status), age, size classification, and minority ownership status. Each observation is a contract-establishment pair, weighted by the natural logarithm of contract value to reflect economic significance. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	Log(Emp.)	Log(Sales)	Paydex	Public	Age	Small	Minority
Antitrust lawsuit	0.192***	0.192**	0.276	0.014	0.133	-0.015	-0.002
	(0.065)	(0.082)	(0.212)	(0.018)	(0.204)	(0.009)	(0.007)
Product Market FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,458,175	22,458,175	21,508,969	22,458,175	22,458,175	22,458,175	22,458,175
R-squared	0.557	0.591	0.477	0.709	0.646	0.354	0.410
Average Outcome	4.435	10.06	72.88	0.423	22.70	0.0804	0.0737

Table 4.8: **Effects of Antitrust Enforcement on Contract Structure**

Table 4.8 examines whether antitrust enforcement affects government procurement costs through changes in contract structure and renegotiation patterns. Following previous literature on government overspending drivers (Bajari & Tadelis, 2001; Crocker & Reynolds, 1993), we investigate whether antitrust lawsuits impact contract renegotiations (Column 1) and the use of fixed-price contracts (Column 2). Each observation is weighted by contract value to reflect economic significance. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)
Variables	Renegotiation	Fixed Price
Antitrust Lawsuit	0.095**	0.006
	(0.042)	(0.014)
Product Market FE	Yes	Yes
County × Year FE	Yes	Yes
Observations	22,458,175	22,458,175
R-squared	0.341	0.395
Average Outcome	1.303	0.973

Table 4.A1: **Variable Definition**

Variable	Definition	Source
Establishment level variables		
Antitrust Lawsuit	An indicator equal to after the first antitrust lawsuits in a product market.	Wolters Kluwer's VitalLaw & the Department of Justice (DOJ) Antitrust Division
Log(Employment)	The natural logarithm of the number of employees in a given establishment and year.	NETS
Log(Sales)	The natural logarithm of the dollar amount ($ thousand) of sales in a given establishment and year.	NETS
Paydex	A measure of the financial health of the establishment. It reflects the establishment's payment behavior, typically based on its history of paying invoices on time. It is a numerical value that ranges between 0 and 100. A higher score indicates timely or early payments, while lower scores suggest delays, signaling potential financial distress or poor credit management.	NETS
Government Contracts	The dollar amount ($ thousand) of federal obligation awarded to an establishment in a given year.	USASpending
Age	The difference between the year and the first year establishment was active in the database.	NETS
Public	An indicator variable equal to one if the establishment belong to a public corporation.	NETS
Small	An indicator variable equal to one if the establishment has less than 5 employees.	NETS
Minority	An indicator variable equal to one if the establishment is minority owned.	NETS
Women Owned	An indicator variable equal to one if the establishment is owned by a woman.	NETS
Industry level variables		
# Unique Firms	The number of different firms that won a procurement contract in a product market (6-digit NAICS code) and year. When multiple transactions are associated with a contract, we identify the main transaction as the one with the highest transaction value.	USASpending
HHI	HHI index in a product market (6-digit NAICS code) and year. It measures the market concentration of firms in a product market. When multiple transactions are associated with a contract, we identify the main transaction as the one with the highest transaction value.	USASpending
Average Value	The total dollar amount of government spending ($ million) in a product market (6-digit NAICS code) and year divided by that total number of contracts. When multiple transactions are associated with a contract, we identify the main transaction as the one with the highest transaction value.	USASpending
Contract level variables		
Renegatation Contracts	The average number of times a contract has been renegotiated, based on the number of transactions.	USASpending
Competition	The share of contracts that have been awarded under a competitive bidding process. When multiple transactions are associated with a contract, we identify the main transaction as the one with the highest transaction value.	USASpending
Fixed price	The share of contracts that have been awarded under a competitive bidding process. When multiple transactions are associated with a contract, we identify the main transaction as the one with the highest transaction value.	USASpending

Table 4.A2: **Normalized Differences Between Treatment and Control Groups**

Table 4.A2 compares observable characteristics between treatment and control groups at the beginning of the analyzed period to assess the validity of our identification strategy. Panel A reports industry-level characteristics, while Panel B reports establishment-level characteristics. Normalized differences (ND) are calculated following Imbens and Wooldridge (2009), with values within ±0.25 indicating similar distributions between groups. As shown in Panel B, establishments in exposed and non-exposed product markets appear similar in observable characteristics, though the markets themselves (Panel A) show some differences in concentration and size. These market-level differences are accounted for by including fixed effects in our empirical specifications.

Panel A: Industry characteristics

	Treated		Untreated		
	Mean	SD	Mean	SD	ND
Log(HHI)	7.44	1.16	8.04	0.95	-0.40
Log(# firms)	3.54	1.72	2.48	1.36	0.48
Average Renegotiation	1.38	0.53	1.29	0.71	0.11
Competition	0.72	0.26	0.70	0.30	0.04
Average Contract Value	445,148.70	976,219.44	260,241.87	587,804.05	0.16
Share Fixed Price	0.84	0.26	0.90	0.20	-0.18

Panel B: Establishment characteristics

	Treated		Untreated		
	Mean	SD	Mean	SD	ND
Age	9.12	4.17	9.61	3.95	-0.09
Women Owned	0.21	0.40	0.17	0.38	0.06
Minority	0.16	0.37	0.09	0.29	0.15
Dummy Women	0.11	0.32	0.13	0.34	-0.04
Log(Employment)	2.69	1.59	2.77	1.71	-0.03
Log(Sales)	7.41	1.87	7.52	1.96	-0.04
Log(Productivity)	4.73	0.84	4.75	0.86	-0.02
Paydex	71.42	8.37	71.23	8.25	0.02
Federal obligation	161.66	720.83	83.00	502.21	0.09

Table 4.A3: **Propensity Score Matching**

Table 4.A3 presents regression results using propensity score matching to address potential selection concerns in our difference-in-differences approach. We match treated and untreated product markets based on market size, concentration, and competitive structure using one-to-three (column 1) and one-to-five (column 2) nearest-neighbor matching algorithms. This approach ensures that both industries and establishments in our treatment and control groups are similar in terms of observable characteristics. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)
Variables	IHS Government Contracts (1-to-3 Matching)	IHS Government Contracts (1-to-5 Matching)
Antitrust Lawsuits	0.179*	0.182*
	(0.104)	(0.102)
Establishment FE	Yes	Yes
County-Year FE	Yes	Yes
Observations	543,990	633,166
R-squared	0.575	0.576

Table 4.A4: **Alternative Clustering Approaches**

Table 4.A4 shows regression results with various clustering approaches to validate the robustness of our standard error estimates. Panel A presents results with single-level clustering, while Panel B shows results with two-way clustering. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at different levels as indicated in each column. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

Panel A: Alternative clustering

	(1)	(3)	(5)
Variables	IHS Government	IHS Government	IHS Government
Antitrust Lawsuits	0.227**	0.227***	0.227***
	(0.092)	(0.021)	(0.029)
Establishment FE	Yes	Yes	Yes
County-Year FE	Yes	Yes	Yes
SE Clustering	PM	Establ.	County
Observations	1,678,543	1,678,543	1,678,543
R-squared	0.561	0.561	0.561

Panel B: Alternative double clustering

	(2)	(4)	(6)
Variables	IHS Government	IHS Government	IHS Government
Antitrust Lawsuits	0.227**	0.227***	0.227***
	(0.087)	(0.026)	(0.033)
Establishment FE	Yes	Yes	Yes
County-Year FE	Yes	Yes	Yes
SE Clustering	PM and Year	Establ. and Year	County and Year
Observations	1,678,543	1,678,543	1,678,543
R-squared	0.561	0.561	0.561

Table 4.A5: **Alternative Government Awards Measures**

Table 4.A5 presents regression results using different specifications of the government contract award variable to test the robustness of our findings. We examine both level measures and inverse hyperbolic sine (IHS) transformations, as well as multi-year averaging approaches. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

Panel A: Baseline and single-period measures

	(1)	(2)	(3)
Variables	Baseline	Level	Level (3 years average)
Antitrust Lawsuit	0.227**	452.990***	442.044***
	(0.092)	(118.916)	(125.548)
Establishment FE	Yes	Yes	Yes
County-Year FE	Yes	Yes	Yes
Observations	1,678,543	1,678,543	1,422,945
R-squared	0.561	0.640	0.729
Average Outcome	2.422	866.2	986.9

Panel B: Multi-period average measures

	(4)	(5)	(6)
Variables	IHS (3 years average)	Level (2 years average)	IHS (2 years average)
Antitrust Lawsuit	0.191**	435.012***	0.210**
	(0.088)	(120.344)	(0.087)
Establishment FE	Yes	Yes	Yes
County-Year FE	Yes	Yes	Yes
Observations	1,422,945	1,549,440	1,549,440
R-squared	0.679	0.691	0.626
Average Outcome	3.142	938.3	2.843

Table 4.A6: **Controlling for Industry Trends**

Table 4.A6 presents regression results with various controls for industry-specific and product market-specific trends to address potential confounding factors. We include time-variant product market characteristics, industry-by-year fixed effects, and product market-specific time trends. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)	(3)	(4)
Variables	IHS Government	IHS Government	IHS Government	IHS Government
Antitrust Lawsuit	0.213**	0.192**	0.240***	0.240***
	(0.086)	(0.077)	(0.089)	(0.089)
Product Market Controls	Yes			
Industry × Year		Yes		
Product Market Linear Trend			Yes	Yes
Product Market Quadratic Trend				Yes
Observations	1,456,591	1,678,543	1,678,543	1,678,543
R-squared	0.563	0.565	0.561	0.561
Average Outcome	2.484	2.422	2.422	2.422

Table 4.A7: **Spillover Effects of Antitrust Lawsuits**

Table 4.A7 tests whether antitrust enforcement in one product market affects related markets through competitive spillovers or resource reallocation. The variable *ntitrust Lawsuit Spillover* is a binary indicator equal to one if the establishment operates in a broader 4-digit NAICS sector containing a product market exposed to an antitrust lawsuit (but not in the specific 6-digit NAICS code directly affected). Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)
Variables	IHS Government Contracts	IHS Government Contracts
Antitrust Lawsuit	0.250**	0.229***
	(0.108)	(0.094)
Antitrust Lawsuit Spillover	-0.023	0.004
	(0.070)	(0.063)
Establishment FE	Yes	Yes
Year FE	Yes	Subsumed
County-Year FE	No	Yes
Observations	1,678,543	1,678,543
R-squared	0.547	0.561
Average Outcome	2.422	2.422

Table 4.A8: **Placebo Test: Non-Government Contractors**

Table 4.A8 presents regression results from a placebo test using establishments that are not government contractors but are located in the same counties and operate in the same product markets as our treatment sample. This analysis tests whether the effects we identify are specific to government contractors or reflect broader market dynamics affecting all establishments. We estimate our baseline specification with employment, sales, and financial health as outcomes using this substantially larger sample of 185,158,511 establishment-year observations. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)	(3)
Variables	Log(Employment)	Log(Sales)	Paydex
Antitrust Lawsuit	0.012	0.025	-0.277
	(0.012)	(0.029)	(0.360)
Establishment FE	Yes	Yes	Yes
County-Year FE	Yes	Yes	Yes
Observations	185,463,304	185,158,511	39,273,724
R-squared	0.910	0.918	0.587

Table 4.A9: **Triple Difference-in-Differences Analysis**

Table 4.A9 presents results from a triple difference-in-differences framework that compares government contractors to non-government contractors operating in the same product markets and locations. This approach includes both product market-by-year and county-by-year fixed effects, allowing us to control for time-varying product market characteristics while isolating the differential effect of antitrust lawsuits on government contractors. The main effects of Government Contractor and Antitrust Lawsuit variables are subsumed by the establishment and product market-by-year fixed effects, respectively. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)	(3)
Variables	Paydex	Log(Employment)	Log(Sales)
Antitrust lawsuit × Government Contractor	1.598***	0.181***	0.307***
	(0.155)	(0.028)	(0.037)
Establishment FE	Yes	Yes	Yes
Product Market × Year FE	Yes	Yes	Yes
County × Year FE	Yes	Yes	Yes
Observations	40,483,362	187,162,559	186,845,138
R-squared	0.650	0.927	0.935
Average Outcome	71.895	0.890	12.023

Table 4.A10: **Antitrust Lawsuits Effects by Product Market Size**

Table 4.A10 examines whether the effects of antitrust lawsuits vary with the size of the affected product market. As suggested by Equation (4.4), the benefits of antitrust enforcement may be more significant when the exposed market is larger. We test this hypothesis by dividing our sample based on the median size of the procurement market, proxied by total government spending. Column (1) reports results for establishments in larger markets (above median size), while column (2) shows results for those in smaller markets (below median size). Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)
Variables	IHS Government Contracts (Size Above the Median)	IHS Government Contracts (Size Below the Median)
Antitrust lawsuit	0.313***	-0.016
	(0.103)	(0.069)
Establishment FE	Yes	Yes
County-Year FE	Yes	Yes
Observations	716,893	716,000
R-squared	0.582	0.588

Table 4.A11: **Antitrust Lawsuits and Defendant Establishment Outcomes**

Table 4.A11 examines how antitrust enforcement affects the performance of establishments belonging to defendant firms. We estimate Equation (4.6) using three different outcome variables measuring establishment size (Log(Employment)), performance (Log(Sales)), and financial health (Paydex). This analysis complements our main findings on non-defendant establishments and helps quantify the market redistribution effects following antitrust enforcement. Robust standard errors, reported in parentheses below the coefficient estimates, are clustered at the product market level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Definitions of the variables are provided in Table 4.A1 in the Appendix.

	(1)	(2)	(3)
Variables	Log(Employment)	Log(Sales)	Paydex
Antitrust Lawsuit	-0.233***	-0.036	-0.530
	(0.073)	(0.090)	(0.538)
Establishment FE	Yes	Yes	Yes
County-Year FE	Yes	Yes	Yes
Observations	1,339,546	1,339,546	934,124
R-squared	0.941	0.948	0.535
Average Outcome	2.444	7.165	72.51

Figure 4.A1: **Word Cloud of Government Procurement-Related Antitrust Cases**

Figure 4.A1 presents textual analysis of government procurement-related antitrust lawsuits from 2001-2021, highlighting the frequency and prominence of specific terms in antitrust litigation. Term size corresponds to relative frequency in the corpus. Panel (a) displays significant terms from case documents after removing common legal terminology, stopwords, and generic business language. Panel (b) illustrates procurement-specific terminology identified through large language model analysis of the same corpus. These visualizations provide intuitive insights into the textual dimensions of our dataset and reveal prominent themes across government procurement-related antitrust cases during the study period.

(a) Word Clouds of Antitrust Cases

(b) Procurement-Specific Words

Figure 4.A2: **Distribution of Antitrust Violations by Product Market**

Figure 4.A2 shows the distribution of antitrust lawsuits related to government procurement activities across different product markets from 2001-2021. We define specific product markets using the 6-digit NAICS code classification system. When information about the product market is unavailable from the Antitrust Case Filings database, we utilize large language models (LLMs) to extract this information from case summaries. The figure displays both the broad industry sectors (2-digit NAICS codes) and the specific product markets (6-digit NAICS codes) within each sector affected by antitrust enforcement, revealing concentration patterns across the U.S. economy. VitalLaw case summaries serve as our primary source of information for identifying and classifying these violations.

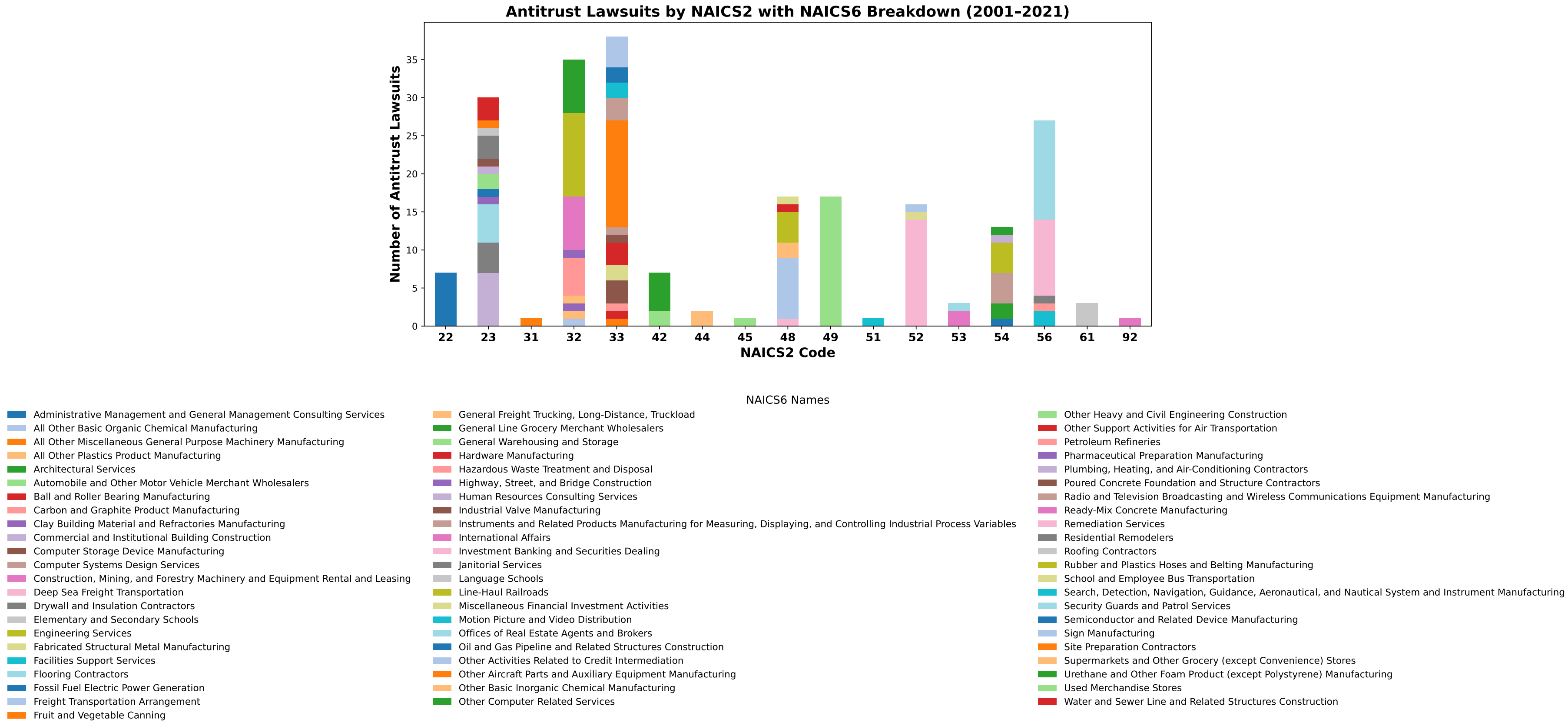
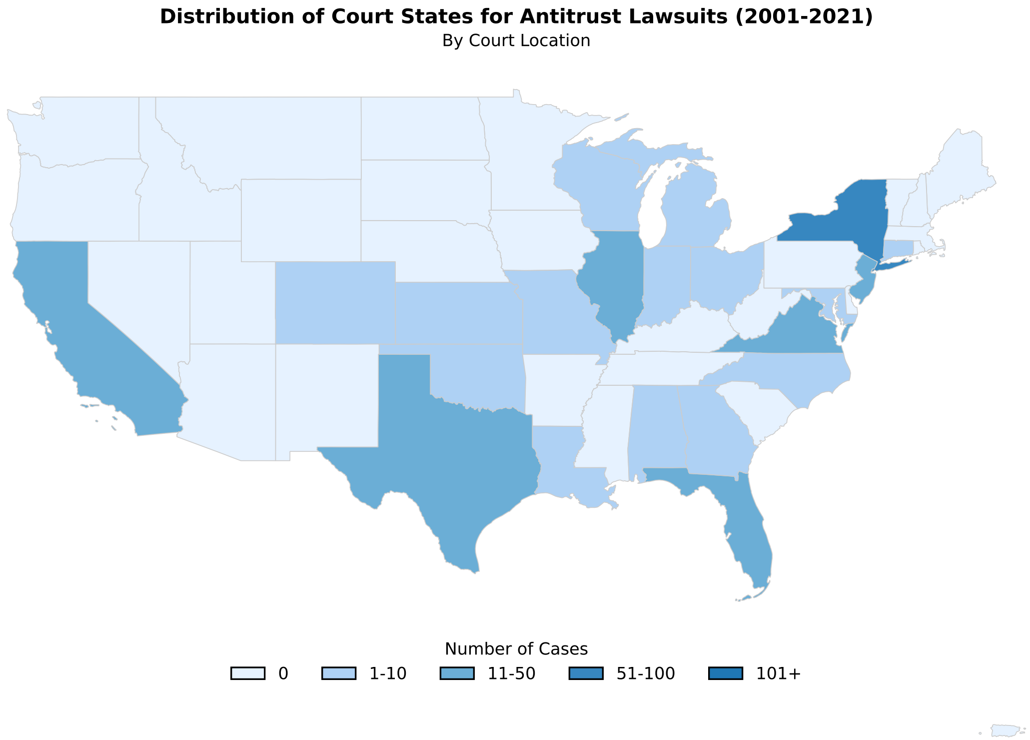
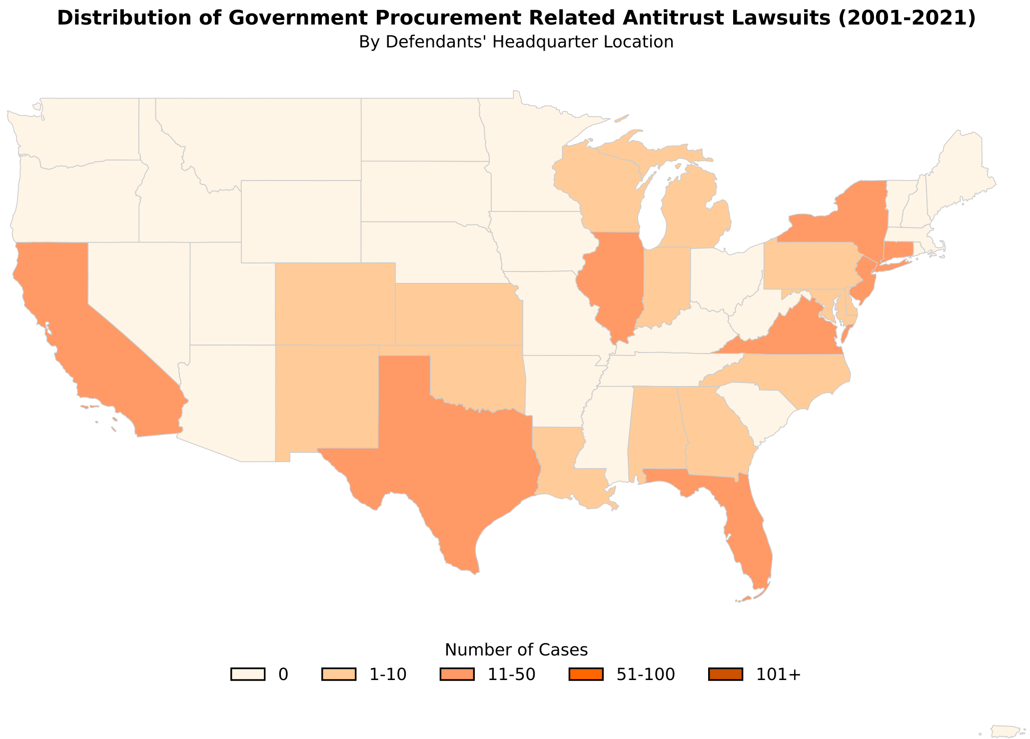


Figure 4.A3: **Antitrust Lawsuits Geographical Distribution (Defendants & Court)**

Figure 4.A3 illustrates the geographic distribution of government procurement-related antitrust lawsuits from 2001-2021. Panel (a) shows the location of courts where cases were filed, while Panel (b) displays the headquarters locations of defendant firms.



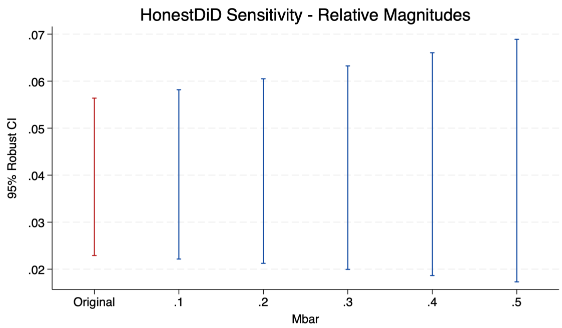
(a) Antitrust Lawsuits Court Locations



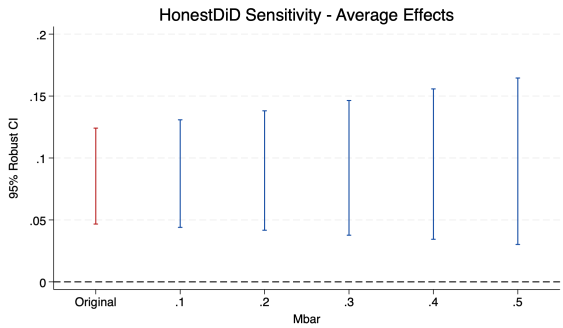
(b) Defendant Headquarters

Figure 4.A4: **Parallel Trend Test**

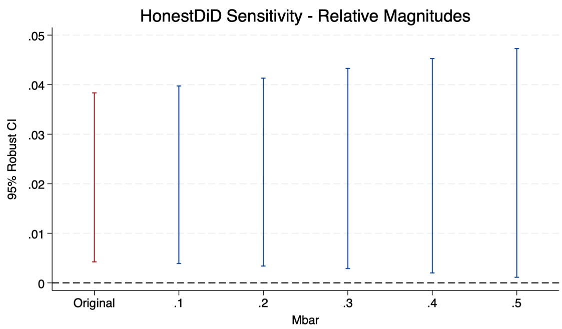
Figure 4.A4 reports the results of the sensitivity test proposed by Rambachan and Roth (2023) to test whether the parallel trend assumption holds more formally. This procedure examines the extent to which our main results remain robust against potential nonlinearities with varying magnitudes in the counterfactual trend. We calculate 95% confidence intervals for our main estimators under different assumptions of the value M, the upper limit for the change between two consecutive periods in the slope of the underlying linear trend. A value of M equal to 0 on the x-axis corresponds with allowing for linear violations of parallel trends, while larger values of M allow for more significant deviations from linearity. Left panels (a,c,e) show period-specific effects, while right panels (b,d,f) display average treatment effects across post-treatment periods, providing a more comprehensive assessment of the parallel trends assumption by focusing on aggregate impacts rather than individual time points.



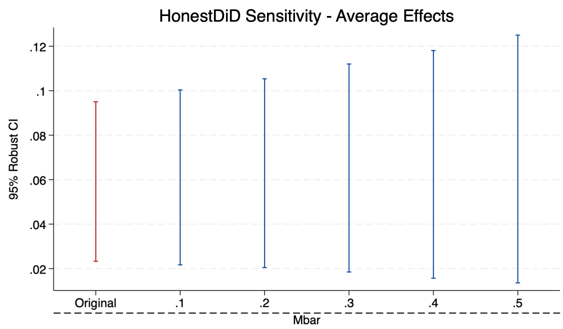
(a) Log(Sales)



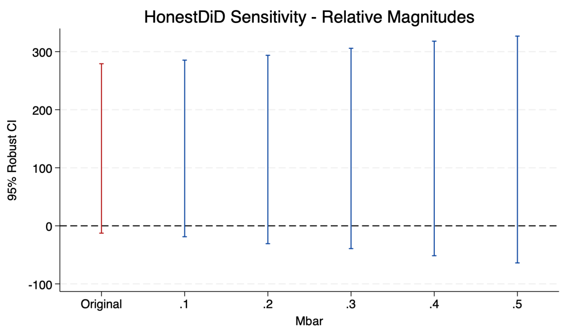
(b) Log(Sales) Average Effects



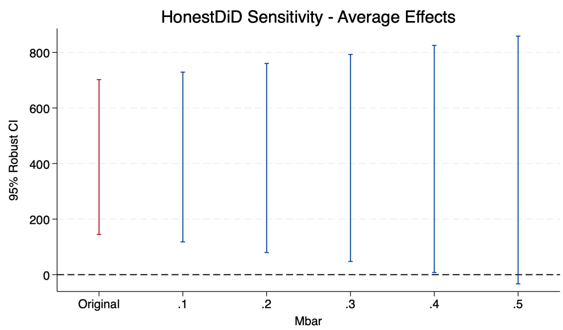
(c) Log(Emp)



(d) Log(Emp) Average Effects



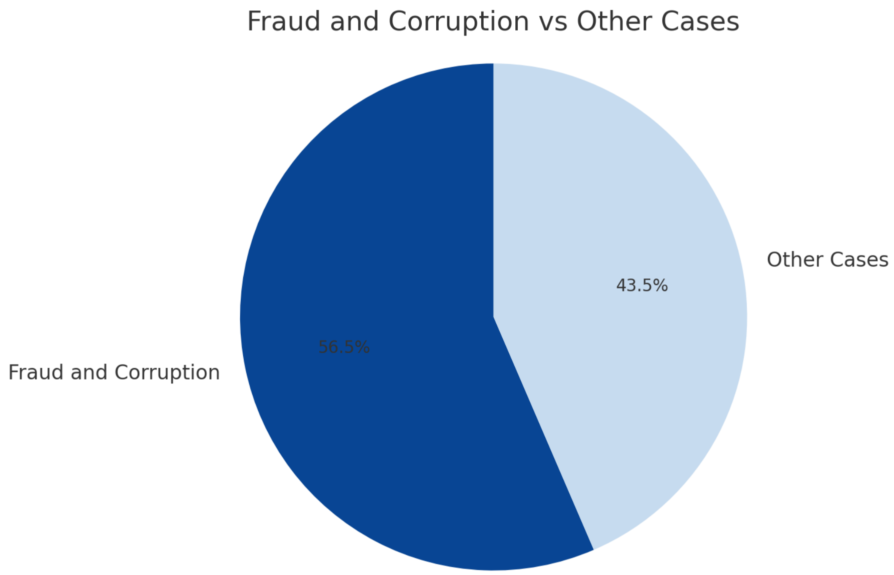
(e) Federal Obligation



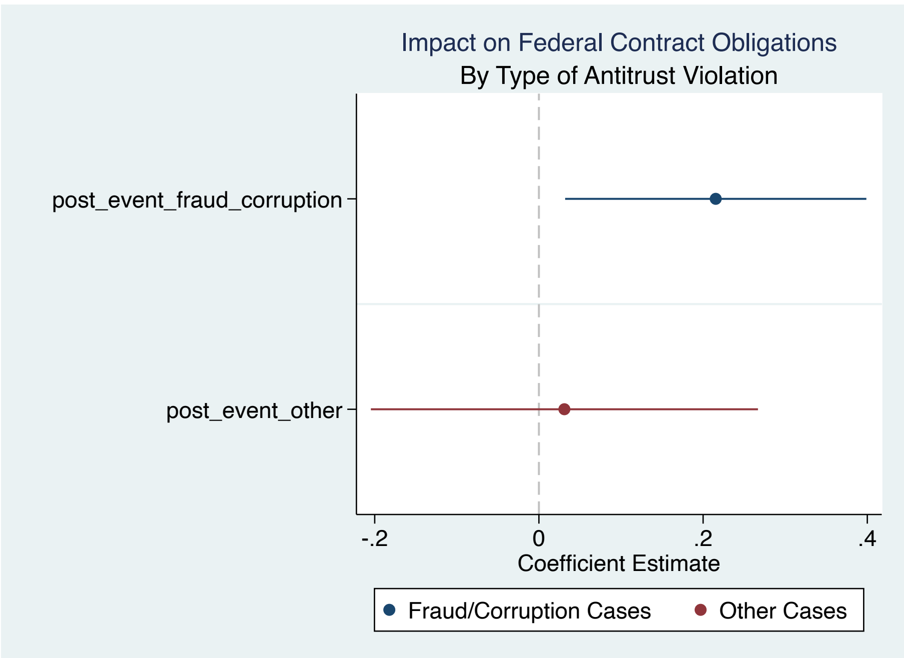
(f) Federal Obligation Average Effects

Figure 4.A5: **Types of Antitrust Violations and Their Impact**

Figure 4.A5 examines how different categories of antitrust violations affect market access outcomes. As reported in Equation (4.4), we expect businesses in exposed product markets to benefit from antitrust lawsuits primarily when colluding businesses lose market share. Panel (a) categorizes violations into two broad areas: anti-competitive practices (including bid rigging, customer allocation, and price fixing) and fraud/corruption (encompassing bribery, government fraud, mail fraud, wire fraud, tax evasion, money laundering, and obstruction of justice). Panel (b) presents coefficient estimates from Equation (4.2) for two categories of violations: *post_event_fraud_corruption* is a dummy variable equal to one for the period after an antitrust lawsuit involving fraud or corruption violations, and *post_event_other* is a dummy variable equal to one for the period after antitrust lawsuits involving other types of violations (primarily anti-competitive practices without fraud elements).



(a) Share of type of violations



(b) Heterogeneous impact

Figure 4.A6: **Dynamic Effects on Defendant Establishments' Procurement**

Figure 4.A6 displays the temporal pattern of how antitrust enforcement affects government contract awards to defendant establishments over time. Using the Callaway and Sant'Anna (2021) estimator, we track the evolution of federal contract obligations (measured using inverse hyperbolic sine transformation) for establishments belonging to defendant firms relative to the filing date of antitrust lawsuits. The x-axis represents event time in periods relative to the enforcement action, with negative values indicating pre-treatment periods and positive values representing post-treatment periods. The y-axis shows the average treatment effect on the treated (ATT). The shaded areas represent 95% confidence intervals. This dynamic analysis provides insights into both the magnitude and persistence of market exclusion effects following antitrust enforcement.

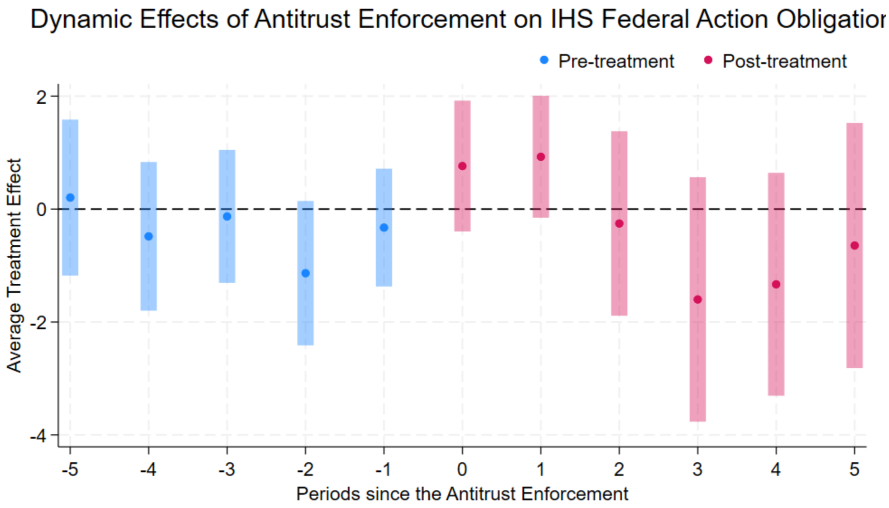
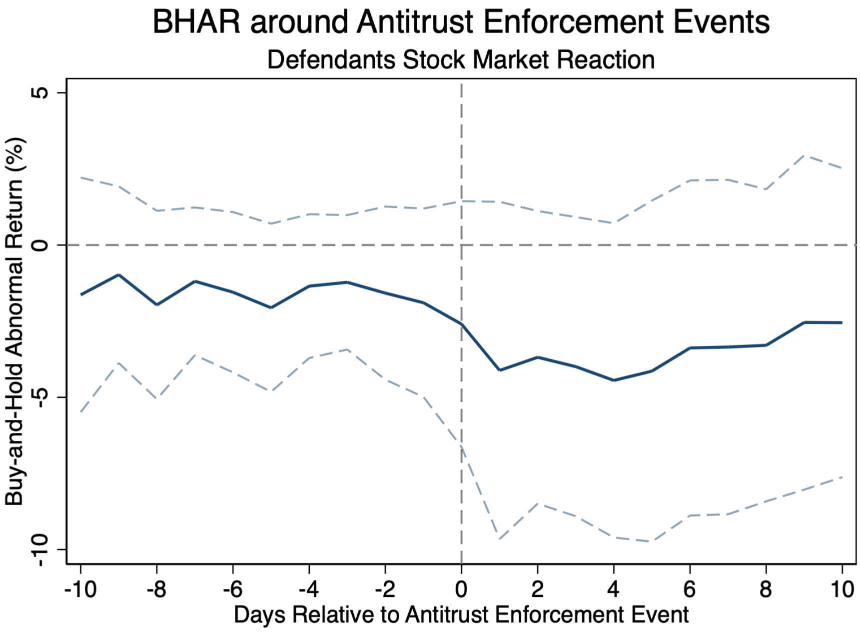


Figure 4.A7: **Stock Market Reactions to Antitrust Lawsuits - Defendant Firms**

Figure 4.A7 examines financial market reactions to antitrust enforcement actions through an event study analysis of stock returns for defendant firms. The figure tracks buy-and-hold abnormal returns (BHAR) for 9 identified publicly-traded defendant firms operating in product markets subject to antitrust enforcement. The analysis centers on a 21-day window surrounding the filing date of antitrust lawsuits (from day -10 to day +10). The BHAR methodology measures the difference between actual returns and expected returns based on a market model, capturing the abnormal performance attributable to the antitrust event. The x-axis represents trading days relative to the filing date (day 0), while the y-axis shows cumulative BHAR in percentage terms. The solid line represents the average BHAR across the defendant firms, with dashed lines indicating 90% confidence intervals.



Chapter 5

Conclusion

This thesis explores three novel dimensions of technology and competition in bank lending and corporate finance. It contributes to our understanding of how technological expertise shapes financial intermediation, how risk propagates through interconnected banking networks, and how regulatory interventions affect market dynamics. This study contributes to the banking and corporate finance literature by utilizing advanced empirical methods such as structural matching models, graph neural networks, and large language models to examine some of the fundamental questions pertaining to information asymmetry, specialized technological knowledge, and competition in financial markets. Together, these studies show how value is created economically through technological specialization, how network topologies transmit financial risk, and how antitrust interventions restructure competitive environments, with far-reaching implications for the efficiency of credit allocation, financial stability, and regulatory policy.

In the first study, we investigate whether banks develop specialized technical knowledge of corporate innovation through their lending experience and how such knowledge affects the terms of loan contracts. Using a large sample of U.S. syndicated loans from 1990 to 2020 combined with patent information, we show that companies with technology profiles similar to those of a bank's past borrowers receive significantly cheaper loans. The results are robust to various specifications, including different industry classifications, different constructs of technological similarity, controls for bank industry specialization, and a number of placebo tests. Moreover, we show that technology similarity contains useful information regarding borrower creditworthiness. In particular, banks lending more to borrowers with similar technologies have higher future returns on assets and larger distances to default, indicating that technological specialization improves bank performance. To identify causality, we use a structural model of bank-borrower matching that shows how matching banks with borrowers of similar technological capabilities maximizes economic surplus. We also use bank M&As as quasi-natural experiments that exogenously increase banks' technological expertise. Further analysis shows that the effect of technology similarity on loan spreads is more pronounced for opaque borrowers with lower technological obsolescence, less risky firms, and smaller or capital-constrained banks. These findings contribute to our understanding of loan pricing by showing that banks' technological expertise transcends traditional industry boundaries and represents a previously unexplored dimension of bank specialization that influences credit allocation.

In the second study, we develop a novel AI-driven Co-Lending Graph Neural Network (CoLGNN) model to capture the dynamics of risk spillovers in syndicated lending markets. Our approach combines bank and loan characteristics with network topology to create a robust spillover risk measure (*CLN score*) that predicts future bank risk and performance up to two years ahead. Using quasi-natural experiments on credit rating downgrades and Lehman Brothers' collapse, we validate that risk indeed propagates through lending connections rather than through common exposures. The co-lending network risk score is an effective early warning indicator for forecasting loan loss provisions, non-performing loans, and profitability. Specifically, our semi-supervised learning framework enhances the risk monitoring capability for private banks that are typically outside the scope of conventional market-based risk measurement methodologies. Further, we demonstrate that risk transmission predominantly takes place through revolving credit facilities, where lead arrangers have ongoing monitoring responsibilities. Our empirical study demonstrates that the predictive power of our measure is especially pronounced for more vulnerable banks that are smaller in size, with lower capital adequacy, and with greater organizational complexity. In addition, our deep learning method significantly outperforms standard network centrality metrics and accounting-based Z-scores, with the predictive power of the *CLN score* remaining persistently significant even when including stock returns and bank-specific fixed effects. By showing previously undisclosed banking vulnerabilities, our deep learning framework provides regulators and market players with an effective early-warning system for spillover risk management in highly interconnected banking systems.

In the third study, we measure the impact of antitrust enforcement on firm performance using government procurement as a natural laboratory. We apply large language models to classify antitrust procurement-related cases from 2001 to 2021 and employ difference-in-differences estimates to study the impact of antitrust lawsuits on firms in the procurement sector. The results indicate that non-defendant firms in affected product markets gain significantly improved access to government contracts, grow their workforce, and increase their sales following antitrust lawsuits. Public non-defendant firms also face higher stock market valuations, with investors anticipating improved future performance. In contrast, defendant firms experience a significant reduction in government contracts, reflecting a significant reallocation of market shares. In product markets, we show increased entry and decreased concentration following antitrust enforcement action, with the largest gains accruing to larger procurement markets and cases of fraud, corruption, and bribery, in which defendant companies are excluded systematically. Contrary to the traditional argument that removing regulations on anti-competitive behavior would primarily help small firms and new entrants, we find that larger, established firms capture most of the benefits, whereas smaller, financially constrained businesses struggle to gain a meaningful market share in the newly competitive environment. Despite increased competition, government procurement costs remain largely unchanged, possibly because more competitive bidding leads to more costly contract renegotiation down the line. These findings offer important insights into when and how antitrust enforcement most effectively reshapes market structures.

More broadly, this study advances the understanding of technological innovation, banking, market competition, machine learning, finance, and government procurement. Through the application of advanced techniques such as graph neural networks and combining large language models, this study reveals the underlying mechanisms in the economy that are important to financial markets and corporate performance. The first paper demonstrates how banks' technological specialization creates real economic value by minimizing information frictions and enhancing the efficiency and availability of credit to innovative firms. The second study employs a deep learning approach to capture new risk transmission channels in previously unknown lending networks to equip regulators with advanced monitoring tools for private banks that have traditionally been beyond the reach of conventional risk model assessment frameworks. The third study uses an ideal natural laboratory to analyze how antitrust enforcement transforms market structures, revealing unanticipated patterns in the distribution of competitive advantages after regulation. This thesis offers substantial practical value to various stakeholders in financial markets. For financial institutions and banking regulators, our findings on technological specialization suggest potential mechanisms to alleviate the underfunding of innovative firms by recognizing how specialized knowledge accumulation can improve screening efficiency, potentially justifying targeted support during the initial adoption phases of emerging technologies. For central banks and financial stability authorities, our deep learning framework for co-lending networks provides an AI-driven approach to monitoring risk transmission and is particularly valuable for detecting vulnerabilities for private banks. For antitrust authorities and procurement agencies, our evidence on the distributional consequences of competition enforcement reveals important market reallocation frictions that may necessitate complementary policies to support smaller firms when implementing interventions in government procurement markets. Overall, these empirical contributions advance both academic understanding and practical approaches to strengthen lending efficiency, risk management precision, and competitive market functioning in an increasingly complex economic environment.

In summary, this study introduces several possible avenues for future research. First, the established economic importance of technological specialization in bank lending motivates future research on how traditional banks compete with emerging FinTech lenders, which can have different technological expertise. Extending our conceptual model and empirical results to explore how banks adapt their technological expertise to FinTech competition could provide important insights into the digital transformation of the banking industry. Meanwhile, future research could investigate the degree to which bank specialization extends to other intangible assets and unpatented innovative activities, which are increasingly key drivers of firm value but are difficult for lenders to assess. Second, our graph neural network approach opens the door to studying the effects of regulatory interventions on risk transmission processes in financial networks and could be applied to other interconnected systems, such as digital payment systems. Third, the antitrust results also call for further investigation into the potential crowding-out effects of enforcement activity, particularly whether gains for some market participants come at the expense of others in related markets. Furthermore, the intersection of three areas of research—technological expertise, network risk transmission, and competitive market dynamics—calls for exploration in the context of digital platforms and the entry of large technology firms into financial services. Future research could shed light on how the unique technological advantages of alternative data may alter competitive dynamics, shift risk transmission channels, and require different regulatory approaches to support an increasingly interconnected global capital market. The three empirical studies in this thesis provide important first steps in these directions, establishing evidence on technological expertise in lending, network-based risk transmission, and competition dynamics that future scholars can build upon.

Appendix A

Internet Appendix for Chapter Two: A structural matching model

A.1 A structural bank-borrower matching model

In this section, we present and estimate a structural model of bank-borrower matching building on Fox (2017, 2018), Fox et al. (2018), and Schwert (2018) to show that the technology similarity between a borrower and bank's prior borrowers is a major determinant of the bank lending decision that results in a bank-borrower match (i.e., loan origination). Specifically, we are interested in how bank's technology knowledge of the borrower, as measured by the similarity measure at the time of loan origination, determines loan spread. However, we cannot observe any counterfactual bank-borrower matching assignments so that we do not have the loan spread a firm would pay if it borrowed from a different bank. The Fox (2018) model addresses this challenge by modelling transfer payments (e.g., the loan spread a firm pays to the bank) as unobservable in the equilibrium condition, and provides a way to estimate the model without data on transfer payments. The model thus enables us to identify the drivers of observed bank-borrower matching assignments in the absence of unobservable non-matching assignments.

In our context, applying the Fox (2018) model treats the observed bank-borrower matches (loans) as outcomes to be explained by a latent match value function. Estimating the function relies on the concept of pairwise stability in equilibrium, which states that neither the bank nor the borrower in an existing match would see an advantage in dissolving their current match in favor of matching with other firms or banks. Such pairwise stability condition leads to full stability under substitutable preferences (Schwert, 2018).

Formally, let there be a space of loans Ω, $\Psi \subseteq \Omega$ the set of borrowing activities for firm f and $\Phi \subseteq \Omega$ the set of lending activities for bank b. Given a value function $V_f(\Psi)$ for firm f, $V_b(\Phi)$ for bank b and total transfer payment (e.g., interests, fees, and other benefits) r_ι for loan $\iota \in \Omega$, the surplus for firm f borrowing loans Ψ is $V_f(\Psi) - \sum_{\iota \in \Psi} r_\iota$, and the surplus for bank b lending loans Φ is $V_b(\Phi) + \sum_{\iota \in \Phi} r_\iota$. Firm f and bank b search for Ψ and Φ, respectively, that maximize their own surpluses.

Consider two actual bank-borrower matches (b_1, f_1) and (b_2, f_2). From the bank's perspective, the pairwise stability condition states that for each bank-borrower match, the bank lending to the firm yields a higher value than to the other firm:

$$\begin{aligned} V_b(b_1, f_1) + r(b_1, f_1) &\geq V_b(b_1, f_2) + \underbrace{r(b_2, f_2) + [V_f(b_1, f_2) - V_f(b_2, f_2)]}_{\text{maximum } f_2 \text{ would pay } b_1 \text{ to switch from } b_2} \\ V_b(b_2, f_2) + r(b_2, f_2) &\geq V_b(b_2, f_1) + \underbrace{r(b_1, f_1) + [V_f(b_2, f_1) - V_f(b_1, f_1)]}_{\text{maximum } f_1 \text{ would pay } b_2 \text{ to switch from } b_1} \end{aligned} \tag{A.A1}$$

Summing these pairwise stability conditions yields a condition without the transfer payments r, which is unobservable for counterfactual matches (b_1, f_2) and (b_2, f_1):[1]

$$V(b_1, f_1) + V(b_2, f_2) \geq V(b_2, f_1) + V(b_1, f_2) \tag{A.A2}$$

where $V = V_b + V_f$, representing the total economic surplus for banks and borrowers (Schwert, 2018). Intuitively, such a condition implies that the actual matching assignments should lead to higher total surplus than counterfactual matches.[2] As a result, it shows that the value function is driven by the match characteristics rather than factors specific to banks or borrowers. Because the model considers all possible matching assignments, all bank and borrower characteristics enter the inequality on both sides and are hence canceled out.

To estimate the model, we follow Fox (2018) and Schwert (2018) and parameterize $V(b, f)$ as a linear function:

$$V(b, f) = X'_{b\times f}\theta + \varepsilon_{b,f} \tag{A.A3}$$

where $X_{b\times f}$ represents the vector of bank-borrower match characteristics. The objective function for estimating the parameter vector θ developed by Fox (2018) is the sum of the indicators of all pairwise matching maximum score inequality (i.e., Equation A.A2), which takes the following form with the linear parameterization of $V(b, f)$:

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \sum_{(b_m, f_n)\in G_t} \mathbb{1}[X'_{b_1\times f_1}\theta + X'_{b_2\times f_2}\theta \geq X'_{b_1\times f_2}\theta + X'_{b_2\times f_1}\theta] \tag{A.A4}$$

[1] The inequality condition in Equation A.A2 is at the core of Fox (2017) and Schwert (2018). Counterfactual bank-borrower matches (loans) have no observable transfer payments like loan spreads. If transfer payments remain in the inequality condition, the model cannot be estimated.

[2] Noticeably, the model uses a subset of all possible matching cases as it excludes counterfactuals such as a bank lending to both firms. However, Bajari et al. (2007) and Fox (2007) show that parameter estimates are consistent as long as more valuable matches are more likely to occur.

where G_t denotes the set of all possible pairwise matching assignments, factual and counterfactual, in year t.[3] Following Schwert (2018), we restrict the samples to the loans with only one lead bank to avoid many-to-many matching complications.[4] Intuitively, maximizing the objective function aims to find the parameter θ that yields the higher occurrence of observed factual matching assignments. We solve for the maximum score estimator θ using the Particle Swarm Optimization (PSO).[5]

We note that the bank-borrower match characteristics $X_{b \times f}$ are observable even for counterfactual matching assignments. For example, assuming that bank A never lends to firm B, the counterfactual A-B match characteristics such as their geographical distance, the technology similarity (between firm B and bank A's prior borrowers) are still known. However, loan characteristics unobservable for counterfactual matches are necessarily excluded.[6] Following Schwert (2018), our match characteristic vector consists of a series of bank-borrower joint characteristics, including the borrower's bank-dependence, the bank-borrower geographical distance, prior lending relationship and the interactions of characteristics of the bank and the borrower. More importantly, we include the borrower's technology similarity with the bank's prior borrowers, and additionally, the borrower product market rivalry effect using the Hoberg and Phillips (2016) Herfindahl-Hirschman Index.

Table A.A1 presents the semi-parametric matching results and the point estimation of parameter vector θ. Given that the p-value is not obtainable in the parameter estimation of such an inequity condition, we compute confidence intervals following the Schwert (2018) bootstrapping method. The positive coefficients of technology similarity across all model specifications indicate that more value is generated by matching banks and borrowers with technology profiles that are similar to the banks' prior borrowers. Consistent with Schwert (2018), we find that well-capitalized banks are more likely to match with bank-dependent firms and that banks and firms located closer to each other or have prior lending relationships are also more likely to match. Similarly, we document a positive assortative matching by size, albeit statistically insignificant, possibly due to the reduced sample size as a result of requiring patent data. We further find

[3] Consistent with Schwert (2018), we consider each year as a separate market in which we construct counterfactual matches. Specifically, within a calendar year. More specifically, counterfactual matches are those bank-firm pairs that do not have a loan in the year.

[4] Schwert (2018) argue that many-to-many matching estimators are complicated to interpret in the case of bank-firm joint characteristics. He shows that all common panel multivariate regressions provide similar results for the sub-sample with only one lead arranger. Fox (2007) provide the theoretical foundation that maximum score estimators are consistent with the sub-sample analysis.

[5] The PSO method (Eberhart & Kennedy, 1995) uses a population (Swarm) of possible solutions (Particles), where possible solutions move around the search space guided by their own best-known positions as well as the whole population's optimal position (Bonyadi & Michalewicz, 2017). According to Fox (2018), the differential evolution (DE) method is an alternative option for solving the Equation A.A4. A comparison between PSO and DE could be found in Das et al. (2008). PSO does not use the gradient of the objective function and is less likely to end in a locally optimal point via searching a large space of candidate solutions, which is helpful in our setting with a large-size counterfactual matching sample. We appreciate the Mathematica code provided by Jeremy Fox.

[6] For example, we cannot observe the spread of a loan that never exists and would have to derive a pricing function should we attempt to include such loan characteristics into $X_{b \times f}$.

Table A.A1: **Structural Bank-Borrower Matching Model Estimation**

Table A.A1 shows the result of the semi-parametric bank-borrower matching model following the Fox (2018) framework. We follow Schwert (2018) to create a series of bank-firm joint characteristics. The key variable of interest is borrowers' technology similarity. Technology similarity measurement is specific at the lender-borrower level each year (or each independent market in our semi-parametric matching setting). Following Abrevaya and Huang (2005), we define the significance of the point estimate using the 95% confidence interval which is generated by drawing 1,000 sub-samples with replacement. We present 95% confidence interval in parentheses below the corresponding coefficient with statistical significance denoted as follows: ** (if the point estimate is within the 95 % confidence interval).

	Point Estimation of the Parametric Vector				
	(1)	(2)	(3)	(4)	(5)
Technology Similarity	7.972**	7.990**	7.881**	4.239**	4.081**
	[7.042, 12.374]	[7.068, 16.250]	[6.909, 16.091]	[1.593, 9.634]	[1.363, 7.822]
Borrower Bank-Dependent × Bank Capital	5.249**	4.839**	8.089**	9.442**	8.100**
	[3.068, 11.120]	[7.749, 11.651]	[7.211, 15.965]	[9.323, 18.368]	[7.228, 13.861]
ln(Geographic Distance)	-3.281**	-5.014**	-6.952**	-3.673**	
	[-4.267, -0.196]	[-6.370, -2.725]	[-9.587, -5.552]	[-4.530, -0.768]	
Borrower Size × Bank Size	-0.150	-0.243**	0.227		-0.094
	[-2.591, 0.242]	[-1.573, 0.040]	[-1.946, 0.915]		[-0.955, 0.153]
Borrower Product Market HHI × Bank Size	5.615**	3.436**		6.962**	2.205
	[3.602, 11.430]	[0.422, 6.770]		[5.914, 12.185]	[-0.819, 5.557]
Borrower In Bank Top Industries	1.115**		1.003**	4.559**	3.823**
	[1.032, 4.759]		[0.989, 6.055]	[4.312, 9.784]	[3.329, 7.517]
Prior Relationship	3.567**	5.627**	6.443**	7.467**	9.302**
	[0.614, 5.076]	[3.619, 7.376]	[4.810, 9.169]	[6.304, 10.615]	[8.982, 13.489]
Number of Inequalities	616,542	616,542	616,542	616,542	616,542
Satisfied Inequalities	0.99	0.99	0.98	0.96	0.96

that borrowers with more market power tend to match with larger or well-capitalized banks. Statistically, the fit of the model is excellent with over 96% of pairwise stability conditions satisfied by our estimated parameters, which is comparable to or better than the fit reported by Schwert (2018) and other earlier papers.[7]

Overall, our estimation of the bank-borrower matching model suggests an equilibrium market outcome where the total economic surplus for banks and borrower firms can be enhanced by matching banks with firms sharing similar technology profiles with banks' prior borrowers. This result provides strong evidence supporting our Hypothesis 1 that the borrower's technology similarity with bank's prior borrowers is informative and a significant factor in the bank's lending decision-making process.

[7]Schwert (2018) reports a fit of model where over 90% of pairwise stability conditions satisfied.

Appendix B

Internet Appendix for Chapter Three: Additional Technical Details

B.1 Co-Lending Graph Neural Network (CoLGNN) Details

B.1.1 From Message-Passing to Mapping Risk Spillover

Conventional econometric methods are not effective in capturing the rich topological information inherent in financial networks. Network centrality measures are frequently used to analyze the role of focal entities in networks (e.g., El-Khatib et al., 2015; Richmond, 2019; Rossi et al., 2018). However, these approaches may not be able to fully capture the multidimensional relationships and dynamic interactions that characterize complex financial systems. Similarly, traditional deep learning (DL) frameworks, while effective at handling high-dimensional datasets, are primarily designed for Euclidean data structures. For instance, convolutional neural networks (CNNs) use hidden layers to identify spatially localized features, but they are limited to fixed-dimension data, requiring feature engineering to incorporate network structures. This adaptation can be challenging as it demands embedding representations that approximate network structure.

To address these challenges, we introduce the **Co**-**L**ending **G**raph **N**eural **N**etwork framework (CoLGNN), a customized deep learning framework built on Graph Neural Networks (GNNs), which have emerged as a powerful class of methods for handling graph-structured data. GNNs apply optimizable transformations across all graph components—nodes, edges, and global contexts—enabling the model to learn intricate patterns directly from the graph's structure. Recent advancements in GNN-related methods exhibit significant out-performance in a wide range of tasks such as node classification, link prediction, and clustering (Bronstein et al., 2017; Wu et al., 2020; Xu et al., 2019). GNN captures the locality of each node, aggregates the neighborhood information, and stacks multiple layers to estimate the coefficient and perform prediction tasks.

CoLGNN is specifically tailored to study risk dynamics in co-lending networks as defined in Section 3.3.1. It employs a message-passing mechanism, which iteratively aggregates information from each node's neighbors, allowing the model to learn and adapt to the topological characteristics unique to co-lending structures.

Message-Passing Paradigm: The Message-Passing paradigm follows a multi-layer scheme of updating node representations based on neighborhood aggregation. Let $\mathbf{h}_i^{(l)}$ represent node i's embedding at the l-th layer. The Message-Passing scheme is defined as:

$$\mathbf{h}_i^{(l)} = \mathrm{UP}\left(\mathbf{h}_i^{(l-1)}, \mathrm{AGGR}(\{\mathbf{h}_j^{(l-1)} : j \in \mathcal{N}(i)\})\right), \tag{B.A1}$$

where the $\mathrm{UP}(\cdot)$ and $\mathrm{AGGR}(\cdot)$ are aggregation function and update function, respectively (Hamilton et al., 2017). Starting from initial node features $\mathbf{H}^{(0)} = \mathbf{X}$ [1], CoLGNN learns neighbor-aggregated representations $\mathbf{H}$ after multiple layers, effectively capturing how risk signals propagate through the network.[2]

The message-passing mechanism in GNNs is particularly well-suited for modeling risk spillover, as it mimics how risk spreads from lead arrangers to other banks within a co-lending network. In syndicated lending markets, lead arrangers—responsible for loan origination, due diligence, and monitoring—not only allocate lending shares but also potentially transmit risks to participant banks. Through message-passing, information (or "messages") about a lead bank's risk level is iteratively propagated along directed edges to other banks. Each message encapsulates data on risk characteristics, enabling the model to aggregate and refine risk signals across the network. This iterative process effectively captures the diffusion of risk throughout the syndicate, accounting for both direct and indirect transmission pathways.

B.1.2 CoLGNN Framework

To model risk spillover in the co-lending network, we propose the **Co-L**ending **G**raph **N**eural **N**etwork framework (CoLGNN), which includes a directional diffusion convolution with edge consideration. To capture the diffusion process of risk across the co-lending network, we utilize the graph diffusion convolution operation first introduced by Li et al. (2017). The directional diffusion convolution over the l-th layer graph embedding $\mathbf{H}^{(l)}$ is defined as:

$$\mathbf{H}^{(l)} = \sum_{k=0}^{K-1} \left(\widetilde{\mathbf{A}}_{out}^{k}\mathbf{H}^{(l-1)}\mathbf{W}_{k,1} + \widetilde{\mathbf{A}}_{in}^{k}\mathbf{H}^{(l-1)}\mathbf{W}_{k,2}\right), \tag{B.A2}$$

[1] $\mathbf{X} = \left[\mathbf{x}_i \in \mathbb{R}^F : i = 1, ..., n\right]$

[2] $\mathbf{H}^{(l)} = \left[\mathbf{h}_i^{(l)} \in \mathbb{R}^D : i = 1, ..., n\right]$ for $l \in \{1, \ldots, L\}$

where K represents the diffusion step, $\widetilde{\mathbf{A}}_{out} = \mathbf{D}_{out}^{-1}\mathbf{A}$ and $\widetilde{\mathbf{A}}_{in} = \mathbf{D}_{in}^{-1}\mathbf{A}^{\top}$, where $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{D}_{out/in} \in \mathbb{R}^{n\times n}$ are graph adjacency matrix and out/in degree matrix respectively, and $\mathbf{W}_{k,out/in}$ are trainable weights for out/in flow components. In particular, the $\widetilde{\mathbf{A}}_{out/in}$ are dual state transition matrices, particularly in the context of a random walk on the graph. This matrix represents the probabilities of risk transmission moving out/in from one node to another in a one-step random walk.[3]

Edge (loan) attributes in co-lending networks carry significant implications for banks' risk dynamics.[4] Therefore, we incorporate both numerical and categorical loan features as edge attributions $\mathbf{e}_{i,t}$. Categorical attributes are represented through one-hot vector encoding. For multiple co-lending relationships between two nodes within a co-lending network, we employ loan amount weighted averages to consolidate edge attributes between nodes and modify the graph diffusion propagation scheme of Equation (B.A2) by merged edge attributions.

Consider node i on the graph, with $\mathcal{N}_{out}(i)$ and $\mathcal{N}_{in}(i)$ denoting the out and in direction neighbors of node i, respectively. From the node features $\mathbf{H}^{(l-1)}$ on the $(l-1)$-layer, we define the following node-wise edge attribution merging scheme:

$$\begin{aligned}\tilde{\mathbf{h}}_{out,i}^{(l-1)} &= \mathbf{h}_i^{(l-1)} + \sum_{j\in\mathcal{N}_{out}(i)} \mathrm{ReLU}(\mathbf{h}_j^{(l-1)} + \mathbf{e}_{i,j}\mathbf{W});\\ \tilde{\mathbf{h}}_{in,i}^{(l-1)} &= \mathbf{h}_i^{(l-1)} + \sum_{r\in\mathcal{N}_{in}(i)} \mathrm{ReLU}(\mathbf{h}_r^{(l-1)} + \mathbf{e}_{r,i}\mathbf{W}).\end{aligned} \tag{B.A3}$$

In this scheme, we merge two directions' edge attributions into the node features from layer $(l-1)$ to become two directions' edge augmented node features $\tilde{\mathbf{h}}_{i,out}$ and $\tilde{\mathbf{h}}_{i,in}$. We use the rectified linear unit (ReLU) as the activation function to introduce the non-linearity and sparsity. We leverage edge-augmented features to give diffusion graph convolution the capability for capturing edge information. To align with the message-passing paradigm, the new edge-augmented graph diffusion convolution can be expressed by the node-wise representation as:

$$\mathbf{h}_i^{(l)} = \sum_{k=0}^{K-1}\left(\sum_i \tilde{\mathbf{a}}_{out,i}^k\,\tilde{\mathbf{h}}_{out,i}^{(l-1)}\mathbf{W}_{k,out} + \sum_i \tilde{\mathbf{a}}_{in,i}^k\,\tilde{\mathbf{h}}_{in,i}^{(l-1)}\mathbf{W}_{k,in}\right), \tag{B.A4}$$

where $\tilde{\mathbf{a}}_{out/in,i} = 1/d_{out/in,i}$ are out/in direction transition probability, with $d_{out/in,i} = \mathbf{D}_{out/in,ii}$. In both schemes (B.A3) and (B.A4), the weight matrices $\mathbf{W}$, $\mathbf{W}_{k,in}$ and $\mathbf{W}_{k,in}$ are layer dependent. For simplicity, we omit the layer index l in the notation.

[3]Under the directional convolution framework, the risk spread-over process is considered to occur in two different transition states, reflecting both the leading and participating roles of banks in the syndicated loan market. These two transition matrices are designed to capture distinct patterns.

[4]Incorporating loan features as edge attributes is supported by existing research on lending, which identifies various loan characteristics related to bank risks, such as the syndicated structure (Ivashina, 2009), loan concentration (Gao et al., 2023), and loan covenants (Demiroglu & James, 2010).

We measure the bank-level CLN score using the output from the last layer of CoLGNN. For bank i in a given co-lending network G_t, the CLN score is estimated as:

$$CLN\ score_i = \sigma\left(\sum_{k=0}^{K-1}\left(\sum_i \tilde{\mathbf{a}}^k_{out,i}\,\tilde{\mathbf{h}}^{(L)}_{out,i}\mathbf{W}_{k,out} + \sum_i \tilde{\mathbf{a}}^k_{in,i}\,\tilde{\mathbf{h}}^{(L)}_{in,i}\mathbf{W}_{k,in}\right)\right) \tag{B.A5}$$

where $\sigma(\cdot)$ represents the sigmoid activation function. The terms $\tilde{\mathbf{a}}^k_{out,i}$ and $\tilde{\mathbf{a}}^k_{in,i}$ are the out/in direction transition probabilities for bank i, while $\tilde{\mathbf{h}}^{(L)}_{out,i}$ and $\tilde{\mathbf{h}}^{(L)}_{in,i}$ represent the edge-augmented node features from the final layer L of the network. These edge-augmented features incorporate both bank characteristics (node attributes) and loan relationship characteristics (edge attributes), enabling the model to capture the risk transmission through syndicated loan relationships. This calculation is performed for each bank within each network, yielding a vector of risk scores across all banks. By applying this procedure to each quarterly co-lending network G_t, we obtain the time series of CLN scores, denoted as $CLN\ score_{i,t}$ for bank i at year-quarter t.

B.1.3 Estimating Bank Co-Lending Network Risk

The CLN score is estimated using the CoLGNN framework in Section B.1.2 on a rich sample of bank characteristics from Form FR Y-9C and loan features from Dealscan. This appendix details our semi-supervised estimation process. As shown in *stage 1* of Figure 3.1, we calculate the quarterly buy-and-hold returns for all public banks. Banks are classified according to their performance, labeled as "safe" and "risky".[5] The remaining public banks and all private banks are treated as unlabeled samples.

Stage 2 of Figure 3.1 displays the design of co-lending network. For each year-quarter t, we construct a co-lending network using syndicated loans originated in the past five years (20 quarters) from $t-1$ to $t-20$. For example, the co-lending network $G_{t=1996Q1}$ includes the syndicated loan originated from 1991Q1 to 1995Q4. In total, we construct 100 co-lending networks ($\mathcal{G} = \{G_{t=1996Q1}, ..., G_{t=2020Q4}\}$). Each co-lending network is an attributed graph $G_t = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{E})$ where $\mathcal{V} = \{v_1, \ldots, v_{|\mathcal{V}|}\}$ is the set of bank holding companies (nodes), and $\mathbf{X}$ are the bank characteristics at year-quarter $t-1$. $\mathcal{E} \subset \mathcal{V}\times\mathcal{V}$ is the set of co-lending relationships (edges). Each edge $e_{i,j} \in \mathcal{E}$ is the directed edge points from bank i to bank j. $\mathbf{E}$ utilize the loan characteristics on the origination date. We aggregate multiple co-lending relationships between two banks within the rolling window into one co-lending relationship. The summary statistics of our co-lending network series are shown in Table B.A1.

[5]The classification is tested for robustness ranging from 20% to 40%, as detailed in the Supplementary Materials.

Next, we implement the CoLGNN model separately for each co-lending network G_t. Figure 3.2 illustrates the framework of the graph diffusion convolution module in CoLGNN. As shown in *Section 3.1* of Figure 3.2, we include the common bank characteristics from FR Y9-C as the node features and high-dimensional edge characteristics from DealScan. Both numerical and categorical data types are utilized. For categorical features, we apply one-hot vector encoding to transform attributes such as 'Seniority Type', 'Primary Purpose', and 'Repayment Type' into numerical vectors. When multiple co-lending relationships exist within the specified time window, we combine these by calculating the loan amount weighted average of individual features from each loan. Since the graph neural network framework requires complete feature matrices without missing values, we fill missing features appropriately: numerical variables are filled with mean or zero, and categorical variables are filled with "No" or "unknown".[6] [7]

We employ a semi-supervised learning framework to leverage the available stock market data from public banks to infer risk dynamics in private banks, which typically lack transparent financial data. We transform the risk estimation into a binary node classification task, labelling the safe group as risky group as $Y_{i,t} = 0$ and the risky group as $Y_{i,t} = 1$, with remaining banks unlabeled ($Y_{i,t} = NaN$). For model training, we split the labeled data into training, validation, and testing subsets in proportions of 70%, 20%, and 10%, respectively.

The training is executed over 200 epochs, employing an Adam optimizer with a learning rate and weight decay as specified in the model configuration. This optimizer is known for its effective handling of sparse gradients and adaptive learning rate adjustments. We also incorporate a learning rate scheduler to adjust the learning rate based on the validation loss, enhancing model convergence. Our CoLGNN model architecture includes two diffusion graph convolution modules, with each node passing through two layers of diffusion graph convolution. These layers are designed to incorporate both node features and edge attributes effectively, allowing the model to capture complex dependencies in the data:

[6] We fill the numerical variables with either mean or zero and categorical variables with either category "No" or category "unknown", according to the nature of variable characteristics. 1) We fill in zero for variables such as "All in Spread Undrawn", "Assignment Minimum", variables related to fees, variables related to participation structures, etc. 2) We fill the loan-amount-weighted mean for the variables such as "All In Spread Drawn bps", "Maturity", etc. 3) We fill the category "No" for variables such as "Secured Type", "Collateral Type", "Secondary Purpose", etc. 4) We fill the category "unknown" for variables such as "Distribution Method", "Tranche Type", "Repayment Type", etc.

[7] To drop the redundant categories and maintain a clear feature importance illustration, we aggregate related loan features from DealScan while preserving their information content. Our aggregation approach follows established syndicated loan market conventions: (1) We consolidate revolving credit facilities into a single "Revolver" indicator; (2) We distinguish between "Term Loan A" (typically held by banks) and "Institutional Term Loan" (Term Loans B through K, typically held by non-bank institutional investors) to reflect their distinct risk characteristics and monitoring intensity; (3) We group specialized facilities by economic function, including "Letter of Credit" facilities, "Acquisition Bridge" facilities, "Capex Construction" facilities, "Liquidity" facilities, and "Multi-Option" facilities; (4) We organize loan seniority into four meaningful categories: Senior Secured, Senior Unsecured, Subordinated, and Mezzanine, reflecting the debt priority structure; and (5) We consolidate distribution methods into Club Deal, General Syndication, and Undisclosed categories. To enhance economic interpretation, we use common generic categories (e.g., general purpose loans) as reference categories.

1. **Initial Feature Transformation:** Each node feature vector undergoes a transformation via the first diffusion graph convolution layer, which integrates incoming features from connected nodes, applying a Rectified Linear Unit (ReLU) activation and dropout regularization to prevent overfitting.

2. **Feature Aggregation and Classification:** The transformed features are then processed through a second diffusion graph convolution layer, aggregating further neighborhood information and passing through a softmax function to yield a two-dimensional output vector per node. This vector represents the probability of each bank being "safe" or "risky".

The highest performing model is selected based on the combined accuracy across the training and validation sets, ensuring the model not only fits well to the training data but also generalizes effectively to unseen data. The CoLGNN model's output is a two-dimensional probability vector showing the probability associated with the "risky" classification. We use the probability associated with the "risky" class as the estimated bank CLN score. Overall, by iterating through 200 epochs, the model dynamically adjusts its weights and biases to minimize prediction errors, refining its ability to distinguish between "safe" and "risky" banks based on their embedded features and topological structure within the co-lending network.

Algorithm 1: CoLGNN module

input : Graph $\mathcal{G}(\mathcal{E}, \mathcal{V})$; node features $\{\mathbf{x}_i, \forall i \in \mathcal{V}\}$; edge features $\{\mathbf{e}_{ij}, \forall ij \in \mathcal{E}\}$; out/in transition probabilities $\{\tilde{\mathbf{a}}_{out/in,i}, \forall i \in \mathcal{V}\}$; layers L; random-walk depth K; weight matrices $\mathbf{W}$; activation $\mathrm{ReLU}(\cdot)$; sigmoid activation $\sigma(\cdot)$

$\mathbf{h}_i^{(0)} \leftarrow \mathbf{x}_i, \forall i \in \mathcal{V}$;

for $l = 1, \ldots, L$ **do**

  for $i \in \mathcal{V}$ **do**

    $\tilde{\mathbf{h}}_{out,i}^{(l-1)} \leftarrow \mathbf{h}_i^{(l-1)} + \sum_{j \in \mathcal{N}_{out}(i)} \mathrm{ReLU}(\mathbf{h}_j^{(l-1)} + \mathbf{e}_{i,j}\mathbf{W})$;

    ; `// Edge fusion (out)`

    $\tilde{\mathbf{h}}_{in,i}^{(l-1)} \leftarrow \mathbf{h}_i^{(l-1)} + \sum_{r \in \mathcal{N}_{in}(i)} \mathrm{ReLU}(\mathbf{h}_r^{(l-1)} + \mathbf{e}_{r,i}\mathbf{W})$;

    ; `// Edge fusion (in)`

    $\mathbf{h}_i^{(l)} \leftarrow \sum_{k=0}^{K-1} \left(\sum_i \tilde{\mathbf{a}}_{out,i}^k \, \tilde{\mathbf{h}}_{out,i}^{(l-1)} \mathbf{W}_{k,out} + \sum_i \tilde{\mathbf{a}}_{in,i}^k \, \tilde{\mathbf{h}}_{in,i}^{(l-1)} \mathbf{W}_{k,in} \right)$;

    ; `// Diffusion conv`

  end

end

$\mathrm{CLN\_score}_i \leftarrow \sigma(\mathbf{h}_i^{(L)})$;

output: $\mathrm{CLN\_score}_i, \forall i \in \mathcal{V}$

B.1.4 Model Hyperparameter and Robustness

Our CoLGNN model employs a fixed set of hyperparameters across all rolling-window co-lending networks. Unlike machine learning models that tune hyperparameters to optimize performance for each specific dataset or sample period, we intentionally maintain the same configuration throughout the full sample period. This design choice shows that our results are driven by the underlying economic structure of co-lending relationships, rather than by statistical overfitting or period-specific optimization. The complete hyperparameter definition is as follows:

Parameter	**Value**	**Description**
Architecture Parameters		
Diffusion steps (K)	2	Number of graph convolution hops capturing risk propagation depth
Hidden channels	12	Dimension of hidden node representations
Output classes	2	Binary classification (safe/risky)
Epsilon (ϵ)	0.01	Numerical stability parameter for graph normalization
Training Parameters		
Learning rate	0.01	Adam optimizer step size
Weight decay	0.001	L2 regularization parameter to prevent overfitting
Dropout ratio	0.1	Dropout probability for additional regularization
Epochs	200	Number of complete passes through training data per quarterly network
Optimizer	Adam	Adaptive moment estimation optimizer
Data Processing		
Scaler type	StandardScaler	Feature normalization method (zero mean, unit variance)

The main findings of our study are robust to a range of commonly used hyperparameter values, including hidden channels (8, 10, 12), diffusion steps (1, 2, 3), and learning rates (0.001, 0.01, 0.1). The untabulated results show that, in all configurations, the predictive performance of the CLN score remains qualitatively consistent. This robustness, together with our use of fixed parameters throughout the sample period, supports the view that the CoLGNN model effectively captures network-based risk spillovers in syndicated lending, rather than simply fitting noise or adapting to specific market conditions.

Table B.A1: **Directed Bank Co-Lending Networks (1991-2020)**

Table B.A1 presents the summary statistics of bank co-lending networks from 1991Q1 to 2020Q4.

1991-2000			2001-2010			2011-2020		
Time Frame	**Nodes**	**Edges**	**Time Frame**	**Nodes**	**Edges**	**Time Frame**	**Nodes**	**Edges**
1991Q1-1995Q4	77	530	2001Q1-2005Q4	164	1306	2011Q1-2015Q4	179	2279
1991Q2-1996Q1	80	543	2001Q2-2006Q1	164	1313	2011Q2-2016Q1	179	2300
1991Q3-1996Q2	81	587	2001Q3-2006Q2	162	1329	2011Q3-2016Q2	178	2296
1991Q4-1996Q3	80	596	2001Q4-2006Q3	163	1347	2011Q4-2016Q3	189	2485
1992Q1-1996Q4	81	631	2002Q1-2006Q4	165	1366	2012Q1-2016Q4	189	2574
1992Q2-1997Q1	84	657	2002Q2-2007Q1	167	1380	2012Q2-2017Q1	187	2625
1992Q3-1997Q2	85	707	2002Q3-2007Q2	167	1392	2012Q3-2017Q2	189	2764
1992Q4-1997Q3	86	730	2002Q4-2007Q3	169	1433	2012Q4-2017Q3	186	2790
1993Q1-1997Q4	87	751	2003Q1-2007Q4	170	1450	2013Q1-2017Q4	187	2838
1993Q2-1998Q1	85	734	2003Q2-2008Q1	171	1465	2013Q2-2018Q1	186	2859
1993Q3-1998Q2	84	760	2003Q3-2008Q2	172	1473	2013Q3-2018Q2	182	2925
1993Q4-1998Q3	87	780	2003Q4-2008Q3	173	1474	2013Q4-2018Q3	181	2945
1994Q1-1998Q4	91	811	2004Q1-2008Q4	171	1457	2014Q1-2018Q4	181	2999
1994Q2-1999Q1	98	861	2004Q2-2009Q1	171	1443	2014Q2-2019Q1	180	3059
1994Q3-1999Q2	106	973	2004Q3-2009Q2	171	1461	2014Q3-2019Q2	179	3092
1994Q4-1999Q3	112	1031	2004Q4-2009Q3	168	1466	2014Q4-2019Q3	178	3128
1995Q1-1999Q4	116	1095	2005Q1-2009Q4	165	1476	2015Q1-2019Q4	175	3164
1995Q2-2000Q1	118	1121	2005Q2-2010Q1	160	1458	2015Q2-2020Q1	174	3149
1995Q3-2000Q2	122	1187	2005Q3-2010Q2	157	1495	2015Q3-2020Q2	171	3111
1995Q4-2000Q3	121	1212	2005Q4-2010Q3	157	1494	2015Q4-2020Q3	170	3072
1996Q1-2000Q4	122	1223	2006Q1-2010Q4	158	1563	2016Q1-2020Q4	166	3028
1996Q2-2001Q1	124	1222	2006Q2-2011Q1	156	1576			
1996Q3-2001Q2	126	1233	2006Q3-2011Q2	150	1609			
1996Q4-2001Q3	127	1231	2006Q4-2011Q3	149	1615			
1997Q1-2001Q4	126	1242	2007Q1-2011Q4	145	1619			
1997Q2-2002Q1	124	1219	2007Q2-2012Q1	150	1618			
1997Q3-2002Q2	126	1202	2007Q3-2012Q2	154	1689			
1997Q4-2002Q3	135	1194	2007Q4-2012Q3	154	1725			
1998Q1-2002Q4	132	1210	2008Q1-2012Q4	156	1757			
1998Q2-2003Q1	140	1235	2008Q2-2013Q1	157	1789			
1998Q3-2003Q2	142	1244	2008Q3-2013Q2	155	1786			
1998Q4-2003Q3	148	1264	2008Q4-2013Q3	159	1838			
1999Q1-2003Q4	148	1268	2009Q1-2013Q4	160	1903			
1999Q2-2004Q1	151	1270	2009Q2-2014Q1	160	1952			
1999Q3-2004Q2	152	1247	2009Q3-2014Q2	162	2009			
1999Q4-2004Q3	156	1208	2009Q4-2014Q3	168	2073			
2000Q1-2004Q4	158	1257	2010Q1-2014Q4	171	2095			
2000Q2-2005Q1	161	1244	2010Q2-2015Q1	173	2127			
2000Q3-2005Q2	162	1289	2010Q3-2015Q2	174	2154			
2000Q4-2005Q3	164	1284	2010Q4-2015Q3	178	2240			

B.2 Network Centrality in the Bank Co-Lending Network

Traditional measures such as eigenvector centrality, closeness centrality, and eigenvector centrality do not account for the directionality of relationships, which is crucial in co-lending networks. To address this, we calculate both in-degree and out-degree versions of centrality measures. In the syndicated loan market, high in-degree centrality indicates a bank that frequently participates in loans arranged by others. Such banks may be more exposed to risks originating from multiple lead arrangers but may also benefit from diversification. High out-degree centrality suggests a bank that often acts as a lead arranger, initiating and structuring syndicated loans. These banks may have more control over loan terms but also bear greater responsibility for due diligence and potentially higher reputational risk.

Table B.A2: **Summary Statistics of Bank-level Network Centrality**

Table B.A2 presents the summary statistics bank-level directed network centrality at year-quarter level for each bank.

	Observations	Mean	10^{th} Percentile	Median	90^{th} Percentile	Standard Deviation
Bank-level Samples						
In-degree centrality	11688	0.080	0.006	0.044	0.215	0.082
Out-degree centrality	11688	0.084	0.000	0.000	0.335	0.153
Closeness centrality (in)	11688	0.209	0.155	0.201	0.276	0.047
Closeness centrality (out)	11688	0.228	0.000	0.000	0.594	0.268
Betweenness centrality (in)	11688	0.003	0.000	0.000	0.009	0.009
Betweenness centrality (out)	11688	0.003	0.000	0.000	0.009	0.009
Eigenvector centrality (in)	11688	0.065	0.007	0.043	0.160	0.060
Eigenvector centrality (out)	11688	0.048	0.000	0.000	0.183	0.074
Katz centrality	11688	0.032	-0.079	0.058	0.105	0.073
Katz centrality (reverse)	11688	0.016	-0.126	0.053	0.072	0.082
PageRank centrality	11688	0.007	0.003	0.005	0.015	0.005
PageRank centrality (reverse)	11688	0.008	0.001	0.001	0.025	0.011

We estimate the network centrality using the `NetworkX` package for directed graphs[8].

In-Degree Centrality $C_D^{in}(v_i)$ measures the number of incoming edges to a node, normalized by the maximum possible in-degree. **Out-Degree Centrality** $C_D^{out}(v_i)$ measures the number of outgoing edges from a node, normalized by the maximum possible out-degree.

$$C_D^{in}(v_i) = \frac{\deg^{in}(v_i)}{|\mathcal{V}| - 1} \tag{B.A6}$$

$$C_D^{out}(v_i) = \frac{\deg^{out}(v_i)}{|\mathcal{V}| - 1} \tag{B.A7}$$

Closeness Centrality (In) $C_C^{in}(v_i)$ assesses how close a node is to all other nodes based on incoming paths. **Closeness Centrality (Out)** $C_C^{out}(v_i)$ assesses how close a node is to all

[8] https://networkx.org/

other nodes based on outgoing paths.

$$C_C^{in}(v_i) = \frac{|\mathcal{V}| - 1}{\sum_{v_j \in \mathcal{V} \setminus \{v_i\}} d(v_j, v_i)} \tag{B.A8}$$

$$C_C^{out}(v_i) = \frac{|\mathcal{V}| - 1}{\sum_{v_j \in \mathcal{V} \setminus \{v_i\}} d(v_i, v_j)} \tag{B.A9}$$

Eigenvector Centrality (In) $C_{EV}^{in}(v_i)$ evaluates a node's influence based on the influence of its incoming neighbors. **Eigenvector Centrality (Out)** $C_{EV}^{out}(v_i)$ evaluates a node's influence based on the influence of its outgoing neighbors.

$$C_{EV}^{in}(v_i) = \frac{1}{\lambda} \sum_{v_j \in \mathcal{N}_{in}(v_i)} C_{EV}^{in}(v_j) \tag{B.A10}$$

$$C_{EV}^{out}(v_i) = \frac{1}{\lambda} \sum_{v_j \in \mathcal{N}_{out}(v_i)} C_{EV}^{out}(v_j) \tag{B.A11}$$

Betweenness Centrality (In) $C_B^{in}(v_i)$ quantifies the number of times a node acts as a bridge along the shortest paths directed towards it. **Betweenness Centrality (Out)** $C_B^{out}(v_i)$ quantifies the number of times a node acts as a bridge along the shortest paths originating from it.

$$C_B^{in}(v_i) = \sum_{s \neq v_i \neq t} \frac{\sigma_{st}(v_i)}{\sigma_{st}} \tag{B.A12}$$

$$C_B^{out}(v_i) = \sum_{s \neq v_i \neq t} \frac{\sigma_{st}(v_i)}{\sigma_{st}} \tag{B.A13}$$

Katz Centrality $C_K^{in}(v_i)$ considers all incoming walks to a node, with longer walks exponentially damped by a factor α. **Katz Centrality (reverse)** $C_K^{reverse}(v_i)$ considers all outgoing walks from a node, with longer walks exponentially damped by a factor α.

$$C_K^{in}(v_i) = \sum_{k=1}^{\infty} \alpha^k \cdot (\mathbf{A}^k)_{ji} \tag{B.A14}$$

$$C_K^{reverse}(v_i) = \sum_{k=1}^{\infty} \alpha^k \cdot (\mathbf{A}^k)_{ik} \tag{B.A15}$$

PageRank Centrality $C_{PR}^{reverse}(v_i)$ measures the probability of arriving at a node through a random walk that follows outgoing edges, incorporating a damping factor α.

$$C_{PR}(v_i) = \frac{1 - \alpha}{|\mathcal{V}|} + \alpha \sum_{v_j \in \mathcal{N}_{in}(v_i)} \frac{C_{PR}(v_j)}{\deg^{out}(v_j)} \tag{B.A16}$$

$$C_{PR}^{reverse}(v_i) = \frac{1-\alpha}{|\mathcal{V}|} + \alpha \sum_{v_j \in \mathcal{N}_{out}(v_i)} \frac{C_{PR}^{reverse}(v_j)}{\deg^{in}(v_j)} \tag{B.A17}$$

Notes: $\mathcal{V}$ denotes the set of all nodes in the graph. $\deg^{in}(v_i)$ and $\deg^{out}(v_i)$ represent the in-degree and out-degree of node v_i, respectively. $d(v_j, v_i)$ is the shortest path distance from node v_j to node v_i. $\mathcal{N}_{in}(v_i)$ and $\mathcal{N}_{out}(v_i)$ denote the sets of nodes with edges directed towards and away from node v_i, respectively. σ_{st} is the total number of shortest paths from node s to node t, and $\sigma_{st}(v_i)$ is the number of those paths that pass through node v_i. $\mathbf{A}$ is the adjacency matrix of the graph. λ is the largest eigenvalue of the adjacency matrix $\mathbf{A}$. α is the damping factor, typically set to 0.85 in PageRank calculations.

We estimate the following regression similar as Eq 2.3 with an additional network centrality control variable.

$$LLP_{i,t+h} = \beta_1 \, CLN\ score_{i,t} + \beta_1 \, Centrality_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + \varepsilon_{i,t+h} \tag{B.A18}$$

where $LLP_{i,t+h}$ is the loan loss provisions for bank i at time $t+h$, $CLN\ score_{i,t}$ refers to the bank-level co-lending network risk score of bank i at time t. $Centrality_{i,t}$ is one of the network centrality measures defined above $Centrality_{i,t} \in \{\ C_D^{in}(v_i),\ C_D^{out}(v_i), ...,\ C_{PR}(v_i),\ C_{PR}^{reverse}(v_i)\}$.

Appendix C

Internet Appendix for Chapter Four: Additional Technical Details

C.1 Appendix: Antitrust Data and Large Language Model

Antitrust Data

In this appendix, we provide the details on the construction of our comprehensive database of Department of Justice antitrust lawsuits. Our data collection relies on case summaries provided by Wolters Kluwer's Vital Law[1] (formerly known as Commerce Clearing House (CCH) Trade Regulation Reporter), which serves as the authoritative source for legal professionals and scholars in antitrust law. Our data source is identical to that used in Babina et al. (2023), as we access the same CCH Trade Regulation Reporter database through Vital Law, which is Wolters Kluwer's enhanced digital platform for delivering this content.[2]

The Department of Justice (DOJ) antitrust case summaries available through VitalLaw provide detailed information covering several key dimensions of each enforcement action: (1) legal identifiers, including case numbers, case names, filing dates, and docket information; (2) case details, describing alleged violations, legal proceedings, and final outcomes; (3) geographic information about where violations occurred; (4) temporal information about when violations began and ended; and (5) market information identifying affected industries and firms. Notably, these summaries serve as the authoritative source for several reasons. First, unlike the DOJ's official

[1] Wolters Kluwer rebranded its legal research platform, Cheetah, as VitalLaw on November 1, 2021. This transition expanded the platform's content and introduced new features to enhance legal research capabilities, including improved search functionality and digital accessibility. The CCH Trade Regulation Reporter, a key resource for antitrust case data, was integrated into VitalLaw during this rebranding, ensuring continued access to its comprehensive content while adding modern digital research tools.

[2] The Commerce Clearing House (CCH) Trade Regulation Reporter used by Babina et al. (2023) continues to be maintained and updated as part of VitalLaw following Wolters Kluwer's acquisition of CCH. The integration of the Trade Regulation Reporter into VitalLaw represents a technological advancement in content delivery while maintaining the underlying data source, enabling more efficient data collection while ensuring continuity and consistency with previous research using this database.

website, which primarily contains recent cases[3], VitalLaw maintains comprehensive coverage across all periods. Second, the summaries provide standardized information in a consistent format, facilitating systematic data collection. Third, the summaries include crucial details about geographic scope and industry classification that are often difficult to extract from raw court documents.

For systematic collection of these case summaries, we develop and implement an automated data collection procedure using Python and Selenium WebDriver.[4] To ensure data quality and consistency, our automated process implements multiple verification steps, including duplicate detection, automated cross-validation of key fields, and systematic logging of all extracted information.

For a systematic collection of case summaries, we employ a two-stage approach combining structural extraction and machine learning techniques. In the first stage, we develop Python-based automated extraction procedures to systematically collect and clean key identifying information from the case summaries, including filing dates, case names, and citation numbers. Our data extraction algorithm incorporates specialized error correction for common typographical errors and standardizes date formats across all entries. Case name extraction employs a sophisticated regular expression system that handles multiple formats and sources, combining information from both file names and document headlines while accounting for various legal naming conventions and potential inconsistencies.

Additional Details of Using LLM (ChatGPT)

Methodological Framework for Implementing LLM

The extraction of detailed information from antitrust case summaries presents unique challenges that limit the effectiveness of traditional rule-based extraction methods or manual collection. While basic information like dates and case numbers can be reliably extracted using regular expressions and pattern matching, many crucial aspects of antitrust cases require sophisticated contextual understanding and legal expertise. Legal documents often contain complex syntax, specialized terminology, and implicit references that create significant cognitive burden even for trained researchers, leading to inconsistent interpretations and potential oversight of critical details during manual review. For instance, distinguishing between different types of government contracts and accurately classifying federal procurement activities requires understanding both explicit and implicit references in the legal text, as well as the ability to interpret the broader

[3] The DOJ website's coverage is particularly limited before the mid-1990s, with significant gaps in historical case documentation(Babina et al., 2023).

[4] Our automated collection process ensures consistency and efficiency while maintaining data quality. The code systematically accesses and downloads case summaries from VitalLaw's DOJ Antitrust Division Case Summaries database, with built-in verification steps and logging mechanisms to prevent duplication and ensure completeness.

context of the case. To address these challenges, we employ LLMs to extract key fields related to each case.

In the second stage of our data collection process, we employ OpenAI's GPT-4 Omni to extract detailed case information requiring contextual understanding and legal expertise. For each legal text X, we produce a structured output Y comprising 24 key fields (e.g., merger indicators, contract involvement, defendant identities):

$Y = \{y_1, y_2, \ldots, y_{24}\}$

We model the conditional probability as:

$P(Y \mid X) = \prod_{i=1}^{24} P(y_i \mid X, y_1, \ldots, y_{i-1})$

Our extraction task selects $Y^* = \arg\max_Y P(Y \mid X)$, representing our best estimate of the case's structured information.

To enhance accuracy and capture the nuances inherent in legal texts, we implement a multi-role analysis system. Each case summary is independently processed from three distinct expert perspectives:

- *Economic professor role (temperature = 0.0, top_p = 0.1)*: ensures highly deterministic and precise economic classifications with minimal variation between runs, prioritizing consistent identification of economic concepts and relationships.
- *Data scientist role (temperature = 0.1, top_p = 0.2)*: allows for moderate creativity in identifying structured patterns while maintaining consistency, balancing between strict adherence to known patterns and flexibility in detecting novel relationships.
- *Legal specialist role (temperature = 0.15, top_p = 0.15)*: provides a middle ground that balances authoritative interpretation with the flexibility needed to recognize varied legal phrasings across different case documents.

For each role, the extraction process is executed R times per field (with $R = 3$ for most fields and $R = 4$ for key fields such as government procurement-related information), yielding a total of $N = 3R$ independent outputs for each field. Denote the outputs for field y_i as $y_i^{(j,k)} : j \in \{1, 2, 3\}$, $k \in \{1, \ldots, R\}$. We then determine the final accepted value y_i^* using a consensus criterion:

$$y_i^* = \text{mode}\left(\{y_i^{(j,k)}\}\right) \quad \text{if} \quad \frac{f\,(\text{mode})}{N} \geq \frac{2}{3},$$

where $f\,(\text{mode})$ is the frequency of the most common output. If this condition is not met, the field is flagged for manual review.

This consensus-based methodology efficiently replics the work of 9 to 12 independent human research assistants—a technique theoretically anticipated to exceed the accuracy of individual human extractions, as shown in works including Babina et al. While useful for extracting simple data (e.g., dates and case numbers), traditional rule-based approaches usually fail to capture complicated contextual signals such implicit references to controlling interests, joint ventures, and corporate restructuring language (e.g., Hart-Scott-Rodino Act references), which are critical for identifying M&A-related cases.

Furthermore, we utilize a robust JSON schema to define our output structure, and we employ a JSON structural formatting prompt (as detailed in He et al. (2024)) to optimize LLM performance. This multi-run, consensus-based approach ensures that our final dataset reliably reflects the true substance of the legal documents while mitigating the occasional ”hallucinations” or omissions inherent in LLM outputs.

Prompt Design

Our prompt design is tailored to extract precise information from legal documents by leveraging role-based expertise. The prompts are structured to maximize the contextual understanding capabilities of the large language model while minimizing hallucinations or factual inconsistencies.

The core system prompt varies based on the expert role:

```
# Economic Professor Role
"You are an economic professor specialized in antitrust enforcement."

# Data Scientist Role
"You are a data scientist specialized in extracting structured data from case summaries."

# Legal Specialist Role
"You are a professional lawyer specialized in antitrust lawsuits."
```

For each case analysis, we provide a consistent user message structure:

```
# User message with case identification
User: "Analyze the case ‘[CASE_NAME]’ and provide structured information as outlined."

# Case text follows
User: "[FULL_CASE_TEXT]"
```

The model is configured to return responses in a JSON format that strictly adheres to our predefined schema. This ensures consistency and facilitates subsequent automated analysis.

Example Prompt Implementation

Below is a complete example of how a prompt is constructed and sent to the API for the economic professor role as an example. The prompt consists of a system message establishing expertise, a user message identifying the analysis task, and the case text from regulatory documents: (only showing the first few sentences of the full case document):

```
# System message
System: "You are an economic professor specialized in antitrust enforcement."
# User message with case identification
User: "Analyze the case ‘United States v. MCC Construction Corp.’ and provide structured
information as outlined."
# Case text follows
User: "MCC Construction Corporation, of Greenwood Village, Colorado, was charged on January 5,
2016, in a one- count information filed in the federal district court in Washington, D.C. with
conspiring to defraud the U.S. government. According to the charge, MCC partnered with other
companies to gain access to GOVERNMENT CONTRACTS that were awarded through the Small Business
Administration’s 8(a) program for businesses controlled by a socially or economically
disadvantaged U.S. citizens, even though MCC was not eligible. The Justice Department announced
[on February 2, 2016] that MCC Construction Company (MCC) has agreed to pay $1,769,294 in
criminal penalties and forfeiture for conspiring to commit fraud on the United States by
illegally obtaining government contracts that were intended for small, disadvantaged businesses.
 The court agreement was announced [on February 2, 2016] by Assistant Attorney General William J
. Baer of the Justice Department’s Antitrust Division, U.S. Attorney Channing D. Phillips of the
 District of Columbia, Assistant Director in Charge Paul M. Abbate of the FBI’s Washington Field
 Office, Inspector General Peggy E. Gustafson of the Small Business Administration (SBA),
Inspector General Carol Fortine Ochoa of the U.S. General Services Administration (GSA), Special
 Agent in Charge Brian J. Reihms of the Defense Criminal Investigative Service’s (DCIS) Central
Field Office and Director Frank Robey of the U.S. Army Criminal Investigation Command’s Major
Procurement Fraud Unit (MPFU). [...]"
```

This classification task of government procurement related cases is crucial for our empirical analysis. Based on our structured schema requirements, the model returns a JSON response with the following key determinations:

```
{
  "Government_Contract_Indicator": "Yes",
  "Government_Contract_Keywords": "federal contracts, small, disadvantaged businesses, U.S.
    Department of Justice, [...]",
  "Federal_Procurement_Activities": "Yes",
  "Federal_Contract_Keywords": "federal contracts set aside for small, disadvantaged businesses,
     [...]",
  "Filing_Date": "2016-01-01"
}
```

We implement a comprehensive JSON schema that defines all relevant fields. This schema provides explicit guidance to the model regarding the expected output format and classification criteria. The schema definition below shows key fields related to government procurement identification, though our full implementation includes additional fields for comprehensive case analysis:

```
schema = {
    "name": "AntitrustCaseSummary",
    "schema": {
        "type": "object",
        "properties": {
            # Government Procurement Classification
            "Government_Contract_Indicator": {"type": "string", "enum": ["Yes", "No", "Unclear"]},
            "Government_Contract_Keywords": {"type": "string"},
            "Federal_Procurement_Activities": {"type": "string", "enum": ["Yes", "No", "Unclear"]},
            "Federal_Contract_Keywords": {"type": "string"},
            # Case Identification and Classification
            "Filing_Date": {"type": "string"},
            "M_A_Indicator": {"type": "string", "enum": ["Yes", "No", "Unsure"]},
            "Geographic_scope": {"type": "string", "enum": ["City", "State", "Several States", "National",
 "International", "No Information"]},
            # Industry Classification
            "NAICS4": {"type": "string"},
            "NAICS6": {"type": "string"},
            # Defendants Information
            "Defendants_Individual": {"type": "string"},
            "Defendants_Company": {"type": "string"},
            "Seller_state": {"type": "string"},
            "Product_state": {"type": "string"},
            # Legal Classification
            "Legal_code": {"type": "string", "enum": ["Sherman Act", "Clayton Act", "Robinson-Patman Act",
 "Hart-Scott-Rodino Act", "Other", "No Information"]},
            "Legal_outcome": {"type": "string", "enum": ["Pleaded Guilty", "Nolo Contendere", "Dismissed",
 "Dropped", "Enjoined", "Plea Agreement", "Found Guilty", "Found Not Guilty", "Consent Decree", "Other", "
No Information"]},
            "Types_of_violations": {"type": "string"},
            # Penalties and Appeals
            "Fine_imposed": {"type": "string"},
            "Jail_sentence_imposed": {"type": "string"},
            "Probation_sentence_imposed": {"type": "string"},
            "District_court_appeal": {"type": "string", "enum": ["Yes", "No"]},
            "Appellate_cout_appeal": {"type": "string", "enum": ["Yes", "No"]},
            "Supreme_court_appeal": {"type": "string", "enum": ["Yes", "No"]},
            # Timeline Information
            "Date_of_Plea": {"type": "string"},
            "Date_of_Sentencing": {"type": "string"},
            "Date_of_beginning_of_conspiracy": {"type": "string"},
            "Date_of_beginning_of_violation": {"type": "string"}
            [Additional fields omitted for brevity] }}}
```

This schema-guided approach ensures that the output is both comprehensive and consistent across all analyzed cases. For each antitrust lawsuit, we collect a complete structured output that contains all relevant fields. Below is an example of the full JSON output:


```
{
  "Government_Contract_Indicator": "Yes",
  "Government_Contract_Keywords": "federal contracts, Small Business Administration's 8(a) program,
    government contracts, disadvantaged businesses",
  "Federal_Procurement_Activities": "Yes",
  "Federal_Contract_Keywords": "federal contracts set aside for small, disadvantaged businesses",
  "Filing_Date": "2016-01-05",
  "Seller_state": "CO",
  "Product_state": "DC",
  "Defendant_Location": "CO",
  "NAICS4": "2362",
  "NAICS6": "236220",
  "M_A_Indicator": "No",
  "Defendants_Individual": "No Information",
  "Defendants_Company": "MCC Construction Corporation",
  "Geographic_scope": "National",
  "Legal_code": "Other",
  "Legal_outcome": "Plea Agreement",
  "Types_of_violations": "Government Fraud;Bid Rigging",
  "Fine_imposed": "$500,000;$1,269,294",
  "Jail_sentence_imposed": "No Information",
  "Probation_sentence_imposed": "No Information",
  "District_court_appeal": "No",
  "Appellate_cout_appeal": "No",
  "Supreme_court_appeal": "No",
  "Date_of_Plea": "2016-02-02",
  "Date_of_Sentencing": "2016-03-15",
  "Date_of_beginning_of_conspiracy": "2008-01-01",
  "Date_of_beginning_of_violation": "2008-01-01",
  [Additional fields omitted for brevity]
}
```


This methodological rigor is especially important for accurately identifying the subset of antitrust cases that involve government procurement activities—a critical distinction for our research focus. Additionally, to ensure maximum accuracy, all LLM outputs were independently validated by two human research assistants with professional expertise in antitrust law, using primary sources including Wolters Kluwer's Vital Law database, the Department of Justice's official website, and contemporaneous news reports. This dual validation process provides further confidence in the reliability of our final dataset.

Antitrust Violation Types

Additionally, we classify the antitrust violations into the following types by utilizing information from the DOJ Antitrust Division Manual and Vital Law's legal summaries. Our classification system reflects the established taxonomic framework employed by the Department of Justice, capturing both statutory authority and enforcement priorities.[5] For each antitrust violation classification, we restricted our large language model to select only from this predefined dictionary of violation types, applying a consensus-based methodology requiring agreement across multiple runs. All classifications were subsequently verified by human researchers with expertise in antitrust law, providing an additional validation layer to ensure classification accuracy.

```
schema = {
  "name": "Antitrust_Violation_Classification",
  "schema": {
    "properties": {
      "Broad_Category": {
        "enum": ["Horizontal Restraints", "Monopolization", "Vertical Restraints", "Merger Violations", "
     Unknown"]
      },
      "Specific_Violation": {
        "enum": ["Price Fixing", "Bid Rigging", "Market Allocation", "Group Boycott", "Information
     Exchange",
                "Monopolization", "Attempted Monopolization", "Predatory Pricing", "Exclusive Dealing",
                "Refusal to Deal", "Tying Arrangement", "Resale Price Maintenance", "Territorial
     Restriction",
                "Merger to Monopoly", "Horizontal Merger", "Vertical Merger", "Conglomerate Merger",
                "HSR Violation", "Other", "Unknown"]
      },
      "Collusion_Tools": {
        "enum": ["Bribery", "Wire Fraud", "Mail Fraud", "Government Fraud", "Tax Evasion",
                "Money Laundering", "Obstruction of Justice", "None", "Unknown"]
      }
    },
    "required": ["Broad_Category", "Specific_Violation", "Collusion_Tools"]
  }
}
```

Horizontal restraints involve agreements between competitors at the same level of production that directly restrict competition, typically prosecuted as per se violations under Section 1 of the Sherman Act and representing the DOJ's highest criminal enforcement priority. Monopolization captures conduct by a dominant firm that maintains or acquires monopoly power through exclusionary practices, prosecuted under Section 2 of the Sherman Act and requiring both market power and anticompetitive conduct elements. Vertical Restraints encompass restrictions imposed by firms at different levels of the supply chain that may unreasonably restrain trade, typically evaluated under the rule of reason and balancing procompetitive efficiencies against anticompetitive effects.[6] Merger Violations include both substantive challenges to transactions that may substantially lessen competition and procedural violations of Hart-Scott-Rodino Act

[5] This classification follows the DOJ Antitrust Division Manual (5th edition), which structures enforcement actions along these primary categories. See U.S. Department of Justice, Antitrust Division Manual (5th ed.), available at https://www.justice.gov/atr/division-manual.

[6] This categorization follows both DOJ enforcement practice and Supreme Court precedent, which has increasingly recognized the potential procompetitive benefits of vertical arrangements.

premerger notification requirements, with the latter constituting a distinct category of enforcement activity.[7] Furthermore, we classify the collusion mechanisms, capturing the tools used to facilitate anticompetitive conduct.

C.2 Example Antitrust Lawsuits

Example 1: United States v. MCC Construction Corp.

Case Summary: In January 2016, MCC Construction Company, a construction management and general contractor, was charged with one count of conspiracy to commit major fraud against the United States (Case No. 1:16-cr-00004). The U.S. Department of Justice (DOJ) alleged that MCC conspired with two companies eligible for federal contracts set aside for small, disadvantaged businesses, with MCC performing the work while the eligible companies submitted the bids. MCC waived indictment, agreed to the filing of the information, and accepted responsibility for its actions.

Relation to Procurement Activities: MCC Construction Company conspired with two small, disadvantaged businesses to illegally win federal contracts. These contracts were set aside under programs like the Small Business Administration (SBA) 8(a) Business Development Program, which reserves certain government contracts for eligible small businesses to promote fairness and opportunity. This case ties directly to government procurement, involving federal contracts set aside for small, disadvantaged businesses under the SBA 8(a) Business Development Program. MCC's fraud undermined these processes, potentially raising taxpayer costs and blocking legitimate small businesses from contract opportunities.

Field	Information
Case Filed	January 5, 2016
Defendants	MCC Construction Company
Industry (NAICS6)	Commercial and Institutional Building Construction (236220)
Court	United States District Court for the District of Columbia (Case No. 1:16-cr-00004)
Legal Basis	18 U.S.C. § 371 (conspiracy to defraud the United States) and 18 U.S.C. § 1031 (major fraud against the United States)
Key Events	- January 2016: Criminal information filed by the DOJ.
	- February 2, 2016: U.S. District Judge Ketanji B. Jackson accepts MCC's guilty plea.
	- March 15, 2016: Sentencing hearing; judgment entered by Judge Jackson.
Outcome	MCC pleaded guilty and agreed to pay $1,769,294 in criminal penalties and forfeiture.

[7] This distinction between substantive and procedural merger violations reflects the DOJ's operational approach, where HSR violations are pursued independently of competitive effect determinations.

Example 2: United States v. SG Interests I, Ltd., SG Interests VII, Ltd. and Gunnison Energy Corporation

Case Summary: SG Interests I, Ltd., SG Interests VII, Ltd. (collectively "SGI"), and Gunnison Energy Corporation (GEC) were charged with violating Section 1 of the Sherman Act for their agreement not to compete in bidding for natural gas leases auctioned by the Bureau of Land Management (BLM). On February 8, 2005, just days before a BLM auction, the companies executed a Memorandum of Understanding (MOU) agreeing that only SGI would bid as nominee for both parties at the February and May 2005 auctions, and if successful, would assign a 50% interest in the acquired leases to GEC at cost. As a result of this agreement, the United States received less revenue than it would have had the companies competed. The Department of Justice filed the civil antitrust complaint on February 15, 2012, marking the first time the DOJ challenged an anticompetitive bidding agreement for mineral rights leases.

Relation to Procurement Activities: This case directly relates to government procurement activities as it involved federal natural gas leases sold at auction by the U.S. Department of Interior's Bureau of Land Management. The bid-rigging scheme undermined the competitive bidding process designed to ensure the government receives fair market value for public resources. The case is particularly significant as it represents the first time the Department of Justice challenged anticompetitive bidding for federal mineral rights leases. The collusive agreement prevented the government from receiving the full competitive value for these natural resources, effectively depriving American taxpayers of revenue that should have been generated through a fair auction process.

Field	**Information**
Case Filed	February 15, 2012
Defendants	SG Interests I, Ltd., SG Interests VII, Ltd., and Gunnison Energy Corporation
Industry (NAICS6)	Crude Petroleum and Natural Gas Extraction (211111)
Court	United States District Court for the District of Colorado (Civil Action No. 12-cv-00395-RPM)
Legal Basis	Section 1 of the Sherman Act (15 U.S.C. § 1) and False Claims Act
Key Events	- February 8, 2005: SGI and GEC executed a MOU - February and May 2005: BLM auctions with collusive bidding - October 2009: Former GEC VP filed qui tam whistleblower complaint - February 15, 2012: DOJ filed civil antitrust complaint - December 12, 2012: Judge rejected initial settlement - April 22, 2013: Final judgment entered
Outcome	Companies paid $550,000 to settle antitrust and False Claims Act violations, with required advance notice of future joint bidding practices for five years.

Example 3: United States v. Scott "Max" Anthony Walker and Ryan Scott McMonigle

Case Summary: In 2009, Scott "Max" Anthony Walker and Ryan Scott McMonigle were charged with conspiring to violate the Anti-Kickback Act of 1986 by soliciting kickbacks from security vendors in connection with a subcontract under a $1.4 billion USAID contract for the Afghanistan Infrastructure Rehabilitation Project (AIRP). The scheme aimed to influence the bidding process for security services, with kickbacks initially set at $250,000 and later negotiated to 1.8% of the contract value. Both individuals pleaded guilty, with Walker entering a plea agreement in November 2009 and McMonigle in January 2010.

Relation to Procurement Activities: This case is directly tied to government procurement activities, as it involves a subcontract under a major USAID project funded by U.S. taxpayers. The kickback scheme sought to manipulate the competitive bidding process for security services, potentially increasing costs and undermining the integrity of the procurement process. By distorting fair competition, the conspiracy could have led to higher expenses for the government and compromised the effectiveness of international development efforts.

Field	Information
Case Filed	August 4, 2009
Defendants	Scott "Max" Anthony Walker, Ryan Scott McMonigle
Industry (NAICS Code)	Security Guards and Patrol Services (561612)
Court	United States District Court for the Eastern District of Virginia, Alexandria Division (Case No. 1:09-CR-478)
Legal Basis	18 U.S.C. § 371 (conspiracy) and 41 U.S.C. § 53 (Anti-Kickback Act)
Key Events	- August 4, 2009: Information filed under seal
	- August 26, 2009: Seal lifted
	- November 16, 2009: Plea agreement for Scott Anthony Walker
	- January 26, 2010: Guilty plea for Ryan Scott McMonigle
Outcome	Both defendants pleaded guilty to the conspiracy charges.

Example 4: United States v. Peter W. Schmidt

Case Summary: In 2001, Peter W. Schmidt, president of Schmidt Construction Company, was charged with conspiracy to defraud the United States by rigging bids for federally funded construction projects in Alabama. He was convicted and sentenced to 21 months in prison.

Relation to Procurement Activities: This case directly relates to government procurement, as it involved manipulating the bidding process for federal construction projects, potentially increasing costs and undermining fairness.

Example 5: United States v. Woodson & Associates Inc.

Case Summary: On September 29, 2005, Woodson & Associates Inc. agreed to plead guilty to bid rigging on electrical construction contracts at Cape Canaveral Air Force Station, violating

Field	Information
Case Filed	July 25, 2001
Defendants	Peter W. Schmidt
Industry (NAICS6)	Water, Sewer, and Pipeline Construction (237110)
Court	United States District Court for the Northern District of Alabama
Legal Basis	18 U.S.C. § 371 (conspiracy to defraud the United States)
Outcome	Convicted and sentenced to 21 months in prison.

Section 1 of the Sherman Act. The conspiracy, spanning from March 1998 to June 2002, involved projects for the Evolved Expendable Launch Vehicle program. Woodson was fined $175,000, with the plea subject to court approval. The investigation was conducted by the Antitrust Division, NASA, and the Air Force.

Relation to Procurement Activities: This case directly impacts government procurement as it involved bid rigging on military contracts at Cape Canaveral Air Force Station. The conspiracy likely resulted in higher costs and potentially lower quality for the government, affecting the Evolved Expendable Launch Vehicle program, which is crucial for national security. Maintaining fair competition in such procurement processes is essential to ensure cost-effectiveness and quality in defense and space projects.

Field	Information
Case Filed	September 29, 2005
Defendants	Woodson & Associates Inc.
Industry (NAICS6)	Electrical Contractors and Other Wiring Installation Contractors (238210)
Court	U.S. District Court for the Middle District of Florida
Legal Basis	Section 1 of the Sherman Act (15 U.S.C. § 1)
Key Events	- March 1998 to June 2002: Conspiracy to rig bids on electrical construction contracts at CCAFS
	- September 29, 2005: Woodson & Associates Inc. agrees to plead guilty and pay a $175,000 fine
Outcome	Woodson & Associates Inc. pleaded guilty to bid rigging and was fined $175,000

Bibliography

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, *138*(1), 1–35.

Abrevaya, J., & Huang, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica*, *73*(4), 1175–1204.

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, *105*(2), 564–608.

Acharya, V. V., & Mora, N. (2015). A crisis of banks as liquidity providers. *The Journal of Finance*, *70*(1), 1–43.

Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies*, *30*(1), 2–47.

Adelino, M., & Ferreira, M. A. (2016). Bank ratings and lending supply: Evidence from sovereign downgrades. *The Review of Financial Studies*, *29*(7), 1709–1746.

Agarwal, S., & Ben-David, I. (2018). Loan prospecting and the loss of soft information. *Journal of Financial Economics*, *129*(3), 608–628.

Agarwal, S., & Hauswald, R. (2010). Distance and private information in lending. *The Review of Financial Studies*, *23*(7), 2757–2788.

Aguzzoni, L., Langus, G., & Motta, M. (2013). The effect of eu antitrust investigations and fines on a firm's valuation. *The Journal of Industrial Economics*, *61*(2), 290–338.

Allen, F., & Gale, D. (2000). Financial contagion. *Journal of political economy*, *108*(1), 1–33.

Anderson, H., Paddrik, M., & Wang, J. J. (2019). Bank networks and systemic risk: Evidence from the national banking acts. *American Economic Review*, *109*(9), 3125–61.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, *111*(12), 4088–4118.

Armour, J., Mayer, C., & Polo, A. (2017). Regulatory sanctions and reputational damage in financial markets. *Journal of Financial and Quantitative Analysis*, *52*(4), 1429–1448.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, *135*(2), 645–709.

Babina, T., Barkai, S., Jeffers, J., Karger, E., & Volkova, E. (2023). *Antitrust enforcement increases economic activity* (tech. rep.). National Bureau of Economic Research.

Badoer, D. C., Demiroglu, C., & James, C. M. (2019). Ratings quality and borrowing choice. *The Journal of Finance*, *74*(5), 2619–2665.

Bahar, D., & Rapoport, H. (2018). Migration, knowledge diffusion and the comparative advantage of nations. *The Economic Journal*, *128*(612), F273–F305.

Bajari, P., Fox, J. T., & Ryan, S. P. (2007). Linear regression estimation of discrete choice models with nonparametric distributions of random coefficients. *American Economic Review*, *97*(2), 459–463.

Bajari, P., & Tadelis, S. (2001). Incentives versus transaction costs: A theory of procurement contracts. *Rand journal of Economics*, 387–407.

Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, *144*(2), 370–395.

Baker, J. B. (2002). The case for antitrust enforcement. *Journal of Economic perspectives*, *17*(4), 27–50.

Baker, J. B. (2019). *The antitrust paradigm: Restoring a competitive economy*. Harvard University Press.

Barkai, S. (2020). Declining labor and capital shares. *The Journal of Finance*, *75*(5), 2421–2463.

Barkley, A. (2023). The human cost of collusion: Health effects of a mexican insulin cartel. *Journal of the European Economic Association*, *21*(5), 1865–1904.

Barnatchez, K., Crane, L. D., & Decker, R. A. (2017). An assessment of the national establishment time series (nets) database.

Barrage, L., Chyn, E., & Hastings, J. (2020). Advertising and environmental stewardship: Evidence from the bp oil spill. *American Economic Journal: Economic Policy*, *12*(1), 33–61.

Barrot, J.-N., & Nanda, R. (2020). The employment effects of faster payment: Evidence from the federal quickpay reform. *The Journal of Finance*, *75*(6), 3139–3173.

Barwick, P. J., Liu, Y., Patacchini, E., & Wu, Q. (2023). Information, mobile communication, and referral effects. *American Economic Review*, *113*(5), 1170–1207.

Battiston, S., Puliga, M., Kaushik, R., Tasca, P., & Caldarelli, G. (2012). Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific Reports*, *2*(1), 541.

Beck, T., De Jonghe, O., & Mulier, K. (2022). Bank sectoral concentration and risk: Evidence from a worldwide sample of banks. *Journal of Money, Credit and Banking*, *54*(6), 1705–1739.

Benmelech, E., Dlugosz, J., & Ivashina, V. (2012). Securitization without adverse selection: The case of clos. *Journal of Financial Economics*, *106*(1), 91–113.

Bereskin, F., Byun, S. K., & Oh, J.-M. (2023). Technological fit and the market for managerial talent. *Journal of Financial and Quantitative Analysis*, *58*(2), 837–874.

Berg, T., Saunders, A., & Steffen, S. (2016). The total cost of corporate borrowing in the loan market: Don't ignore the fees. *The Journal of Finance*, *71*(3), 1357–1392.

Berger, A. N., & Bouwman, C. H. (2013). How does capital affect bank performance during financial crises? *Journal of Financial Economics*, *109*(1), 146–176.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, *119*(1), 249–275.

Besley, T., Fontana, N., & Limodio, N. (2021). Antitrust policies and profitability in nontradable sectors. *American Economic Review: Insights*, *3*(2), 251–265.

Bessen, J. (2020). Information technology and industry concentration. *Journal of Law & Economics*, *63*(3), 531.

Bessen, J. E. (2016). Accounting for rising corporate profits: Intangibles or regulatory rents? *Boston Univ. School of Law, Law and Economics Research Paper*, (16-18).

Bharath, S. T., Dahiya, S., Saunders, A., & Srinivasan, A. (2011). Lending relationships and loan contract terms. *The Review of Financial Studies*, *24*(4), 1141–1203.

Bharath, S. T., & Shumway, T. (2008). Forecasting default with the Merton distance to default model. *The Review of Financial Studies*, *21*(3), 1339–1369.

Bittlingmayer, G. (1993). The stock market and early antitrust enforcement. *The Journal of Law and Economics*, *36*(1, Part 1), 1–32.

Bittlingmayer, G., & Hazlett, T. W. (2000). Dos kapital: Has antitrust action against microsoft created value in the computer industry? *Journal of Financial Economics*, *55*(3), 329–359.

Bizjak, J. M., & Coles, J. L. (1995). The effect of private antitrust litigation on the stock-market valuation of the firm. *American Economic Review*, 436–461.

Blickle, K., Fleckenstein, Q., Hillenbrand, S., & Saunders, A. (2020). The myth of the lead arranger's share. *FRB of New York Staff Report*, (922).

Blickle, K., Fleckenstein, Q., Hillenbrand, S., & Saunders, A. (2022). The myth of the lead arranger's share. *Federal Reserve Bank of New York Staff Reports*, (922).

Blickle, K., He, Z., Huang, J., & Parlatore, C. (2024). Information-based pricing in specialized lending. *Working Paper*.

Blickle, K., Parlatore, C., & Saunders, A. (2023). Specialization in banking. *Working Paper*.

Bloom, N., Schankerman, M., & Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, *81*(4), 1347–1393.

Bonyadi, M. R., & Michalewicz, Z. (2017). Particle swarm optimization for single objective continuous space problems: A review. *Evolutionary Computation*, *25*(1), 1–54.

Botsch, M., & Vanasco, V. (2019). Learning by lending. *Journal of Financial Intermediation*, *37*, 1–14.

Boyd, J. H., & Prescott, E. C. (1986). Financial intermediary-coalitions. *Journal of Economic Theory*, *38*(2), 211–232.

Brogaard, J., Denes, M., & Duchin, R. (2021). Political influence and the renegotiation of government contracts. *The Review of Financial Studies*, *34*(6), 3095–3137.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine*, *34*(4), 18–42.

Brot, Z., Cooper, Z., Craig, S. V., & Klarnet, L. (2024). Is there too little antitrust enforcement in the us hospital sector? *American Economic Review: Insights*, *6*(4), 526–542.

Brunnermeier, M. K., & Oehmke, M. (2013). Bubbles, financial crises, and systemic risk. *Handbook of the Economics of Finance*, *2*, 1221–1288.

Bu, D., Keloharju, M., Liao, Y., & Ongena, S. (2023). Can lending hierarchies balance bias? the role of personal environmental values in credit to green firms. *Available at SSRN 4457192*.

Buccirossi, P., Ciari, L., Duso, T., Spagnolo, G., & Vitale, C. (2013). Competition policy and productivity growth: An empirical assessment. *Review of Economics and Statistics*, *95*(4), 1324–1336.

Byun, S. K., Oh, J.-M., & Xia, H. (2021). Incremental vs. breakthrough innovation: The role of technology spillovers. *Management Science*, *67*(3), 1779–1802.

Cabrales, A., Gottardi, P., & Vega-Redondo, F. (2017). Risk sharing and contagion in networks. *The Review of Financial Studies*, *30*(9), 3086–3127.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, *63*(6), 2899–2939.

Campello, M., & Gao, J. (2017). Customer concentration and loan contract terms. *Journal of Financial Economics*, *123*(1), 108–136.

Canayaz, M. I., Cornaggia, J., & Cornaggia, K. (2025). Choose your battles wisely: The consequences of protesting government procurement contracts. *Management Science*.

Cao, S., Jiang, W., Yang, B., & Zhang, A. L. (2023). How to talk when a machine is listening: Corporate disclosure in the age of ai. *The Review of Financial Studies*, *36*(9), 3603–3642.

Cao, S. S., Ma, G., Tucker, J. W., & Wan, C. (2018). Technological peer pressure and product disclosure. *The Accounting Review*, *93*(6), 95–126.

Carril, R., & Duggan, M. (2020). The impact of industry consolidation on government procurement: Evidence from department of defense contracting. *Journal of Public Economics*, *184*, 104141.

Carroll, C. D., Dynan, K. E., & Krane, S. D. (2003). Unemployment risk and precautionary wealth: Evidence from households' balance sheets. *Review of Economics and Statistics*, *85*(3), 586–604.

Carvalho, D. R., Gao, J., & Ma, P. (2023). Loan spreads and credit cycles: The role of lenders' personal economic experiences. *Journal of Financial Economics*, (2), 118–149.

Cestone, G., Li, J., & Volpin, P. F. (2021). Trading off stakeholder interests: Evidence from antitrust investigations. *Available at SSRN 3773448*.

Chami, I., Abu-El-Haija, S., Perozzi, B., Ré, C., & Murphy, K. (2022). Machine learning on graphs: A model and comprehensive taxonomy. *The Journal of Machine Learning Research*, *23*(1), 3840–3903.

Chava, S., Nanda, V., & Xiao, S. C. (2017). Lending to innovative firms. *The Review of Corporate Finance Studies*, *6*(2), 234–289.

Chava, S., Oettl, A., & Singh, M. (2023). Does a one-size-fits-all minimum wage cause financial stress for small businesses? *Management Science*, *69*(11), 7095–7117.

Chava, S., Oettl, A., Subramanian, A., & Subramanian, K. V. (2013). Banking deregulation and innovation. *Journal of Financial Economics*, *109*(3), 759–774.

Chava, S., & Roberts, M. R. (2008). How does financing impact investment? The role of debt covenants. *The Journal of Finance*, *63*(5), 2085–2121.

Chen, S., Miao, B., & Shevlin, T. (2015). A new measure of disclosure quality: The level of disaggregation of accounting data in annual reports. *Journal of Accounting Research*, *53*(5), 1017–1054.

Cheung, Y.-L., Rau, P. R., & Stouraitis, A. (2021). What determines the return to bribery? evidence from corruption cases worldwide. *Management Science*, *67*(10), 6235–6265.

Chodorow-Reich, G. (2014). The employment effects of credit market disruptions: Firm-level evidence from the 2008–9 financial crisis. *The Quarterly Journal of Economics*, *129*(1), 1–59.

Christensen, H. B., Maffett, M., & Rauter, T. (2024). Reversing the resource curse: Foreign corruption regulation and the local economic benefits of resource extraction. *American Economic Journal: Applied Economics*, *16*(1), 90–120.

Clark, R., Coviello, D., Gauthier, J.-F., & Shneyerov, A. (2018). Bid rigging and entry deterrence in public procurement: Evidence from an investigation into collusion and corruption in quebec. *The Journal of Law, Economics, and Organization*, *34*(3), 301–363.

Colonnelli, E., Lagaras, S., Ponticelli, J., Prem, M., & Tsoutsoura, M. (2022). Revealing corruption: Firm and worker level evidence from brazil. *Journal of Financial Economics*, *143*(3), 1097–1119.

Cong, L. W., Tang, K., Wang, J., & Zhang, Y. (2021). Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai. *Available at SSRN 3554486*.

Cornaggia, J., Mao, Y., Tian, X., & Wolfe, B. (2015). Does banking competition affect innovation? *Journal of Financial Economics*, *115*(1), 189–209.

Croci, E., Degl'Innocenti, M., & Zhou, S. (2021). Large customer-supplier links and syndicate loan structure. *Journal of Corporate Finance*, *66*, 101844.

Crocker, K. J., & Reynolds, K. J. (1993). The efficiency of incomplete contracts: An empirical analysis of air force engine procurement. *The RAND Journal of Economics*, 126–146.

Dannhauser, C. D. (2017). The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs). *Journal of Financial Economics*, *125*(3), 537–560.

Das, S., Abraham, A., & Konar, A. (2008). Particle swarm optimization and differential evolution algorithms: Technical analysis, applications and hybridization perspectives. In Y. Liu, A. Sun, H. T. Loh, W. F. Lu, & E.-P. Lim (Eds.), *Advances of Computational Intelligence in Industrial Systems* (pp. 1–38). Springer Berlin Heidelberg.

Dasgupta, S., & Žaldokas, A. (2019). Anticollusion enforcement: Justice for consumers and equity for firms. *The Review of Financial Studies*, *32*(7), 2587–2624.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–2996.

Demiroglu, C., James, C., & Velioglu, G. (2021). Why are commercial loan rates so sticky? The effect of private information on loan spreads. *Journal of Financial Economics*, *143*, 959–972.

Demiroglu, C., & James, C. M. (2010). The information content of bank loan covenants. *The Review of Financial Studies*, *23*(10), 3700–3737.

Deng, S., Willis, R. H., & Xu, L. (2014). Shareholder litigation, reputational loss, and bank loan contracting. *Journal of Financial and Quantitative Analysis*, *49*(4), 1101–1132.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The Review of Economic Studies*, *51*(3), 393–414.

Donelson, D. C., Flam, R. W., & Yust, C. G. (2022). Spillover effects in disclosure-related securities litigation. *The Accounting Review*, *97*(5), 275–299.

Dou, Y., Ryan, S. G., & Zou, Y. (2018). The effect of credit competition on banks' loan-loss provisions. *Journal of Financial and Quantitative Analysis*, *53*(3), 1195–1226.

Eberhart, R., & Kennedy, J. (1995). Particle swarm optimization. *Proceedings of the IEEE International Conference on Neural Networks*, *4*, 1942–1948.

El-Khatib, R., Fogel, K., & Jandik, T. (2015). Ceo network centrality and merger performance. *Journal of Financial Economics*, *116*(2), 349–382.

Elliott, M., Golub, B., & Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, *104*(10), 3115–3153.

Ellul, A., & Yerramilli, V. (2013). Stronger risk controls, lower risk: Evidence from us bank holding companies. *The Journal of Finance*, *68*(5), 1757–1803.

Erel, I. (2011). The effect of bank mergers on loan prices: Evidence from the United States. *The Review of Financial Studies*, *24*(4), 1068–1101.

Faccio, M., & McConnell, J. J. (2020). *Impediments to the schumpeterian process in the replacement of large firms* (tech. rep.). National Bureau of Economic Research.

Faccio, M., & Zingales, L. (2022). Political determinants of competition in the mobile telecommunication industry. *The Review of Financial Studies*, *35*(4), 1983–2018.

Fahlenbrach, R., Prilmeier, R., & Stulz, R. M. (2018). Why does fast loan growth predict poor performance for banks? *The Review of Financial Studies*, *31*(3), 1014–1063.

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

Fama, E. F. (1991). Efficient capital markets: Ii. *The Journal of Finance*, *46*(5), 1575–1617.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fazio, D., & Zaldokas, A. (2024). Kamikazes in public procurements: Bid-rigging and real non-market outcomes. *Available at SSRN 5043930*.

Fox, J. T. (2007). Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *The RAND Journal of Economics*, *38*(4), 1002–1019.

Fox, J. T. (2017). Specifying a structural matching game of trading networks with transferable utility. *American Economic Review*, *107*(5), 256–60.

Fox, J. T. (2018). Estimating matching games with transfers. *Quantitative Economics*, *9*(1), 1–38.

Fox, J. T., Yang, C., & Hsu, D. H. (2018). Unobserved heterogeneity in matching games. *Journal of Political Economy*, *126*(4), 1339–1373.

Frydman, C., & Papanikolaou, D. (2018). In search of ideas: Technological innovation and executive pay inequality. *Journal of Financial Economics*, *130*(1), 1–24.

Furman, J. (2016). Benefits of competition and indicators of market power. *Council of Economic Advisers (Issue Brief), The White House, Washington, April.*

Gallo, J. C., Dau-Schmidt, K. G., Craycraft, J. L., & Parker, C. J. (1994). Criminal penalties under the sherman act: A study of law and economics. *Rsch. in L. & Econ.*, *16*, 25.

Ganapati, S. (2018). Oligopolies, prices, output, and productivity. *Available at SSRN 3030966*.

Gande, A., & Lewis, C. M. (2009). Shareholder-initiated class action lawsuits: Shareholder wealth effects and industry spillovers. *Journal of Financial and Quantitative Analysis*, *44*(4), 823–850.

Gao, M., Hasan, I., Qiu, B., & Wu, E. (2023). Lone (loan) wolf pack risk. *SSRN Working Paper*.

Giannetti, M., & Saidi, F. (2019). Shock propagation and banking structure. *The Review of Financial Studies*, *32*(7), 2499–2540.

Giometti, M., & Pietrosanti, S. (2022). Bank specialization and the design of loan contracts. *FDIC Center for Financial Research Paper*, (2022-14).

Glasserman, P., & Young, H. P. (2015). How likely is contagion in financial networks? *Journal of Banking & Finance*, *50*, 383–399.

Glasso, A., & Schankerman, M. (2013). Patents and cumulative innovation: Causal evidence. *The Quarterly Journal of Economics*, *125*, 549–589.

Gleason, C. A., Jenkins, N. T., & Johnson, W. B. (2008). The contagion effects of accounting restatements. *The Accounting Review*, *83*(1), 83–110.

Goldman, J., & Zeume, S. (2023). Who benefits from anti-corruption enforcement? *Available at SSRN 3745751*.

Golub, B., & Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, *127*(3), 1287–1338.

Gopal, M. (2021). How collateral affects small business lending: The role of lender specialization. *Working Paper*.

Gopalan, R., Nanda, V., & Yerramilli, V. (2011). Does poor performance damage the reputation of financial intermediaries? evidence from the loan syndication market. *The Journal of Finance*, *66*(6), 2083–2120.

Greenwood, J., Sanchez, J. M., & Wang, C. (2010). Financing development: The role of information costs. *American Economic Review*, *100*(4), 1875–1891.

Gropp, R., & Heider, F. (2010). The determinants of bank capital structure. *Review of Finance*, *14*(4), 587–622.

Gruber, J. (1994). The incidence of mandated maternity benefits. *American Economic Review*, 622–641.

Grullon, G., Larkin, Y., & Michaely, R. (2019). Are US industries becoming more concentrated? *Review of Finance*, *23*(4), 697–743.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Gustafson, M. T., Ivanov, I. T., & Meisenzahl, R. R. (2021). Bank monitoring: Evidence from syndicated loans. *Journal of Financial Economics*, *139*(2), 452–477.

Gutiérrez, G., & Philippon, T. (2017). *Declining competition and investment in the us* (tech. rep.). National Bureau of Economic Research.

Haldane, A. G., & May, R. M. (2011). Systemic risk in banking ecosystems. *Nature*, *469*(7330), 351–355.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). *The NBER patent citation data file: Lessons, insights and methodological tools* (Working Paper No. 8498). National Bureau of Economic Research.

Hall, B. H., & Lerner, J. (2010). The financing of R&D and innovation. In B. H. Hall & N. Rosenberg (Eds.), *Handbook of the Economics of Innovation* (pp. 609–639, Vol. 1). Elsevier.

Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Representation learning on graphs: Methods and applications. *IEEE Data Eng. Bull.*, *40*(3), 52–74.

Hasan, I., Hoi, C. K., Wu, Q., & Zhang, H. (2014). Beauty is in the eye of the beholder: The effect of corporate tax avoidance on the cost of bank loans. *Journal of Financial Economics*, *113*(1), 109–130.

Hasan, I., Hoi, C. K., Wu, Q., & Zhang, H. (2017). Social capital and debt contracting: Evidence from bank loans and public bonds. *Journal of Financial and Quantitative Analysis*, *52*(3), 1017–1047.

Hasan, I., Minnick, K., & Raman, K. (2020). Credit allocation when borrowers are economically linked: An empirical analysis of bank loans to corporate customers. *Journal of Corporate Finance*, *62*, 101605.

Hasan, I., Shen, Y., & Yuan, X. (2021). Local product market competition and bank loans. *Journal of Corporate Finance*, *70*, 102054.

He, J., Rungta, M., Koleczek, D., Sekhon, A., Wang, F. X., & Hasan, S. (2024). Does prompt formatting have any impact on llm performance? *arXiv preprint arXiv:2411.10541*.

He, J., & Tian, X. (2020). Institutions and innovation. *Annual Review of Financial Economics*, *12*, 377–398.

He, J. J., & Tian, X. (2013). The dark side of analyst coverage: The case of innovation. *Journal of Financial Economics*, *109*(3), 856–878.

Heese, J., & Pérez-Cavazos, G. (2019). Fraud allegations and government contracting. *Journal of Accounting Research*, *57*(3), 675–719.

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, *124*(5), 1423–1465.

Hoberg, G., Phillips, G., & Prabhala, N. (2014). Product market threats, payouts, and financial flexibility. *The Journal of Finance*, *69*(1), 293–324.

Hollander, S., & Verriest, A. (2016). Bridging the gap: The design of bank loan contracts and distance. *Journal of Financial Economics*, *119*(2), 399–419.

Holmstrom, B., & Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica: Journal of the Econometric Society*, 303–328.

Hong, H., & Shum, M. (2002). Increasing competition and the winner's curse: Evidence from procurement. *The Review of Economic Studies*, *69*(4), 871–898.

Houston, J. F., Lin, C., Lin, P., & Ma, Y. (2010). Creditor rights, information sharing, and bank risk taking. *Journal of Financial Economics*, *96*(3), 485–512.

Hsu, P.-H., Lee, H.-H., Peng, S.-C., & Yi, L. (2018). Natural disasters, technology diversity, and operating performance. *The Review of Economics and Statistics*, *100*(4), 619–630.

Hsu, P.-H., Tian, X., & Xu, Y. (2014). Financial development and innovation: Cross-country evidence. *Journal of Financial Economics*, *112*(1), 116–135.

Igami, M., & Sugaya, T. (2022). Measuring the incentive to collude: The vitamin cartels, 1990–99. *The Review of Economic Studies*, *89*(3), 1460–1494.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Ioannidou, V., & Ongena, S. (2010). Time for a change: Loan conditions and bank behavior when firms switch banks. *The Journal of Finance*, *65*(5), 1847–1877.

Ivashina, V. (2009). Asymmetric information effects on loan spreads. *Journal of Financial Economics*, *92*(2), 300–319.

Ivashina, V., & Scharfstein, D. (2010a). Bank lending during the financial crisis of 2008. *Journal of Financial Economics*, *97*(3), 319–338.

Ivashina, V., & Scharfstein, D. (2010b). Loan syndication and credit cycles. *American Economic Review*, *100*(2), 57–61.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: Evidence from firms, patents, profits and market value. *American Economic Review*, *76*, 984–1001.

Jiang, J., Kelly, B., & Xiu, D. (2023). (re-) imag (in) ing price trends. *The Journal of Finance*, *78*(6), 3193–3249.

Jiang, L., Levine, R., & Lin, C. (2016). Competition and bank opacity. *The Review of Financial Studies*, *29*(7), 1911–1942.

Johnson, W. C., Xie, W., & Yi, S. (2014). Corporate fraud and the value of reputations in the product market. *Journal of Corporate Finance*, *25*, 16–39.

Kaminsky, G., & Schmukler, S. L. (2002). Emerging market instability: Do sovereign ratings affect country risk and stock returns? *The World Bank Economic Review*, *16*(2), 171–195.

Kang, H. (2025). Price-fixing cartels and firm innovation. *Management Science*.

Karolyi, S. A. (2018). Personal lending relationships. *The Journal of Finance*, *73*(1), 5–49.

Karpoff, J. M., Lee, D. S., & Vendrzyk, V. P. (1999). Defense procurement fraud, penalties, and contractor influence. *Journal of Political Economy*, *107*(4), 809–842.

Karpoff, J. M., Lott, J. R., Jr, & Wehrly, E. W. (2005). The reputational penalties for environmental violations: Empirical evidence. *The Journal of Law and Economics*, *48*(2), 653–675.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, *132*(2), 665–712.

Laeven, L., & Levine, R. (2009). Bank governance, regulation and risk taking. *Journal of Financial Economics*, *93*(2), 259–275.

Lancieri, F., Posner, E. A., & Zingales, L. (2022). *The political economy of the decline of antitrust enforcement in the united states* (tech. rep.). National Bureau of Economic Research.

Lee, C. M., Sun, S. T., Wang, R., & Zhang, R. (2019). Technological links and predictable returns. *Journal of Financial Economics*, *132*(3), 76–96.

Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. *The Review of Financial Studies*, *34*(7), 3265–3315.

Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2017). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*.

Lim, J., Minton, B. A., & Weisbach, M. S. (2014). Syndicated loan spreads and the composition of the syndicate. *Journal of Financial Economics*, *111*(1), 45–69.

Liu, C., Low, A., & Putnins, T. (2024). The real impacts of public short campaigns: Evidence from stakeholders. *Journal of Corporate Finance*, 102624.

Ma, S. (2021). *Technological obsolescence* (Working Paper No. 29504). National Bureau of Economic Research.

Mann, W. (2018). Creditor rights and innovation: Evidence from patent collateral. *Journal of Financial Economics*, *130*(1), 25–47.

Manso, G. (2011). Motivating innovation. *The Journal of Finance*, *66*(5), 1823–1860.

Matray, A. (2021). The local innovation spillovers of listed firms. *Journal of Financial Economics*, *141*(2), 395–412.

Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, *110*(3), 681–712.

McLemore, P., Sias, R. W., Wan, C., & Yuksel, H. Z. (2022). Active technological similarity and mutual fund performance. *Journal of Financial and Quantitative Analysis*, *57*(5), 1862–1884.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, *29*(2), 449–470.

Murfin, J. (2012). The supply-side determinants of loan contract strictness. *The Journal of Finance*, *67*(5), 1565–1601.

Murphy, D. L., Shrieves, R. E., & Tibbs, S. L. (2009). Understanding the penalties associated with corporate misconduct: An empirical examination of earnings and risk. *Journal of Financial and Quantitative Analysis*, *44*(1), 55–83.

Nagel, S., & Purnanandam, A. (2020). Banks' risk dynamics and distance to default. *The Review of Financial Studies*, *33*(6), 2421–2467.

Naumovska, I., & Lavie, D. (2021). When an industry peer is accused of financial misconduct: Stigma versus competition effects on non-accused firms. *Administrative Science Quarterly*, *66*(4), 1130–1172.

Ng, A. C., & Rezaee, Z. (2015). Business sustainability performance and cost of equity capital. *Journal of Corporate Finance*, *34*, 128–149.

Pichler, P., & Wilhelm, W. (2001). A theory of the syndicate: Form follows function. *The Journal of Finance*, *56*(6), 2237–2264.

Qiu, J., & Wan, C. (2015). Technology spillovers and corporate cash holdings. *Journal of Financial Economics*, *115*(3), 558–573.

Rajan, R. G. (1992). Insiders and outsiders: The choice between informed and arm's-length debt. *The Journal of Finance*, *47*(4), 1367–1400.

Rajan, U., Seru, A., & Vig, V. (2015). The failure of models that predict failure: Distance, incentives, and defaults. *Journal of Financial Economics*, *115*(2), 237–260.

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, *90*(5), 2555–2591.

Reed, T., Ana, U. A., Iacovone, L., & Mariana, P. L. (2022). *Cartels, antitrust enforcement, and industry performance: Evidence from mexico*. World Bank.

Reisen, H., & Von Maltzan, J. (1999). Boom and bust and sovereign ratings. *International Finance*, *2*(2), 273–293.

Richmond, R. J. (2019). Trade network centrality and currency risk premia. *The Journal of Finance*, *74*(3), 1315–1361.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, *98*(5, Part 2), S71–S102.

Rossi, A. G., Blake, D., Timmermann, A., Tonks, I., & Wermers, R. (2018). Network centrality and delegated investment performance. *Journal of Financial Economics*, *128*(1), 183–206.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, *4*(3), 305–322.

Saidi, F., & Žaldokas, A. (2021). How does firms' innovation disclosure affect their banking relationships? *Management Science*, *67*(2), 742–768.

Sapienza, P. (2002). The effects of banking mergers on loan contracts. *The Journal of Finance*, *57*(1), 329–367.

Sawyer, L. P. (2019). *US antitrust law and policy in historical perspective.* Harvard Business School.

Schenone, C. (2010). Lending relationships and information rents: Do banks exploit their information advantages? *The Review of Financial Studies*, *23*(3), 1149–1199.

Schwert, M. (2018). Bank capital and lending relationships. *The Journal of Finance*, *73*(2), 787–830.

Shapiro, C. (2019). Protecting competition in the american economy: Merger control, tech titans, labor markets. *Journal of Economic Perspectives*, *33*(3), 69–93.

Sherman, J. (1890). Speech to the United States Senate [Congressional Record, 51st Congress, 1st Session].

Shleifer, A., & Vishny, R. W. (1993). Corruption. *The Quarterly Journal of Economics*, *108*(3), 599–617.

Sproul, M. F. (1993). Antitrust and prices. *Journal of Political Economy*, *101*(4), 741–754.

Stoffman, N., Woeppel, M., & Yavuz, M. D. (2020). Small innovators: No risk, no return. *Kelley School of Business Research Paper*, (19-5).

Sufi, A. (2007). Information asymmetry and financing arrangements: Evidence from syndicated loans. *The Journal of Finance*, *62*(2), 629–668.

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. *International conference on machine learning*, 3319–3328.

Svensson, J. (2005). Eight questions about corruption. *Journal of Economic Perspectives*, *19*(3), 19–42.

Syverson, C. (2004). Product substitutability and productivity dispersion. *Review of Economics and Statistics*, *86*(2), 534–550.

Thaler, R. H. (1988). Anomalies: The winner's curse. *Journal of Economic Perspectives*, *2*(1), 191–202.

Tian, X., & Wang, T. Y. (2014). Tolerance for failure and corporate innovation. *The Review of Financial Studies*, *27*(1), 211–255.

Tirole, J. (1988). *The theory of industrial organization.* MIT press.

Tobias, A., & Brunnermeier, M. K. (2016). Covar. *The American Economic Review*, *106*(7), 1705.

Valta, P. (2012). Competition and the cost of debt. *Journal of Financial Economics*, *105*(3), 661–682.

Valukas, A. R. (2010). Lehman brothers holding inc. chapter 11 proceedings examiner's report, volume 2.

Van Reenen, J. (2018). Increasing differences between firms: Market power and the macro-economy. *CEP Discussion Papers*.

Wier, P. (1983). The costs of antimerger lawsuits: Evidence from the stock market. *Journal of Financial Economics*, *11*(1-4), 207–224.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*.

Xia, H. (2014). Can investor-paid credit rating agencies improve the information quality of issuer-paid rating agencies? *Journal of Financial Economics*, *111*(2), 450–468.

Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? *International Conference on Learning Representations*.

Yellen, J. (2013). *Interconnectedness and systemic risk: Lessons from the financial crisis and policy implications: A speech at the american economic association/american finance association joint luncheon, san diego, california, january 4, 2013* (tech. rep.). Board of Governors of the Federal Reserve System (US).