Investing in Lending Technology: IT Spending in Banking*

Zhiguo He, Sheila Jiang, Douglas Xu, Xiao Yin

October 8, 2023

Abstract

Banks' lending technology hinges on their handling of soft and hard information in dealing with different types of credit demand. Through assembling a novel dataset on banks' investment in information technologies (IT), this paper provides concrete empirical evidence on how banks adapt their lending technologies. We find investment in communication IT is associated with improving banks' ability to produce and transmit soft information, while investment in software IT helps enhance banks' hard information processing capacity. We exploit policies that affect geographic regions differentially to show causally that banks respond to an increased demand for small business credit (mortgage refinance) by increasing their spending on communication (software) IT spending. We also find that the entry of fintech induces commercial banks to increase their investment in IT—more so in the software IT category.

Keywords: Information Technology, Small Business Lending, Mortgage Refinance, Communication Equipment, Software, Hard and Soft Information

JEL codes: G21, G51, O12, O32

*He: Booth School of Business, University of Chicago and NBER, zhiguo.he@chicagobooth.edu; Jiang: Warrington College of Business, University of Florida, sheila.jiang@warrington.ufl.edu; Xu: Warrington College of Business, University of Florida, douglas.xu@warrigton.ufl.edu; Yin: Department of Economics and School of Management, University College London, xiao.yin@ucl.ac.uk. For helpful comments, we thank Doug Diamond, Francesco D'Acunto (discussant), Isil Erel (discussant), Mark Flannery, Yueran Ma, Jun Pan (discussant), Nitzan Tzur-Ilan (discussant), João Granja, Nicola Pierri, Manju Puri, and Raghu Rajan. We are thankful for valuable comments from seminar and conference participants at the Richmond Fed, University of Connecticut, University of Florida, University of Chicago, City University of New York, 5th IMF Annual Macro-Financial Conference, EFA, SFS Calvacade, and the 22nd Annual Conference in Digital Economics, and AFA 2023 in New Orleans. Zhiguo He acknowledges financial support from the John E. Jeuck Endowment at the University of Chicago Booth School of Business. All errors are our own.

1 Introduction

Commercial banks have long relied on cutting-edge technology to deliver innovative products such as ATMs and online banking, streamline loan making processes, and improve back-office efficiency. According to a 2012 Mckinsey Report, across the globe commercial banks spend about 4.7% to 9.4% of their operating income on information technology (IT); for comparison, insurance companies and airlines only spend 3.3 percent and 2.6 percent of their income on IT, respectively. Recently, the impact of information technology on the banking sector and on financial stability has been a headline topic in policy discussions (Banna and Alam, 2021; Pierri and Timmer, 2020).

Although the financial services industry—especially the banking industry—is increasingly becoming a tech-like industry, the academic literature lags behind in understanding the economics of IT spending in banking. Which banks, large or small, have invested more in IT? Do banks adapt their information technologies in response to different credit demand shocks? How do traditional banks react to the entry of fintech in recent years? We take the first step toward understanding the key empirical patterns on these issues, and further explore the mechanism that underlies the connection between these expenditures and the core functioning of banking.

To place our research in the literature, think about the information transmission between a loan officer and a borrower, or between layers of loan officers within a bank organization. As highlighted by Stein (2002), a less hierarchical structure within a bank facilitates the effective transmission of "soft" information. At the same time, fast-developing technologies in recent decades provide more options for the banking sector to cope with such problems.[1] So, can information technologies reduce frictions in communicating soft information and potentially improve banks' credit approval decisions? Likewise, with the explosive development of big data analytics which combine "hard" information such as credit scores with other alternative

[1]For instance, First Citizens National Bank implemented its employee intranet to strengthen internal communications in February 2019. For details, see this article.

data, have traditional banks started adopting these technologies?

Our study relies on a comprehensive dataset, the Harte Hanks Market Intelligence Computer Intelligence Technology database, which has been used in the literature on the economic implications of technology in non-financial sectors (e.g., Bloom et al., 2014; Forman et al., 2012). This dataset, which aligns well with the regulatory Y-9C dataset in measuring total IT spending, provides detailed branch-level information on specific spending categories. We focus on two major categories in banks' IT spending:[2] *software* and *communication*. *Software* IT products mainly aim to improve information processing accuracy and speed through automation, specialized programming and AI technologies, etc. *Communication* IT products facilitate smoother exchanges of information within bank branching networks, across banks, and with borrowers.

In Section 3 we start by documenting that IT expenditure in the U.S. banking sector has been growing rapidly over the last decade. Growth in IT spending varies by bank size: large banks' IT spending increased steadily, while there was almost no growth in IT spending for the smallest banks. Another noticeable distinction between large and small banks is that the latter, who presumably engage in more small business lending, consistently allocate a higher share of their IT budget towards communication technology than the former. As we will elaborate, this pattern points to the role communication IT plays in conducting small business lending.

We then examine the relationship between banks' IT spending and their lending activities. Among the three main categories of loan types in Call Reports, the share in commercial and industrial (C&I) loans is positively associated with the lenders' communication spending, but uncorrelated with their software spending. In contrast, the share of personal loans is positively associated with the lenders' software spending, but not with communication

[2]Section 2.2 explains in detail the four major categories of IT expenditure in the Harte Hanks data set—hardware, software, communication, services—in the context of the banking industry. Representative examples of software include desktop applications (e.g., Microsoft Office), information management software, and risk and payment management software. Examples of communication technology include radio and TV transmitters, private branch exchanges, video conferencing, etc.

spending. Going one step further, within C&I loans, we show that small business lending stands out as a sub-category that drives the overall positive association with communication IT spending, whereas within personal loans, mortgage refinance is the main contributor to the positive correlation between personal loans and software spending. As different types of loans often require different technologies in dealing with relevant information, these positive associations (or the lack thereof) offer important guidance in understanding banks' IT spending profiles from the perspective of lending technology.

Aside from broad credit categories of loan portfolios, we also explore how banks' IT investment is shaped by other factors affecting their business operations. Regarding the complexity of internal hierarchical structure, banks with more internal layers tend to have a higher communication spending. Further, hierarchical complexity has an impact on the responsiveness of banks' IT spending to their loan profiles—a more complex hierarchical structure makes banks' communication spending respond more to their small business lending, but displays no systematic effect on the relation between software spending and mortgage refinancing.[3] Finally, in the context of the syndicated lending market, frequent lead lenders spend significantly more on communication than participant lenders, as lead banks take more direct responsibility in interacting with borrowers.

In Section 4 we delve deeper into the underlying economics behind the connection between banks' IT investment and their lending activities. Conceptually, we distinguish two fundamentally different types of lending technologies. The first heavily relies on the gathering and augmentation of soft information from borrowers; in the context of Berger and Udell (2002), "relationship lending" is a concrete example of the first type. The second type of lending technology relies primarily on the processing and quantification of hard information; leading examples include "transactions lending," i.e., loans that are based on a specific credit scoring system and quantified financial statement metrics (Berger and Udell, 2006).

[3] This asymmetric pattern is consistent with the notion of "hierarchical friction" in Stein (2002): A lower level of hierarchical complexity helps facilitate the within-organization transmission of soft information, which is more relevant for small business lending than mortgage refinancing.

We formulate our first hypothesis along the dimension of soft information. Increased demand for loans that involve intensive soft information production/transmission (e.g., small business loans) should lead banks to invest more in communication technologies, say video conferencing, as they not only enable banks to more effectively collect soft information from entrepreneurs but also allow for a smoother transmission of such otherwise hard-to-verify soft information within a bank organization. Taking advantage of an arguably exogenous demand shifter, we show an increase in banks' small business credit demand—due to a higher ex-ante exposure of local counties to the policy shock exploited in our analysis—leads to a positive and significant growth in banks' communication spending, without much impact on the bank's software spending.[4] Furthermore, we find that such responses are more pronounced for banks operating in regions with more young firms, for whom effective transmission of soft information is particularly important.

In our second hypothesis, a positive demand shock for loans that rely heavily on hard information processing (e.g., mortgage refinancing) should push banks to engage in more IT investment in software (which facilitates lenders in processing such existing data). For causal identification, we utilize the cross-county variation in interest savings of outstanding mortgages to construct a shifter to the mortgage refinance demand across different regions.[5] We show that an increase in mortgage refinancing by a bank (due to its local exposure to high refinance savings) results in a higher software spending, without any significant impact on its communication spending.

The last part of our analysis concerns how the entry of fintech lenders into local credit markets affects banks' IT spending. In the past decade, we have witnessed a growing pene-

[4] Our empirical identification is based on the "Small Business Health Care Tax Credit," which was introduced in 2010 and then experienced a significant policy change in 2014. We construct the instrumental variable using counties' exposure to policy change in 2014, with details explained in Section 4.2.2.

[5] We take advantage of the low-interest episode from 2011 to 2015, during which nationwide average mortgage interest rates decreased from 6.5% to 3.5%. When interest rates drop, the mortgage prepayment option is in the money (Eichenbaum et al., 2022; He and Song, 2022), implying a greater mortgage refinance demand by local households. In addition to mortgage repayment saving, we also consider an alternative instrumental variable (IV) constructed as the weighted mortgage rate gap for outstanding mortgages in a given county.

tration of fintech into the traditional banking sector. Utilizing the staggered entry of Lending Club into seven states after 2010 as an experimental setting, we investigate how the traditional banking sector reacted to the penetrating fintechs. Right after the regulatory approval of Lending Club's operation in a state, banks operating in that state saw a significant increase in their IT investment. Importantly, shocked banks' software spending experienced an economically and statistically significant growth (around 7%), whereas the change in their communication spending was insignificant.

There is also significant heterogeneity across bank size groups in their technology spending reactions in response to fintech entry. In particular, the increased IT investment is predominantly observed among large banks, whereas small banks barely respond.[6] Our findings suggest an overall "competition reaction" from the traditional banking sector, in that banks—particularly larger ones—are catching up with fintech challengers. Consistent with this competition interpretation, such "catching up" behavior by commercial banks is especially noticeable in improving their automating and information processing technology through increased software spending, which is precisely the domain of lending technology in which fintech lenders have a comparative advantage.

Related Literature

Bank lending technology and the nature of information. Berger and Udell (2006) provide a comprehensive framework of the two fundamental types of bank lending technology, i.e., relationship lending and transactions lending, in the SME lending market; see also related work by Bolton et al. (2016). A key difference between these two types of lending is related to the role played by information as highlighted by Stein (2002), who provides an explanation for why soft information production favors an organizational structure with

[6] Relatedly, consistent with large banks' dominance in the personal loan lending market, we also find that the response to Lending Club's entry is particularly pronounced for banks more specialized in personal loan lending.

fewer hierarchical layers.[7]

We contribute to this literature by linking banks' IT spending to their lending technology, especially with regard to the distinction between soft information production/transmission and hard information processing. We further establish causal linkages from the informational components in credit demand to banks' IT spending. It is, to our knowledge, the first attempt in the literature to show how credit demand shocks drive banks' investment in their information-driven lending technologies.[8]

Information technology in the banking industry. Our paper belongs to the literature on the interaction between the development of information technology and the evolution of the banking industry. For instance, Berger (2003) shows that progress in both information and financial technologies led to significant improvement in banking services, and Petersen and Rajan (2002) document that communication technology greatly increased the lending distance of small business loans. Using the number of computers per employee as a measurement for IT adoption, two recent papers show that IT adoption helps banks weather financial crisis (Pierri and Timmer, 2022) and spur entrepreneurial activities (Ahnert et al., 2021). Our paper, with the aid of detailed IT spending data, studies the specific economic mechanisms that connect banks' lending technology with their IT spending.[9]

Our paper is closely related to Modi et al. (2022) who construct IT spending data using the Call Reports; we compare our sample with them in Section 2.1. While their analysis also investigates the linkages between banks' IT spending and their lending behaviors (e.g., mortgage lending and reactions to monetary policies), our analysis differs in our focus on

[7] Along these lines, Liberti and Mian (2009) find empirically that greater hierarchical distance leads to less reliance on subjective information and more on objective information. Paravisini and Schoar (2016) document that credit scores, which serve as "hard information," improve the productivity of credit committees, reduce managerial involvement in the loan approval process, and increase the profitability of lending.

[8] Previous literature has shown that credit supply positively affects non-financial firms' technology adoption or innovation (Amore et al. (2013), Chava et al. (2013), Bircan and De Haas (2019)).

[9] There is also a vast theoretical literature (Hauswald and Marquez, 2003, 2006; Vives and Ye, 2021) on the interactions among information technology and credit market competition; see Freixas and Rochet (2008) for a review. For instance, in the framework of credit market competition where the specialized lender acquires additional "soft" information, He et al. (2023a) study the role of information span where loan quality is determined by multi-dimensional characteristics.

linking different categories of IT investment to banks' lending activities associated with different types of information nature (i.e., soft information vs. hard information). Furthermore, our analysis aims to establish a causal linkage between banks' adaptation of their lending technology and credit demand shocks, which is not the focus of Modi et al. (2022).

Fintech entry and banks' IT spending. The emergence of fintech reflects the recent developments in information technologies.[10] Our study aligns closely with studies on how the emergence of the fintech industry is affecting (or has affected) the traditional banking sector.[11] While a common theme of this research has mostly focused on bank-fintech competition during which traditional banks are largely viewed as a *passive* player, little attention has been paid to the banks' *active* responses; we take the latter angle by studying whether and how traditional banks are catching up with penetrating fintech lenders. Along a similar line, Modi et al. (2022) also document that banks with more fintech exposure in the mortgage market tend to spend more on IT, and that their lending behaviors are also likely to resemble fintech companies.

Micro-level evidence on technology adoption. Our paper also broadly contributes to the literature studying firms' technology adoption behavior using micro-level data. Using the same IT spending data as this paper, Forman et al. (2012) study the impact of firms' technology adoption on regional wage inequality, Bloom et al. (2014) investigate the effect of information technology on firms' internal control, and Ridder (2019) explores how software adoption explains the decline in business dynamism and the rise of market power.[12]

[10] Related works include but are not limited to Jagtiani and Lemieux (2017), Buchak et al. (2018a), Fuster et al. (2019), Frost et al. (2019), Hughes et al. (2019), Stulz (2019), and Di Maggio and Yao (2020).

[11] This fast-growing literature includes Lorente et al. (2018); Hornuf et al. (2018); Calebe de Roure and Thakor (2019); Tang (2019); Erel and Liebersohn (2020); Aiello et al. (2020); Gopal and Schnabl (2022); He et al. (2023b), and Huang (2022).

[12] While we use detailed IT "budget" data from Harte Hanks, several papers use its IT "installment" data that report firm-level IT product installment; see Section 2.1 for details of the differences between these two datasets. For instance, Charoenwong et al. (2022) study installment of IT products catering to compliance requirements, and Pierri and Timmer (2022) investigate whether banks installed with more PCs per employee can better survive a financial crisis. We use the budget data which reports detailed dollar amounts for various IT categories, which are crucial for our study.

2 Data and Background

We explain our main data sources in this section, together with detailed descriptions of various categories of IT spending.

2.1 Data Source for Bank IT Spending and Sample Construction

The data on banks' IT spending comes from the Harte Hanks Market Intelligence Computer Intelligence Technology database, which covers over three million establishment-level observations from 2010 to 2019 obtained while conducting IT-related consulting for firms. Harte Hanks collects and sells this information to technology companies, who then use it for marketing purposes or to better serve their clients. Firms have incentives to report their IT spending data truthfully to Harte Hanks as they also want to receive tailored advice for better IT services in the future.

Our paper focuses on commercial banks.[13] The sample consists of 1,450 commercial banks in the U.S., which covers more than 80% of the U.S. banking sector in terms of asset size (Figure A1). The sample is more representative for large banks, as shown in Table 1 which reports our coverage by bank asset size group. For the three groups of relatively large banks (with assets above $1 billion), the coverage in frequency and assets are both over 80%. However, for small banks with size below $100 million, our sample covers only 9.47% (11.36%) of the total number (assets) of commercial banks in the U.S. system.

Table 2 displays the summary statistics of banks' IT spending. In our sample, total IT spending as a share of net income ranges from 1.7% (25^{th} percentile) to 8.5% (75^{th} percentile), suggesting a large cross-sectional variation across banks. Median IT spending as a share of net income is 5.2%, consistent with a 2012 McKinsey survey (Figure A5) reporting that banks' IT spending as a share of net operating income ranges from 4.7% to 9.4%.

[13] The Harte Hanks dataset has been utilized by a broad literature of economic studies. For instance, Forman et al. (2012) investigate firms' IT adoption and regional wage inequality; Bloom et al. (2014) study the impact of information communication technology on firms' internal control; and Tuzel and Zhang (2021) study labor-technology substitution at establishment level, based on this dataset.

Matching procedure and matching quality. We provide detailed descriptions of the matching algorithm for our sample construction in Appendix B.1.1. This matching mainly involves mapping sites in Harte Hanks to bank branches, where we take bank names from the Call Report. To evaluate the matching quality, we conduct several cross-checks of our sample with other data sources, especially Call Reports, regarding certain key bank-level variables. More specifically, Appendix B.1.2 shows a close alignment of our sample with the Call Report in terms of the total number of banks' branches at both the bank level and bank-county level, as well as banks' total revenue and total number of employees.

Cross-validation of bank IT expense measure. To confirm the reliability of Harte Hanks data in measuring banks' IT spending, we conduct thorough validity checks against various alternative sources, including the "FR Y-9C Consolidated Financial Statements for Holding Companies" which contain regulatory data on IT spending at the bank holding company (BHC)-year level. As explained in Appendix B.2, we follow Kovner et al. (2014) to construct BHCs' IT expenses by summing up two standardized "other noninterest expenses" in the Y-9C dataset, the "Data processing expenses" and the "Telecommunication," together with the unstandardized write-in items reported in "other noninterest expenses" containing IT-related keywords. For the top 50 BHCs sorted by assets, Figure B9 offers a BHC-by-BHC comparison of IT spending between Y-9C and Harte Hanks (adjusted for BHC subsidiaries), at years that Y-9C IT expenses are not missing or zero.[14] Further, for the overall matched sample, Table B4 reports that a regression of the logarithm of the IT spending in Harte Hanks on that in Y-9C has a slope coefficient close to one (0.935) and a constant close to zero (0.037). Overall, we find a decent match between Y-9C regulatory filings and Harte Hanks data (adjusted for BHC subsidiaries), suggesting a high quality of the Harte Hanks data in measuring banks' IT expenses.

[14] An important feature of the regulatory Y-9C and Call Report data is that the reporting of banks' IT expenses is censored at certain threshold, i.e., IT expenses below that threshold are often reported as zero or missing values; this contributes to a higher IT spending measure in our sample than either Kovner et al. (2014) or Modi et al. (2022). More details regarding the reporting rules of these regulatory data are provided in Data Appendix B.2.1.

We also compare our IT spending measure with that in Modi et al. (2022), who construct the IT spending using only IT-related write-in expenses in Call Reports. We find a decent correlation ($\rho = 0.77$) between the two IT expense measures. Furthermore, Figure B12 separates banks into different size groups as in Modi et al. (2022) and shows similar time trends in banks' IT expenses during 2010-2019 across bank size groups in these two data sets.[15] Data Appendix B.3 provides several further comparisons with our sources on the empirical measures for banks' IT investment, from which overall consistent results are obtained.[16]

Data collection practice by Harte Hanks. Our analysis uses the "IT budget data" offered by Harte Hanks.[17] According to the official description provided by the data collection team of Harte Hanks, the construction of this data set is mainly based on data collected from surveys conducted at the site-year level; in addition, the IT budget data reflects the purchases of ready-to-use IT products or services, and does not include expenditures on IT-related R&D activities which might take a longer time to accomplish. Furthermore, since the usage of IT products or services (e.g., software programs) is often licence-based, the IT expense is therefore likely to be spread across branches of a given bank based on branch-level usage, rather than concentrated at the headquarters of the bank. As one piece of supporting evidence, we find no significant differences between the IT expenses at bank headquarters and local bank branches. Data Appendix B.4 provides more details for supplemental materials and analysis regarding the IT data collection practices of Harte Hanks.

[15] It is worth noting that our IT expense measure, which aligns well (in dollar amounts) with the aggregated IT spending in Y-9C following the method in Kovner et al. (2014), is systemically higher than that in Modi et al. (2022). This is because Kovner et al. (2014) sum the two standardized "other noninterest expenses" and unstandardized write-in items, while Modi et al. (2022) only account for the unstandardized write-in items of "other noninterest expenses." Furthermore, the censored reporting in Call Report (and in Y-9C) also contributes to a smaller IT expense measure in Modi et al. (2022) compared to the measure offered by Harte Hanks.

[16] These comparisons include i) the industry-level Bureau of Economic Analysis (BEA) data, with the focus being the detailed composition of bank' IT spending; and ii) the data storage cost in Feyen et al. (2021).

[17] The Harte Hanks data set has two major parts—the "IT installment" data that contains information on whether a firm installs certain IT product and the earliest installment date, and the "IT budget data" which contains information on the dollar amounts of detailed categories of IT investment. While the data vendor gets some of the installment information using algorithms extracting installment information from firms' public reports, job postings, etc., the IT budget data is based primarily on surveys. Our analysis in this paper only uses the IT budget data by Harte Hanks.

2.2 IT Investment Categorization

Our dataset offers a detailed decomposition of banks' IT investments in four major categories specified by Harte Hanks: *hardware*, *software*, *communication*, and *services*. We now explain these categories, with formal definitions given in (a) to (d) of Figure A6.

Software is defined as software programs purchased from third parties, including those offered as an SaaS from a multi-tenant shared-license server accessible by a browser. More specifically, the category of software covers desktop applications, information management software, processing software, and risk and payment management software. For desktop applications, one representative example is Microsoft Office.[18] Processing software specializes in automatically processing information from loan applicants' document packets through specialized programming and AI technologies with improved accuracy and speed, which would otherwise be done manually by loan officers.[19] Risk management software provides ongoing risk assessment after loans have been issued, through augmenting borrowers' repayment status as well as real-time industrial and economic conditions.[20]

Communication is defined as the network equipment that banks operate to support their communication needs. It includes routers, switches, private branch exchanges, radio and TV transmitters, Wi-Fi transmitters, desktop telephone sets, wide-area networks, local-area network equipment, video conferencing systems, and mobile phone devices. For effective project evaluation, these machines allow bankers to conveniently talk to and see borrowers who seek credit. In addition, communication equipment such as private branch exchanges facilitates the exchange of information, opinions, and decisions within the bank branching

[18] These software products are easy to grasp by bank employees who are then able to conduct basic calculations and visualizations of data associated with lending businesses. For example, on Mendeley.com, the job postings for loan officers or project managers by many banks require applicants to be proficient with Microsoft Office.

[19] Examples of processing software include Trapeze Mortgage Analytics, Treeno Software, and Kofax. These software products feature document assembly enhancement, digitization, and information classification.

[20] These software products, e.g. Actico, ZenGRC, Equifax, and Oracle ERP, allow banks to better monitor loans in progress. Other software products include security trading systems and operating systems that are typically bundled with the specific software products.

networks.

Hardware as a form of IT investment includes classic computer hardware such as PCs, monitors, printers, keyboards, USB devices, storage devices, servers, and mainframes. In terms of lending services, hardware investment can complement and facilitate both the gathering of borrower information and the processing of that information. This is because hardware devices, such as PCs and servers, help provide storage and transmission of data, in addition to serving as the carriers of software and toolboxes.

Services are defined as project-based consulting services (including, say, IT strategy and security assessments) or systems integration services that vendors provide to banks, which are often provided by IT outsourcing companies on a contractual basis. Similar to hardware, services work as complements to other categories of information technology investment to facilitate banks' lending. Examples include Aquiety, a Chicago-based IT service company that provides cybersecurity services to banks and other firms; and Iconic IT, a New-York based IT service company that provides software and hardware procurement, together with installment and upgrade services.

Table 2 reports summary statistics on the detailed structure of banks' IT spending profiles. By size, software and services are the top two among all categories of IT spending, each constituting around 33% of total IT budget; while hardware (communication) constitutes about 17% (9%) of total IT budget. We conduct analysis on banks' IT spending at bank-year level in Section 3, while in Section 4 the analysis is at bank-county-year level, in which we aggregate the branch-level spending information of each bank at the county level.

2.3 Other Datasets

To supplement our study on banks' lending technologies and their relation to IT spending, we combine loan-level information from multiple sources.

Bank balance sheet. We obtain bank-level balance sheet information from Call Reports; for detailed matching procedures, see Appendix C.1.1. In our OLS analysis investigating the correlation between IT spending and loan portfolios, we calculate the dependent variable as IT spending in Harte Hanks as a share of "Revenue" (RIAD4000 of Call Report) at bank-level. In the identification part that involves bank-county analysis (Section 4 and 5), since there is no bank-county revenue in regulatory data, we use "revenue" scaled by number of employees in Harte Hanks to control for the profitability at the bank-county level.

Loans and local characteristics. We obtain syndicated loan information on the frequency of a bank acting as lead bank in syndicated loan packages from LPC Dealscan. Small business loan origination data are from the Community Reinvestment Act (CRA), which is at the bank-county-year level covering the sample period of 2010–2019. Mortgage refinance information is available through the Home Mortgage Disclosure Act (HMDA) from 2010–2019, and we use the county-level average mortgage interest rate before 2010 obtained from Freddie Mac as the demand shifter for mortgage refinancing.

Bank hierarchical structure. We obtain banks' hierarchical structure information from Mergent Intellect platform, which covers 97 million public and private businesses including their locations and industry classifications.[21] We restrict our sample to entities with the two-digit SIC code of "60," which designates "Depository Institutions."

The database provides the complete family trees of the companies, with detailed information on its family members. Importantly, this database classifies each family member of a company into one of the three categories of location types: "Headquarters," "Single Location," and "Branch." We define a bank as having n layers of hierarchical structure if the bank has n types of locations in the family tree, where $n \in \{1, 2, 3\}$. To give some concrete examples, Wells Fargo has all three location types and hence is classified as three hierarchical layers; North Valley Bank with headquarters located in Corning (OH) and seven branches is

[21] Huvaj and Johnson (2019) use this database to study the impact of firms' organizational structure on their innovation activities.

classified as two layers; and First Place Bank located in Warren (OH) with one single location is classified as only one layer. For each bank, we match the banks in Mergent Intellect with banks in our sample based on bank names and the city where the banks' headquarters are located (see Appendix C.1.1 for more details). While the number of distinct "location types" in the Mergent Intellect dataset can provide information on the hierarchical complexity of a bank, it is admittedly a somewhat coarse empirical measure and could underestimate the hierarchical complexity, especially for large banks.

3 Empirical Patterns of Banks' IT Spending

We start our analysis by reporting some basic statistics of banks' investment in IT over the last decade as well as across bank size. We further show that banks' IT investment is related to their lending activities and organization structures.

3.1 Banks' IT Investment: Trend and Cross-Section

In Panel a) of Figure 1, we plot the evolution of IT spending as a share of total revenue of banks in our sample, in the manufacturing sector (2-digit SIC "20-39"), and in all other non-depository sectors (2-digit SIC "not 60"). As is evident from the figure, the IT investment in the U.S. banking sector (represented by our sample) has witnessed faster growth during the past two decades compared to other industries. Also, banks' IT spending saw faster growth starting in 2016, which could be potentially driven by the release of a "white paper" by the Office of the Comptroller of the Currency on March 16, 2016. This white paper set forth the regulators' perspective on supporting responsible innovation across all sizes of banks,[22] which might have pushed banks to be more aggressive in embracing technology investment as part of their strategic planning (see this article).

[22] In this white paper, the Office of the Comptroller of the Currency defines "responsible innovation" as "the use of new or improved financial products, services, and processes to meet the evolving needs of consumers, businesses, and communities in a manner that is consistent with sound risk management and is aligned with the bank's overall business strategy."

Banks of different sizes often behave differently in systematic ways. In Panel b) of Figure 1, we follow FDIC bank size classifications to break banks into five size groups, and present the growth trend for banks' IT investments for each group as a fraction of non-interest expenses.[23] We observe that large banks invest more in IT than their small peers do (for more detailed statistics, see Table A2);[24] and IT spending in large banks (with asset size \$10–250 billion and above \$250 billion) has been steadily growing, though there is also an apparent growth in the smaller groups (with asset size below \$10 billion).[25] The study in Modi et al. (2022) also confirms the empirical pattern that larger banks tend to invest more in IT and experience higher rates of growth in IT spending. While we do not have a conclusive answer for why such heterogeneity exists, our analysis of how banks (of different sizes) react to the entry of fintech in Section 5 touches on this issue directly.

Another noticeable pattern in Table A2 is that small banks tend to allocate a higher fraction of their IT budget towards communication technology than large banks do: the average communication spending over total spending decreases from 12.6% for the smallest group (below \$100 million) to 6.2% for mega banks (above \$250 billion). For software spending, however, there are no significant differences across bank size groups. We will come back to this contrast in Section 4, where we connect banks' IT spending categories to their lending activities that involve different ways of handling information.

[23]The magnitude of IT budget as a share of non-interest expenses in this figure is also in line with Hitt et al. (1999), who report banks' IT spending could be as high as 15% of non-interest expenses in their survey. The trend of IT spending as a share of total revenue, as is shown by the solid line in Figure 1, shares a consistent pattern with IT spending as a share of non-interest expenses.

[24]Table A2 shows a robust cross-sectional pattern that IT spending (say, scaled by non-interest income) increases with bank size. This could be due to the fact that small banks often cannot afford IT purchases with significant lump-sum costs.

[25]Medium- and smaller-size banks (asset size bins bellow \$10 billion) saw growth from 2010 to 2013, slowed down in 2015, and then have picked up again since 2016. One possible explanation for the temporary slowdown in IT spending in 2015 is that banks chose to "wait and see" in 2015 before the release of the white paper in 2016 (see first paragraph in Section 3.1).

3.2 Empirical Patterns of Bank IT Investment

We now present the first set of empirical results that relate banks' IT investment to their operations, from three angles: 1) specialization in loan making; 2) the role that a bank plays in syndicated loans; and 3) the complexity of bank's internal hierarchical structure.

3.2.1 Loan Specialization

Banks provide three major types of loans: commercial and industrial (C&I) loans, personal loans, and agricultural loans. Lending to different types of borrowers often involves distinct ways of dealing with borrower-specific information. Therefore, if banks specialize in different types of loans, one should expect them to differ in their IT investment profiles. Specifically, we run the following bank-level regression:

$$\frac{\text{Type S IT spending}}{\text{Revenue}}_{i,\text{10-19}} = \alpha_i + \beta \frac{\text{Type L loan}}{\text{Total loan}}_{i,\text{10-19}} + \gamma \mathbf{X}_i + \epsilon_i. \tag{1}$$

Here, the outcome variable of interest is $\frac{\text{Type S IT spending}}{\text{Revenue}}_{i,\text{10-19}}$, which is the average investment in a specific type of IT spending as a share of bank i's revenue over 2010-2019.[26] The "Revenue" in the denominator of the dependent variables is the total income in Call Report (RIAD4000). The main explanatory variable $\frac{\text{Type L loan}}{\text{Total loan}}_{i,\text{10-19}}$, which is the average share of a specific type of loan relative to bank i's total loan size, captures bank i's loan specialization. Control variables, which are measured over 2010-2019 at the bank level, include net income, total deposits, total equity, total salaries (all scaled by total assets), and revenue per employee. In all of our regression analysis (except for Section 5, where the main independent variable is the dummy variable indicating post-entrance), we standardize the dependent variables and the regressors.

Table 3 reports the estimation results of (1) for C&I loans, together with detailed regression outcomes for control variables. We apply the same methodology for personal loans and

[26] For robustness, we also conduct analyses with an alternative measure of banks' IT spending intensity scaled by banks' deposits, with qualitatively similar results (Table A3).

agricultural loans. For exposition purposes, Table 4 reports key regression coefficients (i.e., those of specific IT spending shares).

A. Commercial and Industrial (C&I) loans

Specialization in C&I loans is most positively associated with banks' spending in communication technology (Table 3 column 2). A one standard deviation (13 percentage points) increase in loan portfolio share allocated to C&I loans predicts a $0.24 million higher expenditure on communication per year. For detailed calculation, see Table A6. Our economic magnitude calculation for Table 3 follows the same method.

A higher degree of specialization in C&I loans also predicts more spending on hardware (column 3), though the magnitude is slightly smaller than that of communication spending. The coefficient of software spending, however, is insignificant (column 1).

Within C&I loans. Rows 2-3 of Table 4 further decompose C&I loans into "Small Business Loans," which are measured by a bank's small business lending reported in the CRA, and "other C&I loans." While the share of small business loans in a bank's portfolio is positively associated with communication spending, it is negatively related to the bank's software spending. In contrast, "other C&I loans" (e.g., loans to large firms) are positively associated with software spending, but not with communication spending. Panel A in Table 5 further shows that this positive association between small business loans and communication spending is more statistically significant for small banks.

B. Personal loans

Row 5 of Table 4 reports the associations between shares in personal loans and banks' IT spending. Contrary to the pattern we observe for C&I loans, a higher share of loan portfolio allocated to personal loans appears to predict more spending on software only. Quantitatively, a one standard deviation increase in personal loans share (an increase of about 7 percentage points) predicts a 0.0617 standard deviation increase in software spending as a share of total revenue; in dollar terms, this amounts to an increase of $1.53 million in software spending per year. On the other hand, a higher personal loans share does not have qualitatively significant predictive power on communication, hardware, or services budgets.

Within personal loans. Paralleling our analysis of small business loans within C&I loans, we decompose personal loans into two subcategories: mortgage refinancing and everything else. It is mortgage refinancing—but not other kinds of personal loans—that positively correlates with banks' software spending. This finding motivates our study in Section 4 to pay particular attention to mortgage refinancing as a specific type of lending activity in which the processing of hard information plays a critical role.

Additionally, the richness of mortgage data allows us to gain further insights by distinguishing "refinancing an existing loan" from "originating a new loan," with results reported in Row 5 and 7 of Table 4. We postpone more detailed discussion to Section 4.3.

C. Agricultural loans

As shown in Row 8 of Table 4, the agriculture loan presence seems to be positively associated with all categories of IT spending, although there is no statistically significant correlation between agriculture loan proportion and a particular type of IT investment.

3.2.2 Complexity of Hierarchical Structure

Another important factor that may affect a bank's efficacy in handling information is the internal organization structure of a bank (Stein, 2002). In the first row of Panel B in Table 4, we use the measure of hierarchical layers defined in Section 2.3 as our main proxy for hierarchical complexity. When the number of banks' hierarchical layers increases, banks spend more across all IT categories, especially on communication. Increasing from 1 hierarchical layer to 3 layers implies \$0.31 million more in communication spending each year. This result is under the specification with bank-size group fixed effects included, implying that hierarchical complexity predicts higher communication spending beyond bank size.[27]

[27] Recall that in Section 3.1 we show that smaller banks tend to allocate a larger portion of their IT budget to communication spending. Our finding therefore suggests that despite its high correlation with bank size, the complexity of a bank's internal hierarchical structure has an additional impact on its IT spending on

For robustness, we also proxy banks' hierarchical complexity using the logarithm of the total number of offices, with qualitatively similar results reported in the second row of Panel B.

As we will explain in Section 4.2.1, one can relate these findings to Stein (2002), from the perspective of within-organization transmission of information that is difficult to verify and relay. Despite a crude empirical measure of hierarchical complexity, our paper establishes a direct link between hierarchical complexity and banks' IT investment for information production and transmission.

3.2.3 Role in Syndicated Lending

Aside from specialization in different types of loans or having different levels of hierarchical complexity, banks may also differ in the role they play in dealing with information when conducting lending. For instance, in the context of syndicated lending, lead lenders and participant lenders perform drastically different tasks. Table 4 Panel C presents the same regression as in Eq. (1), except we replace the right-hand side variable with "%Lead bank/Total syndicate," defined as the percentage frequency that a bank shows up as lead bank in the syndicated loan market. We find that communication, hardware, and services show a strong positive correlation with changes in lead bank frequency in the syndicated loan market, with communication spending having the largest magnitude. A one standard deviation increase in the lead bank frequency is associated with \$0.27 million more in the bank's annual communication budget. These findings, as we will elaborate in Section 4.2, can be attributed to the distinct responsibilities for handling information assumed by lead and participant banks.

top of the bank size effect. Put differently, one cannot simply use the size of a bank as an empirical proxy for its hierarchical complexity.

4 Economics of Banks' IT Investment

Having demonstrated the basic patterns of IT investment in the U.S. banking sector and its interaction with various banking business operations, we now move on to our central question: What are the economics behind banks' IT spending decisions, and how do they relate to—and contribute to—the development of banks' lending technology? We start with a conceptual discussion of lending technologies based on the nature of information handling. By mapping different types of IT investment onto various dimensions of lending technology, this framework helps us understand various empirical patterns shown in Section 3. We then study two credit demand shocks that involve different kinds of information handling, and establish their causal impact on banks' lending technology adoption behaviors.

4.1 Lending Technology, Information Handling and IT Spending

We view a bank's lending technology as its ability to deal with borrower-specific information throughout the lending process. Broadly speaking, banks engage in two types/stages of activities in loan making: information *production/transmission* and information *processing*. More specifically, information *production/transmission*—broadly related to *soft* information in Stein (2002)—refers to the stage in which information on borrowers is gathered and then relayed to those who make decisions later. On the other hand, information *processing*—broadly related to *hard* information—is more about the stage in which lenders assemble and examine existing information on borrowers for better decision making.

Communication IT and soft information production/transmission. When facing borrowers with whom lenders have never dealt, or whose information is relatively opaque, for effective information gathering bankers often need to communicate with their borrowers—either through face-to-face meetings, or seeing borrowers' projects for themselves. Once this first-hand information has been gathered, which often can be subjective and thus difficult to convey to others, effective transmission of such information within the organization can also affect banks' lending efficiency.

One concrete example of how communication technology can help in the two aforementioned dimensions is video conferencing, which has become an important means for loan officers to interact with customers and colleagues during the past decade. In the past, banks opened new checking accounts and originated loans only through brick-and-mortar branches and in-person visits; now, they also use video conferencing, as it makes the direct—yet virtual—contact between loan officers and borrowers more efficient.[28] Moreover, video conferencing within an institution has also been welcomed by the banking sector for its advantages in facilitating effective internal communication and collaboration among employees.[29]

Software IT and hard information processing. Once information has been produced (by the lender itself) or is readily accessible (via a third party), the next concern for the lender is how to use this information. In the context of credit allocation, banks need to properly evaluate the borrowers' creditworthiness to determine loan amounts and rates. For borrowers whom bankers already know from previous interactions or with transparent information, lending decisions simply boil down to efficient processing of the existing information.

Accurate evaluations of borrowers' credit risk often require complicated modeling and simulations, which are impossible without modern software tools. Nowadays banks have actively adopted new software-based technologies to store, organize, and analyze large chunks of loan applicants' data.[30] One popular form of software technology product is credit scoring software for banks making *refinancing* decisions,[31] which primarily involve the processing and assessment of *existing* information that lenders already possesses through past interactions. In fact, the recent penetration of fintech companies, which specialize in utilizing software

[28]See "Liveoak" for a real world example of a communication tool designed for banking services.

[29]See this article from Bankingdive for a detailed description of how video conferencing helps within-bank communication.

[30]For example, "nCino" is operating system software that allows financial institutions to replace manual collection of loan/account applications with automated and AI-based solutions. "Finaxtra" and "Turnkey" are both comprehensive loan origination systems that offer solutions for the whole lending process.

[31]Some concrete examples of credit scoring software include SAS Credit Scoring, GinieMachine, and RND-Point. To use such software, banks usually just need to import borrowers' demographic and historical data, based on which the software calculates credit scores and conducts statistical tests using AI and machine learning methodologies, saving banks from tedious manual work and expediting the processing.

and algorithm-driven lending approaches, has been particularly pronounced in the mortgage refinancing market (Buchak et al., 2018b; Fuster et al., 2019).

In the next two sections we will explore in detail the lending technology adoption along two dimensions—those targeting the production and transmission of soft information (Section 4.2), and those targeting hard information processing (Section 4.3). In short, communication devices facilitate the gathering and dissemination of soft information, whereas software is for efficiently utilizing "hard" information. From this point on, we focus on two particular categories—*communication* and *software*—when examining banks' IT investment behavior.[32]

4.2 Bank IT Spending and Soft Information

4.2.1 Soft Information Production/Transmission in Bank Lending

Small business lending. Lending to small business borrowers is one concrete example in which the efficient production and transmission of soft information is essential. Sahar and Anis (2016) document that in the context of lending to small- and medium-size enterprises, direct contact with borrowers and frequent visits to their work sites allow loan officers to collect and produce soft information. Agarwal et al. (2011) highlight that soft information, such as what the borrower plans to do with the loan proceeds, is always the product of multiple rounds of lender-borrower interactions.

That small business lending involves intensive soft information production and transmission is consistent with Section 3.2.1, where we show that banks which specialize in small business lending (as measured by small business loans over total loans) incur more spending on communication IT. As smaller banks generally extend more loans to small businesses (Berger and Udell, 2006; Chen et al., 2017), this helps explain the robust pattern that smaller

[32] We will shortly show in Section 4.2 and 4.3 that these two categories of banks' IT spending have a more direct link to banks' dealing with different types of borrower-specific information, a fact already hinted at by the empirical patterns of bank IT spending documented in Section 3.2.

banks have higher fractions of communication IT spending shown in Table A2. Indeed, Table 5 shows that small banks' communication spending is significantly more associated with their small business loans compared with large banks.[33]

Hierarchical complexity. Recall that in Section 3.2.2 we find banks with a more complex hierarchical structure tend to have higher communication IT spending. This is in line with Stein (2002), who argues that a low hierarchical complexity facilitates the within-organization transmission of soft information, making it easier to issue loans requiring soft information (e.g., small business loans). Digging one step further, Table 5 Panel B shows that, given the same percentage increase in small business loans, banks with a more complex hierarchical structure respond with a greater increase in their communication spending. This is consistent with "hierarchical frictions" in soft information transmission: When banks face a need (or choose) to make more small business loans, which implies a demand for improving their soft information handling capability, those with a more complex internal hierarchical structure have to spend more on communication IT so as to overcome such frictions.[34]

Finally, as a placebo test, one should expect no systematic impact of banks' hierarchical complexity on the correlation between their software spending and mortgage refinancing activities, which is indeed confirmed in Table 5 Panel B. Overall, our empirical findings on banks' hierarchical complexity corroborate previous works studying banking organization structure and information production (Degryse et al., 2008; Levine et al., 2020; Skrastins and Vig, 2018), and more research needs to be done on this topic.

[33] The relatively lower communication spending by large banks is also consistent with recent empirical findings that large banks, who have deeper pockets than small banks, more frequently invest in or acquire fintech startups (Hornuf et al., 2021; Cornelli et al., 2022). As fintech businesses specialize in transforming the soft information embedded in the alternative data of consumers into credit scores (a form of hard information), large banks' reliance on communication technology in small business lending is lower.

[34] Our finding echoes previous work on credit decision making. For instance, Paravisini and Schoar (2016) document that business loan decisions are often made by committee; when decisions cannot be made after committee discussions, the committee will refer to managers in an upper layer, say regional managers. The greater the hierarchical complexity, the higher the "transaction cost" involved for loan decisions.

Lead lender in syndicated loans. The syndicated loan market also provides a special environment to explore the relationship between communication technology and soft information production/transmission. In syndicated lending, the nature of interactions between lenders and borrowers depends crucially on whether the lender is a lead bank or a participant bank (Sufi, 2007). Compared to participant banks, the lead bank is mandated by borrowers to organize other lending participants, conduct compliance reports, and negotiate loan terms. After the loan is issued, it also has the responsibility to conduct monitoring, distribute repayments, and provide overall reporting among all lenders within the deal.[35] In this regard, performing the job of lead bank involves significantly heavier effort in information generation and sharing as well as coordinating negotiations, during which effective communication plays a central role. These differences between lead and participant banks are empirically verified in Section 3.2.1: There is a strong correlation between the frequency of a bank serving as a lead arranger in syndicated loans and its communication IT spending (Table 4 row 4).

4.2.2 Banks IT Spending and Demand Shock on Small Business Loans

We now present the first piece of causal evidence on banks' adaptation of their lending technology by studying their IT investment responses when hit by a positive demand shock in small business loans. As small business lending is associated with intensive soft information production/transmission, we predict that banks will increase their spending on communication technology (soft information), but not on software (hard information).

Our identification strategy relies on a policy shock that affected small businesses' credit demand, which hit the U.S. banking sector heterogeneously across different regions. The "Small Business Health Care Tax Credit" was initially enacted in 2010 as part of the "Affordable Care Act." The program, whose details are available here, offers a tax credit to small business employers who pay health insurance premium on behalf of employees. From 2010-

[35] Due to the vast reporting and coordination efforts, lead banks often charge an initiation fee, which can be as high as 10% (Ivashina, 2005).

2013 (the first phase), the tax credit was up to 35% for qualified small businesses (QSBs); and in 2014, the tax credit increased from the 35% to 50% for QSBs (the second phase). To qualify, the employer needed to i) have 25 or fewer employees; ii) pay average wages less than $50,000 a year per full-time equivalent; iii) pay at least 50% of its full-time employees' premium costs; and finally, iv) have provided a health plan to employees that is qualified under SHOP requirements.

In addition to raising the tax credit from 35% to 50%, in 2014 the government also launched the Small Business Health Options Program (SHOP) Marketplace to offer small business owners a transparent and convenient platform/exchange to compare and shop for insurance packages. Qualified employers were required to purchase insurance packages via the Marketplace, which directly lists health plan choices certified for the tax credit in which the employers could enroll their employees. The Marketplace was initially planned to be launched at the end of 2013; however, there was a delay in the launch of the marketplace till November of 2014 so that the tax credit could be applied to coverage starting from 2015.[36]

We utilize the tax credit hike in 2014 (i.e., the second phase) to identify the impact of soft information demand on banks' technology adoption. Because 2010 is right after the implementation of "the Recovery Act" in 2009, during which numerous other stimulus policies were launched to aid in post-recession economic recovery, this proximity in timing may contaminate the identification. Perhaps more importantly, several surveys revealed that the first phase of the tax credit was not well implemented; some small businesses think the tax credit in the first phase is not high enough, or were not even aware of the policy after its implementation.[37] On the other hand, after the of tax credit hike and the launch of the SHOP Marketplace in 2014, there is a significant decrease in the number of uninsured small

[36] See the IRS's FAQ regarding the tax claim rules, and see this report for the delay of the Marketplace.

[37] This summary explains the low participation of small businesses in the first few years after 2010: "SHOP programs were operational nationwide, but many features were not initially available, and enrollment had been lower than anticipated. Many small businesses did not enroll because they were apprehensive about joining an unestablished program." Relatedly, the 2012 GAO Report summarized that "the small employers do not likely view the credit as a big enough incentive to begin offering health insurance and to make a credit claim." Regarding small businesses' awareness of the policy, this survey posted in 2011 summarized small businesses' unfamiliarity with the policy.

business employees.[38]

There are many channels through which this program could boost credit demand from small businesses. First of all, the policy is economically significant: the increased tax credit on average can induce a 14% of savings in terms of total net profit.[39] Thanks to the increased program subsidy, some small business owners who previously could not afford employee health coverage were now likely to provide it; and some may even have chosen to expand their businesses by hiring more employees given the lower effective labor cost.[40] More importantly, the nature of the timing for the tax "rebate" incentivizes small businesses to apply for extra business loans, as all business owners would need to borrow in advance to cover employee health packages and then repay the loan once they have claimed the credit the following year. In turn, banks would be handling additional soft information—such as the employee hiring, health plans, etc.—to screen for genuinely credit-worthy borrowers.

The key to our identification is that the fraction of total establishments that are qualified for the tax policy right before the program launch date varies substantially across different counties. Since the qualified small business share is a key determinant for credit demand from local small businesses, such variation thus helps us identify the impact of the small business credit demand shock on local banks' behavior. As the policy only explicitly targets small businesses, its impact on other types of local credit demand would be indirect or limited.

[38]This report finds that uninsured small business employees decreased by around 30% during 2014-2016 compared to 2013.

[39]We provide further evidence for the positive impact of the tax credit hike on QSBs by studying their growth in number of both establishments and employees after the tax policy. The details of these analyses on the mechanism and economic magnitude of the policy impact are provided in Appendix Table A7, Table A8, and Table A9.

[40]Similar expansion of factor input is also documented in Agrawal et al. (2020), who show that following an R&D tax credit to small businesses, which resembles the health insurance tax credit in our paper, firms responded by increasing their R&D spending significantly. Further, Gao et al. (2023) show that following insurance premium increase, firms reduce employment. More broadly, for reactions from small businesses after the implementation of corporate tax cuts or the launch of subsidies, see Cerqua and Pellegrini (2014), Rotemberg (2019), and Ivanov et al. (2021).

Empirical design: 2SLS regression We run the following 2SLS regression:

$$\begin{aligned}\Delta \ln(\text{CRA})_{i,c,post} &= \tilde{\alpha}_i + \mu_1 \left(\frac{\#\text{ Qualified small business est}}{\text{Total }\#\text{ of establishments}}\right)_{c,pre} + \mu_2 \mathbf{X}_{i,c} + \epsilon_{i,c} \\ \Delta \ln \text{IT}_{i,c,post} &= \alpha_i + \beta \Delta \widehat{\ln(\text{CRA})}_{i,c,post} + \gamma \mathbf{X}_{i,c} + \epsilon_{i,c}.\end{aligned} \tag{2}$$

In the first-stage regression, the outcome variable $\Delta\ln(\text{CRA}_{i,c,post})$ is the change in the logarithm of bank i's small business loans in county c in the three-year time window, before and after the policy change in 2014. The instrumental variable $\frac{\#\text{ Qualified small business est}}{\text{Total }\#\text{ of establishments}}_{c,pre}$ is the proportion of total business establishments that have fewer than or equal to 20 employees, averaged between 2011 and 2013 before the shock.[41] In the second stage, we regress $\Delta\ln \text{IT}_{i,c,post}$, which is the change in logarithm of a specific type of IT spending of bank i in county c during 2014-2017 compared to the period of 2011-2013, on the fitted value from the first stage.[42]

The instrument in (2), i.e., the QSB share before the policy shock, is a slow-moving object that reflects the status of the local economy. Our identification assumption is that, conditional on the control variables, the QSB share affects the cross-county growth rate in banks' IT spending around the policy shock only through affecting the small business loans extended in the local economy. Table A8 shows that the growth in small business around the policy year was mostly concentrated in those businesses that were qualified for

[41]Recall that only employers with 25 or fewer employees are qualified for this program. However, the "County Business Pattern" database provides categorization of small businesses sizes (number of employees) based on the following cut-offs: ≤ 5, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-1000, and ≥ 1000. Due to this data limitation, we chose the closest cut-off, which is "fewer than or equal to 20."

[42]We make two points. First, the QSB share, together with the growth rate of bank IT spending and local small business loans, are all independent of scale; this helps alleviate the concern that the heterogeneity in policy exposure might be correlated with the size of the local economy. Second, we use ln(Spending) and ln(Loan) in all of our regression analyses, by removing all observations with zero budget or zero lending. Since we aggregate branch-level observations to bank-county level, the occurrence of zero budget/lending is low—the total amount of observations with zero budget/lending is only 0.6% in our sample. One could use $\ln(1 + x)$ instead of $\ln(x)$ to include observations with zeros; but because the aggregated IT spending (quoted in USD) and CRA loans (quoted in 1K USD) range from a couple of thousands to millions (software spending and CRA lending have medians of 25,000 and 1,400, respectively), which are much larger than 1, $\ln(1 + x)$ and $\ln(x)$ are close to each other. Indeed, these two specifications yield quantitatively similar results (in the second stage, we get 0.67 as opposed to 0.76.)

the tax policy, which corroborates our first stage results that counties with a higher "QSB share" experienced faster growth in small business credit around 2014. Finally, the parallel trend assumption requires that heterogeneity in the qualified small business share explains divergent paths in local banks' IT spending only after the policy, which we empirically verify shortly.

We have included a rich set of pre-shock control variables in regression (2). Bank fixed effects absorb any unobserved heterogeneity that may also induce banks in areas with more qualified small businesses to be on a higher IT spending growth path. Revenue per employee at the bank-county level proxies for investment opportunity of a bank in the local economy. We also add a set of county-level economic characteristics, which include county size (proxied by the logarithm of total number of establishments) and local economic situation (proxied by population growth rate, changes in unemployment, labor force participation ratio, the share of non-tradable sector business establishments, and real GDP per capita).

Besides adding controls and fixed effects, we also perform several placebo tests along various dimensions. In Appendix Table A11, we hypothetically postulate the tax policy event to take place in year 2018 and then examine the effect "QSB share" of the dynamics of small business credit around these pseudo event years. In Appendix Table A12, we test whether the variation in the "QSB share" also drives differences in the growth rate in other types of credit (e.g., mortgage origination or refinancing) around the tax policy time. In either test, we find little effect from the variation in "QSB share," in contrast to its significant impact on the small business credit growth around the policy year of 2014.

Estimation results We report the estimation results of (2) in the first three columns of Table 6. Standard errors are clustered at the county level. Column (1) shows the regression estimates in the first-stage regression with a strong first stage result: the F-statistic of 13.71 is above the conventional threshold for weak instruments (Stock and Yogo, 2005).

We find a positive and statistically significant response in banks' communication investment across counties in the second stage. In particular, banks who were facing a one standard deviation higher growth in their small business loan making—due to a higher policy exposure captured by QSB share—experienced a 0.67 std higher growth in their communication spending; in dollar terms, this translates into an increase of $40,298 in communication IT spending.[43] On the other hand, one standard deviation higher growth in small business loans lead to 0.057 standard deviation slower growth in software spending and is statistically insignificant, suggesting that banks did not respond in increasing their software spending (which is more pertinent to dealing with ready-to-use hard information). Note, by including bank fixed effects, our results come from "within-bank but cross-county" variations. Overall, this asymmetric impact on banks' IT adoption behavior is consistent with our hypothesis that small business lending relies more on soft information handling rather than on processing hard information.

Bank responses and "young firm share" With the premise that young small businesses often lack credit records and thus need extra interaction for loan officers to gather relevant soft information, we test whether the tax policy leads to a larger impact on banks' IT spending response in counties with more young small business borrowers. Specifically, we expand the 2SLS regression in Eq. (2) by introducing an interaction term $\Delta \ln(\text{CRA})_{i,c,post} \times$ High young, where the dummy "High young" takes value 1 for counties whose proportion of small businesses younger than 1 year was above median among all counties in 2013. Table 7 shows that communication spending for banks in the "High young share" counties is the main driver of the overall positive causal impact, while the response of banks in "Low young share" counties is statistically insignificant.

[43] For detailed calculations, see Appendix Table A6. To put this number into perspective, according to our estimation banks can earn around $167K extra revenue (given the $8.366 million extra increase in CRA loans from the first stage), which makes the increased communication IT spending seemingly small. This is expected because soft information relies not only on IT products, but also loan officers who gather and transmit information using the communication IT. Accompanying the communication IT spending, banks will also need to hire extra loan officers (or compensate more hours) when they increase their labor input to deal with the increased small business lending. Since our data does not contain compensation to employees, our calculation only provides a lower bound of the estimation of banks' expenses in response to small business lending growth.

Dynamic treatment effects We now study the dynamics of bank IT spending responses to the policy shock; this also helps us evaluate the validity of the IV by examining the pre-trend patterns of banks' IT spending. We run the following regression with observations of bank i at county c in year t:

$$\ln IT_{i,c,t} = \alpha_{i,t} + \alpha_{i,c} + \sum_{s\in[-3,3],s\neq -1} \beta_s \times \mathbb{1}_{\{\text{t-2014}=s\}} \times \text{High QSB exposure}_{pre} + \Pi_t \times \mathbf{X}_{i,c,t} + \epsilon_{i,c,t}$$

where $\alpha_{i,t}$ and $\alpha_{i,c}$ are bank-year and bank-county fixed effects, and "High QSB exposure" is an indicator variable which equals 1 (0) if the average $\frac{\text{\# Qualified small business est}}{\text{Total \# of establishments}}_{c,pre}$ between 2011 and 2013 sits in the top (bottom) tercile. Note here we allow the coefficients on control variables to be time-varying.

Figure 2 plots the set of estimated coefficients $\left\{\hat{\beta}_s\right\}$, which measures the intent-to-treat (ITT) effects of the policy change on $\ln IT_{i,c,t}$ through heterogeneous exposure as captured by QSB share, with the base year as 2013. Prior to the policy shock, the time trends of both types of IT spending display no significant differences for banks located in high-exposure counties versus those in low-exposure counties. Since the tax credit hike in 2014, the communication spending of banks located in high-exposure counties see a continual growth for two consecutive years (left panel), while the software spending of banks (right panel) in high-exposure counties and low-exposure counties demonstrates no difference before 2014 and remain similar after. Finally, $\left\{\hat{\beta}_s\right\}$ for communication IT spending (left panel) starts to decrease around 2016; this might be due to the "capital" nature of IT investment.[44]

Comparison: OLS estimates We report the OLS estimates in Columns (4)–(5) of Table 6. Qualitatively, OLS estimates are similar to those obtained from the 2SLS method; but in terms of magnitude, they are significantly smaller. One explanation for such a downward bias in OLS estimators could be a potential "omitted variable" problem, in which counties

[44]That is, having built up their "IT capital" stock after two years of high "flow" spending right after the policy shock, bank branches no longer need to install more IT equipment even if the demand for small business credit remains high in these high-exposure regions. This would then translate into a reduction in the flow of IT spending.

experiencing faster growth in small business loans are those with even faster growth in some unobservable economic variables—say, mortgage refinancing demand—that drive local banks to spend less on communication, leading the OLS estimator to be downward biased.

4.3 Bank IT Spending and Hard Information

4.3.1 Hard Information Processing in Bank Lending: Mortgage Refinancing

Unlike the lending activities analyzed in Section 4.2 where soft information handling is key, in other situations banks' ability to extend profitable credit is determined by how efficiently they can deal with hard information. As mentioned earlier, mortgage refinancing is the stereotypical type of loan that relies heavily on efficient processing of readily accessible hard information. The discussion in Section 4.1 suggests that banks' software spending should be positively correlated with mortgage refinancing, an empirical fact that we have shown in Table 4 row 5 in Section 3.2.1.

We move one step further and conduct a similar analysis within the mortgage lending business, by splitting it into mortgage origination and mortgage refinancing. For each bank we construct "Refinance/Origination" over the period of 2010-2019, and Table 4 row 7 shows that banks with a greater "Refinance/Origination" spend more on software, while there is no significant effect on communication spending.

The close linkages between banks' software spending and their engagement in mortgage refinancing is also consistent with a recent strand of literature studying fintech lenders' penetration into credit markets. As documented in Fuster et al. (2019), the expansion of fintech lenders—who often serve as the suppliers of new banking software products and typically rely on readily available hard information—is particularly pronounced in the refinancing segment of the mortgage, auto loans, and student loan markets. Later in Section 5, we confirm that software indeed stands out as the major category of IT spending in which commercial banks respond to the entry of fintech lenders.

4.3.2 Bank IT Spending and Demand Shock on Mortgage Refinancing

Paralleling Section 4.2.2, we ask: How would banks respond when hit by credit demand shocks that mostly involve processing hard information, say mortgage refinancing? We expect banks to increase their spending on software (hard information), but not on communication (soft information).

For exogenous sources of cross-sectional variation in mortgage refinance demand, following Di Maggio et al. (2017) and Eichenbaum et al. (2022) we construct an IV for county-level refinancing propensity by utilizing the post-crisis low interest rate period. The nationwide mortgage rate decrease prompted existing homeowners to refinance their mortgages, and an important determinant of homeowners' refinancing propensity was the pre-crisis mortgage characteristics in place before the low-interest episode kicked in.[45]

We consider two ways to construct the instrumental variable. The first follows Eichenbaum et al. (2022) by constructing the IV as the "dollar amount difference." Specifically, for each loan j in county c with unmatured balance in year t between 2011-2016, we calculate the interest savings under the new mortgage rate compared to the old mortgage rate:

$$
\begin{aligned}
\Delta \text{Payment}_{j,c,t} &= (\text{Total Interest Payment}|\text{mortgage rate}_j) \\
&- (\text{Total Interest Payment}|\text{new mortgage rate}\ _t^{\text{FICO, maturity, zip}})
\end{aligned}
$$

where the total interest payments are calculated from the amortization schedule with the remaining loan balance as principal. The new mortgage rates are constructed by the bucket of "zip $\times$ maturity $\times$ FICO" based on new origination in year t, and then matched to each loan j. We then calculate the average $\Delta\text{Payment}_c$ by taking the average of all savings of unmatured loans j and over years 2011-2016.[46] In words, we calculate the average total

[45] Berger et al. (2021) show that effectiveness of monetary policy is crucially dependent upon the previous levels of mortgage rates.

[46] We construct the payment savings based on the 2011-2016 sample, because the Federal Funds rate and mortgage rate remained at the low level till 2016 (Figure A2).

remaining mortgage savings under old versus new interest rates at the county level.[47] The variation in local homeowners' refinancing savings thus serves as an exogenous shifter on the mortgage refinance demand faced by local banks,

Though frequently utilized by the previous literature, $\Delta\text{Payments}_c$ is likely correlated with the remaining loan balances of a county, which are in turn correlated with the average loan size or house price level of a county. Therefore $\Delta\text{Payments}_c$ may correlate with local banks' IT spending due to other channels beyond mortgage refinance demand. As an alternative IV, we construct the average mortgage rate gap between the rates at origination and the current rate for the unmatured existing mortgages in a county c:

$$\begin{aligned}\Delta\text{Mortgage rate}_{c,t} &= \sum_j(\text{mortgage rate}_{j,c} - \text{new mortgage rate}_t^{(\text{FICO, maturity, zip})}) \\ &\quad \times \frac{\text{Total loan amount}_{j,c}}{\text{Total loan amount during 1999-2010}_c}.\end{aligned}$$

We then take the average of $\Delta\text{Mortgage rate}_{c,t}$—which captures the average mortgage *rate* savings instead of *dollars* at year t—for mortgage borrowers across years 2011-2016. In words, for a given county we compute the weighted average of mortgage interest rate gaps, with weights as the loan volume at the initiation.

Empirical design and estimation results We aim to identify whether banks' software investment increases given a greater mortgage refinance demand compared with mortgage origination, with the following standard 2SLS specification:

$$\begin{aligned}\ln(\text{Refinance/Origination})_{i,c} &= \tilde{\alpha}_i + \mu_1\Delta\text{Payments}_c/\Delta\text{Mortgage rate}_c + \mu_2\mathbf{X}_{i,c} + \tilde{\epsilon}_{i,c}, \\ \ln(\text{Software})_{i,c} \text{ or } \ln(\text{Communication})_{i,c} &= \alpha_i + \beta\ln(\widehat{\text{Refinance/Origination}})_{i,c} + \gamma\mathbf{X}_{i,c} + \epsilon_{i,c}.\end{aligned} \tag{3}$$

Similar to before, our control variables include banks' revenue per employee and deposit market share of the bank in a county. County level control variables include the unemploy-

[47] We remove loans that were defaulted on or prepaid to ensure that the measure captures only refinance propensity from local households with outstanding loans.

ment rate, labor force participation rate, population growth rate, logarithm of number of establishments, and logarithm of small business loans. We include bank fixed effects and cluster standard error at county level.

Table 8 reports our estimation results. In the first stage of column (1), the instrumental variable "ΔMortgage rate$_c$" predicts mortgage refinancing activities quite well, with a high F-statistics (10.81). For the second-stage, Columns (2) and (3) show that a one standard deviation increase in mortgage refinancing relative to mortgage origination—driven by its local exposure to high refinance interest savings—leads to a 0.315 standard deviation increase in software spending. In dollar terms, this translates to an increased software spending of \$133.26K. This increase is of a similar magnitude, though smaller than, the corresponding \$455.5K revenue increase from mortgage refinancing,[48] and it is worth emphasizing that our current data does not provide a comprehensive estimate of costs across other dimensions.

Columns (4)-(6) show the results using ΔPayment$_c$ as the instrumental variable. Note, while the difference between coefficient estimates with these two IVs is statistically insignificant (consistent with the premise that both IVs give us unbiased estimates), we believe ΔMortgage rate$_c$ satisfies the exclusion restriction condition better. In Appendix Table A15, we tabulate the correlations between ΔPayment$_c$ and ΔMortgage rate$_c$ with major county-level economic variables. As shown, ΔMortgage rate$_c$ exhibits statistically and economically insignificant correlations with most of the county characteristic variables, while ΔPayment$_c$ has positive correlations with some of them (though of relatively small magnitude).

Finally, by including bank fixed effects, our result is identified from within-bank-cross-county variations. In addition, communication spending does not demonstrate statistically significant changes in response to the refinancing demand shocks, which supports our premise that mortgage refinancing is a stereotypical lending activity that hinges on efficient processing of readily accessible hard information instead of producing new information.

[48]This increase of revenue is implied by a one standard deviation increase in ln(Refinance/Origination); detailed calculation is provided in Appendix Table A6.

Comparison: IV estimates and OLS estimates We conduct the OLS version of the 2SLS regression in Eq. (3) and report the results in the last two columns of Table 8, with quantitatively smaller OLS estimators.[49] Similar to our analysis of small business credit demand in Section 4.2.2, an "omitted variable" issue can explain such downward biases in OLS estimators. Here, counties seeing more mortgage refinances issued by local banks might also have other loan demands that recovered more significantly during the post-crisis period (say, small business loans), which might then tilt local banks' IT budget towards other types of IT spending (say, communication as shown in Section 4.2), lowering their spending on software. Our instrumental variable used in the 2SLS method addresses this issue.

5 Bank IT Spending and Fintech Entry

In recent years, the emergence and expansion of fintech lenders have drawn heightened public attention to the competition between fintech lenders and traditional banks. Via the angle of examining commercial banks' IT spending, we aim to study a widely debated question: Has the traditional banking sector started reacting to the fast-growing fintech industry? If yes, how?

5.1 How Should Banks React to Fintech Entry?

Existing studies suggest that fintech lenders' services involve better use of technology and little human interaction. This tech-intensive feature improves customer experience and likely reduces lending-associated costs (Buchak et al., 2018a; Fuster et al., 2019).

While fintech lenders have been quickly gaining market share in various markets over the past decade, it remains unclear how the incumbent commercial banks should react. For instance, when banks and non-bank lenders offer complementary services, it is possible for banks to strategically shift investment towards areas with fewer activities from fintech lenders. Furthermore, from an information channel, the emergence of fintech lenders who

[49] Table A4 shows the results of the same OLS specification with bank, year and county fixed effects and bank×year and county fixed effects.

have comparative advantages in information handling in certain markets would render traditional bank lenders more adversely selected in these markets. Both would imply a "falling back" of traditional banks from the markets with fintech entry and a lowered investment in the IT category that fintech lenders have comparative advantages in.

On the other hand, incumbent banks might instead choose to protect their market share and compete against these new fintech entrants, suggesting a potential "catching-up" behavior of the traditional banking sector. Which economic force dominates is an empirical question that we now aim to answer.

5.2 Entry of Lending Club and Local Bank IT Investment

To causally identify banks' response in their IT spending towards the increasing presence of fintech lenders, we employ a difference-in-difference strategy that relies on the staggered entrance of Lending Club into different states.

Staggered entry of Lending Club As one of the leading players in the fintech industry, Lending Club launched its platform in 2007. Since 2008, Lending Club has been pursuing regulatory approval to conduct peer-to-peer lending in all 50 states. By October 2008, 41 states moved relatively fast to approve its entry; and between 2010 and 2016, another nine states approved Lending Club's entrance at different times.[50] Table 9 summarizes the timing of Lending Club's staggered entrance into different states.

Following Wang and Overby (2017) and Kim and Stähler (2020), we first drop the 41 states who approved Lending Club's entry in 2008.[51] For Kansas and North Carolina, the

[50] As explained by Wang and Overby (2017), Lending Club launched its platform in 2007. In April 2008, Lending Club entered a "quiet" period, in which it suspended peer-to-peer lending until it registered with federal and state regulators as a licensed lender (or loan broker). During this quiet period, Lending Club funded some loans with its own money, and pursued regulatory approval to resume peer-to-peer lending in all 50 states. Six months later, it had received approval in 40 states, plus the District of Columbia by October 2008. For nine states, it received approval at different times between 2010 and 2016. For one state (Iowa), it had not received approval as of February 2021.

[51] Given that a majority of states approved Lending Club around the same time period (2008-Q4), a potential concern of endogeneity arises: as these approvals occurred shortly after applications by Lending Club who might have seen a rising opportunity from entering, these approvals might coincide with some unobserved changes in economic conditions happening during the same time.

actual approval was in 2010Q4. Since 2010 is the starting year of our Harte Hanks dataset, 2010 as a pre-treatment period is contaminated for these two states. We hence also exclude these two states, leaving us with a total of seven states for our staggered entrance analysis.

Importantly for our identification purpose, the variation in the approval time since 2010—presumably due to variations in administrative efficiency and potential political issues across states—allows us to get around several major endogeneity concerns regarding the entry of Lending Club. For instance, if Lending Club were to chose to enter the local markets with rising credit demand, then any observed change in local commercial banks' IT investment behavior could not have been convincingly attributed to the entry of their fintech challenger.

For the states in our sample, after its entrance, the personal loan issuance market share of Lending Club across states has a median of 4.85%, with 1.79% (10.29%) being its 25^{th} (75^{th}) percentile (Appendix Table A17). These statistics suggest that i) at the state level, Lending Club's presence in the personal loan market features significant variations; and ii) in states where Lending Club actively operates, it makes a nontrivial contribution to the local personal loan market. Both of these two facts are important for our empirical identification, in which the key variation (driven by differences in approval time) operates at the state level.

From the perspective of incumbent banks, Table A18 shows that personal loans represent a significant portion of banks' interest income among all categories of loans, especially for larger banks (banks with more than \$10 billion in assets): around 20% of interest income comes from personal loans. This indicate that banks have a compelling reason to react when fintech lenders emerge in one of their most profit-generating loan segments.

Empirical design and results Our empirical method follows the staggered difference-in-difference design as in Wang and Overby (2017). The regression specification is

$$\ln(\text{IT Spending})_{i,c,t} = \alpha_{i,t} + \alpha_c + \beta \times \text{LC}_{i,c,t} + \mu_t X_{i,t} + \epsilon_{i,c,t}, \tag{4}$$

where IT Spending $\in$ {Software, Communication}. We include the bank-year and county fixed effects, denoted by $\alpha_{i,t}$ and α_c respectively; and controls $X_{i,t}$ are in the caption of Table 10. $\mathrm{LC}_{i,c,t}$ is a dummy variable that is equal to one if Lending Club entered the state where county c is located in year t for bank i; β hence measures the average treatment effect of Lending Club entry on bank technology spending. Estimations are weighted by Lending Club loan volume after entry, and the standard error is clustered at county level.

Columns (1) to (2) in Table 10 Panel A report the results for software and communication spending, respectively. Consistent with the "catching-up" story, column (1) shows that, after Lending Club entered country c, banks on average increase their software IT spending in county c by around 7.6 percentage points, and this estimate is statistically significant. In contrast, communication spending right after Lending Club's entry displays no statistically significant changes compared to pre-entrance.

Figure 3 graphically explores the dynamics of banks' IT spending within the 3-year time window around the fintech entrance year, from the following estimation:

$$\ln \mathrm{IT}_{i,c,t} = \alpha_{i,c} + \mu_t + \sum_{s\in[-3,3],s\neq-1} \beta_s \times \mathbb{1}_{s=t-\text{entrance year}} + \Pi_t \mathbf{X}_{i,c,t} + \epsilon_{i,c,t},$$

where fixed effects and controls are the same as in Eq. (4) and the coefficients on the controls are allowed to be time varying. The estimated $\{\hat{\beta}_s\}$ and the 95% confidence intervals are plotted. Importantly for our identification, there is no statistically significant pre-trend in either type of IT spending before the fintech entrance, which allows us to plausibly attribute changes in banks' IT spending to the penetration of fintech into the local economy. Consistent with Table 10, a bank's software spending displayed a significantly sharper increase than communication IT spending after the fintech entry.

Recent literature points out the bias in a staggered two-way fixed effects (TWFE) setting, even if the assumption of parallel trends holds. For robustness, we use the interacted TWFE design as in Callaway and Sant'Anna (2021).[52] As shown in columns (4) to (5) of Table 10, the estimates are similar to, albeit a little larger than, those in columns (1) and (2).

Heterogeneity in responses across bank sizes In Panel B of Table 10, we explore whether banks of different sizes respond differently to fintech entry. Similar to our specification in Table 5, large (small) banks are defined as lenders with asset size above (below) the median size in our sample. We find that large banks increased software spending by 6.2 percentage points more compared to small banks after Lending Club's entry, and the difference is statistically significant. On the other hand, large banks cut their communication spending by 5.8 percentage points compared to small banks following the fintech entry, which is statistically significant. Note, via a different instrument variable, Modi et al. (2022) also document that large banks increase their IT spending when facing competition from fintech lenders, but our data allow us to speak to the underlying mechanism of such response by separating different categories of IT spending, showing that it is mainly driven by "hard" information considerations.

The asymmetric impact on the IT spending reactions by different sized banks is intriguing, and suggests that the specialty (regarding information handling) of the newly entered fintech is more relevant for the market segments served by large banks. This is consistent with Balyuk et al. (2020), who find that fintech lending often substitutes lending made by large banks rather than small banks. Given that small banks engage more in relationship-based small business lending, the entry of Lending Club—who is equipped with superior hard information processing capacity—will not strongly affect these banks' profit making. Furthermore, that large banks cut their communication spending is also consistent with the recent literature studying how fintech entry affects credit market outcomes. For instance, as documented by Balyuk et al. (2020), credit extended by fintech entrants often substi-

[52] In this method, we run separate regressions in (4) for each group of states that are treated at the same year, with the not-yet-treated as the comparison group, and then aggregate β to form the aggregated average treatment effect of the treated (ATT). For aggregation, we weight the cohort-specific treatment effect by the total volume of loans made through lending club within the three years after Lending Club's entry. Standard errors are based on Bootstrapping with 50 draws.

tutes for loans by out-of-market banks (which are often large ones), as opposed to those by small/in-market banks. As a consequence of large banks' retracted engagement in out-of-market lending, which often rely on the support of communication IT, one should naturally expect them to reduce their communication IT spending.

That banks' IT spending responses are size-dependent is also consistent with the notion that the entry of fintech lenders helps convert soft information to hard information.[53] Linking this "hardening soft-information" effect to our analysis where the focus is placed on bank lenders' decision making, one should expect large banks—rather than small ones who specialize more in relationship-based soft information handling—to reallocate their investment away from communication to software due to a decreased (increased) need of dealing with soft (hard) information.

Finally, recall that in Section 3.2.1 we document a strong correlation between banks' software IT spending and their specialization in the personal loan lending, which is what Lending Club mainly focuses on. Consistent with this, Table 10 Panel C shows that banks with higher personal loan shares respond more significantly to the entry of Lending Club.

Summary and discussion We find that the fintech entry induces banks—especially large ones—to "catch up" and invest to adapt their lending technology. To the best of our knowledge, this is the first piece of direct evidence that the entry of fintech lenders spurs incumbent banks to invest more in their technology to catch up. Furthermore, consistent with existing literature (say, Berg et al., 2021) that highlights the comparative advantage of fintech lenders in processing hard information and making prompt decisions, we show that most "catching up" from traditional banks takes the form of ramping up their *software* IT spending.

We have discussed in Section 5.1 the potential channels through which the entry of fintech lenders affects local commercial banks' IT investment decisions. Our empirical findings support a competition story that, following fintech entry, large banks respond by increasing their IT spending in the relevant categories, presumably to protect their market share.

[53] For instance, Beaumont et al. (2019) show that borrowers with better fintech-access are more likely to purchase and pledge hard-information-heavy assets as collateral to obtain new bank credit.

Behind this increased investment in IT could be a "winner's curse" channel that banks need to upgrade their lending technology for fear of being adversely selected by the newly entered fintech competitors, once they have decided to continue operating in the same market segment. However, to fully assess this channel one would need to investigate the composition change of banks' customers induced by the entry of fintech lenders, as well as the dynamics of market share composition. We leave these endeavors to future research.

6 Conclusion

Development of information technologies over the past several decades has dramatically revolutionized the way lending is conducted by the banking sector. In this paper, we provide the first comprehensive study of banks' IT spending, which we view as banks' investment to improve their lending technology, especially their ability to deal with soft information and hard information.

The detailed IT spending profiles available in our unique dataset enable us to uncover several novel findings. First, at the aggregate level, we document an overall fast-growing trend in banks' IT spending in the last decade. Second, as a key step in linking banks' IT spending to the development of their lending technology, we show that different types of information technology are closely related to the nature of information embedded in different types of lending activities. More specifically, the production and transmission of "soft" information, which plays a crucial role in conducting small business lending or performing the role of a "lead" bank in syndicated lending, is strongly associated with banks' communication spending. By contrast, "hard" information processing, which is most relevant for conducting mortgage refinancing, is strongly associated with banks' software spending.

We conduct a set of event-based analyses whose answers inform us of how banks adapt their lending technologies in response to economic shocks on their operating environment, including credit demand shocks and the entry of fintech. These causal analyses, to the best of our knowledge, provide the first piece of evidence on the endogenous lending technology adoption in the banking literature.

Our findings open up several important follow-up questions. How does endogenous technology adoption in the banking sector transform the banking/credit market structure? How do technology upgrades in the banking sector affect banks' deposit-taking activities, loan outcomes, properties of the credit cycle, and monetary policy transmission? We leave these questions to future research.

References

Agarwal, S., Ambrose, B., Chomsisengphet, S., and Liu, C. (2011). The role of soft information in a dynamic contract setting: Evidence from the home equity credit market. *Journal of Money, Credit and Banking*, 43:633–655.

Agrawal, A., Rosell, C., and Simcoe, T. (2020). Tax credits and small firm r&d spending. *American Economic Journal: Economic Policy*, 12(2):1–21.

Ahnert, T., Doerr, S., Pierri, N., and Timmer, Y. (2021). Does IT Help? Information Technology in Banking and Entrepreneurship. *IMF Working Paper, 21/214*.

Aiello, D., Garmaise, M. J., and Natividad, G. (2020). Competing for deal flow in local mortgage markets. *Working Paper*.

Amore, M. D., Schneider, C., and Žaldokas, A. (2013). Credit supply and corporate innovation. *Journal of Financial Economics*, 109(3):835–855.

Avraham, D., Selvaggi, P., and Vickery, J. (2012). A structural view of u.s. bank holding companies. *Economic Policy Review*, (07):65–81.

Balyuk, T., Berger, A., and Hackney, J. (2020). What is fueling fintech lending? the role of banking market structure. *Working Paper*.

Banna, H. and Alam, M. R. (2021). Is digital financial inclusion good for bank stability and sustainable economic development? evidence from emerging asia. *Working Paper*.

Barkai, S. (2020). Declining labor and capital shares. *The Journal of Finance*, 75(5):2421–2463.

Beaumont, P., Tang, H., and Vansteenberghe, E. (2019). The role of fintech in small business lending. *Working Paper*.

Berg, T., Fuster, A., and Puri, M. (2021). Fintech lending. Working Paper 29421, National Bureau of Economic Research.

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

Berger, A. N. and Udell, G. F. (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *The Economic Journal*, 112(477):F32–F53.

Berger, A. N. and Udell, G. F. (2006). A more complete conceptual framework for sme finance. *Journal of Banking and Finance*, 30(11):2945 – 2966.

Berger, D., Milbradt, K., Tourre, F., and Vavra, J. (2021). Mortgage prepayment and path-dependent effects of monetary policy. *American Economic Review*, 111(9):2829–78.

Bircan, C. and De Haas, R. (2019). The Limits of Lending? Banks and Technology Adoption across Russia. *The Review of Financial Studies*, 33(2):536–609.

Bloom, N., Garicano, L., Sadun, R., and Reenen, J. V. (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science*, 60(12).

Bolton, P., Freixas, X., Gambacorta, L., and Mistrulli, P. E. (2016). Relationship and Transaction Lending in a Crisis. *The Review of Financial Studies*, 29(10):2643–2676.

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

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

Calebe de Roure, L. P. and Thakor, A. V. (2019). P2P Lenders versus Banks: Cream Skimming or Bottom Fishing? *SAFE Working Paper No. 206*.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230. Themed Issue: Treatment Effect 1.

Cerqua, A. and Pellegrini, G. (2014). Do subsidies to private capital boost firms' growth? A multiple regression discontinuity design approach. *Journal of Public Economics*, 109:114–126.

Charoenwong, B., Kowaleski, Z. T., Kwan, A., and Sutherland, A. (2022). RegTech. *Working Paper*.

Chava, S., Oettl, A., Subramanian, A., and Subramanian, K. V. (2013). Banking deregulation and innovation. *Journal of Financial Economics*, 109(3):759–774.

Chen, B. S., Hanson, S. G., and Stein., J. C. (2017). The decline of big-bank lending to small business: Dynamic impacts on local credit and labor markets. *NBER Working Paper Series, No. 23843*.

Cornelli, G., Doerr, S., Franco, L., and Frost, J. (2022). Funding for fintechs: patterns and drivers. *Working Paper*.

Degryse, H., Laeven, L., and Ongena, S. (2008). The Impact of Organizational Structure and Lending Technology on Banking Competition. *Review of Finance*, 13(2):225–259.

Di Maggio, M., Kermani, A., Keys, B. J., Piskorski, T., Ramcharan, R., Seru, A., and Yao, V. (2017). Interest rate pass-through: Mortgage rates, household consumption, and voluntary deleveraging. *American Economic Review*, 107(11):3550–88.

Di Maggio, M. and Yao, V. (2020). Fintech borrowers: lax Screening or cream-skimming? *The Review of Financial Studies*.

Eichenbaum, M., Rebelo, S., and Wong, A. (2022). State-dependent effects of monetary policy: The refinancing channel. *American Economic Review*, 112(3):721–61.

Erel, I. and Liebersohn, J. (2020). Does finTech substitute for banks? Evidence from the paycheck protection program. *Working Paper*.

Feyen, E., Frost, J., Gambacorta, L., Natarajan, H., and Saal, M. (2021). Fintech and the digital transformation of financial services: implications for market structure and public policy. *BIS Working Paper 117*.

Forman, C., Goldfarb, A., and Greenstein, S. (2012). The internet and local wages: A puzzle. *American Economic Review*, 102(1):556–75.

Freixas, X. and Rochet, J.-C. (2008). *Microeconomics of Banking, 2nd Edition*. MIT Press Books. The MIT Press.

Frost, J., Gambacorta, L., Huang, Y., Shin, H. S., and Zbinden, P. (2019). Investment in ict, productivity, and labor demand : The case of argentina. *BIS Working Papers*.

Fuster, A., Plosser, M., Schnabl, P., and Vickery, J. (2019). The role of technology in mortgage lending. *The Review of Financial Studies*, 32(5).

Gao, J., Ge, S., Schmidt, L., and Tello-Trillo, C. (2023). How do health insurance costs affect firm labor composition and technology investment? *Working Paper*.

Gopal, M. and Schnabl, P. (2022). The Rise of Finance Companies and FinTech Lenders in Small Business Lending. *The Review of Financial Studies*, 35(11):4859–4901.

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

Hauswald, R. and Marquez, R. (2006). Competition and Strategic Information Acquisition in Credit Markets. *The Review of Financial Studies*, 19(3):967–1000.

He, Z., Huang, J., and Parlatore, C. (2023a). Multi-dimensional information with specialized lenders. *Available at SSRN 4557168*.

He, Z., Huang, J., and Zhou, J. (2023b). Open banking: Credit market competition when borrowers own the data. *Journal of Financial Economics*, 147(2):449–74.

He, Z. and Song, Z. (2022). Agency mbs as safe assets. Technical report, National Bureau of Economic Research.

Hitt, L., Frei, F., and Harker, P. (1999). How financial firms decide on technology. *Brookings Wharton Papers on Financial Services*.

Hornuf, L., Klus, M. F., Lohwasser, T. S., and Schwienbacher, A. (2018). How do banks interact with fintechs? forms of alliances and their impact on bank value. *CESifo Working Paper*.

Hornuf, L., Klus, M. F., Lohwasser, T. S., and Schwienbacher, A. (2021). How do banks interact with fintech startups? *Small Business Economics*, 57(3):1505–1526.

Huang, J. (2022). Fintech expansion. *Available at SSRN 3957688*.

Hughes, J., Jagtiani, J., and Moon, C.-G. (2019). Consumer lending efficiency:commercial banks versus a fintech lender. *FRB of Philadelphia Working Paper No. 19-22*.

Huvaj, M. N. and Johnson, W. C. (2019). Organizational complexity and innovation portfolio decisions: Evidence from a quasi-natural experiment. *Journal of Business Research*, 98:153–165.

Ivanov, I., Pettit, M. L., and Whited, T. (2021). Taxes depress corporate borrowing: Evidence from private firms. *Working Paper*.

Ivashina, V. (2005). Structure and pricing of syndicated loans. *Working Paper*.

Jagtiani, J. and Lemieux, C. (2017). Fintech lending: Financial inclusion, risk pricing, and alternative information. *FRB of Philadelphia Working Paper No. 17-17*.

Kim, J.-H. and Stähler, F. (2020). The impact of peer-to-peer lending on small business loans. *Working Paper*.

Kovner, A., Vickery, J., and Zhou, L. (2014). Do big banks have lower operating costs? *Economic Policy Review*, 20(2).

Kwon, S., Ma, Y., and Zimmermann, K. (2022). 100 years of rising corporate concentration. *Working Paper*.

Lerner, J., Seru, A., Short, N., and Sun, Y. (2021). Financial innovation in the 21st century: Evidence from u.s. patents. Working Paper 28980, National Bureau of Economic Research.

Levine, R., Lin, C., Peng, Q., and Xie, W. (2020). Communication within Banking Organizations and Small Business Lending. *The Review of Financial Studies*, 33(12):5750–5783.

Liberti, J. M. and Mian, A. R. (2009). Estimating the effect of hierarchies on information use. *The Review of Financial Studies*, 22(10):4057–4090.

Lorente, C., Jose, J., and Schmukler, S. L. (2018). The fintech revolution: A threat to global banking? Research and Policy Briefs 125038, The World Bank.

McGrattan, E. R. (2017). Intangible capital and measured productivity. Working Paper 23233, National Bureau of Economic Research.

Mian, A. and Sufi, A. (2014). What explains the 2007–2009 drop in employment? *Econometrica*, 82(6):2197–2223.

Modi, K., Pierri, N., Timmer, Y., and Pería, M. S. M. (2022). The anatomy of banks' it investments: Drivers and implications. *IMF working paper*.

Paravisini, D. and Schoar, A. (2016). The incentive effect of scores: Randomized evidence from credit committees,. *NBER Working Paper, 19303*.

Petersen, M. A. and Rajan, R. G. (2002). Does distance still matter? the information revolution in small business lending. *The Journal of Finance*, 57(6):2533–2570.

Pierri, N. and Timmer, Y. (2020). Tech in fin before fintech: Blessing or curse for financial stability? *IMF Working Paper, No. 20/14*.

Pierri, N. and Timmer, Y. (2022). The importance of technology in banking during a crisis. *Journal of Monetary Economics*.

Ridder, M. D. (2019). Market Power and Innovation in the Intangible Economy. *Working Paper*.

Rotemberg, M. (2019). Equilibrium effects of firm subsidies. *American Economic Review*, 109(10):3475–3513.

Sahar, L. and Anis, J. (2016). Loan officers and soft information production. *Cogent Business & Management*, 3(1):1199521.

Skrastins, J. and Vig, V. (2018). How Organizational Hierarchy Affects Information Production. *The Review of Financial Studies*, 32(2):564–604.

Stein, J. C. (2002). Information production and capital allocation: Decentralized versus hierarchical firms. *The Journal of Finance*, 57(5):1891–1921.

Stock, J. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, pages 80–108. Cambridge University Press, New York.

Stulz, R. M. (2019). FinTech, BigTech, and the future of banks. *Journal of Applied Corporate Finance*, 31(4):86–97.

Sufi, A. (2007). Information asymmetry and financing arrangements: Evidence from syndicated loans. *The Journal of Finance*, 62(2):629–668.

Tang, H. (2019). Peer-to-Peer lenders versus banks: Substitutes or complements? *The Review of Financial Studies*, 32(5):1900–1938.

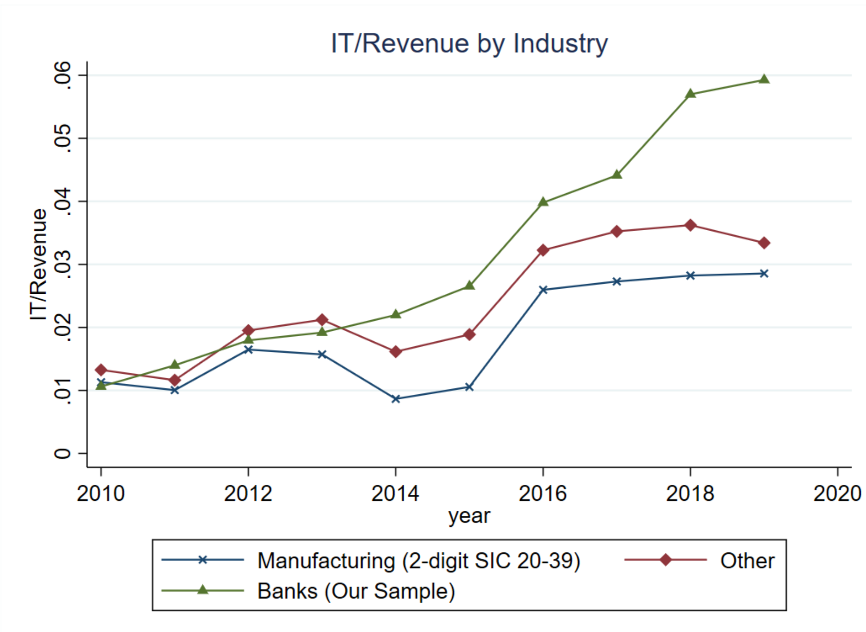
Tuzel, S. and Zhang, M. B. (2021). Economic stimulus at the expense of routine-task jobs. *The Journal of Finance*, 76(6):3347–3399.

Vives, X. and Ye, Z. (2021). Information technology and bank competition. *Working Paper*.

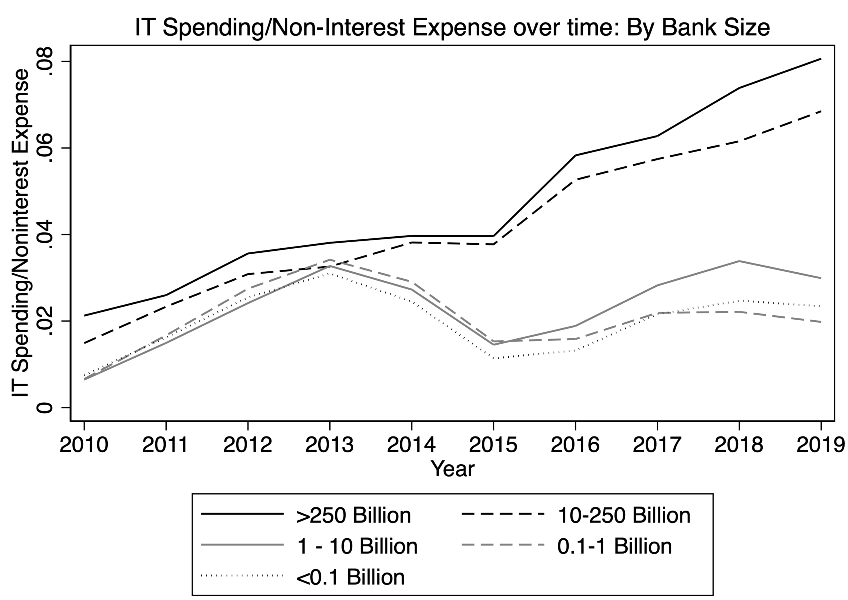
Wang, H. and Overby, E. (2017). How does online lending influence bankruptcy filings? evidence from a natural experiment. *Management Science*, 2022:15937.

Figure 1. IT Spending: Time Trend

Panel A shows the evolution of IT spending as a share of total revenue of banks in our sample, the manufacturing sector, and all other industries constructed using IT budget and revenue from Harte Hanks. "Manufacturing" sector is defined as establishments with 2-digit SIC code 20-39. "Other" sector is defined as all industries other than "Depository institutions" (2-digit SIC code=60). "Banks (our sample)" refers to banks in our sample. These ratios are calculated by aggregating total IT spending and then scaling it against the total revenues sourced from Harte Hanks. In Panel B, the vertical axis is banks' total IT spending scaled by non-interest expenses. The asset size groups are categorized based on a bank's average asset size during 2010 and 2019. Non-interest expenses are calculated using banks' balance sheet item "RIAD4093" in the Call Report.



(a) Panel A



(b) Panel B

Figure 2. Bank IT Spending Around Small Business Tax Credit Policy

This figure reports the event studies of IT spending around the small business tax credit event. The specification is

$$\ln IT_{i,c,t} = \alpha_{i,t} + \alpha_{i,c} + \sum_{s\in[-3,3],s\neq-1} \beta_s \times \mathbb{1}_{\{\text{t-2014}=s\}} \times \text{High QSB exposure}_{pre} + \Pi_t \times \mathbf{X}_{i,c,t} + \epsilon_{i,c,t}$$

where for bank i at county c in year t, $\alpha_{i,t}$ are the bank-year fixed effects, $\alpha_{i,c}$ are the bank-county fixed effects. $1_{\{t-2014=s\}}$ is a dummy variable that is equal to one if the distance between year t and the event year (2014) is s. "High exposure$_{pre}$" is equal to one if the average $\left(\frac{\text{\# Qualified small business est}}{\text{Total \# of establishments}}\right)_{c,pre}$ is within the top tercile between 2011-2013; "High exposure" is equal to zero if the average $\left(\frac{\text{\# Qualified small business est}}{\text{Total \# of establishments}}\right)_{c,pre}$ in the bottom tercile between 2011-2013. Bank control variables include banks' revenue per employee of the bank in a county. County control variables include unemployment rate, labor force participation rate, population growth rate, logarithmic of total number of establishments, share of small businesses in non-tradable sector, and GDP per capita. Shaded regions are the 95% confidence interval of the estimated β_s. Standard errors are clustered at the county level.

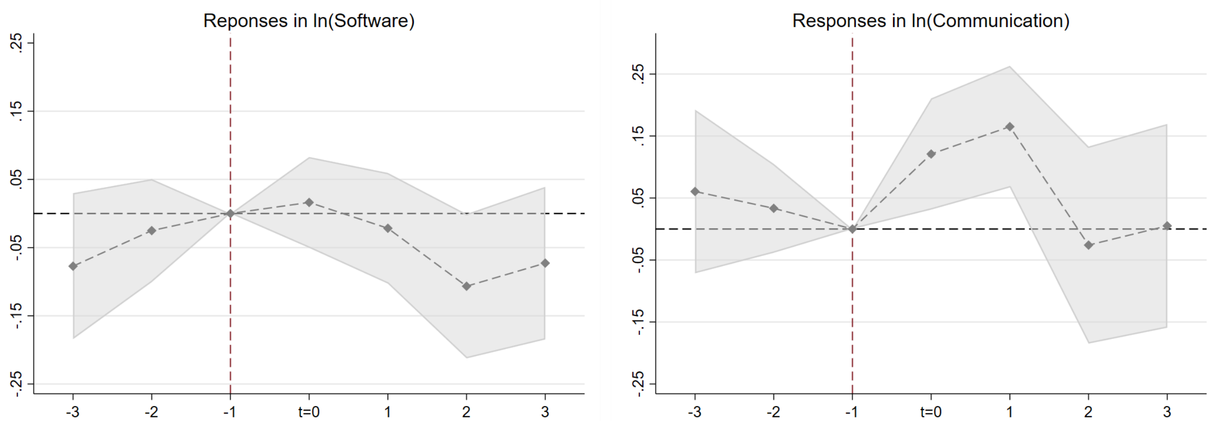


Figure 3. Bank IT Spending Around Fintech Entrance

This figure reports the event studies of IT spending around the entrance of Lending Club. The specification is

$$\ln IT_{i,c,t} = \alpha_{i,c} + \mu_t + \sum_{s\in[-3,3],s\neq-1} \beta_s \times \mathbb{1}_{t-\text{entrance year}=s} + \Pi_t \mathbf{X}_{i,c,t} + \epsilon_{i,c,t}$$

where for bank i at county c in year t, $\alpha_{i,c}$ are the bank-county fixed effects, μ_t are the year fixed effects. $\mathbb{1}_{t-\text{entrance year}=s}$ is a dummy variable that is equal to one if the number of years between the observation year t and the Fintech entrance year into the state where county c is located is s. For the left panel, the left-hand-side variable is logarithmic spending on software IT. For the right panel, the left-hand-side variable is logarithmic spending on communication IT. Bank control variables include banks' revenue per employee of the bank in a county. County control variables include unemployment rate, labor force participation rate, population growth rate, logarithmic of total number of establishments, share of small businesses in non-tradable sector, and GDP per capita. Shaded regions are the 95% confidence interval of the estimated β_s. Shaded regions are the 95% confidence interval of the estimated β_s. Standard errors are clustered at the county level.

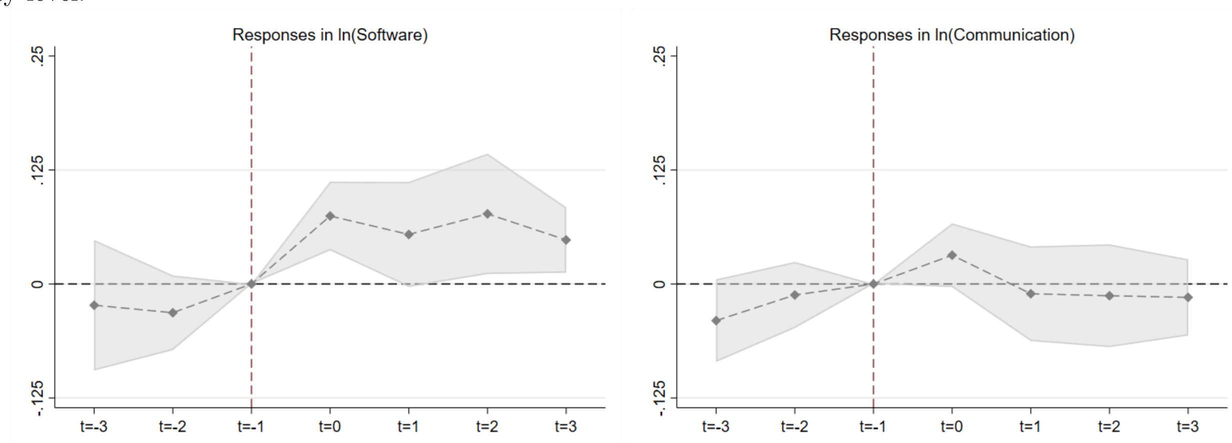


Table 1. Sample Coverage

This table demonstrates the sample coverage of banks across five categories of banks' size groups. The Call Report bank population is constructed by applying the commercial bank restriction ("Charter Type" being 200) following FFIEC definition. The first two columns show the number of banks and the average asset sizes of banks in our sample, across five size groups. Column 3 and column 4 show the total number of banks and average asset sizes of all banks in the Call Report. Column 5 shows the percentage of sample coverage in terms of frequency compared with the population in Call Report, and column 6 shows the percentage of sample coverage in terms of total asset size compared with the population in Call Report.

Coverage of data	Sample		Call Report		Freq %	Asset %
Average Assets 2010-2019 (Billion)	Num banks	Ave Assets	Num banks	Ave Assets		
>$250 Billion	6	1196.15	6	1196.15	100%	100%
$10 Billion–$250 Billion	98	43.82	106	43.69	92.45%	92.72%
$1 Billion–$10 Billion	418	2.95	590	2.78	70.85%	85.62%
$100 Million–$1 Billion	734	0.42	4161	0.32	17.64%	23.43%
<$100 Million	194	0.06	2048	0.05	9.47%	11.36%

Table 2. IT Spending Summary Statistics

This table shows the summary statistics of banks' IT Spending. Total IT Spending is the sum of all types of IT spending in millions of dollars. "IT Spending/Revenue" is total IT Spending scaled by banks' total gross income ("Revenue" is RIAD4000 of Call Report); "IT Spending/Non-interest expense" is total IT spending scaled by non-interest expenses ("Non-interest expenses" is RIAD4093 in Call Report); "IT spending/Net income" is total IT spending scaled by total income minus the gross total expenses ("Net income" is total income minus the sum of interest expenses and non-interest expenses, or the sum of RIAD4073 and RIAD4093 in Call Report). The different categories of IT spending are the four categories of IT spending scaled by total IT spending.

	Mean	S.d.	p(25)	Median	p(75)
Total IT Spending (Million)	11.125	160.239	0.024	0.159	0.796
No. of IT Employees	178.756	1828.766	5.000	20.682	56.912
IT Spending/Income	0.020	0.039	0.006	0.012	0.021
IT Spending/Net income	0.068	0.113	0.017	0.037	0.084
IT Spending/Expenses	0.022	0.027	0.008	0.014	0.026
IT Spending/Noninterest expense	0.051	0.036	0.009	0.018	0.035
Communication/Total	0.089	0.108	0.028	0.051	0.110
Software/Total	0.334	0.172	0.219	0.315	0.468
Hardware/Total	0.172	0.111	0.066	0.161	0.235
Services/Total	0.327	0.129	0.243	0.329	0.415
Other/Total	0.066	0.104	0.009	0.022	0.111

Table 3. C&I Loans and Banks' IT Spending

This table presents the results of the regression of banks' C&I loan on the four major categories of banks' IT spending and a vector of control variables at bank-year level. The sample period is 2010 to 2019.

$$\frac{\text{Type S IT spending}}{\text{Revenue}}_{i,10\text{-}19} = \alpha + \beta \frac{\text{C\&I Loan}}{\text{Total loan}}_{i,10\text{-}19} + \gamma \mathbf{X} + \epsilon_i$$

C&I Loan/Total Loan is commercial and industrial loan of bank i scaled by total loan between 2010-2019, Software/Rev is software spending scaled by total revenue, Communication/Rev is communication spending scaled by total revenue, Hardware/Rev is Hardware spending scaled by total Revenue, Services/Rev. Control variables include net income scaled by total assets, deposits scaled by total assets, revenue per employee, salaries scaled by total assets and equity scaled by total assets. Both the left-hand side and the right-hand side variables are taken using the average values across 2010-2019 for each bank i. Fixed effects include bank size group and banks' headquarters state fixed effects. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	Software/Revenue	Communication/Revenue	Harware/Revenue	Services/Revenue
	(1)	(2)	(3)	(4)
C& I loans/Total loan	0.0380	0.0813***	0.107***	0.0386
	(0.0276)	(0.0270)	(0.0273)	(0.0279)
Net income/Total Assets	-0.126***	-0.155***	-0.183***	-0.0904***
	(0.0316)	(0.0308)	(0.0312)	(0.0318)
Deposits/Assets	-0.0928	-0.284*	-0.242	-0.103
	(0.175)	(0.171)	(0.173)	(0.177)
Revenue/Employee	-0.313***	-0.437***	-0.397***	-0.333***
	(0.0542)	(0.0529)	(0.0536)	(0.0546)
Salaries/Assets	0.0683	-0.147***	-0.0550	0.0104
	(0.0452)	(0.0441)	(0.0447)	(0.0455)
Equity/Assets	0.140**	0.0925*	0.0647	0.124**
	(0.0574)	(0.0560)	(0.0567)	(0.0578)
Size group FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
AdR-squared	0.0925	0.127	0.110	0.0649
N	1442	1442	1442	1442

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4. Bank Characteristics and Banks' IT Spending

This table presents the results of correlation between banks' IT spending and banks' characteristics. The regression specification is as follows.

$$\frac{\text{Type S IT spending}}{\text{Revenue}}_{i,10\text{-}19} = \alpha + \beta \frac{\text{Type L loan}}{\text{Total loan}}_{i,10\text{-}19} \text{ or (Bank Char)} + \gamma \mathbf{X} + \epsilon_i$$

Panel A shows how the banks' loan specialization correlates with banks' IT spending. Type L loan/Total Loan is the average of a specific type of loan scaled by total loan. Among them, Personal loan/Total Loan is the sum of personal loans and real estate loans to 1-4 family units scaled by total loan; Agriculture/Total loan is the agricultural loan scaled by total loan; CRA/Total loan is the sum of small business loans reported in CRA scaled by total loan; "Other C&I/Total loan" is the total C&I loan minus small business loans reported in CRA, scaled by total loan; "Mortgage refinance" is the total amount of mortgage refinance reported in HMDA scaled by the bank's total loan; "Other personal loans" is the deduction of "Mortgage refinance" from "Personal and mortgage loans." "Refinance/Origination" is the dollar amount of mortgage refinance scaled by dollar amount mortgage origination of a bank. Software/Rev is software spending scaled by total revenue, Communication/Rev is communication spending scaled by total revenue, Hardware/Rev is Hardware spending scaled by total revenue, Services/Rev is services spending scaled by total revenue. Panel B shows how a bank's hierarchical structure correlates with its IT spending. "Hierarchical layer" is the number of types of its locations as defined in Section 2.3. "ln(num offices)" is the logarithmic of total number of offices. Control variables include net income scaled by total assets, deposits scaled by total assets, revenue per employee, salaries scaled by total assets and equity scaled by total assets. Fixed effects include bank size group, and banks' headquarter state fixed effects. Panel C shows how a bank's role in the syndicated loan market correlates with its IT spending. %Lead bank is the frequency of a bank's showing up as a lead bank in the syndicated loan market as a share of total number of syndicated loans lent out. All of the loan profile variables are calculated as the average of the loan profile of a bank between 2010 and 2019. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	Software/Revenue	Communication/Revenue	Hardware/Revenue	Services/Revenue
	(1)	(2)	(3)	(4)
Panel A: Loan Specialization				
C& I loan/Total loan	0.0380	0.0813***	0.107***	0.0386
	(0.0276)	(0.0270)	(0.0273)	(0.0279)
CRA/Total loan	-0.190***	0.124***	0.0662**	0.0482
	(0.0307)	(0.0303)	(0.0309)	(0.0314)
Other C&I loan/Total loan	0.0645**	0.0653**	0.0995***	0.0330
	(0.0275)	(0.0269)	(0.0273)	(0.0278)
Personal loan	0.0617**	0.0507*	0.0162	-0.000905
	(0.0294)	(0.0287)	(0.0292)	(0.0297)
Refinance/Total loan	0.0763***	0.0369	0.0466	-0.000550
	(0.0311)	(0.0305)	(0.0309)	(0.0314)
Other personal loan/Total loan	0.0530*	0.0522*	0.0103	0.00167
	(0.0292)	(0.0285)	(0.0290)	(0.0295)
Refinance/Origination	0.0682**	0.0257	0.0482	0.0442
	(0.0329)	(0.0315)	(0.0322)	(0.0335)
Agriculture loan/Total loan	0.0238	0.0504	0.00356	0.0150
	(0.0337)	(0.0331)	(0.0336)	(0.0343)
Panel B: Hierarchical Complexity and IT Spending				
Hierarchical layer	0.0244	0.0721**	0.0331	0.0371
	(0.0354)	(0.0342)	(0.0351)	(0.0354)
ln(num of offices)	0.0519	0.0842**	0.0278	0.0487
	(0.0413)	(0.0394)	(0.0406)	(0.0413)
Panel C: Banks' Role in Syndicated Lending				
% Lead bank/Total syndicate	0.0751	0.0932**	0.0475	0.0162
	(0.0468)	(0.0453)	(0.0467)	(0.0473)

Table 5. Bank Characteristics and Banks' IT Spending: Size- and Hierarchical-Dependence

This table presents the results of the dependence of correlation between banks' IT spending with their lending activities on the size and hierarchical complexity of banks. The regression specification is as follows.

$$\frac{\text{Type S IT spending}}{\text{Revenue}}_{i,\text{10-19}} = \alpha + \beta \times (\text{Bank Char.}) \times \left(\frac{\text{CRA}}{\text{Total loan}}_{i,\text{10-19}} \text{or } \frac{\text{Refinance}}{\text{Total loan}}_{i,\text{10-19}} \right) + \gamma \mathbf{X} + \epsilon_i$$

In Panel A, small (large) banks are defined as the banks with asset size below (above) median asset size in our sample. In Panel B, "High layer" is defined equal to 1 if a bank has 2 or 3 hierarchical layers. the number of types of its locations as defined in Section 2.3. "Size group FE" refers to the fixed effects of the five bank asset groups defined in Section 3.1 or Table 1. Control variables include net income scaled by total assets, deposits scaled by total assets, revenue per employee, salaries scaled by total assets and equity scaled by total assets. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Bank Size and IT Spending

	Software/Revenue	Communication/Revenue
	(1)	(2)
Refinance/Total loan	0.0817*	
	(0.0428)	
Small×Refinance/Total loan	0.0315	
	(0.0538)	
CRA/Total loan		0.0282
		(0.0413)
Small×CRA/Total loan		0.162**
		(0.0699)
Small	-0.0310	-0.286***
	(0.0824)	(0.0894)
Size group FE	Y	Y
State group FE	Y	Y
R-squared	0.103	0.133
N	1432	1432

Panel B: Bank Hierarchical Structure and IT Spending

	Software/Revenue	Communication/Revenue
	(1)	(2)
Refinance/Total loan	0.0409	
	(0.0361)	
High layer×Refinance/Total loan	0.0836	
	(0.0516)	
CRA/Total loan		0.301**
		(0.121)
High layer×CRA/Total loan		0.0870*
		(0.0517)
High layer	0.0532	0.0242
	(0.0521)	(0.0506)
Size group FE	Y	Y
State group FE	Y	Y
R-squared	0.0894	0.127
N	1426	1426

Table 6. Soft Information and Banks' IT Spending

This table presents the results of 2SLS and OLS discussed in Section 4.2.2. The first three columns show the results for the following specification:

$$\Delta \ln(\text{CRA})_{i,c,post} = \tilde{\alpha}_i + \mu_1 \times \left(\frac{\#\text{ Qualified small business est}}{\text{Total }\#\text{ of establishments}}\right)_{c,pre} + \mu_2 \mathbf{X}_{i,c} + \epsilon_{i,c}$$

$$\Delta \ln(\text{IT})_{i,c,post} = \alpha_i + \beta \times \Delta \widehat{\ln(\text{CRA})}_{i,c,post} + \gamma \mathbf{X}_{i,c} + \epsilon_{i,c}$$

Column (4) and column (5) show the following OLS specification:

$$\Delta \ln(\text{IT})_{i,c,post} = \alpha_i + \beta \times \Delta \ln(\text{CRA})_{i,c,post} + \mu_c + \gamma \mathbf{X}_{i,c} + \epsilon_{i,c}$$

$\Delta\ln\,(\text{CRA})_{i,c,post}$ is the change in average natural log of small business loans reported in CRA of bank i at county c during the years 2014-2017 compared with 2011-2013. Bank control variables include pre-shock revenue per employee of the bank in a county. County level control variables include the pre-shock unemployment growth rate, labor force participation rate, population growth rate, logarithmic of total number of establishments, share of nontradable sector small business establishments, and GDP per capita. Fixed effects include bank fixed effects. Standard errors are clustered at county level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	First stage	ln(Software)	ln(Communication)	ln(Software)(OLS)	ln(Communication)(OLS)
	(1)	(2)	(3)	(4)	(5)
$\left(\frac{\text{Qualified small businesses establishments}}{\text{Total establishments}}\right)_{c,pre}$	1.032***				
	(0.251)				
$\Delta\ln\widehat{(CRA)}$		-0.057	0.670**		
		(0.305)	(0.328)		
$\Delta\ln(CRA)$				0.004	0.019*
				(0.010)	(0.011)
Bank FE	Y	Y	Y	Y	Y
Clustered	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y
F-stat	13.708				
AdR-squared	0.427	-0.179	-0.522	0.120	0.102
N	19,848	19,848	19,848	19,848	19,848

Table 7. Dependence of Banks' Shock Response on "Young Firm Share"

This table presents the impact of soft information demand and banks' IT spending for counties with differential share of younger small businesses. The 2SLS regression specifications are as follows:

$$\Delta \ln(\text{CRA})_{i,c,post} = \hat{\alpha}_i + \eta_1 \times \left(\frac{\#\text{ Qualified small business est}}{\text{Total }\#\text{ of establishments}}\right)_{c,pre} + \eta_2 \mathbf{X}_{i,c} + \epsilon_{i,c}$$

$$\Delta \ln(\text{CRA})_{i,c,post} \times \text{High young} = \tilde{\alpha}_i + \mu_1 \times \left(\frac{\#\text{ Qualified small business est}}{\text{Total }\#\text{ of establishments}}\right)_{c,pre} + \mu_2 \mathbf{X}_{i,c} + \epsilon_{i,c}$$

$$\Delta \ln(\text{IT})_{i,c,post} = \alpha_i + \beta \times \Delta \ln\widehat{(\text{CRA})}_{i,c,post} + \beta_1 \times \Delta \ln(\text{CRA})_{i,c,post} \widehat{\times \text{High young}} + \beta_2 \text{High young} + \gamma \mathbf{X}_{i,c} + \epsilon_{i,c}$$

Δln $(\text{CRA})_{i,c,post}$ is the change in average natural log of small business loans reported in CRA of bank i at county c during the years 2014-2017 compared with 2011-2013. Bank control variables include pre-shock revenue per employee. "High young" counties are defined as counties whose proportion of small businesses younger than 1 year old was above median among all counties in 2013. County level control variables include the pre-shock unemployment growth rate, labor force participation rate, population growth rate, log of total number of establishments, share of nontradable sector small business establishments, and GDP per capita. Fixed effects include bank fixed effects. Standard errors are clustered at county level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	First stage $\Delta\ln(\text{CRA})$	First stage $\Delta\ln(\text{CRA})\times$ High young	Second stage ln(Software)	Second stage ln(Communication)	OLS ln(Software)	OLS ln(Communication)
	(1)	(2)	(3)	(4)	(5)	(6)
$\%QSB_{pre}$	0.024*	0.028**				
	(0.012)	(0.014)				
$\%QSB_{pre}\times$ High young	-0.021**	0.025**				
	(0.010)	(0.010)				
$\Delta\ln\widehat{(\text{CRA})}$			-0.429	-0.321		
			(0.617)	(0.685)		
$\Delta\ln\widehat{(\text{CRA})}\times$ High young			0.687	1.534*		
			(0.844)	(0.928)		
High young			0.017	0.025	-0.042*	-0.056**
			(0.070)	(0.075)	(0.023)	(0.022)
$\Delta\ln(\text{CRA})$					0.003	0.017
					(0.014)	(0.016)
$\Delta\ln(\text{CRA})\times$ High young					0.004	0.019
					(0.021)	(0.021)
Bank FE	Y	Y	Y	Y	Y	Y
Clustered	Y	Y	Y	Y	Y	Y
F-stat	12.350	12.350				
AdR-squared	0.321	0.291	-0.330	-1.289	-0.179	-0.178
N	19,234	19,234	19,234	19,233	19,234	19,233

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8. Hard Information and Banks' IT Spending

This table presents the results of the regressions discussed in Section 4.3.2.
The first six columns show the results for the 2SLS specification below:

$$\ln(\text{Refinance/Origination})_{i,c} = \tilde{\alpha}_i + \mu_1 \times \Delta\text{Mortgage rate}_c(\text{or } \Delta\text{Payments}_c) + \mu_2\ \mathbf{X}_{i,c} + \epsilon_{i,c}$$

$$\ln(\text{Type S Spending})_{i,c} = \alpha_i + \beta \times \ln(\widehat{\text{Refinance/Origination}})_{i,c} + \gamma\ \mathbf{X}_{i,c} + \epsilon_{i,c}$$

Column (7) and (8) show the results of the OLS specification below:

$$\ln(\text{Type S Spending})_{i,c} = \alpha_i + \beta \times \ln(\text{Refinance})_{i,c} + \gamma \mathbf{X}_{i,c} + \epsilon_{i,c}$$

$\ln(\text{Type S Spending})_{i,c}$ is the average logarithmic of banks' IT spending during 2011 and 2016. $ln(\text{Refinance/Origination})_{i,c}$ is the average logaithmic of amount of mortgage refinance loan relative to mortgage origination issued by bank i in county c during 2011 and 2016. $\Delta\text{Payments}_c$ is the hypothetical amount of interest payments that could be saved due to the interest rate decrease, if local households chose to refinance their mortgages during the year of 2011 and 2016. $\Delta\text{Mortgage rate}_c$ is the average differences of mortgage rate of unmatured loans in 2011-2016 and the prevailing mortgage rates of newly issued mortgages in a county c. Bank control variables include banks' revenue per employee of the bank in a county. County level control variables include unemployment growth rate, labor force participation rate, population growth rate, logarithmic of total number of establishments, logarithmic of total small business loan, share of nontradable sector establishments, and GDP per capita. Fixed effects include bank fixed effects. Standard errors are clustered at county level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	2SLS			2SLS			OLS	
	First stage	ln(Software)	ln(Comm)	First stage	ln(Software)	ln(Comm)	ln(Software)	ln(Commu)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔMortgage rate$_c$	1.824***							
	(0.622)							
ΔPayment$_c$				0.819***				
				(0.237)				
$\ln \widehat{\text{Refinance/Origination}}$		0.315*	0.239		0.373*	0.296		
		(0.167)	(0.150)		(0.225)	(0.211)		
ln(Refinance/Origination)							0.024***	0.025***
							(0.006)	(0.006)
Bank FE	Y	Y	Y	Y	Y	Y	Y	Y
Clustered	Y	Y	Y	Y	Y	Y	Y	Y
F-stat	10.81			13.82				
AdR-squared	0.356	-0.349	0.072	0.423	0.447	0.576	0.449	0.423
N	14,626	14,626	14,626	14,626	14,626	14,626	14,626	14,626

Table 9. Staggered Entry of Lending Club to 9 States after 2010

State	Approval year
All states, except the states listed below	2008
Kansas	2010 Q4
North Carolina	2010 Q4
Indiana	2012 Q4
Tennessee	2013 Q1
Mississippi	2014 Q2
Nebraska	2015 Q2
North Dakota	2015 Q2
Maine	2015 Q3
Idaho	2016 Q1
Iowa	Not approved as of 2022-Q1

Table 10. Fintech Entry and Banks' Lending Technology Adoption

This table presents the effect of Lending Club's entrance on local banks' IT spending. The regression equation is as follows

$$\ln(\text{ITSpending})_{i,c,t} \quad = \quad \alpha_{i,c} + \alpha_t + \beta \times \text{LC}_{i,c,t} + \gamma_t \mathbf{X}_{i,t} + \epsilon_{i,c,t},$$

where $\alpha_{i,c}$ and α_t are the bank-county and year FE, respectively. Column (1) and (2) of Panel A show the baseline results. Bank control variables include banks' revenue per employee of the bank in a county. County level control variables include unemployment growth rate, labor force participation rate, population growth rate, logarithmic of total number of establishments, share of nontradable sector establishments, and GDP per capita. Standard errors are based on 50 Bootstrapped samples. Panel B presents the differential responses to Fintech entrance of banks with different sizes. "Large banks" are defined as banks with asset size above median of all the asset sizes in the sample. Panel C presents the differential responses to Fintech entrance of banks with different personal loan share. "High personal loan" banks are defined as banks for which the personal loan as a share of total loan is above median among all banks in the sample. The estimations in Panel column 3 and column 4 of the three panels are based on the interacted TWFE method as in Callaway and Sant'Anna (2021). Standard errors are in the parentheses and are clustered at the county level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A	Baseline		Callaway and Sant'Anna (2021)	
	ln(Software)	ln(Communication)	ln(Software)	ln(Communication)
	(1)	(2)	(3)	(4)
After	0.076***	0.001	0.080**	0.007
	(0.023)	(0.018)	(0.042)	(0.040)
Fixed Effects	Bank×County, Year, Size group			
AdR-squared	0.808	0.790		
N	13,406	13,406		
Panel B	Baseline		Callaway and Sant'Anna (2021)	
	ln(Software)	ln(Communication)	ln(Software)	ln(Communication)
	(1)	(2)	(3)	(4)
After	0.051	0.037	0.043	0.090
	(0.032)	(0.030)	(0.048)	(0.052)
After× Large	0.062*	-0.058*	0.097***	-0.167***
	(0.032)	(0.035)	(0.040)	(0.061)
Fixed Effects	Bank×County, Year, Size group			
Clustered	Y	Y		
AdR-squared	0.777	0.96		
N	13,406	13,406		
Panel C	Baseline		Callaway and Sant'Anna (2021)	
	ln(Software)	ln(Communication)	ln(Software)	ln(Communication)
	(1)	(2)	(3)	(4)
After	0.054**	0.007	0.058	0.050
	(0.026)	(0.023)	(0.050)	(0.052)
After×High personal loan	0.055**	-0.002	0.078**	-0.041
	(0.025)	(0.026)	(0.043)	(0.054)
Fixed Effects	Bank×County, Year, Size group			
Clustered	Y	Y		
AdR-squared	0.837	0.774		
N	13,406	13,406		