Investing in Lending Technology: IT Spending in Banking *

Zhiguo He, Sheila Jiang, Douglas Xu, Xiao Yin

August 20, 2022

Abstract

This paper studies the economics behind the investment in information technologies (IT) by U.S. commercial banks in the past decade. By linking banks' IT spending to their lending technologies, we analyze the distinctive natures of banks' dealings with information across various lending activities. Investment in communication IT is shown to be associated more with improving banks' ability of soft information production and transmission, while investment in software IT helps enhance banks' hard information processing capacity. We exploit polices 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: PhD Candidate in Finance, Haas School of Business, University of California, Berkeley, yinxiao@berkeley.edu. For helpful comments, we thank Douglas Diamond, Raghuram Rajan, Isil Erel (discussant), Mark Flannery, and Yueran Ma. We are thankful for valuable comments from seminar and conference participants at the Richmond Fed, University of Connecticut, University of Florida, University of Chicago, EFA and SFS Calvacade. 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. This trend has accelerated at an unprecedented pace in recent years, especially after the COVID-19 pandemic, as industry professionals often consider top commercial banks to be more like "technology companies" than actual technology firms by virtue of their enormous IT budgets.[1] Recently, the impact of information technology on the banking sector and 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 business, 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 differ in their IT investment profile given their specialization in different types of loan markets? How do traditional banks react to the entry of fintech in recent years? We take the first step toward understanding the key empirical pattern of these IT expenditures by banks, and further explore the economic mechanism that underlies the connections between these expenditures and the core functioning of the banking system.

To place our research in the established banking 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; and the same idea of "soft"

[1]For instance, this article shows that IT spending by most top banks (say JP Morgan and Goldman Sachs), exceeds 17% of their total operating costs, while Amazon and Alphabet devote 12% and 20% of their operating costs respectively to IT. This article also cautions that the above-mentioned IT spending numbers do not include compensation for IT staff members.

information—which plays an important role in small business lending—may help researchers understand "the consequences of consolidation in the banking industry, particularly the documented tendency for mergers to lead to declines in small business lending."[2] But, the fast developing technologies in recent decades provide more options for the banking sector to cope with such problems, and banks have paid more attention to strengthening their internal communication.[3] So, can information technologies reduce frictions in communicating soft information and potentially improve banks' credit approval decisions? And, have traditional banks started adopting the explosive big data analytics technology—which combines "hard" information like credit scores and other alternative data—in their lending activities?

Our study relies on a comprehensive dataset, the Harte Hanks Market Intelligence Computer Intelligence Technology database, which has been used in the literature for studying the economic implications of technology adoption in the non-financial sector (e.g., Bloom et al., 2014; Forman et al., 2012). This dataset provides detailed branch-level information on specific spending categories, and our paper provides the first set of comprehensive analysis of this dataset in the context of the banking sector.

Our study focuses on two of four major categories in the database, with the first being *Software*.[4] This type of IT product mainly aims to improve information processing accuracy and speed through automation, specialized programming and AI technologies, etc. The second category is *communication*, which facilitates smoother exchanges of information within bank branching networks, as well as across banks and their borrower customers.[5]

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

[2]See Stein (2002). This downward trend in small business lending is also documented in the more recent literature, e.g., Chen et al. (2017a).

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

[4]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.

[5]Examples include radio and TV transmitters, private branch exchanges, and video conferencing, etc.

size. Larger banks' IT spending has been growing more steadily, while there was almost no growth in IT spending for the smallest banks (assets below $0.1 billion). A noticeable distinction between large and small banks is that smaller banks, who presumably engage in more small business lending, consistently allocate a higher share of their IT budget towards communication technology than larger banks do. As we will elaborate, this pattern points to the role communication IT plays in conducting small business lending.

To better understand the economics of IT investment in the banking industry, in Section 3.3 we examine the relationship between banks' IT spending and their lending activities. Among the three main categories of loan types in Call Report, the shares in commercial and industrial (C&I) loans and agricultural loans are 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 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 the personal loans, mortgage refinance is the main contributor to the positive correlation between personal loans and software spending. Given that different types of loans often require different lending technologies to deal with the relevant information, these positive associations (or the lack thereof) thus offer important guidance on how to understand banks' IT spending profiles from the perspective of the lending technology adopted by lenders.

Aside from broad credit categories of loan portfolios extended by lenders, we also explore how banks' IT investment is shaped by other factors affecting their business operations. By examining how the complexity of banks' internal hierarchical structure affects banks' IT investment, we find that banks with more internal layers in their hierarchical structure tend to have a higher communication spending intensity. 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 changes in their intensity of small business lending, but displays no systematic effect on how banks' software spending reacts to their mortgage refinancing activities.[6] Finally, in the context of the syndicated lending market, we show that frequent lead lenders spend significantly more on communication than participant lenders, consistent with the notion that 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 differentiate two fundamentally different types of bank 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, on the other hand, relies primarily on the processing and quantification of hard information; "transactions lending" in Berger and Udell (2006), i.e., loans that are based on a specific credit scoring system and quantified financial statement metrics, are standard examples of the second type.

We formulate our first hypothesis along the dimension of soft information. More specifically, 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. This is because communication technologies—say video conferencing—not only enable banks to more effectively collect soft information from small business borrowers (who often inhabit an opaque information environment), but also allow for a smoother transmission of this otherwise hard-to-verify soft information within a bank organization. Taking advantage of an arguably exogenous demand shifter, we find that 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.[7]

[6] This asymmetric pattern is consistent with the notion of "hierarchical friction" (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.

[7] We construct our instrumental variable for the shock of small business credit demand based on "Small

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 on software. This is because for loans that primarily involve dealing with existing or readily accessible "hard" information, software—a type of IT investment that facilitates analyzing such existing information—should be the more relevant category of information technologies to invest in. For causal identification, we utilize the cross-county variation in the interest payment savings of outstanding mortgages to construct a shifter to the mortgage refinance demand faced by banks across different regions.[8] Thanks to this instrumental variable, we show that a one standard deviation higher in mortgage refinancing lent out by a bank (due to its local exposure to high refinance savings) leads to around 5% higher software spending intensity relative to the sample average, without any significant impact on local banks' communication spending.

The last part of our analysis concerns how the entry of fintech lenders into local credit markets affects banks' IT spending and their associated lending technologies. In the past decade, the growing penetration of fintech—as a special group of tech-intensive (potential) competitors to the traditional banking sector—has drawn much attention in the financial industry. Yet, direct evidences on how the traditional banking sector has been reacting to the penetrating fintechs are lacking in the academic literature. Using the staggered entry of Lending Club into seven states after 2010 as an experimental setting, we find that right after the regulatory approval of Lending Club's operation in a state, banks operating in counties located in that state saw a significant increase in their IT investment. Crucially, the growth in software spending (5.61%) is quantitatively (as well as statistically) more significant compared with the growth in communication spending (1.9%).

Business Health Care Tax Credit." As a part of the Affordable Care Act, this program was enacted between 2014 and 2015 and provided beneficial tax treatment specifically targeted at small business establishments in the US economy.

[8] We take advantage of the low-interest episode from 2011 to 2015, when nationwide average mortgage interest rate 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.

Aside from the asymmetry in the IT category of banks' response to fintech entry, we also find significant heterogeneity across bank size groups in their technology spending reactions. In particular, the increased IT investment intensity is predominantly observed among the large banks, whereas small banks barely respond in their IT spending towards the new entry of fintech players. Our findings suggest an overall "competition reaction" from the traditional banking sector in that banks—particularly bigger banks—tend to catch up with the newly entered fintech lenders. 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—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.[9] A fundamental difference between these two types of lending is related to the role played by information as highlighted by Stein (2002), who provides an explanation as to why soft information production favors an organizational structure with fewer hierarchical layers.[10]

We contribute to this strand of the 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 demonstrate causal linkages from the informational components in credit demand shocks to banks' lending technologies, via their endogenous decisions on IT spending. It is, to our knowledge, the first

[9] Relatedly, Bolton et al. (2016) study the joint determination of relationship lending and transactions lending. They find that firms that rely more on relationship banking are better able to weather a crisis than firms that rely on transactions banking, suggesting a higher capital requirement for relationship banks.

[10] 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.

attempt in the literature to show how credit demand shocks drive banks' investment in their information-driven lending technologies.[11]

Information technology and 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 technological progress in both information technology and financial technology led to significant improvement in banking services, and Petersen and Rajan (2002) document that development in communication technology greatly increased the lending distance of small business loans. Our paper, with the aid of detailed IT spending data in the banking sector, fills in the details of specific economic mechanisms that connect banks' lending technology with their IT spending.[12]

Fintech entry and bank's IT spending. The emergence of fintech is a signature result of recent developments in information technologies.[13] Our study aligns closely with studies on how the emergence of fintech industry is affecting (or has affected) the traditional banking sector.[14] While a common theme of this research has mostly focused on examining how the rising fintech industry is affecting bank-fintech competition, a process in which traditional banks are largely viewed as a *passive* player, little attention has been paid to how banks are *actively* responding to these challengers. Our paper makes the initial step in studying whether and how the traditional banking sector is catching up with penetrating fintech lenders through examining IT investment behavior in the U.S. banking sector.

Micro-level Evidence on Firm Technology Adoption. Our paper also broadly contributes to the literature studying firms' technology adoption behavior using new micro-level

[11]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)).

[12]There is also a vast theoretical literature on the interactions among information technology, banking market competition and bank lending; see Freixas and Rochet (2008) for a review. For instance, Hauswald and Marquez (2003, 2006) analyze the implications of information technology on banking market competition; and more recently, Vives and Ye (2021) study how the diffusion of information technology affects competition in the bank lending market and banking sector stability.

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

[14]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), Schnabl and Gopal (2020), and He et al. (2022).

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, and Bloom et al. (2014) investigate the effect of information technology on firms' internal control. The recent literature starts using the IT installment data;[15] for instance, Charoenwong et al. (2022) study compliance-driven investment in technology, and Pierri and Timmer (2022) show that technology adoption helps banks weather financial crisis.

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 via conducting IT-related consulting for firms.[16] Harte Hanks collects and sells this information to technology companies, who then use it for marketing purposes or to better serve their clients. Firms with IT spending information have incentives to report truthfully to Harte Hanks as they also want to receive tailored advice for better IT services in the future. This dataset has been used in the literature; Forman et al. (2012) investigate firms' IT adoption and regional wage inequality, and Bloom et al. (2014) study the impact of information communication technology on firms' internal control.

Our paper focuses on commercial banks. The sample consists of 1806 commercial banks

[15] This dataset reports firm-level software product installment categorized by type; this is different from the one we use here, which is on the establishment level IT budget by categories. Both datasets are maintained by Harte Hanks.

[16] One important measurement issue is the method of allocating IT costs to branches when the headquarter makes the purchase. According to the data provider, such spending is reflected in the branch's spending rather than in that of the headquarters. This claim is indeed supported by our data: when we compare the headquarters and branch spending scaled by revenue of the largest 100 banks in our sample, we find no statistically significant difference between the IT spending intensity at the headquarters and branches.

in the U.S., which covers more than 80% of the U.S. banking sector in terms of asset size (Figure A1). Our sample is more representative for large banks, as shown in Table 1 which reports the coverage of our sample by bank asset size group. For the three groups of relatively large banks (with asset size above $1 billion), the coverage in frequency and asset are both over 80%. However, for small banks with size less than $100 million, our sample covers only 14.45% (14.23%) 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, bank's total IT spending as a share of its net income ranges from 1.8% (25th percentile) to 13.5% (75th percentile), suggesting a large cross-sectional variation across banks. Median IT spending as a share of net income is 7.1%, consistent with the 2012 survey by McKinsey reporting that banks' IT spending as a share of net operating income ranges from 4.7% to 9.4%.[17]

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 6(a) to 6(d) of Figure A5.

Software is defined as software purchased from third parties, including those offered on 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, risk and payment management software. One representative example of a desktop application is the Microsoft Office software package.[18] Processing software specializes in automatically processing information from loan applicants' paper document packets through specialized programming and AI technologies, which would otherwise be done manu-

[17] If a bank has a negative income that year, we treat its IT spending share as an arbitrary large number. And a screenshot of the McKinsey report is in Appendix Figure A4.

[18] These software products are easy to grasp by bank employees who are then able to conduct basic calculations and visualization 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.

ally by loan officers, improving processing accuracy and speed.[19] Risk management software provides on-going risk assessment after loans have been issued, through augmenting borrowers' repayment information 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. When there is a need for bankers to contact or interact directly with borrowers, a set of advanced communication equipment allows bankers to conveniently talk to and see borrowers, which helps banks' effective evaluation of projects that borrowers seek for finance. In addition, communication equipment such as private branch exchanges facilitates a more effective exchange of information, opinions, and decisions within the bank branching 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 is a fundamental type of tech investment that complements and facilitates 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, and meanwhile they serve as the carriers of software and toolboxes.

Services are defined as project-based consulting services or systems integration services that vendors provide to banks. Specifically, these include consulting services for IT strategy, security assessments, system integration, project services, hardware support and maintenance services. The services are mainly provided by IT outsourcing companies on a contractual basis. Similar to hardware, services work as complements to other categories of

[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, 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.

information technology investment to facilitate banks' lending, although these services are not directly associated with banks' information gathering or processing. 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.

As a summary of this section, Table 2 reports summary statistics on the detailed structure of banks' IT spending profiles. By size, software and services are the largest among all categories of IT spending, each constituting 33% of total IT budget. Hardware constitutes about 17% of total IT budget, and communication is on average 9%.

2.3 Other Datasets

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

Bank Balance Sheet We obtain bank-level balance sheet information from Call Report.[21] Bank-year level control variables constructed from Call Report include Net income, Equity, and Deposits, all as a ratio of Assets. For bank-county or bank-county-year analysis, we utilize information on banks' revenues at the county level from Harte Hanks.[22] Our data cleaning procedure further requires that, at the county-year level, a bank must have non-missing total revenue and total number of employees. To construct the key left-hand side variables "IT spending/Revenue," we aggregate all branches' spending in a specific category of bank i in county c in year t, and scale it by the sum of revenues of all branches of that bank in that county. The control variable "Revenue per employee" is at bank-county-year level, with total revenue and total number of employees both from Harte Hanks. When using

[21]To merge Call Reports with the Harte Hanks data, we merge by bank name using the Levenshtein distance (which assigns a matching score between two string variables based on the minimum number of edits needed to match two strings) after dropping suffixes such as "inc." or "corporation."

[22]Since branch-level revenue information is not available in Call Report, we use branch-level revenue information supplied by Harte Hanks for our bank-county or bank-county-year analyses. For figures plotting aggregate spending trends, we use Call Report information, which is at bank level.

this control variable for bank-level analysis, we aggregate revenue and employees at the bank level across the nation and calculate this ratio.

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. When constructing an instrumental variable that serves as a demand shifter, we use the county-level average mortgage interest rate before 2010 obtained from Freddie Mac.

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.[23] Furthermore, the database provides the complete family trees of the companies (in our studies, bank holding companies), with detailed information on its family members. As we focus on commercial banks, we restrict our sample to entities with the two-digit SIC code of "60," which designates "Depository Institutions."

Importantly, this database classifies each family member of a company into one of the three categories of location types: "Headquarter," "Single Location," and "Branch." We define a bank as having three- (two-) layers of hierarchical structure if the bank has three (two) type of locations in the family tree, and accordingly call the bank a single-layer bank if it has only one type of location. To give some concrete examples, Wells Fargo & Company has more than six thousands offices in U.S., and the location types include all the three categories, so we define its hierarchical layer as 3; North Valley Bank with headquarter located in Corning (OH) is classified as two layers, as it has one headquarter and seven branches; and First Place Bank located in Warren (OH) has only one layer in our sample and the only office it has is a single location.[24] For each bank, we match the banks in

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

[24] In the Mergent Intellect database, if a bank has two layers in its hierarchy, these two types of locations

Mergent Intellect with banks in our sample based on bank names and the city where the banks' headquarters are located.

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 in the U.S. economy and explore how a bank's IT spending profile varies with its size. We further show that banks' IT investment is shaped by their lending activities, by demonstrating several robust correlation patterns between the profiles of banks' IT spending and their loan specializations (e.g., commercial loans versus personal loans). Finally, we explore other dimensions that can relate banks' IT spending to their operations, including the complexity of a bank's internal hierarchical structure and the role a bank normally plays in syndicated lending.

3.1 Time Trends of Banks' IT Investment

Panel A of Figure 1 displays the average IT spending as a share of total expenses as well as total revenue from 2010 to 2019. Overall, banks have drastically increased their investment in information technologies over the last decade. As a share of total expenses, their IT budgets climbed up from nearly zero in 2010 to about 5% after 2015. To put these numbers in context, total IT spending across all banks in our sample is about 40% of their total interest expenses in 2016.

After a slight slowdown in 2015, there was a dramatic pick-up of IT spending in 2016 as shown in Figure 1 Panel A. This could be potentially driven by the release of a "white paper" by the Office of the Comptroller of the Currency on March 16, 2016, which set forth the regulators' perspective on supporting responsible innovation across all-sized banks in the banking system.[25] As mentioned in this article, this white paper might have pushed banks

are "Headquarter" and "Branch;" while if a bank has only one layer in its hierarchy, it is "Single Location."

[25]The Office of the Comptroller of the Currency defines "responsible innovation" as "the use of new or

to be more aggressive in embracing technology investment into their strategic planning.

What is more, the white paper also encouraged banks to collaborate with non-banks in developing responsible financial products that satisfy regulator requirements, and it is widely believed that they have been more actively investing in IT in order to better catch up with their fintech peers.[26] Figure 1 Panel B, which plots IT spending over years as a share of revenue by local banks for regions with high and low fintech presence,[27] offers evidence suggesting a potential "catching-up" behavior of the traditional banking sector. While IT spending in both groups share a common upward trend, the high-fintech-presence group increases their IT spending at a faster rate than the one with low fintech presence. This interesting empirical pattern motivates us to conduct a rigorous analysis on this topic in Section 5.

3.2 Bank IT Spending across Bank Size

Banks of different size often behave differently in systematic ways. We now conduct the same set of analyses as in the last section, but for different bank size groups. Following FDIC bank size classification, we break banks into five size groups. Overall, large banks make more IT investment than their small peers do. As can be seen in Table A2, banks with total assets of less than $0.1 billion have an average IT over revenue ratio of 1.5%, and this ratio monotonically increases to 4.5% for banks with $10–$250 billion asset size. This pattern could be due to the fixed cost nature of IT spending, as small banks often cannot afford IT purchases that require significant lump-sum payments.[28]

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." For more details about the white paper, see here.

[26]This article by McKinsey documents a fintech IPO boom as well as a fintech investment boom by venture capitalists since 2016.

[27]County-level fintech presence measure is based on "fintech lending share in local mortgage market" proposed in Fuster et al. (2019), and we define high (low) fintech presence regions as counties whose fintech lending share is above-median (below-median).

[28]The ratio of IT investment as a share of revenue drops to 1.9% for banks with asset size above $250 billion, presumably due to the economy of scale of IT investment for these "mega" banks.

Panel A of Figure 2 displays the time trend of banks' IT investments, as a fraction of non-interest expense, by each bank size group. Overall, the past decade witnesses an upward trend in IT spending as a share of non-interest expense for all bank size groups.[29] Despite this common upward trend, there are also some noticeable differences across different bank size groups. IT spending in large banks (with asset size $10–250 billion) has been steadily growing, while there is almost no growth in the smallest group (with asset size below $0.1 billion).[30] In contrast, mega banks (with asset size above $250 billion) only picked up their IT spending after 2015. While we do not aim to provide a conclusive answer for why such heterogeneity in the time dynamics of IT spending exists across different bank size groups, our analysis on how banks (of different sizes) react to the entry of fintech in Section 5 touches on this issue directly.

Another noticeable feature revealed by Panel B in Figure 2 is that smaller banks tend to allocate a higher fraction of their IT budget towards communication technology than larger banks do: the average communication over total spending ratio is 15.9% for banks with assets less than $0.1 billion while it falls below 5% for mega banks, and this ratio monotonically decreases with bank size. For software spending, however, there are no significant differences across bank size groups. We will come back to this sharp contrast in Section 4, where we establish linkages between banks' IT spending categories and their lending activities that involve different natures of information handling. A full comparison of the spending on communication and software (as a share of total IT spending) across different bank size groups is shown in Table A2.

[29]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 in Figure A2, shares consistent pattern with IT spending as a share of non-interest expense.

[30]Medium-sized banks (asset size bins $0.1–1 billion and $1–10 billion) saw the fastest growth in their IT spending during 2010–2014 but dramatically slowed down during 2015–2019. One possible explanation for the temporary slowdown in IT spending in 2015 might be that banks chose to pause or "wait and see" in 2015 before the release of the white paper in 2016 (see second paragraph in Section 3.1).

3.3 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 specific angles: 1) relative specialization in loan making; 2) the role that a bank plays in syndicated loans; and 3) complexity of bank's internal hierarchical structure.

3.3.1 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-type specific information. As a consequence, if banks specialize in different types of loan making, one should expect them to differ in their IT investment profiles.

In our benchmark empirical specification, we run the following bank-level regression (we leave the more granular bank-county level analysis for later):

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

Here, i refers to an individual bank and the outcome variable of interest is $\frac{\text{Type S IT spending}}{\text{Revenue}}_{i,10\text{-}19}$, which is the average investment in a specific type of IT spending as a share of bank i's revenue during the period of 2010-2019. The main explanatory variable $\frac{\text{Type L loan}}{\text{Total loan}}_{i,10\text{-}19}$ captures bank i's loan specialization; it is measured by the average share of a specific type of loan relative to bank i's total loan size. Control variables, which are measured over the decade of 2010-2019 at the bank level, include net income, total deposits, total equity, total salaries (all scaled by total assets), and revenue per employee.

Table 3 reports the estimation results of the regression (1) for C&I loans with the detailed regression outcome including control variables and fixed effects, while for exposition purposes Table 4 only reports the key regression coefficients (i.e., those of specific IT spending shares) for C&I loans, personal loans, and agricultural loans using the same methodology.

A. Commercial and Industrial (C&I) Loans

Table 3 shows that specialization in C&I loans is most positively associated with banks' spending in communication technology (column 2). A one standard deviation (8%) increase in loan portfolio share allocated to C&I loans predicts a $0.13 million higher spending on communication per year.[31] A higher degree of specialization in C&I loans also predicts more spending on hardware (column 3), although the magnitude is slightly smaller than that of the impact on communication budget. The coefficient of software spending, however, is insignificant (column 1).

Within C&I Loans Rows 2 and 3 of Table 4 further decompose C&I loans into "Small Business Loans" (measured by a bank's small business lending reported in CRA) and "other C&I loans." While the share of small business loans in a bank's loan 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. Although we delay more detailed discussions to Section 4.2.1, Table 5 shows that our results are robust to bank sizes.

B. Personal Loans

The second major category of loan type we examine is personal loans and mortgages, which is defined as the sum of personal loans in Call Report. Row 5 of Table 4 reports the associations between shares in personal loans and mortgages and banks' IT spending. Contrary to the pattern we observe for C&I loans, a higher share of loan portfolio allocated to personal loans and mortgages appears to predict more spending on software only. Quantitatively, a one standard deviation increase in personal loans and mortgages share (about an increase of 7 percentage points) predicts an increase of $0.65 million on software spending

[31] As is shown in Table 2, the standard deviation of communication/revenue is 0.0075, and the average revenue is $344 million. Given a coefficient of 0.05, the implied increase of communication spending is $0.0075 \times 0.05 \times$ $344 million=$0.13 million. Our economic magnitude calculation follows this way throughout our paper.

per year. On the other hand, a higher personal loans and mortgages share does not have qualitatively significant predictive power on communication, hardware, or services budgets.

Within Personal Loans Paralleling our analysis of CRA loans within C&I loans, we also decompose personal loans and mortgages into two subcategories: mortgage refinancing and everything else. We find that it is mortgage refinancing—but not others—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. The main results are reported in Row 8 of Table 4 and the bank-size dependent results reported in Table 5, and we postpone more detailed discussion to Section 4.3.

C. Agricultural Loans

In the last category of loan types, we examine the association between agricultural loan specialization and banks' IT spending profiles. As shown in Row 9 of Table 4, a higher agricultural loan share positively correlates with the bank's communication spending. A one standard deviation increase in the share of agricultural loans (4.8 percentage points higher) is associated with about $0.11 million more communication spending per year.

3.3.2 Complexity of Bank's Hierarchical Structure

Another important factor that may affect a bank's efficacy in information handling is the internal organization structure of a bank (Stein, 2002). In the first row of Panel B in Table 4, we use the hierarchical layer defined in Section 2.3 as our main measure of banks' hierarchical complexity. We find that when the number of banks' hierarchical layers increases, banks increase all types of their IT spending—and especially their communication spending. A one standard deviation increase in the hierarchical layers predicts about $0.22 million more communication spending each year. Notice that this result is under the specification with bank size group fixed effects included, implying that hierarchical complexity predicts higher communication spending beyond bank size.[32] As a robustness check, we proxy banks' hierarchical complexity using the logarithmic of the total number of its offices, and qualitatively similar results are obtained in the second row of Panel B.

Our results suggest that when the hierarchical complexity within a bank organization increases, the bank will need to spend more on communication. As will be explained in Section 4.2.1, one can relate these findings to the analysis in Stein (2002) regarding the within-organization transmission of information that is difficult to be verified and relayed. 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. We will return to this issue later in Section 4.2.1.

3.3.3 Bank's 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 in terms of dealing with information. A natural empirical test then is to examine whether there exists a systematic difference in IT investment between lead and participant lenders.

In Panel C of Table 4, we present the same regression as in Eq. (1), except we replace the key right-hand side variable with "%Lead bank/Total syndicate," which is 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 syndicated loan market, with communication spending having the

[32] Recall that in Section 3.2 we show that smaller banks actually tend to allocate a larger portion of their IT budget on communication spending. Our findings here therefore suggest that despite its high correlation with bank size, the complexity of banks internal hierarchical structure has an additional impact on banks' IT spending on top of the bank size effect. Put it differently, one cannot simply use the size of a bank as the empirical proxy for its hierarchical complexity.

largest magnitude. Quantitatively, a one standard deviation increase in the lead bank frequency is associated with \$3.34 million higher in the bank's communication budget per year, while \$8.06 million lower in its software budget. These findings, as we will elaborate in Section 4.2, can be attributed to the distinctive natures of information-handling responsibility assumed respectively by lead and participant banks.

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 the central question of our analysis: What are the economics behind these banks' IT spending, and in particular, how can they be related—and contribute—to the development of banks' lending technology? We start with a conceptual discussion on how we distinguish banks' lending technologies based on the nature of information handling. This framework maps different types of IT investment onto different dimensions of banks' lending technology, which helps us understand various empirical patterns we established in Section 3. Finally, we establish the causal impact on banks' (endogenous) lending technology adoption behaviors of two credit demand shocks that involve different kinds of information nature.

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, in conducting their lending businesses, banks engage in two types/stages of activities regarding their borrowers' information: information *production/transmission* and information *processing*. More specifically, information *production/transmission*, which is broadly related to *soft* information in Stein (2002), refers to the stage in which information on borrowers needs to be created or gathered and then relayed to the hands of those who later make decisions based on this information.

On the other hand, information *processing*, which is broadly related to *hard* information in Stein (2002), is more about the stage in which lenders assemble and examine existing (or readily available) 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 structure is relatively opaque, bankers often need to communicate with their borrowers to gather the relevant information—either through face-to-face meeting, or from seeing borrowers' projects by themselves. Once these first-hand information about borrowers has been gathered, which could be subjective and thus difficult to convey to others, effective transmission of the gathered information within the organization can also crucially 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 banks to interact with customers during the past decade. In the past, banks opened new checking accounts, originated loans, and resolved problems through in-person visits to the brick-and-mortar branches, but they now use video conferencing as it makes the direct—yet virtual—contact between loan officers and borrowers more efficient.[33] Moreover, video conferencing within a financial institution has also been welcomed by the banking sector for its advantage in facilitating effective internal communication/collaboration among employees.[34]

Software IT and Hard Information Processing Once information has been produced (by the lender itself) or is readily accessible (via the third party), the next concern for the lender is how to most efficiently utilize this information to make wise decisions. In the context of credit allocation, banks need to properly evaluate the creditworthiness of their borrowers to determine loan amounts and rates. More specifically, when banks are facing borrowers either whose information structure is relatively transparent or who they already know from previous interactions, lending decisions simply boil down to efficient utilization

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

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

and processing of the existing information.

Accurate evaluations of borrowers' credit risk often require complicated modeling and simulations, which are impossible without the support of sophisticated software tools. From the banks' perspective, software differs from communication technology in that communication devices can facilitate the gathering and dissemination of information, whereas software is more targeted at efficiently utilizing "hard" information that is readily available at hand. Nowadays banks have actively adapted new software-based technologies to store, organize, and analyze large chunks of loan applicants' data, or data augmented by other software.[35] One popular form of software technology product is the credit scoring software utilized by banks when making *refinancing* decisions,[36] which primarily involves the processing and assessment of *existing* information that lenders already possesses through past interaction.

In what follows, we explore in detail the lending technology adoption behaviors in the banking sector along these two dimensions—those targeting the production and transmission of soft information (in Section 4.2), and those targeting hard information processing (in Section 4.3). Accordingly, from this point on we will focus on two particular categories—*communication* and *software*—in our examination of banks' IT investment behavior.[37]

4.2 Bank IT Spending and Soft Information

4.2.1 Soft Information Production/Transmission in Bank Lending

In this section, by reviewing certain empirical results on banks' IT investment profile established in earlier sections, we discuss situations where banks' capacity for dealing with

[35] 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.

[36] 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 expedite the processing.

[37] 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 3.3.

soft information is crucial for their profit making in lending.

Small Business Lending The 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 from loan officers to borrowers' 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.[38]

That small business lending involves intensive soft information production and transmission is consistent with our empirical finding in Section 3.3.1 that banks specialized in small business lending (as measured by the ratio of small business loans to total loans) incur more spending on communication IT. As in general smaller banks overall extend more loans to small businesses than larger banks do (Berger and Udell, 2006; Chen et al., 2017b), this helps explain the observation that smaller banks have a higher fraction of communication IT spending shown in Panel B of Figure 2; but the positive relationship between small business lending and communication spending is robust for bank-size subgroups with above- and below-median asset cutoff, as shown in Table 5 Panel A.

Hierarchical Complexity Next, in the context of soft information handling, we revisit our earlier findings in Section 3.3.2 along the dimension of banks' hierarchical complexity. There, we find banks with more complex hierarchical structure tend to have a higher intensity in their IT spending and especially so in the communication categories. These findings are in line with Stein (2002), who argues that a lower level of hierarchical complexity greatly facilitates the within-organization transmission of soft information and thus encourages the institution to engage more in projects requiring soft information generation (e.g., small

[38] As will be analyzed in the next part, the nature of interactions between lenders and borrowers also depends on whether the lender is a lead bank or a participant bank in the context of syndicated lending: being a frequent lead bank requires frequent communication, reporting, coordination among borrowers and peer lenders.

business loans as discussed above).

We now dig one step further. In Table 5 in Panel B, we show that given the same percentage increase in small business loan origination, banks with a more complex hierarchical structure (measured as having more hierarchical layers) respond with a greater increase in their communication spending compared with their less-complex peers. This result is consistent with "hierarchical frictions" in soft information transmission (Stein, 2002): When banks face a need (or choose) to increase their engagement in the small business loan market, which implies a demand for improving their soft information handling capability, those with a more complex internal hierarchical structure often have to incur a larger portion of spending on communication IT so as to overcome such "hierarchical frictions."

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 by 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 promising topic.

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 differs drastically if the lender is a lead bank as opposed to being a participant bank; see, e.g., Sufi (2007). Compared to participant banks, lead banks are mandated by borrowers to acquire other lending participants, conduct compliance reports, and negotiate loan terms. After the loan is issued, they also have the responsibility to conduct monitoring, distribute repayments, and provide overall reporting among all lenders within the deal.[39] In this regard, performing the job of lead bank involves significantly heavier effort in information generation and sharing as well as coordinating negotiations. In short,

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

effective communication plays a more central role in the functioning of lead banks than that of participant banks. These conceptual differences between the roles played by lead and participant banks are empirically verified in Section 3.3.1; there, row 4 in Table 4 shows that the frequency of a bank serving as a lead arranger in syndicated loans exhibits a strong positive association with the communication IT spending by this bank.

4.2.2 Banks IT Spending and Demand Shock on Small Business Loans

This section provides the first piece of causal evidence on banks' lending technology on their IT spending, by studying banks' response in their technology adoption behavior when hit by a positive credit demand shock in small business loans. As small business lending is associated with intensive soft information production/transmission, we predict banks to increase their spending on communication technology (soft information), but not on software (hard information). We formally test this hypothesis now.

Our identification strategy relies on a policy shock that hits the U.S. banking sector heterogeneously across different regions in terms of small businesses' credit demand. The "Small Business Health Care Tax Credit" in the Affordable Care Act enacted between 2014 and 2015 aims to support small businesses by providing health care coverage to their employees. The program offers tax credit to small business employers who pay health insurance premia on behalf of employees. To qualify for the tax credit, the employer needs to (1) have 25 or fewer employees; (2) pay average wages less than \$50,000 a year per full-time equivalent; (3) pay at least 50% of its full-time employees' premium costs; and (4) have provided a health plan to employees that is qualified under the SHOP program coverage requirements.[40]

There are many channels through which this program could boost local small businesses'

[40] See the guidance here for an introduction of the policy. According to the IRS, small business employers should apply for the tax credit by filling in Form 8941. The tax credit can be carried backward or forward to other tax years. Also, since the amount of the health insurance premium payments is more than the total credit, eligible small businesses can still claim a business expense deduction for the premia in excess of the credit, which means both a credit and a deduction for employee premium payments. Small businesses receive credit on a sliding scale based on firm size: the tax credit is highest for small companies with fewer than 10 employees who receive an average annual salary of \$25,000 or less.

credit demand.[41] For instance, small business owners who were unable to provide employee health coverage before this program might borrow to cover the health coverage thanks to the subsidy provided by the program. What is more, small businesses could potentially initiate desired project expansions once the program relaxed their financial constraints.

While this program stimulates credit demand from small business borrowers across the entire U.S. economy, the exposure to this credit demand shock may vary substantially across different regions. In particular, the number of qualified establishments at (or right before) the program launch date, which is a key determinant for the credit demand from small business borrowers in the local area, features a substantial geographical variation across different counties. Such variation in the pre-event number of qualified establishments, together with the policy shock, can thus help us identify small business credit demand shock; recall that since the policy only explicitly targets small businesses, its impact on other types of credit demand in the local area would be indirect or limited.

Empirical Design: 2SLS Regression We run the following 2SLS regression which uses "Qualified Small Businesses" (QSB) in 2013 as an instrumental variable for the local county's exposure to the program:

$$\begin{aligned} \Delta \ln(\mathrm{CRA})_{i,c,post} &= \tilde{\alpha_i} + \mu_1 \ln(1+\mathrm{QSB})_{c,pre} + \mu_2 \mathbf{X}_{i,c} + \epsilon_{i,c} \\ \Delta \ln \mathrm{IT}_{i,c,post} &= \alpha_i + \beta \Delta \widehat{\ln(\mathrm{CRA})}_{i,c,post} + \gamma \mathbf{X}_{i,c} + \epsilon_{i,c}. \end{aligned} \tag{2}$$

The first equation in Eq. (2) is the first-stage regression. The outcome variable $\Delta\ln(\mathrm{CRA}_{i,c,\mathrm{post}})$ 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 program. The instrumental variable $\ln(1+\mathrm{QSB})_{c,pre}$ is the logarithm of one plus the total number of small businesses with fewer than 20 employees in 2013 before the policy shock.[42] In the second stage, we regress $\Delta\ln\ \mathrm{IT}_{i,c,post}$, which is the

[41]Previous literature in economics has documented increases in labor demand, firm expansion, productivity growth, etc., after the implementation of corporate tax cuts or the launch of subsidies (Cerqua and Pellegrini (2014), Rotemberg (2019), and Ivanov et al. (2021)).

[42]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)

change in logarithm of a specific type of IT spending of bank i in county c after the program implementation compared to before, on the fitted value from the first stage. In both stages, X represents a vector of control variables that includes the bank fixed effect together with several bank-county-level control variables.[43]

Estimation Results We report the estimation results of (2) in the first three columns of Table 6. Column (1) shows the regression estimates in the first-stage regression with a strong first stage result: we have an F-statistic of 31.09, which is well above the conventional threshold for weak instruments (Stock and Yogo, 2005).

Columns (2) and (3) show results of the second stage. We find a positive and statistically significant response in banks' communication investment across counties. In particular, banks who saw a one standard deviation higher growth in their small business loan making, due to the higher shock exposure of the counties in which they operated, experienced a 0.771 standard deviation higher growth in their communication spending. This translates to an average of $699,384 more per year. On the other hand, variations in local small business loan growth have no statistically significant impact on banks' software spending after the program launch. Note, our estimation includes bank fixed effects, so the above result applies to "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.

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: within a bank, its branches in counties seeing a higher growth rate in small business loans invest more in communication than other branches, but not in software spending.

In terms of magnitude, the OLS coefficients are significantly smaller compared with

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 20."

[43]The main control variable is "Revenue per Employee" at the bank-county level. The total revenue is at bank-county-level from Harte Hanks, and total number of employees is from Harte Hanks. County level control variables include the labor force, population growth rate, and total number of establishments.

those of the 2SLS estimators. One explanation for such a downward bias in OLS estimators could be a potential "omitted variable" problem, in which counties with faster growth in small business loans are those with even faster growth in some unobservable economic variables—say, mortgage (and in particular refinancing) demand—that drive local banks to spend less on communication. Specifically, if the demand for mortgage positively correlates with that of small business loans, and if banks have a fixed amount of IT budget each year, then they will allocate more IT spending—say, on software—to cater to mortgage demand. The omitted-variable problem then leads 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 the key, in other situations banks' ability to extend profitable credit is more determined by how efficiently they can deal with hard information. As mentioned earlier, one stereotypical type of loan that relies heavily on the efficient processing of readily accessible hard information is mortgage refinancing.

The discussion in Section 4.1 suggests that banks' software spending should be positively correlated with mortgage refinancing, which is consistent with row 5 of Table 4 shown in Section 3.3.1. We can move one step further and conduct a similar analysis within the mortgage lending business, by splitting mortgage business into mortgage origination and mortgage refinancing. As shown in row 7 of Table 4, banks with a larger share of refinancing loans (as a fraction of their total mortgage lending in the HMDA data) spend more on software. As expected, communication spending shows no correlation with activities in mortgage refinance business under both empirical settings.

We investigate further whether bank size plays any role in the positive association between mortgage refinancing and software spending. As reported in Table 5 Panel A, while above-median-size banks strongly positively react in their software spending to changes in their mortgage refinancing intensity, such response is relatively muted (though still statistically significant) among the group with below-median asset size. This is likely driven by the fixed-cost nature of banks' IT spending (as discussed in Section 3.2): to the extreme, tiny banks should be reluctant to adjust their IT spending—which is often lumpy—to changes in their lending activities.

These findings on the close linkages between commercial banks' spending on software IT and their engagement in refinancing is also consistent with a recent strand of literature studying fintech lenders' penetration into the credit markets. As documented in Fuster et al. (2019) and Seru (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 loan, 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 in their technology adoption behavior when hit by credit demand shocks that mostly involve processing of 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 variation in mortgage refinance demand across different regions, we construct an instrumental variable to proxy for county-level mortgage refinance propensity similar to Eichenbaum et al. (2022) and Di Maggio et al. (2017), utilizing the fact that the mortgage interest rate is systemically low during the post-crisis period.

The nationwide mortgage rate decrease has prompted existing homeowners with mortgage balances to refinance their mortgages, and an important determinant of homeowners' refinancing propensity is the pre-crisis mortgage characteristics in place before the low-interest episode kicks in. For the time window between 2011 and 2016, we construct the following county-level measure that captures the heterogeneity of each county's refinance propensity:

$$\Delta\text{Payment}_c = Ave.(\text{Payments}_{\text{old interest rate}} - \text{Payments}_{\text{new interest rate}})_c.$$

In words, we calculate the average total remaining mortgage payment savings under old versus new interest rates at the county level. In constructing this measure, we use information about all local household outstanding mortgage loans and their mortgage rates at issuance by county since 2000. We construct the hypothetical new interest rate using the interest rate of newly issued mortgage in county c matched with the loan maturity and FICO.[44] We remove loans that were defaulted on or prepaid to ensure that the measure captures only refinance propensity from local households with outstanding loans.

Importantly for our identification purpose, the county-level payment savings measure as constructed above features significant variation across regions.[45] This variation in local homeowners' payment savings from interest rate differences allows us to construct an exogenous shifter on the mortgage refinance demand faced by banks operating in the local economy.

Empirical Design: 2SLS Regression We aim to identify whether IT investment specialized in processing existing information increases when there is a higher mortgage refinance demand compared with mortgage origination. The regression specification using Payments gap$_c$ as the instrumental variable is:

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

[44] While this measure shares many similarities with the "interest rate gap" constructed in Eichenbaum et al. (2022), one difference is that we multiply the difference in interest rates of each loan by its remaining loan balance. This better reflects the "refinancing gap" as in Di Maggio et al. (2017) and therefore serve as a better proxy for the propensity of mortgage refinancing.

[45] Eichenbaum et al. (2022) also showed that the long-term low interest rate had a substantial and heterogeneous impact on regional refinancing by households located in different states, depending on how much those households could save on interest expenses by refinancing.

As shown in Eq. (3), in the first stage we regress the average of logarithmic refinance loan relative to mortgage loan origination volume of a bank i in county c during 2011 and 2016 on the instrument $\Delta\text{Payments}_c$. In the second stage, we then regress the average of the logarithmic of IT spending of bank i in county c during 2011 and 2016 on the fitted values of $\ln(\text{Refinance/Origination})_{i,c}$.

Estimation Results Table 7 reports our estimation results. In the first stage, the instrumental variable "$\Delta\text{Payments}_c$" predicts mortgage refinancing activities across different counties quite well, with a high F-statistics (10.71). Columns (2) and (3) show the results of second-stage regressions. By including bank fixed effects, this result is identified from the within-bank-cross-county variations. A one standard deviation increase in mortgage refinancing relative to mortgage origination lent out by a bank—driven by its local exposure to high refinance savings—leads to a 0.529 standard deviation increase in software spending, which amounts to \$1.68 million more budget on software per year.

It is also interesting to observe that, in contrast, communication spending does not demonstrate significant changes in response to the refinancing demand captured by the predetermined refinance propensity. This fact supports our premise that mortgage refinancing is a stereotypical bank lending activity that hinges on efficient processing of readily accessible hard information instead of producing new information.

Comparison: 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 7. Similar to the comparison between OLS and 2SLS estimates we reported in Section 4.2.2, we find qualitatively similar yet quantitatively smaller OLS estimators.[46]

Similar to our analysis on small business credit demand in Section 4.2.2, an "omitted variable" issue can explain such downward biases in the OLS estimators. Here, counties seeing more mortgage refinances issued by local banks might also have other loan demands

[46]Table A3 shows the results of the same OLS specification with bank, year and county fixed effects and bank×year and county fixed effects.

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 commercial banks.[47] In this section, through the angle of examining US commercial banks' IT spending, we seek to provide some answer to a widely debated question: Has the traditional banking sector started reacting to the fast growing fintech industry? If yes, how?

5.1 How do 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. For example, Buchak et al. (2018) provide evidences that fintech lenders offer faster and more convenient services but with higher interest rates; Fuster et al. (2019) find that fintech lenders process mortgage applications 20% faster without incurring a higher default rate.

While fintech lenders have been quickly gaining market share in various loan markets over the past decade, the exact ways through which incumbent commercial banks react to the aggressive fintech entry remain unclear. 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, as fintech services increase the marginal value of such investment due to complementarity. Furthermore, from an information channel,

[47]See link to the article talking about this issue.

the emergence of fintech lenders who have comparative advantages in information handling in certain markets would render traditional bank lenders more adversely selected in these markets. Concerning a more severe "winner's curse," banks might thus respond by decreasing their engagement in these credit 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 with these new fintech entrants. Since fintech lenders often possess superior information processing capacity, this implies that incumbent banks respond by increasing their investment in software. From the information aspect, one explanation for this increased investment by banks is to prevent themselves from getting adversely selected in lending markets. In fact, Figure 2(b) has shown that bank branches in areas with higher fintech presence increase their IT spending at a faster rate than those in areas with low fintech presence, suggesting a potential "catching-up" behavior of the traditional banking sector.

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 diff-in-diff 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 of 2008, forty States and the District of Columbia (DC) moved relatively fast to approve its entry; and between 2010 and 2016, another nine states approved Lending Club's entrance at different times.[48] Following Wang and Overby (2017) and Kim and Stähler (2020), we summarize

[48] 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

Lending Club's staggered entrance into nine states since 2010 in Table 8.[49] For Kansas and North Carolina, the actual approval time was 2010Q4; since 2010 is the starting year of our Harte Hanks dataset, this implies that 2010 is a contaminated year as a pre-treatment period for these two states. We hence exclude these two states, leaving us with a total of seven states for our staggered entrance analysis.

Importantly for our identification purpose, this variation in the approval time—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 chooses to enter the local markets with a rising credit demand, then any observed change in local commercial banks' IT investment behavior cannot be convincingly attributed to the entry of their fintech competitor.

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

$$\ln(\text{IT Spending})_{i,c,t} = \alpha_{i,c} + \alpha_t + \beta \times \text{LC}_{i,c,t} + \mu \mathbf{X} + \epsilon_{i,c,t}, \tag{4}$$

where IT Spending $\in$ {Total, Software, Communication}. We include the bank-times-year and county fixed effects, denoted by $\alpha_{i,t}$ and α_c respectively. $\text{LC}_{i,c,t}$ 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. X is a set of control variables, and the standard error is clustered at county level. Our main parameter of interest is β, which measures the average treatment effect of Lending Club approval on bank technology spending. Estimations are weighted by Lending

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.

[49] 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. Therefore, we drop these 40 states in our analysis. The same treatment is adopted in Wang and Overby (2017).

Club loan volume after the entry.

Columns (1) to (2) of Panel A in Table 9 gives the results respectively for software and communication spending in the baseline setting. Consistent with "catching-up" story, column (1) shows that, after Lending Club enters in country c, banks on average increase their software IT spending in county c by around 7.0%, and this estimate is statistically significant. The growth in communication spending right after the entrance of Lending Club was slightly negative (-1.5%) and the estimation is statistically insignificant.

Recent literature points out the bias in staggered two-way fixed effects (TWFE) setting even if the assumption of parallel trends holds (Callaway and Sant'Anna (2021), Sun and Abraham (2021), and Baker et al. (2022), etc.). For robustness, we adopt the interacted TWFE design as in Callaway and Sant'Anna (2021). The design involves running a separate regression 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 aggregating 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 the Lending Club entry. Standard errors are based on Bootstapping with 50 draws. The results are in columns (4) to (5) of Table 9. The estimates are quite similar to while a little larger than those in columns (1) through (2).

Heterogeneity in Response across Bank Sizes In Panel B of Table 9, we explore whether banks of different sizes respond differently to the fintech entry. Similar to our specification in Table 5, large (small) banks are defined as banks with asset size above (below) the median bank asset size in our sample. We find that large banks increased software spending by 10.9% right after the entry of Lending Club, and the increase is statistically significant; whereas small banks' software spending growth was only about one fifth of the magnitude compared with larger banks, and statistically insignificant. On the other hand, while small banks also barely display any response in their communication spending, large banks actually cut their spending on communication IT by 6.1% (which is statistically significantly) following the fintech entry.

The asymmetric impact on the technology adoption reaction by different sized banks is intriguing, and suggests that the specialty (regarding information handling) of the newly entered fintech is more relevant for the credit market segments that large banks actively engage in. This finding is consistent with Balyuk et al. (2020), who find that fintech lending more often substitutes lending made by large banks rather than smaller banks whose lending technology is more relationship based. Given that smaller banks engage more in relation-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.

Finally, that large banks significantly reduce their communication spending is also consistent with several recent papers studying how fintech entry affects credit market outcomes. For instance, Beaumont et al. (2019) show that borrowers with better fintech-access are more able to purchase and pledge hard-information-heavy assets as collateral to obtain new bank credit. This suggests that fintech lenders, thanks to their advanced data collection and processing technology, help convert soft information (say borrowers' ability to repay) to hard information (say trucks). 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.[50]

Summary and Discussion To sum up, our findings suggest that the entrance of fintech lenders such as Lending Club into the credit markets overall induces banks—especially large ones—to "catch up" and invest to adapt their lending technology. To the best of our

[50]There are several other plausible explanations to the spending cut in communication IT empirically identified by our analysis. For instance, a simple "fixed total budget" story can easily rationalize this finding. Furthermore, as documented by Balyuk et al. (2020), credit extended by newly fintech entrants often substitutes for loans by large/out-of-market banks, as oppose to those by small/in-market banks. As a consequence of large banks' retracted engagement in out-of-market lending activities, one should naturally expect them to reduce their communication IT spending.

knowledge, this is the first piece of direct evidence that the entry of fintech lenders in the credit market can spur banks operating in the same local area to invest more in their lending technology to catch up. Furthermore, consistent with the existing literature that highlights the comparative advantage of fintech lenders in processing hard information and making prompt decisions, our findings also suggest that such "catching up" behavior engaged by the commercial banking sector is more reflected in their investment in *software* IT.[51]

We have discussed in Section 5.1 the potential channels through which the entry of fintech lenders could potentially affect local commercial banks' IT investment decisions. Our empirical findings in this section are consistent with a competition story that following fintech lenders' entry into certain credit markets (e.g., mortgage markets), large banks who originally play an active role in these markets tend to respond by increasing their IT spending in the relevant categories so as to protect their market share. 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 traditional banking sector. In this paper, we provide the first comprehensive study of banks' IT spending, which we view as banks' investment to develop and improve their lending technology.

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

[51] See summary of this point in Berg et al. (2021)

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, during the information era, develop and adapt their lending technology in response to economic shocks on their operating environment, including credit demand shocks as well as fintech entry. These causal analyses, to the best of our knowledge, provide the first piece of evidences on the endogenous lending technology adoption in the banking literature.

Our findings open up several important further questions. For instance, 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 credit cycle, and monetary policy transmission? We leave it to future research to provide answers to these questions.

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.

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.

Baker, A. C., Larcker, D. F., and Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395.

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*.

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.

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. (2018). 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. (2017a). The decline of big-bank lending to small business: Dynamic impacts on local credit and labor markets. Working Paper 23843, National Bureau of Economic Research.

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

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*.

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).

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 Zhou, J. (2022). Open banking: Credit market competition when borrowers own the data. *Working Paper, University of Chicago*.

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*.

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*.

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*.

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.

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*.

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.

Schnabl, P. and Gopal, M. (2020). The rise of finance companies and finTech lenders in small business lending. *Working Paper*.

Seru, A. (2019). Regulating banks in the era of fintech shadow banks. *BIS Speach 2019*.

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.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199. Themed Issue: Treatment Effect 1.

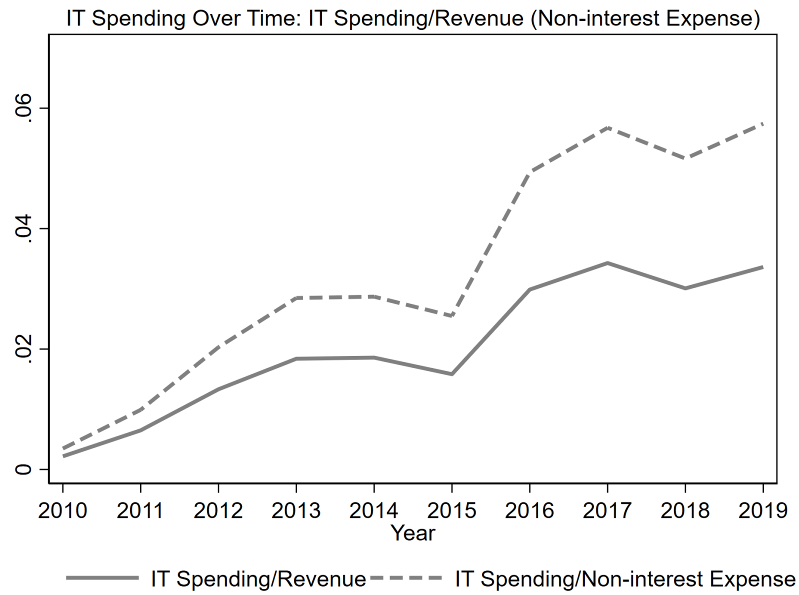
Tang, H. (2019). Peer-to-Peer lenders versus banks: Substitutes or complements? *The Review of Financial Studies*, 32(5):1900–1938.

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

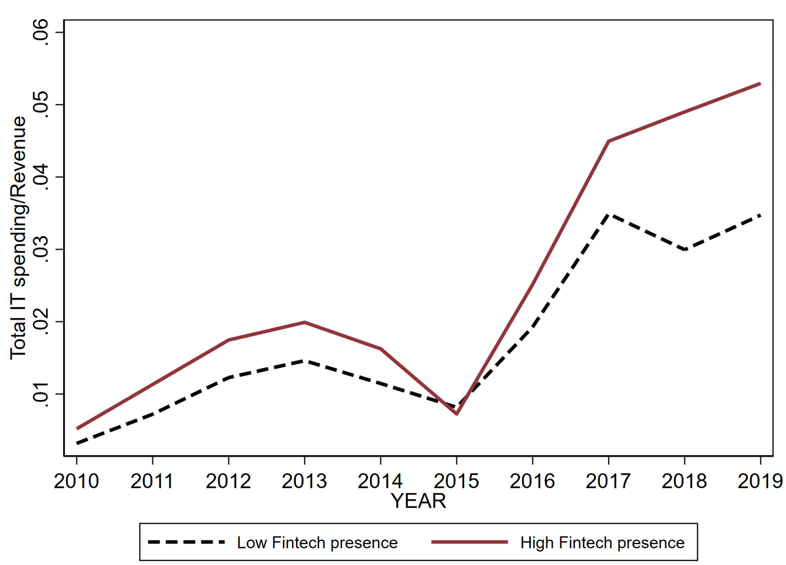
Wang, H. and Overby, E. (2017). How does online lending influence bankruptcy filings? evidence from a natural experiment. *Management Science*, 2022:15937.

Figure 1. IT Spending: Time Trend

Panel A shows the time trend of banks' IT spending from 2010 to 2019. The solid line shows the weighted average of banks' IT Spending as a share of banks' total income; the dashed line shows the weighted average of banks' IT spending as a share of banks' total income. "Revenue" is constructed using the item RIAD4000 in Call Report, and "Non-interest Expense" is the non-interest expenses reported by item RIAD4093 in Call Report. Panel B shows the relationship between banks' IT spending and the presence of Fintech companies in the local economy. The y-axis is the county-level "Total IT spending/Revenue" of local banks. Based on the average Fintech lending share in mortgage market of a county during 2010-2019 used in (Fuster et al. (2019)), we define high and low fintech presence as counties with above-median and below-median fintech lending share in the local mortgage market, respectively.



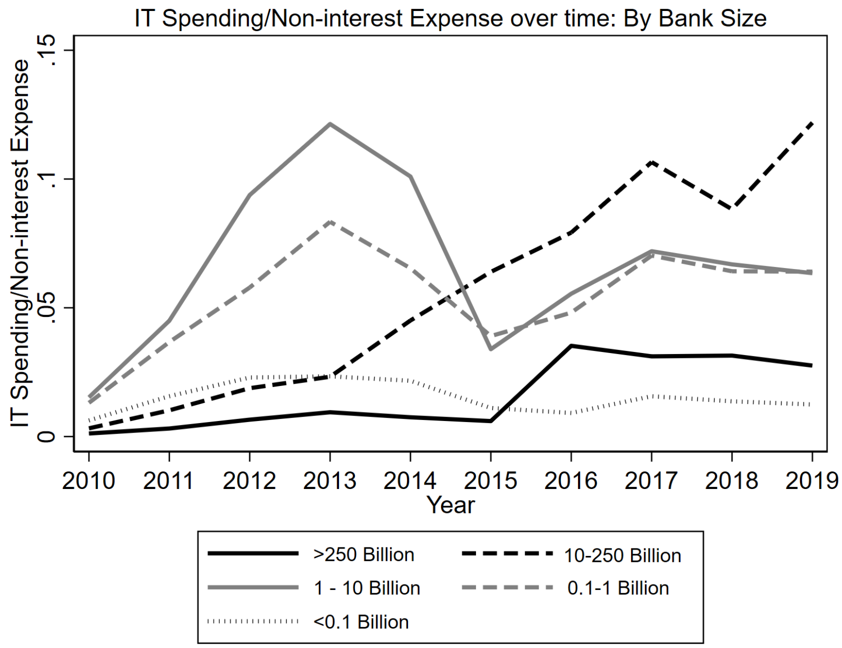
(a) Panel A



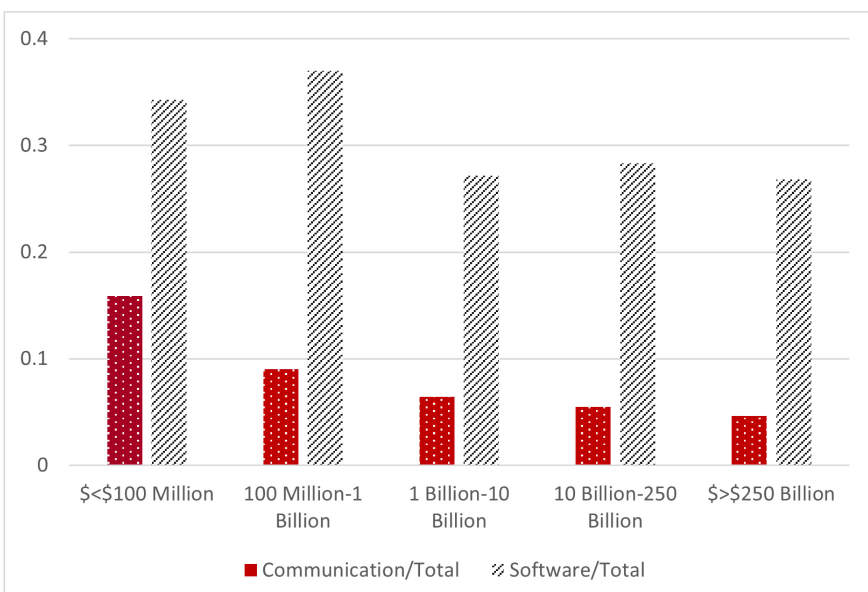
(b) Panel B

Figure 2. IT Spending Time Trend and Composition, by Bank Size

The figures show the time trend of banks' IT spending from 2010 to 2019 by the five categories of bank asset size groups (Panel A) and the differences in composition of IT spending by bank size groups (Panel B). In Panel A, 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 Call Report. In Panel B, the shaded bars show the average proportion of total IT budget spent on communication and software across bank size groups.



(a) Panel A



(b) Panel B

Table 1. Sample Coverage

This table demonstrates the sample coverage of banks across five categories of banks' size groups. All banks in the sample are commercial banks. The Call Report bank population includes only commercial banks (with "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	1184.24	7	787.34	85.71%	96.66%
$10 Billion–$250 Billion	88	42.30	106	43.69	83.02%	73.22%
$1 Billion–$10 Billion	474	2.90	590	2.78	80.34%	89.43%
$100 Million–$1 Billion	942	0.40	4161	0.32	22.64%	29.44%
<$100 Million	296	0.06	2048	0.05	14.45%	14.23%

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. No. of IT employees is the total amount of IT-related employees. IT Spending/Revenue is total IT Spending scaled by banks' total gross income; IT Spending/Non-interest expense is total IT spending scaled by non-interest expenses; IT spending/Net income is total IT spending scaled by total income minus the gross total expenses. 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)	7.311	111.354	0.030	0.215	1.056
No. of IT Employees	133.434	872.102	7.000	21.578	56.400
Storage Amount(PB)	3.517	25.522	0.107	0.476	1.779
IT Spending/Revenue	0.031	0.155	0.003	0.010	0.023
IT Spending/Net income	0.597	18.475	0.018	0.051	0.135
IT Spending/Expenses	0.037	0.191	0.004	0.012	0.028
IT Spending/Non-interest Expenses	0.044	0.213	0.005	0.014	0.034
Communication/Total	0.092	0.117	0.028	0.052	0.105
Communication/Revenue	0.0016	0.0075	0.0001	0.0005	0.0014
Software/Total	0.334	0.161	0.220	0.321	0.474
Software/Revenue	0.011	0.066	0.001	0.003	0.007
Hardware/Total	0.171	0.119	0.063	0.158	0.235
Services/Total	0.327	0.137	0.243	0.323	0.417
Other/Total	0.056	0.099	0.008	0.014	0.062

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

This table presents the results of 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.

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

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 the average values across 2010-2019 within bank i. Fixed effects include bank size and banks' headquarter state fixed effects. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	Software/Revenue	Communication/Revenue	Hardware/Revenue	Services/Revenue
	(1)	(2)	(3)	(4)
C& I loans/Total loan	0.031	0.050**	0.049**	0.043*
	(0.025)	(0.024)	(0.024)	(0.025)
Net income/Total Assets	-0.142***	-0.217***	-0.242***	-0.107***
	(0.028)	(0.028)	(0.028)	(0.029)
Deposits/Assets	-0.013	0.028	0.032	-0.008
	(0.031)	(0.030)	(0.030)	(0.031)
Revenue per Employee	-0.267***	-0.347***	-0.301***	-0.298***
	(0.035)	(0.034)	(0.034)	(0.035)
Salaries/Assets	-0.018	-0.132***	-0.099***	-0.034
	(0.026)	(0.025)	(0.025)	(0.026)
Equity/Assets	0.070**	0.051*	0.046*	0.071***
	(0.028)	(0.027)	(0.027)	(0.028)
Size FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
AdR-squared	0.098	0.162	0.150	0.110
N	1798	1798	1798	1798

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.

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

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 is the frequency of refinance as a percent of total number of mortgage issuances that are reported in HMDA. 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.

Panel A: Loan Specialization

	Software/Revenue	Communication/Revenue	Hardware/Revenue	Services/Revenue
	(1)	(2)	(3)	(4)
C& I loans/Total loan	0.031	0.050**	0.049**	0.043*
	(0.025)	(0.024)	(0.024)	(0.025)
CRA/Total loan	-0.150***	0.113***	0.063**	0.039
	(0.029)	(0.028)	(0.029)	(0.029)
Other C&I/Total loan	0.050**	0.036	0.041*	0.039
	(0.025)	(0.024)	(0.024)	(0.025)
Personal loan/Total loan	0.045*	0.033	-0.014	0.025
	(0.027)	(0.026)	(0.023)	(0.023)
Mortgage refinance/Total loan	0.093***	-0.032	0.033	0.027
	(0.033)	(0.034)	(0.035)	(0.029)
Other personal loans/Total loan	-0.028	0.039	0.019	0.011
	(0.026)	(0.028)	(0.035)	(0.026)
% Refinance/Total Mortgage	0.081***	-0.002	0.041*	0.055**
	(0.024)	(0.024)	(0.024)	(0.023)
Agricultural loans/Total loan	0.026	0.073**	0.048	0.043
	(0.031)	(0.030)	(0.030)	(0.031)

Panel B: Hierarchical Complexity and IT Spending

	(1)	(2)	(3)	(4)
Hierarchical layer	0.00859	0.0484**	0.0147	0.0144
	(0.0233)	(0.0232)	(0.0232)	(0.0231)
ln(num of offices)	-0.00506	0.0528**	0.0166	0.0121
	(0.0239)	(0.0238)	(0.0238)	(0.0237)

Panel C: Banks' Role in Syndicated Lending

	(1)	(2)	(3)	(4)
% Lead bank/Total syndicate	-0.612*	1.759***	1.393***	1.083***
	(0.333)	(0.295)	(0.278)	(0.262)

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.

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

In Panel A, small (large) banks are defined as the banks with asset size below (above) median asset size in our sample. "Size Group FE" refers to the fixed effects of small (large) banks. In Panel B, "Hierarchical layer" is the number of types of its locations as defined in Section 2.3. "Size FE" refers to the fixed effects of the five bank asset groups defined in Section 3.2. 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/Rev	Communication/Rev
	(1)	(2)
Small × Refinance/Total loan	0.0794*	
	(0.0438)	
Large × Refinance/Total loan	0.194***	
	(0.0506)	
Small × CRA/Total loan		0.423***
		(0.142)
Large × CRA/Total loan		0.156***
		(0.0335)
State × Size Group FE	Y	Y
AdR-squared	0.113	0.171
N	1792	1789

Panel B: Bank Hierarchical Structure and IT Spending

	Software/Rev	Communication/Rev
	(1)	(2)
Hierarchical layer=1 × Refinance/Total loan	0.0761	
	(0.0807)	
Hierarchical layer=2 × Refinance/Total loan	0.0870*	
	(0.0523)	
Hierarchical layer=3 × Refinance/Total loan	-0.0160	
	(0.0496)	
Hierarchical layer=1 × CRA/Total loan		0.0558
		(0.0724)
Hierarchical layer=2 × CRA/Total loan		0.0872**
		(0.0406)
Hierarchical layer=3 × CRA/Total loan		0.164***
		(0.0494)
Size× Layer Group FE	Y	Y
State× Layer Group FE	Y	Y
AdR-squared	0.0962	0.177
N	1779	1778

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

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:

$$\Delta \ln(\text{CRA})_{i,c,post} = \tilde{\alpha_i} + \mu_1 \times \ln(1 + \text{Qualified Small Buzs})_{c,pre} + \mu_2 \mathbf{X}_{i,c} + \epsilon_{i,c}$$
$$\Delta \ln(\text{IT})_{i,c,post} = \alpha_i + \beta \times \Delta \widehat{\ln(\text{CRA})}_{i,c,post} + \gamma \mathbf{X}_{i,c} + \epsilon_{i,c}$$

The last two columns show the following OLS specification:

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

$\Delta\ln\ (\text{CRA})_{i,c,post}$ 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 2010-2013. Bank control variables include pre-shock revenue per employee and logarithmic of mortgage refinance over mortgage origination. County level control variables include the pre-shock labor force, population growth rate, and total number of establishments, and county-level mortgage loan HHI. Fixed effects include bank fixed effects. ***, **, 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)
ln(Qualified Small businesses)	0.063***				
	(0.011)				
$\Delta \widehat{\ln(CRA)}$		0.244	0.771***		
		(0.218)	(0.252)		
$\Delta \ln(CRA)$				0.013	0.025***
				(0.009)	(0.009)
Bank FE	Y	Y	Y	Y	Y
F	31.09				
AdR-squared	0.397	-0.219	-0.621	0.101	0.102
N	19,803	19,799	19,798	19,799	19,798

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

Table 7. Hard Information and Banks' IT Spending

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

$$\ln(\text{Refinance/Origination})_{i,c} = \tilde{\alpha_i} + \mu_1 \times \Delta\text{Payments}_c + \mu_2\ \mathbf{X}_{i,c} + \epsilon_{i,c}$$
$$\ln(\text{Type S Spending})_{i,c} = \alpha_i + \beta \times \ln(\widehat{\text{Refinance/Origination}})_{i,c} + \gamma\ \mathbf{X}_{i,c} + \epsilon_{i,c}$$

Column (4) and (5) 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. Payments gap is the hypothetical amount of interest payments that could be saved due to the interest rate gap, if local households chose to refinance their mortgages during the year of 2011 and 2016. Control variables include banks' revenue per employee and logarithmic of small business loans. County level control variables include the pre-shock labor force, population growth rate, total number of establishments, and county-level mortgage loan HHI. Fixed effects include bank fixed effects. ***, **, 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)
$\Delta\text{Payment}_c$	0.654***				
	(0.199)				
$\text{Ln}(\widehat{\text{Refinance/Originagion}})$		0.529**	0.333		
		(0.244)	(0.212)		
Ln(Refinance/Originagion)				0.019***	0.021***
				(0.004)	(0.004)
Bank FE	Y	Y	Y	Y	Y
F	10.71				
AdR-squared	0.367	-0.315	0.029	0.219	0.225
N	23,884	23,884	23,884	23,884	23,884

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

Table 8. 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 9. Fintech Exposure 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

$$\ln(\text{IT Spending})_{i,c,t} = \alpha_{i,c} + \alpha_t + \beta \times \text{LC}_{i,c,t} + \gamma \mathbf{X} + \epsilon_{i,c,t},$$

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. Column (3) and (4) of Panel A use the interacted TWFE method as in Callaway and Sant'Anna (2021). 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. The estimations in Panel B are based on the TWFE method as in Callaway and Sant'Anna (2021). ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A: IT Spending and Fin-tech Entrance

	Baseline		Callaway and Sant'Anna (2021)	
	ln(Software)	ln(Communication)	ln(Software)	ln(Communication)
	(1)	(2)	(3)	(4)
After	0.070***	-0.015	0.073***	-0.031
	(0.021)	(0.019)	(0.029	(0.033)
Fixed Effects	Bank×County, Year, Size group			
AdR-squared	0.799	0.784		
N	13,552	13,552		

Panel B: Responses by bank sizes

	Baseline		Callaway and Sant'Anna (2021)	
	ln(Software)	ln(Communication)	ln(Software)	ln(Communication)
	(1)	(2)	(3)	(4)
Small×After	0.036	0.037	0.034	0.067
	(0.028)	(0.026)	(0.042)	(0.047)
Large×After	0.109***	-0.061***	0.119***	-0.127***
	(0.023)	(0.021)	(0.029)	(0.039)
Fixed Effects	Bank×County, Year, Size group			
AdR-squared	0.786	0.771		
N	13,546	13,546		

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

Appendix

A Figures and Tables

Figure A1. Total Asset of Banks in Sample

This figure shows the sum of total asset size of all banks in our sample from 2010 to 2019 U.S. The red dashed line is the sum of all commercial banks' asset size in U.S., data source is Board of Governors of the Federal Reserve System (US), Total Assets, All Commercial Banks [TLAACBW027SBOG]. The red solid line is the sum of total asset sizes of banks in our sample. The black solid line is the sample bank size out as a share of total nation-wide banks' total asset size.

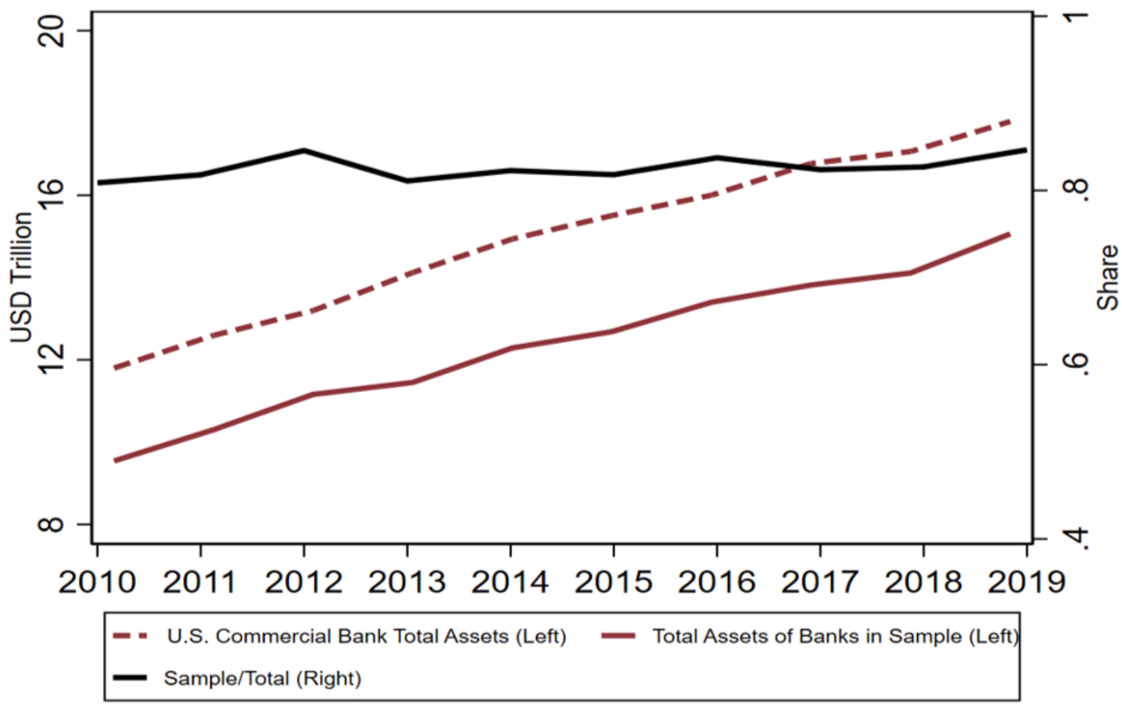


Figure A2. IT Spending Time Trend: IT Spending as a Share of Total Revenue

The figures show the time trend of banks' IT spending from 2010 to 2019 by the five categories of bank asset size groups. The vertical axis is banks' total IT spending scaled by banks' total revenue. The asset size groups are categorized based on a bank's average asset size during 2010 and 2019. Total revenue is calculated as the sum of banks' net interest income (RIAD4074) and non-interest income (RIAD4079).

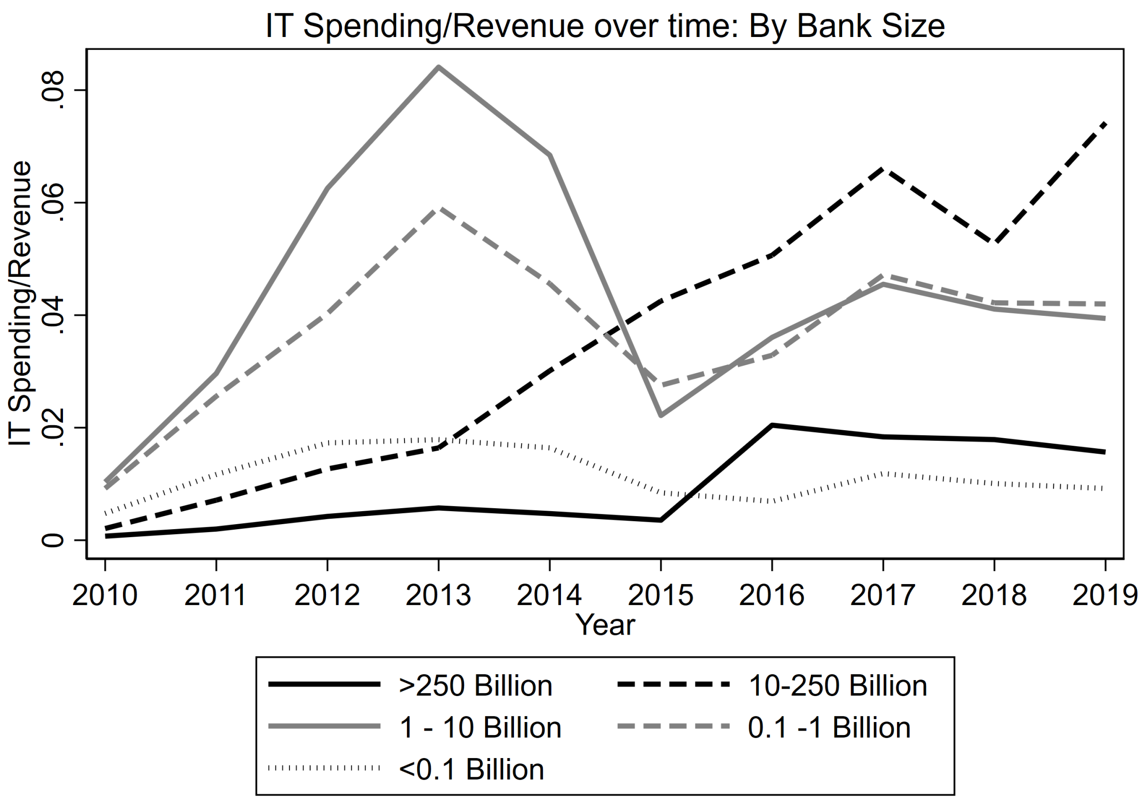
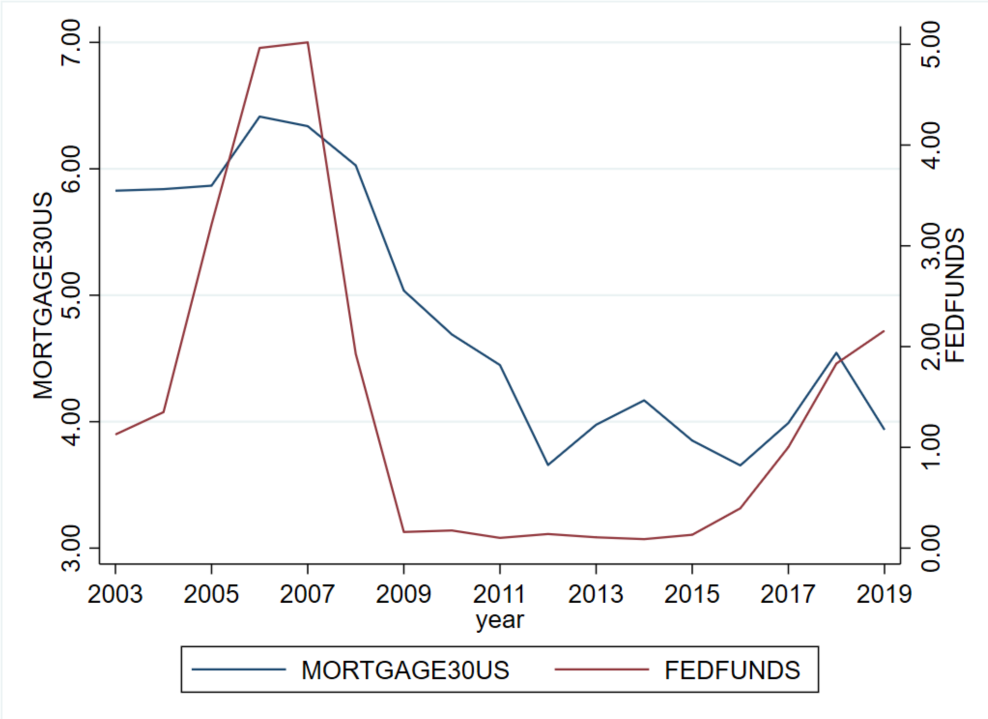


Figure A3. "Low Mortgage Rate Episode"

The figures show the time-series of aggregate mortgage interest rate and the Federal Funds Rates in the upper panel and the time-series thirty-year fixed mortgage rate and refinance activities. "MORTGAGE30US" is the 30-Year Fixed Rate Mortgage Average in the United States from Freddie Mac. "FEDFUNDS" is the effective Federal Funds Rate by Board of Governors of the Federal Reserve System. "Refi Index (SA)" is the Mortgage Bankers Association (MBA) Refinancing Index (seasonally adjusted).



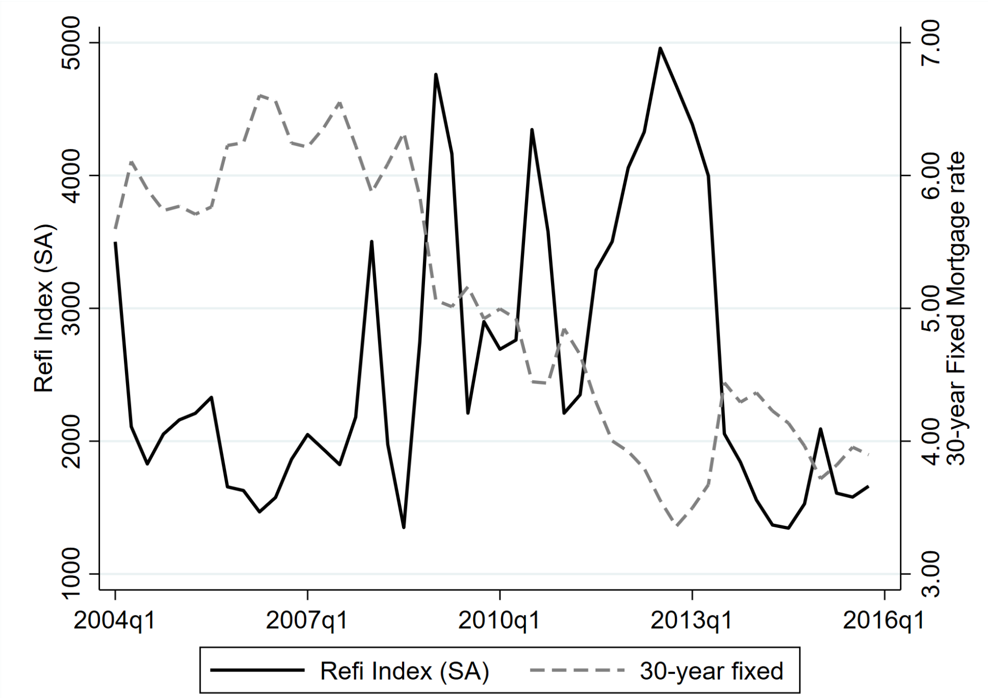


Figure A4. McKinsey (2012) "Breakthrough IT Banking"

Business Technology Office

Breakthrough IT banking

Some Asian banks achieve superior returns despite relatively low IT expenditures. What's their secret?

Sai Gopalan, Gaurav Jain, Gaurav Kalani, and Jessica Tan

Banks have long relied on technology to introduce products such as online banking, ATMs, and mobile payments, and to improve back-office efficiency. But that reliance comes with a price. Globally, the banking sector spends an average of 4.7 percent to 9.4 percent of operating income on IT, while other sectors spend less: insurance companies and airlines, for example, spend 3.3 percent and 2.6 percent of income, respectively.

Our Asian Banking IT Benchmarking Study[1] finds, however, that a bank's high IT expenditures do not always correlate with superior performance. Some banks with large IT budgets often have trouble leveraging investments to generate commensurately high revenue growth and operational efficiency. Survey data show that 66 percent of banks with higher-than-average IT spending relative to income generated lackluster results, with revenue growth 0.4 percentage points lower than the industry standard and a cost-income (C/I) ratio 2.5 percentage points higher.

By contrast, 23 percent of the 44 banks surveyed outperformed the market on both revenue growth (up 10.9 percentage points) and C/I ratio (down 4.6 percentage points) while spending 29 percent less on IT than other banks in our study. These outperforming banks are more likely to view IT as a strategic enabler, and their investments mirror this outlook. Outperformers direct a higher share of spending toward technologies designed to create new business value and a lower share of spending on support operations, such as finance and human resources. These banks are also more likely than the lower performers to promote efficiency through a consolidated IT footprint as well as formal vendor- and demand-management practices.

The common denominator linking high-performing Asian banks is a commitment to strong governance and spending alignment with the needs of the business. This finding supports our experience with bank clients in Europe and the Americas, and prompted us

[1] The 2010 biennial McKinsey Asian Banking IT Benchmarking survey comprised 44 banks across 11 Asia-Pacific countries, with the results tracked against prior year benchmarks from 2006 onward.

Figure A5. Definition of Different Types of IT Spending

- *COMM_BUDGET*

The modeled IT budget for communication services at this site.
It is defined as the network equipment that companies operate to support their communications needs.
It includes:

- routers
- carrier line equipment
- fiber optic equipment
- switches
- private branch exchanges (PBXes)
- radio and TV transmitters
- Wi-Fi transmitters
- desktop telephone sets; wide-area network (WAN) and local-area network (LAN) equipment
- videoconferencing and telepresence equipment
- cable boxes
- other network equipment.
- end-client mobile devices like cell phones/iPhones that are bought by individuals

(a) Figure A

- *SOFTWARE_BUDGET*

The modeled IT budget for software at this site.
It is defined as software from third parties, whether that software is packaged or semipackaged software delivered on CD and installed within the company, hosted by a third party, offered on a SaaS basis from a multitenant shared-instance server accessible by a browser, or custom-created for a company by third-party contractors or consultants.
It includes:

- license, maintenance, subscriptions and software vendor-provided services revenues for all categories of middleware software (including storage management systems, database management systems, IT management systems, security software, application servers and application development software)
- application software such as :
    - desktop applications
    - information management software (like business intelligence and enterprise content management)
    - process applications (like ERP, CRM, SCM or PLM)
    - ePurchasing software
    - risk and payment management software

    . We also include vertical industry applications (like banking management systems, security trading systems, insurance underwriting or claims management software, retail management software, or hospital information systems). Finally, we include computer operating systems software, even though that cost is often bundled
- vertical industry applications (like banking management systems, security trading systems, insurance underwriting or claims management software, retail management software, hospital information systems)
- computer operating systems software (even though that cost is often bundled)

(b) Figure B

- *SERVICES_BUDGET*

The modeled IT budget for IT-related services at this site.
It is defined as project-based consulting or systems integration services that vendors provide to businesses and Governments, whether on or off-site.
It includes:

- contractors, consulting services for IT strategy, security assessments and process change
- systems integration
- project services
- mainframe outsourcing, desktop support outsourcing, distributed systems outsourcing, network outsourcing, application hosting, application management outsourcing and application testing. These applications are single-instance software deployments, generally owned rather than subscribed to, and thus are different from SaaS.
- computer hardware support and maintenance services.

(c) Figure C

- *HARDWARE_BUDGET*

The modeled IT budget for hardware at this site.
It includes the classic computer hardware that IT departments buy and support, regardless of whether the IT department itself operates that equipment (such as servers) or oversees the use of this equipment by employees (such as PCs):

- PCs: personal computers (laptops, desktops, and tablets)
- Servers/Mainframes
- Peripherals: monitors, terminals, printers, keyboards, mice, USB devices, etc...
- Storage: storage devices (NAS, DAS, tape)
- Other hardware: hardware specific to the industry (point-of-sales equipment based on PCs, smart cards, embedded computer chips, etc...)

(d) Figure D

Table A1. Bank Characteristics and IT Spending

This table presents the results of regression of banks' IT spending structure between 2010 and 2019 on banks' loan portfolio on balance sheet before the financial crisis. The regression specification is as follows:

$$\frac{\text{Type S IT spending}}{\text{Total}}_{i,2010-2019} = \alpha + \beta \frac{\text{Type l Loan}}{\text{Total loan}}_{i,2005-2009} + \gamma X + \epsilon_i$$

C&I loan/Total loan is commercial and industrial loan scaled by total loan; Personal loan/Total loan is personal loan and the real estate loan to 1-4 families scaled by total loan; agriculture loan/Total loan is agricultural loan scaled by total loan. All the three types of loans as a share of total loans are the bank-level average loan proportions from 2005 to 2009. IT spending profiles are defined as each type of IT spending scaled by total IT spending, and taking an average at the bank level between 2010 and 2019. 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; control variables are at bank-year level. Fixed effects include bank fixed effects, year fixed effects and county fixed effects. Standard errors are clustered at county and bank level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A

	$\frac{\text{Software}}{\text{Total}}_{10-19}$	$\frac{\text{Communication}}{\text{Total}}_{10-19}$	$\frac{\text{Hardware}}{\text{Total}}_{10-19}$	$\frac{\text{Services}}{\text{Total}}_{10-19}$
	(1)	(2)	(3)	(4)
C&I loan/Total loan(05-09)	-0.001	0.047***	0.041***	-0.011
	(0.022)	(0.015)	(0.015)	(0.016)
Net income/Total Assets	0.217	-0.166	-1.878***	2.017***
	(0.351)	(0.200)	(0.210)	(0.289)
Revenue per Employee	0.051**	0.041**	0.020	-0.026
	(0.025)	(0.017)	(0.014)	(0.028)
Equity/Assets	0.103	-0.008	-0.122*	0.087
	(0.100)	(0.071)	(0.071)	(0.085)
Salaries/Assets	1.304***	-0.334	0.215	0.580**
	(0.238)	(0.331)	(0.616)	(0.268)
Deposits/Assets	-0.044	0.018	0.026	-0.037
	(0.053)	(0.029)	(0.029)	(0.036)
AdR-squared	0.199	0.161	0.110	0.220
N	1649	1649	1649	1649

Panel B

	(1)	(2)	(3)	(4)
Personal loans/Total loan(05-09)	0.052***	-0.058***	-0.038***	-0.010
	(0.014)	(0.011)	(0.009)	(0.009)
Net income/Total Assets	0.382	-0.335*	-1.985***	1.981***
	(0.346)	(0.201)	(0.213)	(0.294)
Revenue per Employee	0.058**	0.041***	0.021	-0.030
	(0.027)	(0.015)	(0.013)	(0.027)
Equity/Assets	0.086	0.006	-0.114	0.092
	(0.101)	(0.071)	(0.071)	(0.085)
Salaries/Assets	1.244***	-0.269	0.257	0.592**
	(0.226)	(0.306)	(0.599)	(0.266)
Deposits/Assets	-0.043	0.023	0.031	-0.038
	(0.053)	(0.029)	(0.029)	(0.036)
AdR-squared	0.205	0.172	0.114	0.220
N	1649	1649	1649	1649

Panel C

	(1)	(2)	(3)	(4)
Agriculture loan/Total loan(05-09)	-0.082***	0.057***	0.028**	0.025**
	(0.015)	(0.015)	(0.013)	(0.011)
Net income/Total Assets	0.441	-0.307	-1.942***	1.945***
	(0.353)	(0.203)	(0.217)	(0.290)
Revenue per Employee	0.050**	0.049**	0.026*	-0.027
	(0.025)	(0.019)	(0.015)	(0.028)
Equity/Assets	0.079	0.003	-0.118*	0.096
	(0.100)	(0.071)	(0.071)	(0.085)
Salaries/Assets	1.211***	-0.271	0.245	0.609**
	(0.213)	(0.298)	(0.601)	(0.259)
Deposits/Assets	-0.055	0.031	0.036	-0.035
	(0.052)	(0.029)	(0.029)	(0.036)
AdR-squared	0.208	0.165	0.109	0.221
N	1649	1649	1649	1649

Table A2. Summary Statistics of Banks' IT Spending by Bank Size Group

This table presents the summary statistics of banks' IT spending by banks' size groups. Banks in the sample are split into five groups. Total IT Spending is the sum of all types of IT spending in millions of dollars. No. of IT employees is the total amount of IT-related employees. IT Spending/Income is total IT Spending scaled by banks' total income, IT Spending/Non-interest expense is total IT spending scaled by non-interest expenses; IT spending/Net income is total IT spending scaled by total income minus the gross total expenses. The different categories of IT spending are the four categories of IT spending scaled by total IT spending.

	Mean	S.d.	Median
< $100 Million			
IT Spending/Revenue	0.015	0.048	0.005
IT Spending/Expenses	0.019	0.075	0.007
Communication/Total	0.159	0.176	0.086
Software/Total	0.343	0.132	0.341
Services/Total	0.330	0.138	0.340
Hardware/Total	0.206	0.133	0.203
PC/Total	0.102	0.132	0.075
Server/Total	0.100	0.127	0.066
Terminal/Total	0.023	0.081	0.004
Printer/Total	0.022	0.081	0.003
Storage/Total	0.092	0.135	0.040
Other/Total	0.098	0.137	0.037

	Mean	S.d.	Median
$100 Million-$1 Billion			
IT Spending/Revenue	0.028	0.140	0.009
IT Spending/Expenses	0.040	0.200	0.014
Communication/Total	0.090	0.109	0.052
Software/Total	0.370	0.155	0.348
Services/Total	0.308	0.124	0.316
Hardware/Total	0.167	0.115	0.158
PC/Total	0.066	0.090	0.053
Server/Total	0.070	0.087	0.052
Terminal/Total	0.012	0.047	0.004
Printer/Total	0.011	0.047	0.003
Storage/Total	0.051	0.091	0.022
Other/Total	0.056	0.097	0.014

	Mean	S.d.	Median
$1 Billion-$10 Billion			
IT Spending/Revenue	0.043	0.193	0.014
IT Spending/Expenses	0.062	0.262	0.021
Communication/Total	0.064	0.078	0.042
Software/Total	0.272	0.166	0.231
Services/Total	0.361	0.152	0.336
Hardware/Total	0.165	0.117	0.140
PC/Total	0.056	0.064	0.043
Server/Total	0.065	0.063	0.050
Terminal/Total	0.007	0.021	0.004
Printer/Total	0.007	0.022	0.003
Storage/Total	0.033	0.061	0.017
Other/Total	0.036	0.072	0.012

	Mean	S.d.	Median
$10 Billion-$250 Billion			
IT Spending/Revenue	0.045	0.249	0.012
IT Spending/Expenses	0.067	0.310	0.019
Communication/Total	0.055	0.052	0.042
Software/Total	0.283	0.161	0.233
Hardware/Total	0.147	0.105	0.108
Services/Total	0.335	0.137	0.293
PC/Total	0.047	0.041	0.029
Server/Total	0.057	0.041	0.038
Terminal/Total	0.006	0.009	0.004
Printer/Total	0.005	0.010	0.003
Storage/Total	0.027	0.036	0.011
Other/Total	0.032	0.057	0.010

	Mean	S.d.	Median
> $250 Billion			
IT Spending/Revenue	0.019	0.049	0.005
IT Spending/Expenses	0.031	0.075	0.008
Communication/Total	0.046	0.041	0.031
Software/Total	0.268	0.137	0.228
Services/Total	0.357	0.149	0.328
Hardware/Total	0.158	0.103	0.138
PC/Total	0.051	0.043	0.039
Server/Total	0.062	0.044	0.050
Terminal/Total	0.007	0.011	0.004
Printer/Total	0.006	0.012	0.003
Storage/Total	0.031	0.039	0.018
Other/Total	0.036	0.061	0.012

Table A3. Small Business Loan, Mortgage Refinance and Bank IT Spending: II

This table presents regression results of banks' new mortgage issuance on the four major categories of banks' IT spending and relevant control variables at bank-county-year level. The sample period is 2010 to 2019.

$$\frac{\text{Type S IT spending}}{\text{Revenue}}_{i,c,t} = \alpha_i + \mu_{c,t} + \beta \ln(\text{refinance})_{i,c,t} \text{ or } \ln(\text{CRA})_{i,c,t} + \gamma X + \epsilon_{i,c,t}$$

ln refinance$_{i,c,t}$ is the natural logarithm of newly issued mortgage refinance of bank i at county c in year t in reported in HMDA, ln CRA$_{i,c,t}$ is the natural logarithm of small business loans issued by bank i in county c and in year t. Software(Communication)/Revenue is software (communication) spending scaled by total revenue, measured at bank-county-year level. 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; control variables are at bank-year level. Fixed effects include bank fixed effects, county × year. Standard errors are clustered at county and bank level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A

	Software/Revenue			Communication/Revenue		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(refinance)	0.027***	0.035***	0.034***	-0.002	0.003	0.002
	(0.006)	(0.007)	(0.010)	(0.008)	(0.007)	(0.007)
Revenue per Employee		-0.213***	-0.225***		-0.228***	-0.236***
		(0.014)	(0.015)		(0.011)	(0.012)
Net income/Assets		0.004	0.004		0.028***	0.033**
		(0.011)	(0.012)		(0.010)	(0.014)
Equity/Assets		0.013	0.003		-0.007	-0.012
		(0.028)	(0.027)		(0.011)	(0.012)
Deposits/Assets		0.104	0.106*		0.022	0.024
		(0.070)	(0.060)		(0.029)	(0.023)
Salaries/Assets		-1.560	-1.382		-0.899	-0.800
		(1.064)	(1.089)		(1.281)	(1.349)
Fixed effects	Bank, **County×Year**					
Bank Controls	Y	Y	Y	Y	Y	Y
AdR-squared	0.449	0.477	0.487	0.468	0.496	0.501
N	179713	176486	159732	179759	176532	159778

Panel B

	Software/Revenue			Communication/Revenue		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(CRA)	0.005	0.006*	0.006	0.014***	0.014***	0.014***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Revenue per Employee		-0.212***	-0.224***		-0.228***	-0.235***
		(0.014)	(0.015)		(0.011)	(0.012)
Net income/Assets		0.003	0.002		0.026**	0.028**
		(0.010)	(0.011)		(0.010)	(0.013)
Equity/Assets		0.011	0.002		-0.006	-0.010
		(0.028)	(0.027)		(0.011)	(0.012)
Deposits/Assets		0.097	0.101*		0.022	0.025
		(0.068)	(0.059)		(0.029)	(0.023)
Salaries/Assets		-0.702	-0.548		-0.394	-0.304
		(0.825)	(0.880)		(0.988)	(1.049)
Fixed effects	Bank, **County×Year**					
Bank Controls	Y	Y	Y	Y	Y	Y
AdR-squared	0.459	0.485	0.497	0.479	0.506	0.511
N	184314	181056	163775	184363	181105	163824

Table A4. IT Spending and IT Employment

This table presents the results of association between banks' IT investment and banks' employment, with a distinction between IT employees and non-IT employees. The regression equations are as follows:

$$\Delta \ln \text{ IT Emp or } \Delta \ln \text{Non-IT employees} = \beta \Delta \ln(\text{Software spending})_{i,c,t} \text{ or } \Delta \ln(\text{Communication spending})_{i,c,t} + \alpha_{i,t} + \mu_{c,t} + \gamma X + \epsilon_{i,c,t}$$

Δlog IT Emp and Δlog Non-IT Emp are the logarithm of IT related employees and Non-IT related employees respectively. Software/Revenue is software spending scaled by total revenue, Communication/Revenue is communication spending scaled by total revenue. The IT spending scaled by revenue is at bank-county-year level. 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; control variables are at bank-year level. Fixed effects include bank-year fixed effects and county-year fixed effects. Standard errors are clustered at county and bank level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	$\Delta \ln$ IT Emp	$\Delta \ln$ Non-IT Emp	$\Delta \ln$ IT Emp	$\Delta \ln$ Non-IT Emp
	(1)	(2)	(3)	(4)
$\Delta \ln$ Software	-0.148***	-0.117***		
	(0.051)	(0.043)		
$\Delta \ln$ Communication			0.043**	0.086***
			(0.022)	(0.022)
County×Year FE	Y	Y	Y	Y
Bank×Year FE	Y	Y	Y	Y
Control	Y	Y	Y	Y
AdR-squared	0.237	0.231	0.236	0.232
N	148617	148617	148617	148617

B Data Construction

This Online Appendix consists of three sections. In Section 1, we describe how we extract and clean IT spending data in Harte Hanks. In Section 2, we provide data-construction details on how to map banks in the IT spending data set with "Summary of Deposits" and "Call Report." In Section 3, we provide the construction details of the other supporting data sets utilized in the paper.

B.1 Construction of IT Data

In this part, we provide details on how to extract the relevant information in the original Harte Hanks IT data sets and create the panel data of banks' characteristics and IT spending at bank-county-year level. Harte Hanks collects the establishment-level (hereafter "site-level") information on IT spending and the characteristic annually. For each given year, the site-level IT spending and site characteristics are saved in two different files, "IT Spend" and "Site Description," respectively. We extract site level variables from the two files and combine them together to get the panel data of site-level IT spending and characteristics. First, for each year, from the site characteristics file, we restrict the data to one-digit SIC code equal to 6. Then we keep "site ID," company name, location (zip code), homepage url, revenue, and number of employees as our site level characteristics variables. "Site ID" is the unique identifier of the site across years in the Harte Hanks data. Second, from the IT Spend file, we get the site-level IT budget data including total budget, communication budget, software budget, services budget, hardware budget, etc, as well as the site ID. Then we merge the site characteristics and site-level IT Spending using "Site ID." Repeating the process for each year gives us a panel data set of site-level IT spending information and site level characteristics.

Next, we aggregate the number of sites, IT spending variables, revenues and employees at the zip code-year-bank level. Most sites include a url variable that labels the homepage website address of the bank. When aggregating site-level variables into county level, we first aggregate the variables by url. For those sites without a url, we aggregate by the cleaned company names. Cleaned company names are defined as the lowercase of company names after removing "national association", "n.a.", "fsb", "s.b." etc. This gives us IT spending profile and revenue and employee profile of a bank at zip code and year level.

Finally, we crosswalk zip codes to fips code (the commonly used county identifier) and aggregate all the variables at the county level, this gives us banks' IT budget and characteristics at bank-county-year level. When mapping zip codes into fips code, we noticed that some zip codes are mapped into multiple different counties. This is because some zip code areas are at the border of multiple different counties and some of the businesses or residents reside in one county while the rest of the zip code's businesses or residents are located in the other counties. For instance, zip code 49963 is mapped into both "Houghton, MI" and "Ontonagon, MI." To correctly account for the IT spending of banks located in zip codes like this into the two counties, we multiply the IT spending of a bank in this zip code with the ratios of addresses in this zip code that belongs to the two counties, before aggregating to county level IT spending. In the above example, 23% of the addresses in zip code 49663 belongs to "Houghton, MI," while 77% of the addresses in zip code 49663 belongs to "Ontonagon, MI." We multiply a bank's IT spending in zip code 49663 by 0.23 and aggregate this adjusted number to "Houghton, MI"; we multiple a bank's IT spending in zip code 49663 by 0.77 before and aggregate the adjusted number to "Ontonagon, MI." We obtain the the information on the ratio of a zip code's addresses that belong to each county for each zip code from the Office of Policy Development and Research. We use "TOTAL RATIO" provided by the Office of Policy Development and Research, which is the ratio of all types of addresses in the zip code that belongs to a county, to adjust for the spending before aggregation.[1]

B.2 Matching Bank Names in IT Data with Bank Names in Summary of Deposits

Matching at Bank-Year Level This subsection describes how do we match bank names in Summary of Deposits (hereafter SOD) and bank names in the IT data at the bank level and construct the panel data containing bank IT and bank characteristics at bank-year level.

To start with, we take the bank names from SOD data set and obtain the banks' homepage from Google. The first step is to extract a smaller set of site names in the site-level bank IT data that are similar to the names of the banks in SOD. We drop the suffixes ", national association", "national association", ", fsb", "fsb", ", n.a.", "n.a.", " f.s.b.", " f. s. b.", ", f. s. b.", ", s.b.", ", s/b", ", s.b.", ", ssb", ", s.s.b.", " (west), fsb", ", fsb", ", fsb", ", a fsb", and ", a federal savings and loan association", "bank", and "national bank", etc in the SOD data. We split the names into at most two key words by spaces. For example, Wells Fargo Bank is labeled as "Wells" and "Fargo." This is because many site names in the IT data set, which is going to be merged later, are written without spaces. In the Wells Fargo Bank case, the site names could be written as "wellsfargo bank" or "thewellsfargobank." Given that most sites in the IT data set also include a url variable that label the website address of the bank's homepage, we conduct the matching using url first, and if matching with url doesn't work, we match using keywords in names constructed above. For those sites with url, we first outer merge the names from the "Site Description" files with the url of the

[1] See the link to Office of Policy Development and Research's webpage for the relevant files.

banks' website address (after dropping “www.” and “.com”), and retain the sites whose url contains the url of the banks' website address. Then we can match sites names with SOD bank names with the url's. For those sites without a url, we outer merge the names from the “Site Description” files with the key words constructed in SOD, and only keep the sites of which the names contain all the keywords from SOD. These above procedures give us the extracted site names whose names as close to names of banks in SOD. In the last step, we assign these extracted sites names with a bank name from the names of banks in SOD that has the largest Levenshtein score. We aggregate the site level IT Spending using the matched bank names in the Bank IT data and merge with SOD through the matched bank names. This gives us the panel data of banks IT spending, total assets and deposits in SOD, matched bank names, and the bank identifier RSSDID in SOD, at bank-year level. The bank identifier RSSDID is also utilized to merge with other data sets such as “Call Reports” and HMDA, etc.

Matching at Bank-County-Year Level In this subsection, we describe how we match the banks in the IT data constructed in Section 1 with banks in Summary of Deposits (SOD) and merge the IT data with bank characteristics in SOD at county level. The output of this matching procedure generates the panel data on banks' IT spending (from the IT data) and bank assets, deposits, and bank identifier (from the SOD) at bank-county-year level.

We get total assets, year, total deposits and bank names from the SOD data set. We convert the bank names in SOD data set to its lower case. We drop the suffixes “, national association”, “national association”, “, fsb”, “fsb”, “, n.a.”, “n.a.”, “ f.s.b.”, “ f. s. b.”, “, f. s. b.”, “, s.b.”, “, s/b”, “, s.b.”, “, ssb”, “, s.s.b.”, “ (west), fsb”, “, fsb”, “, fsb”, “, a fsb”, and “, a federal savings and loan association”. We then collapse the above SOD information by zip code, bank name and year. This gives us a panel of banks in SOD with bank names, location information (zip codes), total assets, total deposits, RSSDID, and year.

To match the SOD panel data with the IT spending panel data, we first merge these two data sets using zip code and year, this gives us all the possible pairs of bank names in SOD and IT spending for each combination of zip code and year. Then for each bank name showing up in IT spending data at the zip code-year level, we calculate the Levenshtein distance of the string names between this bank name string and all the string names showing up in SOD within the same zip code and year. For the merged observations, we keep the RSSDID (the unique bank identifier) from the SOD data set with the highest Levenshtein score and we keep the observations with the calculated highest Levenshtein score larger than 2/3. This gives us a panel data of banks at zip code level that match with banks in SOD at zip code level, other variables include IT spending, bank identifier (RSSDID), total assets and total deposits.

Finally, we employ the same method described in Part I to aggregate the matched bank IT spending panel data at zip code-year level to county-year level by adjusting the ratio of addresses of a zip code that could potentially show up in multiple counties. This gives us a panel data of banks' IT spending and banks' deposits and assets, and RSSDID (bank identifier) at county-year level.

B.3 Construction of Other Data Sets

Call Report In this subsection, we describe how do we construct loan portfolio information and bank-level control variables using "Call Reports." We get the banks' balance sheet information from "Call Reports" quarterly data for the year of 2010-2019. We collapse the key variables by the last quarter of a bank within a year. The linkage between Call Reports and our IT data set is through RSSDID. We define CI loan share as the "ciloans" scaled by "qavgloans," we define personal loan share as (personal loans) "persloans" scaled by "qaveloans," and we define agriculture loan share as "agloans" scaled by "qavgloans." Banks' Profitability ("prof") is defined as net income (netinc) scaled by "qavgassets," Equity/assets is defined as "equity" scaled by "assets," Deposits/assets is defined as "dep" scaled by "assets," Salary/assets is defined as "sal" scaled by "assets," and number of employees per thousand dollar assets is defined as number of employees (nume multiplied by 1000) scaled by assets (this is because number of employees is in the unit of 1000), we define revenue per employee as income scaled by number of employees.

Home Mortgage Disclosure Act Data This subsection describes the construction of refinancing and origination amount for each bank in each year in a county.

We use the panel of "HMDA nationwide records" files to construct origination and refinance volumes. We define loan as origination if "loan purpose" is equal to 1 and define loan as refinance if loan "loan purpose" is equal to 3. We aggregate the total loan amount of each bank (identified by respondent id) at state code-county code and year level, for origination and refinance, respectively. We then construct the fips code (county identifier) by combining the state code with county code. Finally, we crosswalk respondent id to RSSDID provided in the "HMDA institution" files.

Freddie Mac Single-Family Loan-Level Data Set This subsection describe how do we construct the potential mortgage repayment savings using the Freddie Mac Single Family Loan Data Set at the county-year level.

We first use the Historical data to get the average interest rate between 2010 and 2015 at the zip code-maturity-FICO group level. Specifically for each year, we assign loans into 12 FICO bins: $<$620, ..., 780-800, 800-820, and $>$820. We then calculate the average interest rate by year, zip code, FICO group and maturity. We then use loans originated between 1999 and 2009 from the Historical Time Data to get the payment savings. Specifically, for each loan originated between 1999 and 2009, we first keep those that are not pre-paid or defaulted between 2010 and 2015, and the calculate the remain balance separately based on the loans' original interest rate and the hypothetical interest rate as the zip code-maturity-fico group average from 2010 to 2015. We then take the average payment saving by each zip code, and aggregate the data to the county level.

Mergent Intellect This subsection describes how we construct the bank hierarchical structure data using Mergent Intellect.

We download all the information of banks' family trees in Mergent Intellect with two-digit SIC code "60" and "61." We replace "Domestic Parent Name" with "Company Name" if an entity's "Domestic Parent Name" is missing. We then sum up the number of "Headquarters," number of "Single Location" and number of "Branch" offices within the "Domestic Parent Name." Banks with only one type of locations is defined as having 1 layer in their hierarchy; banks with two different types of locations is defined as having 2 layers in their hierarchy and banks with three different types of locations is defined as having 3 layers in their hierarchy.

To match the bank names in the cleaned version of Mergent Intellect as described above, we link bank names in Mergent Intellect with the institution names provided by FDIC and then link the matched results with banks in our sample. Specifically, we first remove the words "Bank", "INC", "National Association", "LLC", "CORPORATION", "COMPANY", "THE", "CORP", "SERVICES" from the names in Mergent Intellect, and unify names of entities within the Mergent Intellect, then we append cities' names where the banks are located to bank names. Next, we repeat the same process with our sample data. Finally, we merge bank names and cities in the Mergent Intellect with bank names and cities in our sample data using Jarodistance algorithm and keep the matched pairs with the highest Jarodistance score.