Bank Technology Adoption and Loan Production in the U.S Mortgage Market*

José Ignacio Cuesta, Sheila Jiang, Adam Jørring, Douglas Xu

June 2023

Abstract

Information technology plays a key role in the consumer credit market, by shaping the way lenders underwrite borrowers. We study how the adoption of information technology by lenders affects approval decisions, pricing, and repayment in the U.S mortgage market. We assemble a loan-level dataset that covers the trajectory of mortgages from application to repayment and combine it with detailed information about information technology investment by lenders. We empirically identify that higher investment in information technology leads lenders to increase approval rates for loan applications, introduce greater granularity in their pricing, and create portfolios with better ex-post performance. A simple model of investment, underwriting and pricing is developed to explain our empirical findings.

Keywords: Bank screening, mortgage market, technology adoption, financial inclusion

JEL codes: G21, D22, G20, G51

*We thank Darren Aiello (discussant), Mark Flannery, Zhiguo He, Chris Palmer, Agustin Hurtado (discussant), Paymon Khorrami, Yueran Ma, Maryam Farboodi and conference and seminar participants at SFS Cavalcade, the University of Florida, Yiran Fan Memorial Conference of Chicago Booth for their comments and suggestions. Cuesta: Department of Economics, Stanford University, jicuesta@stanford.edu; Jiang: Warrington College of Business, University of Florida, sheila.jiang@warrington.ufl.edu; Jørring: Carroll School of Management, Boston College, adam.jorring@bc.edu; Xu: Warrington College of Business, University of Florida, douglas.xu@warrigton.ufl.edu.

1 Introduction

Information technology has brought about a revolutionary transformation in the way businesses operate. The availability of extensive consumer data has paralleled this trend. These trends play a particularly important role in the consumer lending industry due to the prominence of applicant screening. In the 1990s, most of the mortgage screening was done through interviews and involved loan officers had to ask applicants to mail their W2's and pay stubs, and loan officers had to manually go through all of the inspection.[1] The 2000s saw a significant shift in the way borrower screening was conducted in the lending industry, marked by the rise of risk pricing and computerized credit scoring (Edelberg, 2006; Einav et al., 2013). On the credit supply side, the adoption of automated screening technologies also introduces a trade-off by replacing soft information with hard information, such that the overall effects of technology adoption on lender profitability are ambiguous (Stein, 2002).[2] In addition, the adoption of these technologies enables banks to tailor their screening and pricing, which likely has heterogeneous implications across the spectrum of borrower risk.

This paper aims to provide a thorough analysis of the adoption of banking technology and its impact on bank *loan production*. To accomplish this objective, we compile a novel dataset that links mortgage applications, approvals, pricing and performance with the level of IT spending by banks, serving as a measure of their investment in screening technology. To the best of our knowledge, this study is the first to construct an empirical measurement of screening technology investment and integrate it with loan decision-making and performance, thereby making a unique contribution to the existing body of literature.

To guide our empirical investigation into the impact of bank IT investment on their role as mortgage loan producers, we first develop a conceptual framework that links lender screening accuracy to their approval and pricing policies for credit applications. In the model, a lender operates under asymmetric information by making two sequential choices. First, the lender decide on his screening technology investment before receiving any credit application. Subsequently, based on the screening accuracy, the lender make decisions regarding loan approval and pricing for the received applications.

When a borrower submits a loan application with an observable risk level, which influ-

[1] As documented in Straka (2000), traditional labor and paper-intensive mortgage risk assessment typically takes weeks and even months.

[2] As technology develops, mortgage lenders are also increasingly obtaining and embedding various forms of alternative data of applicants to their mortgage originating models. For instance, as introduced in the FT article, mortgage lenders purchase software like Arc Systems and further develop their own software to risk-assess mortgage applicants. More recently, as described in "Business Insider," lenders are also actively embedding alternative data to better assess the credit worthiness of the credit "invisible" and likely getting a higher return from lending (Di Maggio et al., 2022).

ences the lender's prior belief about the borrower's type, the lender could utilize his screening technology to form an updated (posterior) belief about the borrower's true type. Depending on the borrower's observable risk level, the lender chooses to either desk-reject the application without further screening or incur some flow costs, such as labor costs, to conduct screening using the adopted technology. In the latter case, based on the screening results, the lender provides an interest rate quote to the borrower, who then decides whether to accept the offer.

Using this framework, we establish testable hypotheses on how bank IT investment, influencing borrower screening accuracy, can impact mortgage origination decisions. Specifically, we demonstrate that lenders follow a cutoff strategy in approving applications. In equilibrium, lenders reject applications with observable risk level below certain threshold, while conducting screening and providing an interest rate quote for applications above that threshold. Importantly, our analysis reveals that lenders with more accurate screening technology set a lower threshold for their approval decisions, leading them to approve applications that lenders with less accurate screening technologies would reject.

Taking our analysis a step further in the mortgage origination process, we demonstrate that the increased inclination of lenders to approve applications from marginal borrowers is a direct result of their enhanced ability to devise more precise pricing strategies through improved screening accuracy. This capability enables these lenders to extract sufficient surplus from these borrowers, thereby rendering the net payoff from screening to be positive. Building upon this intuition, our model predicts that given an approved application, lenders with superior screening accuracy are more likely to have their offers accepted (by borrowers), consequently leading to a higher conversion rate for loan origination from these lenders.

Furthermore, our analysis suggests that more accurate screening technology leads to increased dispersion in interest rates on originated loans. Our model unveils two forces that account for this relationship between screening technology and pricing granularity. On the extensive margin, lenders with enhanced screening accuracy exhibit a greater willingness to grant approval to applications from marginal borrowers with relatively high observable risk, expanding the range of borrowers they extend loans to compared to less accurate lenders. On the intensive margin, superior screening accuracy enables lenders to devise more granular prices for inframarginal borrowers with similar observable risk, allowing them to extract more surplus. Relatedly, our model predicts that lenders with higher level of technology adoption tend to have a lower delinquency rate on their originated loans, despite their higher willingness to extend loans to riskier borrowers.

To conduction our empirical analysis, we construct a novel data set linking loan production trajectory and the originating banks' IT expenditure as a measurement of investment in screening technology. Specifically, we combine HMDA, which contains a rich set of borrower and loan characteristics at origination, with GSE loan performance data, which contains post-origination loan performance and borrowers' credit profile. Since the GSE loan performance dataset does not have information of the originating banks, we first match the two datasets based on a wide range of borrower, loan and property characteristics. We then match the borrowers in HMDA with lenders' IT expenditure in the Harte Hanks Market Intelligence Computer Intelligence Dataase (CiTDB).

Our empirical analysis begins by examining factors influencing bank IT spending. We investigate a diverse range of variables to discern their correlation with banks' technology adoption decisions. Regarding local credit market characteristics, we find empirical evidences supporting the use of IT spending as informative measure for the required effort or investment in bank screening. Higher average FICO scores of past local mortgage borrowers are associated with lower IT spending by local banks, while the debt-to-income ratio of previous borrowers and the loan delinquency rate are positively linked to IT spending. Additionally, higher proportions of loans being successfully securitized and of loans meeting the conforming loan limit are associated with lower IT spending by local banks.[3] We also find positive correlations between local economic conditions, such as house price level, establishment growth, and the share of population with college degrees, and lenders' IT expenditure. Conversely, unemployment and the share of population aged 65 and above are negatively associated with IT spending.[4] Furthermore, our findings suggest that higher quality technology infrastructure, including broadband service coverage and faster speeds, or shorter distances to warehouses of major tech product producers, leads to increased IT expenditure by lenders. [5] Lastly, lenders' profitability, deposit market share, and mortgage market share in the local market are positively associated with IT spending.[6]

A main objective of our analysis is to empirically identify the impact of lenders' technology adoption, as measured by their IT spending, on their loan production. To address endogeneity concerns, we employ two instrumental variables exploiting non-local factors that drive changes in banks' IT spending. The first instrumental variable leverages variations in banks' exposure to housing price growth at the nationwide level. We utilize this bank-level

[3] These findings are consistent with previous works on securitization and bank screening incentive (Keys et al. (2009), Keys et al. (2010), and Choi and Kim (2021)).

[4] The findings are qualitatively consistent with recent works documenting banking digitization and population profile (Haendler (2022), Jiang et al. (2022b)).

[5] These results align with the recent findings reported in D'Andrea et al. (2021) and underscore the importance of robust technology infrastructure for banks' investment in IT.

[6] These findings are consistent with previous researches studying market structure and technology adoption or innovation (Hannan and McDowell (1984)) and the more recent work by Yannelis and Zhang (2021) who study the U.S. auto loan market.

exposure as an exogenous source of variation that affects the IT adoption of multi-market banks' branches operating in counties with relatively low housing price growth. Additionally, we construct a second instrumental variable exploiting the geographic exogeneity in local bank branches' distance to major IT equipment suppliers' warehouse locations. The variation extracted by both instrumental variables are likely to be independent of local borrower characteristics or local economic conditions.

We formally investigate the impact of bank IT spending on the full trajectory of mortgage loan production in both OLS and 2SLS analysis, incorporating a rich set of control variables and fixed effects. We begin by examining how banks' IT spending influences decision-making at the credit application stage. Overall, we find that banks investing more in IT tend to make more inclusive decisions. Specifically, a one standard deviation increase in a banks' total IT spending in a county, driven by higher bank-level exposure to high housing price in other counties, leads to a 8.62 percentage points decrease in the loan application rejection rates, which is statistically significant after controlling for a wide range of borrower and loan characteristics. Additionally, there is also a 5.20 percentage points decrease in the probability of loan offers being rejected by borrowers, which together leads to a 13.82 percentage points higher likelihood of a loan application being converted into a loan origination.

To investigate the potential heterogeneity in the impact of banks' IT adoption on different borrower groups, we construct a least absolute shrinkage and selection operator (LASSO) estimator of borrowers' default probability as a measure their observable risk level.[7] Our analysis reveals some intriguing patterns. When confronted with borrowers who exhibit relatively high observable risk, banks with higher IT spending are significantly less inclined to reject loan applications compared to those with lower IT spending. This difference becomes much more modest for credit applications from relatively safe borrowers.[8] Conversely, while higher bank IT spending reduces the instances of "approval but denial" for both borrower groups, the effect is more pronounced on applications sent by borrowers who exhibit relatively low observable default risk. These findings underscore the importance of considering the interaction between IT spending and borrower risk characteristics in assessing the impact of technology adoption on banks' loan production.

We then delve into the pricing patterns of originated loans, which is another crucial aspect of banks' loan production. We find that higher level of technology adaptation by

[7]Recent literature utilizing similar LASSO estimator construction approach include Crawford et al. (2018), Ellison and Pathak (2021) and Braghieri et al. (2022).

[8]We also find that banks' technology adoption has a larger impact on affecting rejection rate of the credit application from low-income and minority borrowers. Relatedly, high spending banks have a significantly higher conversion rate in getting the credit applications from low-income and minority borrowers into actual origination.

banks, driven by higher bank-level exposure to high housing price markets, leads to a larger dispersion in the distribution of interest rate charged on originated loans. In particular, a one standard deviation increase in a bank's total IT spending in a county leads to 0.155 percentage points higher interest rate dispersion (standard deviation of interest rates offered by a bank in a county), which means a 32% increase in the interest rate dispersion. Two underlying factors that drive this impact of banks' IT investment on the pricing dispersion, an *extensive margin* effect and an *intensive margin* effect, are further uncovered by our analysis.

Our empirical findings on banks' decision-making for application approval imply an extensive margin effect. Specifically, we observe that banks with higher levels of technology adoption display a greater willingness to accept observably riskier borrowers. Consequently, the variation in banks' IT spending encompasses a compositional effect, leading to loans originated by banks with higher IT investment being distributed among borrowers with a wider range of observable risk profiles. On the intensive margin, our analysis reveals that, on average, banks with high IT spending offer mortgage rates that are 1.1 basis points lower compared to banks with low IT spending.[9] However, when faced with loan applicants in the high-risk segment based on observable characteristics, high IT spending banks offer mortgage rates that are 3.27 basis points higher compared to banks with lower IT spending. On the other hand, these banks offer mortgage rates that are 1.7 basis points lower to observably low-risk borrowers compared with their low spending counterparts.[10] These findings suggest that banks that invest more in IT have a more granular pricing strategy compared to their counterparts with lower spending when dealing with similar distributions of loan applicants. Our findings of banks' loan decisions and pricing echo the findings in Maggio et al. (2021), Jagtiani and Lemieux (2019), and Cornelli et al. (2022).

Finally, we investigate the impact of IT investment on the ex-post performance of originated loans, shedding light on the efficacy of banks' ex-ante decision-making. In our 2SLS regression analysis, we observe that a one standard deviation increase in a bank's total IT spending in a county, driven by greater bank-level exposure to housing prices nationwide, leads to a statistically significant decrease of 11.1 percentage points in the 2-year delinquency rate and a reduction of 12.3 percentage points in the 4-year delinquency rate for newly originated loans. Notably, this effect is particularly pronounced among the high-risk borrower group identified based on observable characteristics. These findings suggest that banks with

[9] A bank is defined as a "high IT spender" in a specific year if the bank's IT spending scaled by total revenue is above median among all banks in the same county in that year.

[10] In our analysis on the intensive margin effect of bank technology adoption, we control a rich set of borrower characteristic to ensure that the comparison is made between loans extended by banks with varying IT spending towards a similar distribution of borrowers.

higher IT spending are more effective in identifying borrowers with relatively high observable risk, allowing them to select borrowers with lower latent default risk and thus have a lower delinquency rate on originated loans despite their higher willingness of extending loans to observably risky borrowers.

As a more comprehensive evaluation of banks' ex-ante pricing decisions on their approved loan applications, we also examine the ex-ante pricing on ex-post defaulted loans. By comparing the pricing patterns of delinquent loans originated by banks with varying levels of IT investment, we find that banks with higher IT spending are able to "correctly" charge higher interest rates. This suggests that a higher level of IT investment, which likely enhances banks' screening accuracy and risk assessment capabilities, enables them to effectively identify potential borrower risks and adjust interest rates accordingly to mitigate those risks.

Related literature. Our work contributes to several strands of literature. First, our paper supplement the theoretical literature studying screening under asymmetric information. Broecker (1990) develops a framework to study lending market outcomes with asymmetric information where lenders have symmetric screening ability and compete. Hauswald and Marquez (2003) study the competition between an inside bank who can conduct credit screening and an outside bank who has no access to screening and shed light on the impact of information technology on lending market.[11] Ruckes (2004) links banks' screening effort to economic prospect and show how the interaction between screening effort and borrower entrance explains the lending standards and lending outcomes over economic cycles. More recently, Yannelis and Zhang (2021) considers a framework with imperfect competition among auto lenders.[12] This large theoretical literature that involves explicitly modeling banks' effort and investment in screening also include but do not limit to Thakor (1996), Cao and Shi (2001), Manove et al. (2001), Hauswald and Marquez (2006), Vanasco (2017), Vives and Ye (2021), He et al. (2022), etc. Our work contributes to this literature by providing a concrete empirical measure for banks' investment in screening technology—a crucial component in these theoretical works yet remains difficult to be empirically tested—and makes causal linkages between the screening effort and loan making behavior. To the best of our knowledge, we are the first to directly measure banks' screening input into their loan production and link it to the loan production output.

Our paper also contributes to a large literature studying the economic impact of technology adoption. In the context of banking sector, Berger (2003) provides evidences that

[11] Relatedly, Marquez (2002) studies how differences in banks' screening intensity results in higher dispersion in the credit market through competition among banks.

[12] The empirical literature on competition among mortgage lenders include Scharfstein and Sunderam (2016) and Buchak and Jørring (2021) who study the effect of competition on interest rates and fees, respectively.

progress in both information technology and financial technology led to improvement in banking services. More related to our study, recent lending technology innovation in the banking industry stimulated researches studying how adoption of specific technology has affected loan making behavior. For instance, exploring the automated credit scoring by a large auto finance company targeting low-income and high-risk borrowers, Einav et al. (2013) find that the adoption of automated credit scoring enabled the lender to better "screen out" high-risk borrowers. More broadly, our work is related to the burgeoning literature investigating how specific types of technology or alternative data reshape lending (e.g., Berg et al. (2019), Agarwal et al. (2019), Blattner and Nelson (2021), Fuster et al. (2022), Bartlett et al. (2019), Maggio et al. (2021), and Jiang et al. (2022a).)[13] By studying a more broadly and comprehensively defined technology adoption measure captured by banks' IT spending and linking it to banks' lending outcomes in the U.S. mortgage market, our analysis complements this literature by providing the first piece of systematic and quantitative evidence for the impact of technology adoption on banks' loan production.

Under the backdrop of recent financial technology development, our work is also related to a growing literature studying the loan-making behavior of Fin-tech lenders, which are practically be viewed as a special group of "tech-savvy" lenders. Fuster et al. (2019) document that Fin-tech lenders process applications 20% faster than traditional bank lender without significantly lifting delinquency rates. Di Maggio and Yao (2020) document that borrowers accepted by Fin-tech lenders tend to have higher default probabilities although they receive a short-lived reduction in cost of credit. Related to our findings, using loan-level data from a Fin-tech lender, Maggio et al. (2021) find that usage of alternative data and algorithmic underwriting to assess borrowers' creditworthiness can result in broader credit access.[14] Our analysis shed light on how the loan production of traditional banking sector, which still plays a predominant role in the credit market, is affected by development and adoption of financial technologies.

Finally, our paper also contribute to a recent literature investigating the determinants of banking technology adoption. Hannan and McDowell (1984) document that market struc-

[13] Berg et al. (2019) show that digital footprint left by borrowers that are easily accessible complements the traditional credit scoring system. Blattner and Nelson (2021) demonstrate that lenders face more uncertainty when assessing default risk of historically under-served borrower groups, especially using the credit scores. Fuster et al. (2022) show that Black and Hispanic borrowers are disproportionately less likely to gain from if banks were to introduce machine learning method to make loan decision rather than traditional credit scoring model. Bartlett et al. (2019) find that algorithmic scoring reduces price discrimination in the lending market. Jiang et al. (2022a) find that the use of machine limits the incorporation of same-race loan officers' soft information. Agarwal et al. (2019) find that mobile footprint outperforms traditional credit scoring in predicting loan approval decision and default probability.

[14] Other related works include Buchak et al. (2018), Cornelli et al. (2021), Hertzberg et al. (2018), Croux et al. (2020), and Jørring (2020).

ture has important impact on the technology adoption in banking sector. D'Andrea et al. (2021) find that banks in areas reached by fast broadband internet increase loan supply, and reduce credit price. Jiang et al. (2022b) show that entrance of 3G into a local area intensified local banking market competition and enabled banks catering non-digital depositors to gain market power. Lin et al. (2021) find that introduction of telegraph significantly expanded banks' branch networks and facilitated the formation of modern banking system in China. He et al. (2021) show that credit demand shocks of different information nature drive banks to invest in different lending technologies to deal with hard information or soft information. Lerner et al. (2021) documents the tightness of local financial regulation affect the geographical allocation of banks' patenting activities. Recent related works include but do not confine to Frost (2020), Ridder (2021), Glode and Ordoñez (2022). Our work contribute to the existing literature by demonstrating that as an investment to sharpen loan screening, how do a wide range of factors, in particular local borrower credit profiles, could shape banks' IT investment.

The rest of the paper is organized as follows. Section 2 provides a detailed description of the combined data sets utilized for our empirical analysis. In Section 3, we develop a simple theoretical framework to guide our empirical analysis. Section 4 explores economic factors potentially influencing lenders' IT spending at the local level. As the main part of our analysis, Section 5 presents the empirical set up and identifies the impact of technology adoption on lenders' loan production. Section 6 concludes the paper.

2 Data and Sample Construction

2.1 Lender IT Spending and Balance Sheets

We measure U.S. mortgage lenders' IT spending using the Harte Hanks Market Intelligence Computer Intelligence Technology data, which cover over three million establishment-level observations from 2010 to 2020. Harte Hanks collects and sells this information to technology companies, which then use it for marketing purposes or to better serve their clients. These data include information on fixed IT capital including computers and communication systems, annual IT spending in software and hardware, and number of employees. All these variables are reported at the branch level, and the data include lender identifiers as well as branch location.

In Panel B of Table 1, we show summary statistics of balance sheet information of the bank lenders in our sample. Panel B shows bank spending and balance sheet information at the annual level. The annual balance sheet information of banks come from Call Report data. Panel C shows summary statistics of lenders' IT spending, lenders' local revenue and number of employees at the lender-county-year level. One noticeable feature of banks in our sample is such that the real estate loans accounts for more than 77% of total loans of banks in our sample, this highlights the importance of mortgage loans of the banking sector, and this also highlights that banks investment into technology should be plausibly dedicated to their main assets, which is mortgage lending.[15] In addition to total IT spending, we also separately use communication spending and software spending to capture banks' technology adoption. Communication spending can be utilized to measure banks' investment in physical communication devices, which facilitate banks to generate soft information through interacting with borrower and transmit the soft information within banks, while software spending captures banks' investment in technology such as data processing, storage, alternative data, etc.

2.2 Mortgage Origination, Pricing and Performance

To measure loan applications, pricing, and repayment in the U.S. mortgage market, we combine three sources of administrative data. First, we use data collected under the Home Mortgage Disclosure Act (HMDA), which covers close to the universe of mortgage applications in the country. These comprehensive data include lender identifiers, application outcomes, property attributes, loan attributes, and borrower attributes and location. Starting from 2018, the data include a particularly rich set of loan characteristics, including interest rates, loan-to-value (LTV) and debt-to-income ratio (DTI). Secondly, we incorporate data on loan performance from Fannie Mae and Freddie Mac, which cover the portfolios that these GSEs hold of 30-year, single-family, conforming, fixed-rate mortgages and provide similar information about the lender, the mortgage and the borrower than the HMDA data, along with performance metrics over time. Finally, we integrate data from CoreLogic, which extends coverage to non-conforming loans. In fact, for our analysis, we focus solely on non-conforming loans within the CoreLogic dataset, while utilizing the data from Fannie Mae and Freddie Mac as the primary source for conforming loans. The CoreLogic data also include similar information about the lender, the mortgage and the borrower than the HMDA data.

To conduct our analysis, we match the HMDA data on applications and approvals to the

[15] A research conducted by StartUs investigated the products of 1926 banking technology startups around the world and uncovered top 10 categories of technology trends such as Artificial Intelligence (AI), Open Banking, Banking of the Thing, and Quantum Computing, etc. Many of the top categories of the top technology trends are related to the automated service processing, data collection, credit score modeling, and computing etc. As another example, a recent report conducted by Mckinsey & Company provides in-depth analysis of how AI and its related technologies being adopted by banking sectors could tremendously transform the credit decision making process.

performance data from the GSEs and CoreLogic. In all datasets, we restrict our attention to 30-year mortgages originated for purchase of primary residence homes. We match originated loans in HMDA to those in the GSEs and CoreLogic data by matching exactly—up to rounding whenever the datasets report rounded variables—on geography, loan amount and interest rate (when available), conforming status and purchaser type. We exclude from our sample all originations that have more than one unique match across datasets.

Our final panel dataset includes both approved and rejected loan applications, as well as information on the performance of approved loans. Panel A of Table 1 shows the summary statistics for 30% of a randomly generated subsample of our whole dataset, which is also the sample utilized in the empirical analysis of this paper. The time frame for our analysis spans from 2010 to 2019, encompassing a total of 10,430,632 loan applications and 2,380,472 originations issued by 3,653 lenders.[16] The average rejection rate in our sample is 0.55, the overall "approval but denied by applicant" rate is 0.15 and the overall loan origination rate is 0.29. The median loan size in our sample is $0.213 million, the median loan-to-value (LTV) ratio is 82.6, the median FICO score of borrowers is 762, and the average 2-year delinquency rate of borrowers is 5.1%.

2.3 Other Data Sources

Local technology infrastructure. We construct two sets of economic variables to capture the local technology infrastructure. The first set of variables are the internet coverage and speed, which we think of as a shifter to the return to IT spending. The second variable is the distance of a bank branch to the main warehouses of IT product producers. we think of this variable as a shifter to the cost of IT investment.

To explore local broadband speed, which captures infrastructure for data transmission, storage, and exchange, we use the "Fixed Broadband Deployment Data" from FCC, which contains all the information on the services and the corresponding speed filed by fixed broadband service providers in census blocks where they offer services.[17] From these data, we get the maximum advertised downstream and upstream speeds offered by local fixed broadband providers, and take the average among of the downstream and upstream speeds of all service providers. The summary statistics of these data are provided in Table A1.

To construct measures of distance from bank branches to major technology product

[16] We restrict our samples to lenders with at least 100 annual applications per year.

[17] The coverage and speed of broadband internet services in a county comes from the Federal Communications Commission (FCC). All facilities-based broadband providers are required to file through Form 477 to the FCC twice a year on where they offer internet access service at speeds exceeding 200 kbps. The FCC organizes the filings and makes the information publicly available on their website.See the link for a detailed explanation.

suppliers, we utilize information of major hardware producers based in U.S. To do this, we take all the listed companies in Compustat with the two-digit SIC code associated with computer equipment, and focus on the six largest companies in the industry between 2010 and 2019. We then locate the distributing warehouses' locations of these producers in U.S, and calculate the distance between each bank branch in our data and each of these distribution centers.[18] Table A2 provides a summary of the six companies and FIPS codes of their warehouse locations. In Table A1, we provide the summary statistics of the average log of the distances between a bank and these firms' warehouses.

Local House Price. We obtain the county-level house price index and annual growth of house price index between 2010 and 2020 from Federal Housing Finance Agency (FHFA). In our 2SLS regression analysis, we construct banks' exposure to house price growth that could potentially drive up the demand of non-conforming loan in the nation-wide geographic areas, thus drive up banks' IT expenditure to deal conduct more careful screening. We postpone this part of analysis to Section 5.2.2. The summary statistics of HPI is provided in Table A1.

Other County level economic characteristics. We incorporate other county-level economic and demographic variables from multiple sources. We obtain county-level population and GDP growth data from the U.S. Census Bureau; local business establishment growth data from County Business Pattern; and data on local population age profile and education attainment profile are from the American Community Survey. Moreover, we construct local bank deposit HHI and bank deposit market shares using information from the Summary of Deposits.

3 A Model of Credit Screening, Approvals and Pricing

This section aims to provide a conceptual framework that serves as a guide for our empirical analysis. The framework focuses on modeling the adoption of technology by creditors and its impact on each stage of the credit process, including credit application, screening, pricing, and the ultimate origination of credit.

3.1 Model Setup

Players and information structure. Consider a setting in which lenders make loans to borrowers under asymmetric information. Borrowers have private information about their

[18]Note that distribution centers are typically located at a different city from their headquarters. For instance, HP Inc is headquartered in Palo Alto, California, while its warehouse is in Des Moines, Iowa.

type. To simplify the analysis, let the borrowers be of one of two types, namely a good type G who will never default, or a bad type B borrower, who always defaults. Let the unconditional probability of a borrower being of type G be p, so that the probability of being of type B is $1 - p$. Practically, one can think of these probabilities as the prior belief the lender has on the borrower. Such prior belief, in turn, can be formed based on certain observable risk factors and is common knowledge among all lenders. Without loss of generality, assume credit applicants' risk profile follows certain distribution on $[\underline{p}, 1]$.

Bank screening. While the true type of a borrower cannot be perfectly observed by the lender, the lender can get a noisy signal $s \in \{b, g\}$ by screening the borrower. Specifically, we assume the following screening technology

$$\Pr(g|G) = \Pr(b|B) = \theta; \quad \Pr(b|G) = \Pr(g|B) = 1 - \theta$$

such that screening of either types of borrower generates correct signals with probability θ and mistakenly generates positive signal with probability $1 - \theta$. Therefore, when facing a borrower with observable risk profile p (which is the lender's prior belief of the borrower's type), the posterior belief given a signal g is

$$\pi_g(p; \theta) \equiv \Pr(G|g) = \frac{p\theta}{p\theta + (1-p)(1-\theta)},$$

and the posterior belief given a signal b is

$$\pi_b(p; \theta) \equiv \Pr(G|b) = \frac{p(1-\theta)}{p(1-\theta) + (1-p)\theta}.$$

Screening is costly, and the lender must incur a cost $\tau > 0$ to screen a borrower. The accuracy parameter $\theta \in [\frac{1}{2}, \bar{\theta}]$ is determined by the lender's investment in screening technology, which involves a convex cost function $c(\theta)$. Under this specification of the screening technology, one can think of the accuracy parameter θ as the degree of information technology adoption by lenders.[19]

Loan application process. We model the process for each loan application towards the potential final origination as depicted in Figure 1. First, a borrower with prior/observable risk profile p sends her loan application to the lender. Having received the loan application,

[19]As a real-world mapping, one can think of the ex-ante investment decision as the decision about how much to invest on IT input such as computers, communication devices, software, and data storage services. The flow-cost can be thought of as whether the loan officer of the lender, given the ex-ante investment into IT services, is going to spend another couple of hours going through the processing using the IT that has been invested.

the lender who is equipped with a screening accuracy θ then makes a binary decision on whether to screen the borrower, or to directly reject the application without doing further screening—in which case the credit application fails.

If the lender chooses to screen the borrower, which requires him to incur the cost $\tau > 0$, a signal $s \in \{g, b\}$ is generated by the lender's screening as specified above. Based on the signal s received, the lender then optimally chooses an interest rate r_s and makes a take-it-or-leave-it offer to the loan applicant. In the final stage, the borrower decides whether to accept the offer (with interest rate r_s) proposed by the lender—the credit application successfully turns into loan origination if she accepts the offer, while the credit application fails if she rejects the offer.

To simplify the analysis, we assume that the lender makes a profit equal to the interest rate spread r if the borrower does not default, and that the lender suffers a loss $l > 0$ if the borrower defaults.[20] Furthermore, we assume that each borrower has a reservation rate such that she will only accept rates below such threshold. Specifically, we assume that the distribution of this reservation interest rate is the same for both borrower types, wit CDF being $F(r)$ and the density function is $f(r) = F'(r)$. Therefore, the borrower accepts an offer at interest rate r with probability $F(r)$. For illustration purpose and algebraic simplicity, in what follows we assume functional form $F(r) = r$ for $r \in [0, 1]$ and $F(r) = 1$ for $r \geq 1$.[21] Furthermore, we impose the following parametric assumption to avoid the complication of potential corner case analysis.[22]

Assumption 1. *Model parameters satisfy:* $\frac{(1-\underline{p})\bar{\theta}l}{\underline{p}(1-\bar{\theta})} \leq 1$

Finally, we assume that each lender receives numerous loan applications drawn from the same pool of borrowers. Having received these applications, lenders then make their decisions over each individual application independently.[23]

3.2 Equilibrium Decision Making

A lender in this model makes two choices in a sequential way. First, a lender makes an investment decision to determine his screening technology θ, before receiving any credit

[20] The interest rate spread r can be thought of as the interest rate net of the lender's funding cost.

[21] In other words, it is assumed that the borrowers' reservation interest rate r has a uniform distribution on $[0, 1]$. More generally, as will be shown in later analysis (details provided in Appendix), a sufficient condition that guarantees the mixed effect of screening accuracy on pricing policy as stated in Proposition 3 is that the hazard rate $h(r) \equiv \frac{f(r)}{1-F(r)}$ of function $F(r)$ is increasing in r.

[22] As will be shown later, this assumption allows us to focus on interior cases where in equilibrium lenders always offer interest rate below 1, which is the upper bound of borrowers' reservation interest rate distribution.

[23] One can think of each lender in the model as having deep pockets and unlimited in their credit supply, which guarantees that the credit decisions are independent across applications.

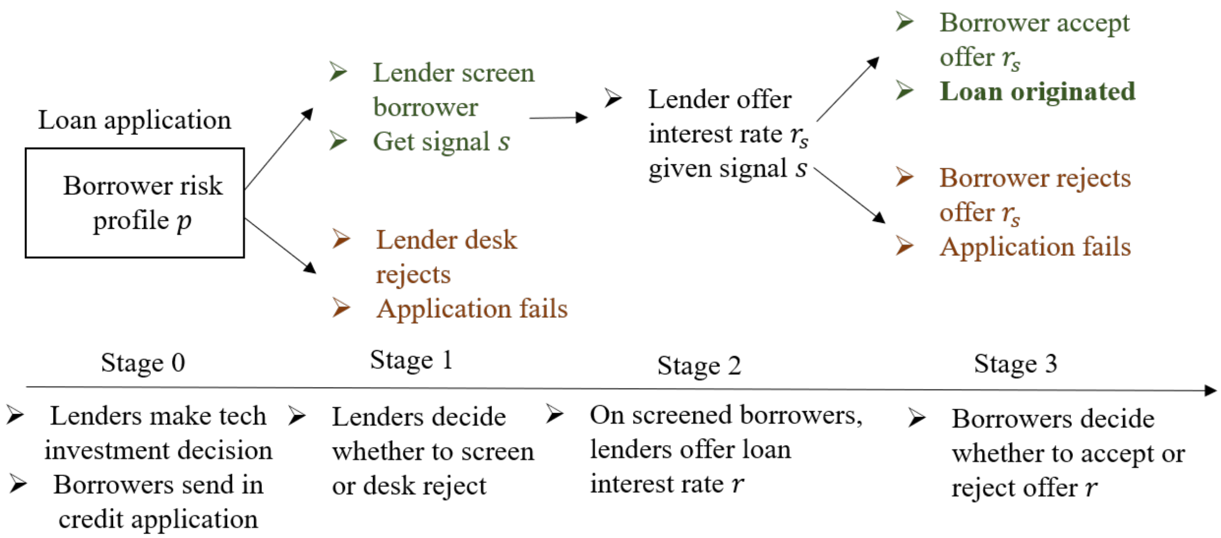


Figure 1. Procedure of credit application, screening, and pricing

applications from borrowers. Second, and given his screening technology θ, the lender then makes his approval/reject and potentially pricing decisions after receiving credit applications, the process of which is described as above. Below, we characterize optimal lender choices in a backward manner. That is, we first solve the lender's approval and pricing problem after receiving an application from a borrower with risk profile p, given a technology θ. Then, we go back and solve for the lender's optimal ex-ante choice of investment in screening technology.

Application approval and pricing. When facing a borrower with risk profile p, a lender with screening technology θ forms a posterior belief $\pi_g(p;\theta)$ if a g signal is obtained. With this belief, the expected payoff to the lender from approving the application and offering a net interest rate r that is taken by the borrower is

$$\Pi_g(r;p,\theta) \equiv \pi_g(p;\theta)r + [1-\pi_g(p;\theta)]\,(-l),$$

where r is the profit the lender obtains if the borrower does not default—which happens with certainty if the borrower is a G type and with probability λ if the borrower is a B type—and l is the net loss to the lender when the borrower defaults—which only happens (with probability $1-\lambda$) when the borrower is a B type. In contrast, if the interest rate offered by the lender is not taken by the applicant, the lender's payoff is simply zero.

Similarly, if a b signal is received after screening, the lender will accordingly form his posterior belief $\pi_b(p;\theta)$. With this posterior belief, the lender's expected payoff from approving the application and offering an interest rate r that is taken by the borrower is

$$\Pi_b(r;p,\theta) \equiv \pi_b(p;\theta)r + [1-\pi_b(p;\theta)](-l)$$

whereas the lender's payoff is zero if the offer is rejected by the borrower.

Once the lender has decided to approve (instead of desk reject) the loan application and screen the borrower, he then makes an optimal choice of the net interest rate r_s to charge contingent on the signal $s \in \{g,b\}$ generated from screening. In making this choice, a lender with technology θ that receives a signal $s \in \{g,b\}$ on a borrower with risk profile p solves

$$V_s(p;\theta) \equiv \max_r \underbrace{(1-F(r))\Pi_s(r;p,\theta)}_{\substack{\text{Borrower}\\ \text{accepts}}} + \underbrace{F(r)\cdot 0}_{\substack{\text{Borrower}\\ \text{rejects}}}, \tag{1}$$

where the first term represents the lender's payoff when the borrower takes the lender offer, which happens with probability $1-F(r)$ if the lender offers an interest rate r; and the second term stands for the lender's payoff when his offer is rejected by the borrower, which happens with probability $F(r)$. The value function, $V_s(p;\theta)$, is thus defined as the optimized expected payoff to a lender with technology θ when getting a signal $s \in \{g,b\}$ on a borrower with risk profile p.

Putting the analysis above together, the value function of a lender with technology θ when facing a borrower with risk profile p, $V(p;\theta)$, can be written as

$$V(p;\theta) \equiv \max\Big\{ \underbrace{\ 0\ }_{\substack{\text{Lender}\\ \text{rejects}}}, \underbrace{\sum_{s=g,b} \Pr(s;p,\theta)V_s(p;\theta) - \tau}_{\substack{\text{Approve and quote } r_s\\ \text{contingent on signal } s}} \Big\}, \tag{2}$$

where the probabilities $\Pr(s;p,\theta)$ of receiving signal $s \in \{g,b\}$ conditional on screening are

$$\Pr(g;p,\theta) \equiv p\theta + (1-p)(1-\theta);\ \Pr(b;p,\theta) \equiv p(1-\theta) + (1-p)\theta.$$

Screening technology investment. Having characterized the lender's approval and pricing behavior, we now describe how the lender decides to invest in screening technology θ. In particular, the lender solves

$$\max_\theta E_p[V(p;\theta)] - c(\theta),$$

where $E_p[V(p;\theta)]$ is the expected value of an application for a lender with screening technology θ; and $c(\theta)$ is the investment cost associated with improving screening accuracy, for which we assume that $c'(\cdot) > 0$ and $c''(\cdot) > 0$. Furthermore, for our later analysis of banks' ex-ante investment in θ, we assume $c(\theta)$ is sufficiently convex on $\theta \in [\frac{1}{2}, \bar{\theta}]$ and satisfies Inada condition $\lim_{\theta \to \frac{1}{2}} c'(\theta) = 0$.

Lender's equilibrium decisions. We now describe the optimal behavior of the lender in terms of screening technology investment, application approval, and pricing.

In terms of pricing, the problem consists of offering a rate that maximizes expected profits in Equation (1), in which the lender seeks to maximize "consumer surplus" under information asymmetry as a producer of loans. For a lender with screening accuracy θ that approves an application from a borrower with risk profile p and received a g signal from screening, then the optimal interest rate $r_g(p;\theta)$ solves

$$[1 - F(r)]\,[\pi_g(p;\theta) + \lambda(1 - \pi_g(p;\theta))] - f(r)\Pi_g(r;p,\theta) = 0,$$

whereas if the same lender receives a b signal from screening such a borrower, then the optimal interest rate $r_b(p;\theta)$ solves

$$[1 - F(r)]\,\lambda - f(r)\Pi_b(r;p,\theta) = 0.$$

Regarding the approval decision upon receiving a loan application, a lender with screening accuracy θ chooses to reject an application from a borrower with risk profile p if and only if

$$\sum_{s=g,b} \Pr(s;p,\theta)V_s(p;\theta) - \tau < 0,$$

where $V_s(p;\theta)$ is the lender's optimized payoff from pricing policy $r_s(p;\theta)$ given a signal $s \in \{g, b\}$. This condition implies that the expected profits from screening the borrower and processing her application—which is the maximum "consumer surplus" can be extracted from this borrower under information asymmetry—must exceed the screening cost for the application not to be rejected upfront.

Finally, the optimal investment in screening technology solves

$$\frac{\partial E_p[V(p;\theta)]}{\partial \theta} - c'(\theta) = 0,$$

such that the lender invests until the marginal return to screening accuracy meets its marginal cost.

3.3 Model Implications and Testable Predictions

This section discusses key implications derived from the model, and based on these model implications several testable hypotheses are developed, which we empirically investigate in subsequent analysis.

Let us begin with lender's ex-ante technology adoption decision, which is characterized by the following proposition.

Proposition 1. *In equilibrium, lender's choice of screening accuracy θ_E increases when loss l associated with borrower default is larger, i.e., $\frac{\partial\theta_E}{\partial l} > 0$.*

Intuitively, lenders' incentive of investing in information technology that improves screening accuracy is higher when the marginal value of information is relatively high—which is the case when lender suffers larger loss in the situation of borrower default. Practically, this maps into the situation where the lender is operating in environments featuring relatively low conforming loan ratio, such that the lenders' own exposure to loan default risk is larger due to difficulty in selling the originated loans.[24] Accordingly, the following hypothesis is developed for our later empirical analysis.

Testable Hypothesis 1. ***Ex-ante technology adoption***
Lenders operating in economies with lower conforming loan ratio or worse borrower pool spend more on IT.

Next, upon receiving the credit application from a borrower with risk profile p, each lender follows a cutoff strategy in determining whether to reject the application directly without further screening. More specifically, the following proposition describes how the rejection policy of a lender is affected by his screening accuracy θ.

Proposition 2. *In equilibrium, a lender with screening accuracy θ adopts a cutoff strategy $\hat{p}(\theta)$ in making approval/reject decision, in which the lender will screen the credit application if and only if the borrower has a risk profile p above threshold $\hat{p}(\theta)$. Furthermore, function $\hat{p}(\theta)$ is decreasing in θ.*

Based on this proposition, we develop the following hypothesis which we empirically test in our subsequent analysis.

[24] Relatedly, recall that it is assumed in our model that type B borrowers always default. In the more realistic setting where "bad type" borrowers default with certain probability, l could be interpreted as the expected loss associated with mistakenly originating a loan to a "bad type" borrower. In this context, larger l could be mapped to environments with relatively poor average borrower quality.

Testable Hypothesis 2. ***Lender approval/rejection decision***
Lenders with higher level in IT spending have a lower rejection rate on loan applications. This effect is stronger for applications from riskier borrowers.

Conditional on not rejecting a credit application associated with risk profile p, a lender with screening accuracy θ conducts screening on the applicant and then chooses the interest rate offer $r_s(p;\theta)$ contingent on the signal $s \in \{g, b\}$ received. We have the following proposition characterizes the key properties of the pricing function $r_s(p;\theta)$ set by lenders in equilibrium.

Proposition 3. *Conditional on a borrower with risk profile p being screened, the interest rate $r_s(p;\theta)$ offered by a lender with screening accuracy θ when receive signal s satisfies:*

1. *fixing p, $r_g(p;\theta)$ decreases in θ;*
2. *fixing p, $r_b(p;\theta)$ increases in θ.*

Proposition 2 and Proposition 3 together imply that in equilibrium lenders with higher screening accuracy offer more dispersed interest rates to their borrowers. There are two channels contributing to the higher loan price dispersion. On the intensive margin (Proposition 3), more accurate screening allows the lender to charge a lower rate for borrowers whose risk profile p is high, i.e., low default risk. On the extensive margin (Proposition 2), with a more accurate screening technology, lenders are more willing to process applications from borrowers with relatively low risk profile p, i.e., high default risk, and charge a relatively high rate on these borrowers. We formalize this test as the following hypothesis.

Testable Hypothesis 3. ***Dispersion in loan pricing***
Lenders with higher IT spending offer more dispersed interest rates on their originated loans.

Furthermore, item 1 of Proposition 3 suggests that a more informative g signal—due to a higher screening accuracy θ—allows the lender to offer a lower interest rate, which is thus more likely to be accepted by the borrower. As such, we develop the following hypothesis for our later empirical study.

Testable Hypothesis 4. ***Borrower accept/reject the offer***
Offers made by lenders with higher IT spending are less likely to be turned down, particularly for credit applications from borrowers with relatively low risk profile.

Finally, since ex-post delinquency only occurs when lenders offer an interest rate that a type B borrower accepts the interest rate offered by the lender, item 2 of Proposition 3 thus suggests that lenders with more accurate screening can reduce the likelihood of ex-post delinquency by "correctly" asking for a relatively high interest rate on borrowers based on a more informative b signal. Accordingly, we propose the following hypothesis for empirical tests.

Testable Hypothesis 5. ***Ex-post loan delinquency***
Lenders with higher IT spending has lower delinquency rate in their originated loans, particularly for credit applications from borrowers with relatively high risk profile. Relatedly, lenders with higher IT spending tend to charge higher interest rates on loans that later become delinquent.

4 Determinants of Bank IT Spending

We begin our empirical analysis by describing how bank technology adoption is related to the environment in which they operate. Specifically, we study four broad categories of factors that potentially affect banks' technology adoption: local technology infrastructure, local credit market characteristics, local economic conditions, and bank balance sheet condition.

4.1 Estimation Results

We estimate the following regression specification

$$\log(\text{Total IT})_{i,c,t} = \alpha_i + \eta_t + \beta \text{Local Characteristics}_{i,c,t} + \pi X_{c,t} + \epsilon_{i,c,t}, \tag{3}$$

where the dependent variable is the log of total IT spending of bank i in county c in year t. The main explanatory variables are the local characteristics mentioned above, whereas $X_{c,t}$ is a set of control variables including real GDP per capita and population growth rate. Tables 2 and 3 show the correlation between local banks' IT spending and these four categories of determinants.

Local technology infrastructure. We start by exploring how local technology infrastructure, as captured by local internet-related infrastructure or transportation distance from IT firms' warehouses, is correlated with bank IT spending. Panel A of Table 2 reports the results from this analysis. A one standard deviation higher maximum downstream speed is associated with 9 percent higher IT spending by local banks, which corresponds to a 0.21 standard deviation.[25] Additionally, a one standard deviation increase in the share of local population with access to broadband internet services is associated with 4.3 percent more

[25] The magnitude is calculated as follows: $0.0913\times$ 2.174 (standard deviation of log(Total IT)) and taking the exponential and minus 1.

IT spending by local banks, equivalent to a 0.13 standard deviation increase. Conversely, an increase of one standard deviation longer average shipping in the distance between the bank branch county and the warehouses of top IT producers is associated with a 2.2 percent lower IT spending.[26]

Local credit market. We then explore how past local credit market characteristics are associated with local banks' IT spending. Panel B of Table 2 shows that a one standard deviation increase in the lagged average FICO score of local borrowers is associated with a 0.7 percent decrease in total IT spending by local banks. This implies that for every 24-point increase in average FICO score of local borrowers, local bank IT spending decreases by 2.6 percentage points. Additionally, we find that a one standard deviation increase in local borrowers' debt-to-income ratio corresponds to a 0.9 percent increase in local bank IT spending. Finally, a one standard deviation increase in past mortgage delinquency rate is associated with 7 percent higher IT spending.

Along another important dimension, a one standard deviation increase in the likelihood of past mortgage securitization is associated with a 34 percent decrease in local bank IT spending, while a one standard deviation increase in the proportion of conforming loans is linked to a 39 percent decrease in local bank IT spending. Our results thus support and confirm the previous literature studying the linkage between loan securitization or the ease of loan sales and bank screening (Keys et al., 2009, 2010; Choi and Kim, 2021). Relatedly, our findings also lend support to the use of IT spending as a proxy for the screening input of banks.[27] These correlations are consistent with our model prediction that banks operating in environments with lower conforming loan ratio or weaker borrower pool tend to make larger investment in IT, as proposed in Hypothesis 1.

Local economic and demographic conditions. Panel A of Table 3 shows the correlation between local economic conditions and local banks' IT spending. Our results show that a one standard deviation higher local house price index is associated with a 6 percent higher IT spending. We also find that a one standard deviation higher proportion of local population with collage degree is associated with 6.6 percent higher IT spending. Moreover, a one standard deviation higher growth in local small business establishment is associated with 1.7 percent higher IT spending by local banks. On the other hand, a one standard deviation increase in the proportion of the local population aged 65 years and above is asso-

[26]To complement this regression analysis, Figure A4 displays the binned scatter plot of the correlations demonstrated in Panel A of Table 2. These figures show consistent relationships between local technology infrastructure and local banks' IT spending: higher broadband services coverage and higher internet speed display strong positive correlations with IT spending by local lenders, while shipping distance from technology manufacturers displays a negative correlation with IT spending by local banks.

[27]Figure A2 provides further support for these correlations by showing the related binned scatter plots.

ciated with a 4 percent lower in IT spending by local banks, and a one standard deviation increase in the local unemployment rate is associated with a 1.7 percent lower IT spending by local banks.[28]

Bank local presence and balance sheet condition. In Panel B of Table 3, we report the correlation between banks' IT spending with their presence in the local economy and their balance sheet condition. The results indicate that a one standard deviation increase in local bank deposit HHI is associated with a substantial 30.8 percentage point decrease in a bank's IT spending in that county. Conversely, a one standard deviation increase in a bank's own deposit market share in a county is associated with a noteworthy 17.1 percentage point increase in its IT spending in that same county, and a one standard deviation increase in a bank's mortgage market share in a county is associated with an 8.24 percentage point increase in IT spending by that bank in the county.

In terms of banks' balance sheet conditions, a one standard deviation increase in the deposit-to-loan ratio corresponds to an 11.1 percentage point increase in IT spending. Additionally, a one standard deviation increase in a bank's profitability, measured by its revenue per employee in that county, is linked to a significant 24.9 percentage point increase in IT spending by that bank in the county.[29]

4.2 Discussion on Constructing Bank IT Spending Shifters

The main goal of our paper is to establish the causal impact of banks' technology adoption on their loan production activities. However, identifying such impact requires us to address the issue of potential omitted variable problems or reverse causality. Banks' adoption of information technology is an endogenous choice, which means that any observed differences in loan market outcomes could be driven by other confounding local economic factors that also influence local banks' decisions regarding their IT investments. More specifically, in addition to the loan production function of local banks, borrower-side factors can also affect local credit market outcomes. For instance, these factors include the composition of local credit applicants and their credit search behavior, which may have important implications for local banks' loan approval and pricing policies. Moreover, the ex-post performance of originated loans could be influenced by hidden economic variables (e.g., condition of the local economy) that also affect local banks' IT investment decisions. In such cases, one cannot attribute these differences solely to variations in banks' technology adoption levels.

[28] Figure A3 displays the binned scattered plots of these correlations, which are qualitatively consistent with the results in Table 3.

[29] Figure A4 presents the binned scatter plots of these correlations, which are consistent with the regression analysis.

To overcome this issue in causal identification, it is necessary to construct plausibly exogenous variation in banks' technology adoption. Our analysis in Section 4 provides useful insights on potential sources that drive variations in banks' IT investment. However, while various factors are shown to be correlated with the IT spending of banks operating in the local area in Table 3, many of these factors are also related to loan origination rates and ex-post loan performance. For instance, local economic factors including local borrowers' credit profile and local demographic characteristics are unlikely to produce the desired exogenous variations in banks' IT spending because they obviously do not satisfy exclusion restrictions. In this regard, the source of variation in banks' IT spending should not be based on local economic factors, such as the local credit market condition or local demographic characteristics.

Among the various factors explored in Section 4, the transportation distance from banks' operating location to major computing device warehouses offers a potential source of plausibly exogenous variation in the banks' level of technology adoption.[30] Indeed, as shown in Panel A of Table 2, longer distance to the warehouses of major IT suppliers—the location choice of which is made at a nationwide level—predicts lower IT spending local bank branches. While this distance measure provides variation that is likely uncorrelated with changes in local demand or creditor profile, it does not enable us to compare differences in lender decisions *within the same location.* In other words, it does not allow us to compare variations in lender decisions driven by differences in IT adoption when faced with the same borrower pool.

To strengthen the identification, we construct our main instrumental variable based on *multi-market* banks' exposure to high housing price markets. The economic rationale behind this instrument is that increasing housing prices are often accompanied with a broader pool of borrowers coming into a local housing market and a decrease in the proportion of conforming loans. As shown in Panel B of Table 2, these changes are associated with higher levels of IT adoption by banks. Importantly, the level of or change in housing price in the other counties are driven by economic conditions in those counties, which are unlikely to be correlated with the demand side factors in a focal county itself. Thus by constructing a bank's exposure to high housing price markets in a given year across the whole nation, we are able to compare loan production among lenders in the focal county with different levels of IT adoption that are driven by differential nation-wide exposure to high housing markets. We will provide

[30] The main advantage of this shipping distance for ticking the exclusion restrictions is that the average shipping distance is averaged among multiple wholesale IT product producers' warehouses in the whole nation, which is mainly driven by the producers geographic optimization considering the nation-wide demand factors as well as the producer-specific cost factors, this means that it unlikely to be affected by a one-time local demand change.

detailed explanation and analysis of this instrumental variable in Section 5.2.2.

5 Technology Adoption and Bank Loan Production

In this section, we study how bank adoption of information technologies impacts their production of loans. Specifically, we focus on three aspects in our examination of the loan market outcomes: (i) approval/rejection decisions in loan application stage, (ii) pricing patterns of originated loans, and (iii) ex-post performance of originated loans.

5.1 Empirical Design

The main goal of our analysis is to investigate whether IT adoption by banks has a causal impact on their loan production. To address the endogeneity concerns discussed in Section 4.2, we employ an instrumental variable strategy based on plausibly exogenous factors driving changes in banks' IT spending through banks' connected operations across counties.

Specifically, for each multi-market bank i in a given year t, we construct the following metric to measure its exposure to house price growth in the nation-wide house market:

$$\text{HP exposure}_{i,t} = \sum_j \text{HPI}_{j,t} \times \frac{\text{Deposits}_{i,j,t}}{\sum_k \text{Deposits}_{i,k,t}}$$

where $\text{HPI}_{j,t}$ is the level of the FHFA house price index of a county j in year t, $\frac{\text{Deposits}_{i,j,t}}{\sum_k \text{Deposits}_{i,k,t}}$ is the share of bank i's deposit in county j among bank i's total deposit in year t across all counties. Hence, the variable HP $\text{exposure}_{i,t}$ captures bank i's exposure to the level of house prices in a year t across all counties in which it operates, weighted by the deposit share of bank i in each county j among its total deposits.

The mechanism at work is that for a bank operating in multiple counties, when some counties where it operates have high house prices (high HPI counties) in a year, the bank's branches in those counties tend to have an increased demand for IT investment. This could be driven by the need to handle the growing demand for housing—which itself drives prices up—or it could be driven by the need for more thorough screening due to relatively high proportion of non-conforming loans in these high HPI counties. The exclusion restriction of this instrumental variable assumes that higher exposure to housing price index affect banks' screening in other counties exclusively through a within-organization technology spillover or technology sharing.

Two alternative channels could potentially lead to the violation of exclusion restriction. First, higher exposure to house price index might be driven by demand growth in some specific locations, and this might induce banks' to expand their operation, say, through hiring more loan officers and opening more branches. Second, higher exposure to house price index might induce banks to spend on other types of expenses to improve operation, say advertising and marketing. In Table A14, we examine the correlation between "HP exposure" and a variety of non-IT factors that could also affect loan decision and outcomes. In particular, we find that the number of branches and number of employee expansion is not significantly correlated with HP exposure, furthermore, among the various non-interest expenses such as printing, advertising and marketing, as well as IT spending, the total IT expenditure stands out as the most responsive spending to the bank-level "HP exposure."

Several categories of IT relevant for mortgage screening, such as software, alternative data, and cloud storage, are often partially shared among all branches within a bank. As a result, a bank with exposure to high HPI counties tends to have higher total IT spending at the *bank level* and the higher IT spending is hence likely to spillover across all its branches, including those in how HPI counties, compared to a bank without such exposure or with lower exposure. In essence, this instrumental variable captures a technology spillover within a bank across its branches operating in different geographic areas. The requirement for the exclusion restriction of this instrumental variable relies on the assumption that mortgage demand in high HPI counties is not perfectly correlated with mortgage demand in low HPI counties.[31]

We also employ a second instrumental variable leveraging the cross-bank variation in their proximity to main IT producers' warehouses.The rationale is that longer shipping distances imply higher potential costs and waiting time for banks to upgrade their hardware and conduct maintenance, hence leading to lower level of IT investment.[32] Importantly, since location decisions of IT product suppliers' warehouse choice is often made at a nationwide basis, variation in banks' proximity to these warehouses is thus unlikely to be correlated with changes in the economic conditions of specific local economies in which banks operate. This allows us to extract variations in banks' IT spending that are presumably uncorrelated with factors influencing the credit demand side of a particular local credit market.[33]

[31]Under the premise that such a within-bank spillover effect of technology adoption is more relevant for banks' investment and purchase in software products, in the Appendix Table A9 to Table A10 we also focus on banks' software IT spending (instead of total IT spending) as an alternative measure of banks' technology adoption.

[32]This correlation has been established in Section 4. In Panel A of Table 2, we find that the farther the transportation distances, the lower total IT spending the local banks are willing to incur to frequently upgrade their IT installments. The effect of geographic barrier or transportation cost on technology adoption and diffusion has also been well established in the literature of international trade (Alfaro and Chen, 2018) and development economics (Storeygard, 2016).

[33]Since shipping distance is more relevant for banks' purchase of communication IT and hardware than it is for software purchase, in Appendix Table A8 to Table A10 we also conduct analysis with this second in-

5.2 Bank IT Spending and Loan Decisions Outcomes

We start our analysis by studying the extensive margin effect of banks' technology adoption. That is, we investigate how banks' IT spending will impact banks' approval/rejection decision-making after receiving loan applications.

5.2.1 Variable Specification and OLS Analysis

In conducting this analysis, we begin with the following OLS regression:

$$\text{Loan decision}_{l,i,c,t} = \alpha_i + \mu_c + \eta_t + \beta \times \log(\text{Total IT})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \text{FE} + \epsilon_{l,i,c,t}, \quad (4)$$

where the dependent variable Loan decision$_{l,i,c,t}$ is a loan-level dummy variable capturing the decision of bank i in county c made on loan l in year t, and the main explanatory variable is the log of bank IT spending is measured at the bank-county-year level. The specification includes bank, county and year fixed effects; along with a set of time-varying controls for local economic factors that includes population, real GDP, unemployment rate, and a house price index; and a set of time-varying bank characteristics including the log of bank total revenue in the county, net revenue over total assets, total deposits over total assets, and the log of bank total assets.

For the dependent variable, we classify application outcomes as follows. First, we classify an application as rejected if the lender receives the loan application, and decides not to make the applicant an offer after screening the applicant. Second, we classify an application as an approval that is declined by the applicant as the scenario where the lender approves the application and makes an offer to the borrower, but the applicant decides not to take the offer. Third, we classify an application as originated if the lender approves the application and makes an offer to the applicant, and the latter takes the offer.

Table 4 reports OLS estimates of equation (4). Our analysis shows that a one standard deviation increase in a bank IT spending is associated with a 0.019 standard deviation decrease in the probability that an application is rejected. This is in line with Hypothesis 2 in our model. Furthermore, a one standard deviation increase in bank IT spending is associated with a 0.01 standard deviation decrease in the likelihood that a lender's offer is declined by the applicant, which is also consistent with Hypothesis 4. On the flipside, a one standard deviation increase in bank IT spending in a particular county is associated with a 0.022 standard deviation increase in the probability that an application leads to an origination.

strumental variable using banks' spending on communication IT and hardware (instead of total IT spending) as measure of their IT adoption intensity.

5.2.2 2SLS Analysis

In this section, we examine whether IT adoption by banks has a causal impact on application outcomes. We follow a Two-Stage least squares (2SLS) approach leveraging the instrumental variables constructed in Section 5.1. Specifically, we estimate the following regression specification:

$$\begin{aligned} \log(\text{Total IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{HP exposure}_{i,t} + \pi_1 \mathbf{X} + \mathbf{FE\text{'}s} + \epsilon_{i,c,t} \\ \text{Loan decision}_{l,i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Total IT}})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE\text{'}s} + \epsilon_{l,i,c,t} \end{aligned} \tag{5}$$

Panel B of Table 4 shows the results from estimating equation (5) by 2SLS. Column (1) shows the result for the first stage, whereas columns (2)–(4) show the IV estimates. Specifically, we find that a one standard deviation increase in HP exposure is associated with a 0.07 standard deviation increase in IT spending of a bank, which amounts to 14 percentage points higher total IT spending for a bank in a county.The F-stat is 450, well above the conventional threshold for weak instruments (Stock and Yogo, 2005).

Our main estimates exhibit the same direction as the OLS estimates but with more pronounced magnitudes. Specifically, a one standard deviation increase in IT spending, driven by additional bank-level exposure to high housing price markets, leads to a 0.11 standard deviation lower rejection rate, equivalent to a substantial 5.4 percentage points reduction.[34] Considering the average rejection rate in our sample stands at 0.553, this effect translates to a noteworthy 10% decrease in the rejection rate. Likewise, a one standard deviation increase in IT spending corresponds to a 0.094 standard deviation decline in the rate of declined approvals, amounting to a 3.36 percentage points reduction. Relative to the sample mean, this represents a 22.4% decrease in declined approvals. Finally, a one standard deviation increase in IT spending leads to a 0.15 standard deviation increase in the loan origination rate, equivalent to a substantial boost of 6.85 percentage points. This increase signifies a 23.1% higher loan origination rate compared to the sample mean.

Overall, our 2SLS analysis, which leverages bank exposure to high house prices in other markets as an instrumental variable, indicates that bank IT spending has strong impacts on application outcomes: banks with higher IT spending reject fewer applications, and have a higher origination rate on approved applications.[35]

[34] Standard deviation of rejection rate in our sample is 0.497, a 0.11 standard deviation lower rejection rates means 0.11×0.497 which is 0.054 or 5.4 percentage points lower rejection rates.

[35] We also conduct a similar analysis using the transportation distances between a bank's operating county and the warehouses of major technology producers. The 2SLS analysis employing this transportation distance as the instrumental variable reveals that, at the cross-section level, bank branches situated closer to major tech product producers tend to allocate higher expenditures towards IT products and services. Consequently,

5.2.3 Heterogeneous Impacts Across Borrower Groups

Our analysis thus far suggests an economically and statistically significant overall impact of banks' IT spending on application outcomes. We now explore how these impacts may differ across applicants with varying observable risk profiles.

Observable riskiness of applicants. We start by constructing measures of applicants' observable default risk using a LASSO estimator based on their observable characteristics. We estimate a default probability model using 2-year delinquency as the depenendent variable, and loan amount, applicant income, ethnicity, gender, age, race and combined-loan-to-value-ratio as covariates. We then predict default risk for the whole sample. In what follows, we define an applicant as high-risk (low-risk) if she falls in the top (bottom) quartile in the distribution of applicant risk in a year.

We start by comparing the distribution of credit applicants that get approval from banks with different levels of IT spending. In Table A4, we show the correlation between estimated default risk of all applicants taken by a bank in a county in a year and the banks' IT spending, as well as the correlation between the dispersion of applicants' default risk and the bank's IT spending. We find that higher IT spending is positively associated with the applicants' average riskiness and with the dispersion of applicants' riskiness.

Figure A6 illustrates the distribution of applicants' observable risk by high IT spenders and low IT spenders, where high (low) IT spenders are defined as banks with IT spending scaled by total revenue above (below) median among all banks in the same county in a specific year. Notably, the distribution of applicant riskiness is more right-skewed for high IT spenders than that for low IT spenders. This implies that, in equilibrium, banks with higher IT spending tend to be more willing to grant approval to observably risky applicants than low IT spending banks do. This observation aligns with our earlier finding in Section 5.2.2, which indicates that higher IT spending leads to reduced loan rejection rates.

Regression analysis. Table 5 presents the results for regression analysis of the heterogeneous impact of bank IT spending on loan decision outcomes. Specifically, we conduct the same set of analysis as in Section 5.2.1 (OLS) and Section 5.2.2 (2SLS), but separately restricting to the sub-sample of loan applications sent by high(low)-risk borrowers as defined above. In Panel A of Table 5, we present the OLS regression estimates on how a bank's IT spending affects its loan-making decision for high-risk borrowers and low-risk borrowers respectively. Panel B of Table 5 reports the 2SLS estimates using banks' exposure to housing price index as the instrument variable, which is constructed in Section 5.1.

these banks exhibit a lower rate of rejection for loan applications and a higher rate of approved applications resulting in loan origination. The results using transportation distances between a county and major IT producers' warehouses are demonstrated in Table A11 to Table A13.

Columns (1) and (2) of Table 5 present estimates of the impact of banks' IT spending on their loan approval rate for borrowers with different risk profiles based on observable characteristics. As anticipated by Hypothesis 2, our OLS estimates show that higher levels of IT investment by banks lead to a statistically significant reduction in lenders' rejection rate for high-risk borrowers, while having no discernible effect on the rejection rate for low-risk borrowers. Moving on to the 2SLS estimates, we find statistical significance for both borrower groups. Nevertheless, the impact of banks' IT spending on loan rejection rates is more pronounced when dealing with risky borrowers. Specifically, a one standard deviation increase in a bank's IT spending in a county, driven by additional bank-level exposure to high housing price markets, results in a 0.203 (0.108) standard deviation reduction in the bank's rejection rate for high-(low-)risk borrowers in that county. This difference is statistically significant, highlighting a stronger effect on loan rejection rates for high-risk borrowers.

Column (3) and column (4) display how the occurrence of "Approval but denial" is impacted banks' IT spending for high-risk borrowers and low-risk borrowers, respectively. We find that following a one standard deviation increase in a bank's total IT spending in a county, driven by increased bank-level exposure to high housing price markets, the "Approval but denial" rate for loan offers the bank grants to low-risk borrowers is reduced by a statistically significant 0.0391 standard deviation. In contrast, for high-risk borrowers, the impact of increased IT spending on the "Approval but denial" rate is insignificant both qualitatively and quantitatively. Qualitatively similar results are also observed on OLS estimates. These findings are consistent with Hypothesis 4, which is proposed based on the reasoning that higher IT spending (hence more accurate screening) allows lenders to set more refined pricing menus for low-risk borrowers, resulting in lower likelihoods of these borrowers rejecting the offered rates.

Bank technology adoption and credit market "discrimination." To shed light on the issue of potential discrimination against certain groups in credit markets, such as low-income or minority borrowers, we also investigate whether banks that invest more in IT are more likely to make different loan decisions when dealing with these vulnerable groups.

Our analysis on the impact of IT spending on lending decisions towards "low income" borrowers is presented in Table A5.[36] Low-income borrowers are significantly more likely (by 11.25 percentage points) to be rejected by lenders. However, this effect is mitigated by 2.54 percentage points when the lender is a high-spender in the same county, as shown in column (1) of Table A5. Additionally, in column (3) we find that low-income borrowers are

[36]We define a borrower as "low income" if their income falls below the median among all borrowers in the county in that year, and a bank as a "high spender" if their IT spending, scaled by total revenue, is above the median among all banks in the same county.

less likely to have their loans finally originated, but if they are dealing with a high-spender lender in their local area, their chances of loan origination increase by 2.48 percentage points. We do not find high IT-spender have significant improvement on "Approval but reject" for low-income borrowers.

Table A6 presents our findings on whether high IT spender banks make different lending decisions when dealing with minority borrowers.[37] Our results show that, overall, minority borrowers are 9.77 percentage points more likely to be rejected, 2.45 percentage points less likely to turn down a lender's offer, and 7.32 percentage points less likely to have their loan application finally originated. However, if the lender is a high spender, the desk rejection rate for minority borrowers decreases by 2.64 percentage points, and their loan origination rate increases by 2.57 percentage points. We do not find a significant change in the likelihood of a borrower rejecting a lender's offer if the lender is a high spender.

In summary, our analysis in this section demonstrates that while banks' technology adoption has an overall significant impact on their decision making towards credit applications, there is significant heterogeneity in its impact for different credit applicants. Borrowers who have a higher risk profile and those who belong to low-income or minority groups benefit more from a larger reduction in the likelihood of being rejected by banks, which we refer to as the extensive margin effect of banks' technology adoption. These findings suggest that technology adoption by lenders can potentially address or at least mitigate inequality in credit markets by improving access to credit for marginal borrowers.

5.3 Bank IT Spending and Mortgage Pricing

We now shift our focus to the pricing patterns of originated loans after examining the decision-making outcomes during the loan application and approval stages. In this section, we begin by demonstrating the substantial impact of banks' IT spending on the dispersion of prices for their originated loans. We then delve into exploring the underlying mechanism behind the linkage between the granularity of banks' loan pricing and the technology adoption.

5.3.1 Dispersion in Mortgage Pricing

In Section 5.2, we observed that banks with higher IT spending demonstrate a greater propensity to grant loans to marginal borrowers that their low IT spending counterparts would reject. Simultaneously, when these banks extend loan offers, borrowers are more

[37] We define a borrower as belonging to a minority ethnicity group if they are Hispanic or Latino.

likely to accept them. Consequently, with a more inclusive decision-making policy at the credit application stage, it is reasonable to anticipate a wider dispersion of loan prices offered by banks with higher IT spending in equilibrium (Hypothesis 3).

Following this logic, in this section we start with examining how banks' investment in IT would affect the interest rate dispersion of the loans they originated. After taking into control the dispersion of FICO score, LTV ratio and a rich set of bank-level and county level control variables, we find a strong positive correlation between the residualized mortgage rates dispersion and the residualized bank IT spending. Figure 2 shows the binned scatter plot of the correlation between the mortgage interest rates dispersion and the originating banks' IT spending. Both the residualized binned scatter plot and the scatter plot display a strong positive correlation between mortgage interest rate dispersion and the originating banks' IT spending.

For more formal investigation, we run the following OLS specification

$$\text{Interest rate dispersion}_{i,c,t} = \alpha_i + \mu_c + \eta_t + \beta \times \log(\text{Total IT})_{i,c,t} + \beta_1 \text{X} + \epsilon_{i,c,t} \qquad (6)$$

The dependent variable "Interest rate dispersion$_{i,c,t}$" is the standard deviation of loan interest rate offered by a bank i in county c in year t. Column (4) of Table 6 shows the OLS regression results. We find that a one standard deviation increase in a bank's total IT spending in a county is associated with 0.0438 standard deviation increase in the dispersion of interest rate charged by a bank. With bank fixed effects, the result is interpreted as a within-bank effect of higher IT spending. This means that for a given bank, in counties where its branches spend more in IT, the interest rate distribution it offers in that county will also have a wider distribution compared with the interest rate menu it offers in other counties.

Paralleling the investigation on the causal linkages between bank IT spending and loan decision making, we run the following 2SLS specification to explore whether IT investment is causally linked with loan pricing dispersion:

$$\begin{aligned} \log(\text{Total IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{HP exposure}_{i,t} + \pi_1 \mathbf{X} + \mathbf{FE's} + \epsilon_{i,c,t} \\ \text{Int rate dispersion}_{i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Total IT}})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE's} + \epsilon_{i,c,t} \end{aligned} \qquad (7)$$

Table 6, columns (1)-(3), show the results for the above 2SLS specification. We find that a one standard deviation increase in the "HP Exposure" of a bank across the nationwide geographic area is associated with 0.0323 standard deviation increase in a bank's total IT spending in that a county. In the second stage, we find that a one standard deviation increase in the a bank's total IT spending in a county lead to 0.756 standard deviation increase in the interest rate dispersion, which means 0.155 percentage point increase in the interest rate dispersion. Compared with the average interest rate dispersion of 0.481, this means a 32% increase expansion of the interest rate dispersion among the interest rates offered by the bank.

5.3.2 Heterogeneous Impact on Loan Pricing for Different Borrowers

In this section, we aim to uncover the underlying mechanism behind the impact of banks' IT investment on the pricing dispersion of their originated loans. Specifically, we examine how lenders' IT spending affect their loan pricing towards high-risk and low-risk borrowers separately. We start with the following OLS regression specification:

$$\begin{aligned}\text{Interest rate}_{l,i,c,t} = \beta \times 1[\text{High spender}]_{i,c,t} \times 1[\text{High/Low Def risk}]_{l,c,t} \\ + \beta_1 \times 1[\text{High spender}]_{l,c,t} + \beta_2 \times 1[\text{High/Low Def risk}]_{l,c,t} \\ \alpha_i + \mu_c + \eta_t + \beta_3 \mathbf{X} + \mathbf{FE}s + \epsilon_{l,i,c,t}\end{aligned} \tag{8}$$

The dependent variable "Interest rate$_{l,i,c,t}$" is the interest rate of loan l made by bank i in county c and in year t. The indicator variables "1[High def risk]$_{l,c,t}$" (regarding borrowers) and "1[High spender]$_{i,c,t}$" (regarding lenders) are the same as defined in Section 5.2. Control variables and the loan characteristics fixed effects are the same as in Equation (4).

Table 7 presents the regression results for the above specification. We find that high IT spenders charge overall low interest rates on originated loans than their lower IT spending counterparts do. Furthermore, as expected, borrowers with high (low) default risk based on their observable characteristics are on average charged 9.61 basis points higher (10.4 basis points lower) interest rates. More importantly, the coefficient estimates on the interaction term reveal a heterogeneous impact of IT spending on lenders' loan pricing towards different borrower groups. Specifically, high-risk borrowers are charged an additional 3.27 basis points when facing lenders with high IT spending. In contrast, low-risk borrowers are able to enjoy a 1.7 basis point further reduction in loan rates when facing high IT spending lenders.

Paralleling our analysis in Section 5.2.3, we also conduct an IV analysis by separately restricting the regression to the sub-samples of loans originated to high-risk and low-risk borrowers, respectively.

$$\begin{aligned}\log(\text{Total IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{HP exposure}_{i,t} + \pi_1 \mathbf{X} + \mathbf{FE\text{'}s} + \epsilon_{i,c,t} \\ \text{Interest rate}^{\text{high/low def risk}}_{l,i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Total IT}})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE\text{'}s} + \epsilon_{i,c,t}\end{aligned}$$

The regression equation is specified above. Table 8 presents the differential impact of IT spending on interest rates charged towards high-risk borrowers and low-risk borrowers using "HP exposure" as instrumental variable. We find that a one standard deviation increase in a bank's total IT spending in a county—due to additional bank-level exposure to high house price across the nationwide housing market—leads to a 2.99 standard deviation increase in the interest rate charged towards a "high def risk" borrowers, which amounts to 1.53 percentage points higher interest rates. On the other hand, a one standard deviation increase in a bank's total IT spending in a county leads to 0.32 standard deviation lower interest rate charged towards "low def risk" borrowers, which amounts to 0.17 percentage points lower interest rate.

These findings corroborate our theoretical prediction that banks with more accurate screening tend to set more granular loan pricing by charging extra higher on risky borrowers while reducing the rates offered to safe borrowers, thereby enabling them to extract more "consumer surplus" from borrowers. Our findings further indicate a welfare reallocation effect of banks' technology adoption across borrowers. When banks have a relatively poor screening capability and cannot generate sufficiently informative signals from screening, the interest rates they offer on originated loans are likely to be relatively pooling. At such pooling interest rates, the safe borrowers who accept their loan offers are effectively subsidizing the risky borrowers. Our findings in Table 8 suggest that higher bank IT investment can mitigate such cross-subsidization between borrowers, by pushing up the interest rates charged on risky borrowers while bringing down the rates offered to safe borrowers. Furthermore, this opposite effect of IT adoption on loan pricing across borrower groups also implies an *intensive margin* effect that contributes to a more dispersed interest rates offered by high IT spending banks, together with the *extensive margin* effect due to their higher willingness to accept marginal borrowers as documented in Section 5.2.3.

Finally, in addition to the observable riskiness of credit applicants as constructed in Section 5.2.3, we also examine how loan pricing patterns vary across different demographic characteristics and the extent to which it depends on banks' IT investment. Table A7 reports the findings of whether high IT spenders set notably different interest rates for low-income and minority groups. Our results show that while high IT spending banks are more likely to approve loans for low-income borrowers, they also charge, on average, 1.94 basis points higher interest rates for this group. For borrowers from minority groups, we find that high IT spenders charge only 0.49 basis points higher interest rates, but this difference is not statistically significant.

In summary, our research suggests that higher IT investment by banks can lead to more accurate screening of borrowers, resulting in a more refined loan pricing menu for borrowers (intensive margin effect). Additionally, higher IT investment enables banks to serve a broader range of credit applicants, resulting in a more dispersed distribution of interest rates charged on originated loans (extensive margin effect). Both these effects together contribute to a more diverse range of interest rates offered by banks to their borrowers.

5.4 Bank IT Spending and Mortgage Repayment

To assess the impact of banks' IT investment on the efficacy of their loan production, it is essential to analyze the actual ex-post performance of originated loans. In this section, we investigate the relationship between banks' IT spending and the performance of their originated loans.

Impact on ex-post loan delinquency. We begin with a graphical representation. Figure 3 displays the residualized binned scatter plot between the 2-year delinquency rates and 4-year delinquency rates of loans originated by a bank in a county and the bank's IT spending in that county. The figure demonstrates a strong negative correlation between banks' IT spending during loan origination and the ex-post delinquency probabilities. This indicates that better and more intensive screening, as proxied by banks' IT spending, is negatively associated with post-origination loan default.

Formally, to analyze the relationship between loan performance and IT spending of originating banks, we estimate the following benchmark OLS regression specification:

$$1[\text{Delinquent in x years}]_{l,i,c,t} = \alpha_i + \mu_c + \eta_t + \beta \times \log(\text{Total IT})_{i,c,t} + \beta_1 \mathbf{X} + \mathbf{FE}s + \epsilon_{l,i,c,t}$$

The dependent variable of this regression is a dummy variable indicating whether a loan l originated by lender i in county c in year t was ever delinquent within 2 years, 4 years and 6 years of origination. The main explanatory variable is the logarithmic of total IT spending of the originating bank i in county c in year t. The control variables and the fixed effects are the same as specified in equation (4). Table 9 panel A shows the results of the above OLS regression specification. In particular, we find that a one standard deviation increase in a bank's total IT spending in a county is associated with 0.0132 (0.0121) standard deviation decrease in the 2-year (4-year) delinquency rate of the loans originated. These findings support Hypothesis 5, which posits that higher IT spending has an overall effect that lowers the delinquency rate for originated loans.

Paralleling the previous sections, we run the following 2SLS regression specification to establish the causal linkage between banking IT adaptation and loan performances:

$$\begin{aligned}
\log(\text{Total IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{HP exposure}_{i,t} + \pi_1 \mathbf{X} + \mathbf{FE's} + \epsilon_{i,c,t} \\
\mathbb{1}_{\text{Delinq x years},l,i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Total IT}})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE's} + \epsilon_{l,i,c,t}
\end{aligned}$$

Panel B of Table 9 presents the results of the analysis on how bank IT spending affects loan outcomes. Columns (2) to (4) show the results of the second-stage regressions using the fitted values of the logarithm of IT spending from the first stage. We find that a one standard deviation increase in total IT spending of a bank in county, driven by its operation in high HP counties through connected market, leads to 0.5157 standard deviation decrease in the 2-year delinquency rate for newly originated loans, which amounts to a 11.1 percentage points decrease in the 2-year delinquency.Similarly, it also results in a 12.3 percentage points reduction in the 4-year delinquency rate for newly originated loans.

It is worth noting that in Section 5.2, we established that banks with higher IT spending tend to exhibit more inclusive decision-making processes when evaluating loan applications. Given this context, our above findings on the performance of originated loans suggest that these banks, despite their willingness to serve borrowers whom banks with low IT spending would typically reject, do not actually experience a higher delinquency rate in originated loans. This implies that the proactive approach to granting loans of these high IT spending banks does not compromise the quality and repayment performance of the loans themselves.

To investigate the potential variation in the impact of bank IT spending on loan outcomes based on the risk level of borrowers, we estimated the following regression specification that includes an interaction term between the high-risk borrower indicator and the indicator variable of high IT spending lender. The results are shown in Table 10.

$$\begin{aligned}
1[\text{Delinquent in x years}]_{l,i,c,t} = \alpha_i + \mu_c + \eta_t + \beta \times 1[\text{High spender}] \times 1[\text{High def risk}] \\
\beta_1 1[\text{High def risk}] + \beta_2 1[\text{High spender}] + \beta_3 \mathbf{X} + \mathbf{FE}s + \epsilon_{l,i,c,t}
\end{aligned}$$

As expected, the coefficients on the high default risk dummy are significantly positive. However, the coefficients on the interaction term are negative and statistically significant. For borrowers who are perceived to have a high delinquency risk based on observable characteristics, they are on average 1.8 percentage points more likely to default within a 2-year time window since the loan is originated, and 2.16 percentage points more likely to default within 4-year and 6-year time windows. However, high IT spending seems to enable banks to screen out the relatively credible borrowers among the high-risk group, resulting in a reduction in the actual default rate of high-risk borrowers lent by high IT spenders. Specifically, high IT spenders are able to bring down the default rate of high-risk borrowers by 1.32 percentage points within a 2-year time window, and by 1.8 percentage points within 4-year and 6-year time windows.

These findings suggest that banks that spend more on IT are more capable of identifying the relatively "safer" borrowers among the marginal credit applicants who are perceived to have a higher delinquency risk. This ability to pick out lower-risk borrowers leads to better ex-post performance outcomes on originated loans.

Ex-ante pricing of ex-post delinquent loans. In addition to examining the delinquency rate of originated loans, it is also important to consider how banks account for or anticipate the associated risks in terms of their loan pricing in assessing the impact of IT investment. In this regard, we analyze the ex-ante pricing patterns of loans that later become delinquent, focusing on the comparison between banks with different levels of IT investment. Our aim is to determine whether banks with greater IT investment demonstrate better proficiency in incorporating or anticipating risk, thereby adjusting their ex-ante loan pricing accordingly to account for risk compensation.

Table 11 presents an analysis of the correlation between loan interest rates and banks' IT spending, with a specific focus on loans that become delinquent following origination. The results, displayed in rows (1) and (4), confirms that loans which experience delinquency within two and four years after origination are indeed associated with higher interest rates. More importantly, rows (3) and (5) provide supporting evidence for Hypothesis 5, indicating that loans originated by banks with high IT spending are linked to an extra increase in interest rates had they become delinquent subsequently. Specifically, for borrowers that become delinquent within two years after origination, high IT spending banks charge an additional 2.84 basis points in interest rate. Similarly, when a borrower becomes delinquent within four years, high IT spending banks charge an extra 4.67 basis points compared to what low IT spending banks do.

These findings imply that banks with higher IT spending possess the capability to better discern borrowers with a higher likelihood of delinquency, ultimately enabling them to adjust interest rates accordingly. This ability to differentiate borrowers who are more likely to default from those who are less likely showcases the value of increased IT investment in enhancing risk management practices within the banking sector.

6 Conclusion

In this paper, we propose a novel proxy for banks' investment in their screening technology during the loan production process—their IT spending. WWe construct a unique dataset that links banks' screening intensity with their loan decision-making, loan pricing, and loan performance in the U.S. mortgage market. Our analysis reveals strong correlations between banks' screening effort, as measured by IT spending, and various factors such as local technology infrastructure, borrower credit risk profiles, easiness of securitization loan sales, and local economic and demographic conditions.

By linking the screening technology investment proxied by IT spending to the loan production process and exploiting two distinct instrumental variable strategies, we uncover several key findings. First, banks with higher IT spending are more inclined to make inclusive loan approval decisions, demonstrating a greater willingness to approve applicants with higher observable risk. Second, these banks are less likely to get denials from borrowers when they grant approvals, especially when faced with low-risk borrowers. As a result, they have a higher proportion of loans that convert to origination, indicating a more efficient loan origination process. Furthermore, we find that banks with higher IT spending exhibit more granular loan pricing strategies—they charge higher mortgage rates for high-risk borrowers and lower mortgage rates for low-risk borrowers—indicating their ability to better assess and price risk. Importantly, the loans originated by banks with higher IT spending exhibit significantly lower overall delinquency rates, particularly among the high-risk borrower group. We provide a simple framework rationalizing all of the empirical findings.

Our analysis in this paper makes contribution to the existing literature in several ways. Firstly, we provide a concrete and measurable proxy for banks' investment in their screening technology, which plays a key role in the existing theoretical literature yet remained to be conceptual. Secondly, we establish a direct link between banks' IT spending and loan-making behavior, shedding light on the relationship between banks' screening technology and their loan production function. Moreover, our research offers novel empirical evidence on how technological advancements impact banks' loan production by examining the adaptation of screening skills as reflected in IT spending. Unlike previous studies that focus on specific technology products and loan segments, we establish a comprehensive connection between technology adoption and a quantitatively significant credit market. Our findings also open up avenues for future research, such as investigating the impact of technology adoption on inequality, the competitiveness of local credit markets, and the changing dynamics of the local economy in response to credit cycles.

References

Agarwal, S., Alok, S., Ghosh, P., and Gupta, S. (2019). Financial inclusion and alternate credit scoring: Role of big data and machine learning in fintech. *Working Paper*.

Alfaro, L. and Chen, M. (2018). Transportation cost and the geography of foreign investment. In *Handbook of International Trade and Transportation*, chapter 12, pages 369–406. Edward Elgar Publishing.

Bartlett, R., Morse, A., Stanton, R., and Wallace, N. (2019). Consumer-lending discrimination in the fintech era. Working Paper 25943, National Bureau of Economic Research.

Berg, T., Burg, V., Gombović, A., and Puri, M. (2019). On the Rise of FinTechs: Credit Scoring Using Digital Footprints. *The Review of Financial Studies*, 33(7):2845–2897.

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.

Blattner, L. and Nelson, S. (2021). How costly is noise? data and disparities in consumer credit. *Working Paper*.

Braghieri, L., Levy, R., and Makarin, A. (2022). Social media and mental health. *American Economic Review*, 112(11):3660–93.

Broecker, T. (1990). Credit-worthiness tests and interbank competition. *Econometrica*, 58(2):429–52.

Buchak, G. and Jørring, A. (2021). Do Mortgage Lenders Compete Locally? Evidence Beyond Interest Rates. *Working Paper*.

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.

Cao, M. and Shi, S. (2001). Screening, Bidding, and the Loan Market Tightness *. *Review of Finance*, 5(1-2):21–61.

Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Feng, Y. (2022). On binscatter. *Federal Reserve Bank of New York Working Paper No.881.*

Choi, D. B. and Kim, J.-E. (2021). Does securitization weaken screening incentives? *Journal of Financial and Quantitative Analysis*, 56(8):2934–2962.

Cornelli, G., Frost, J., Gambacorta, L., and Jagtiani, J. (2021). The impact of fintech lending on credit access for u.s. small businesses. *FRB of Philadelphia Working Paper No. 22-14*.

Cornelli, G., Frost, J., Gambacorta, L., and Jagtiani, J. (2022). The impact of fintech lending on credit access for u.s. small businesses. *FRB of Philadelphia Working Paper No. 22-14*.

Crawford, G. S., Pavanini, N., and Schivardi, F. (2018). Asymmetric information and imperfect competition in lending markets. *American Economic Review*, 108(7):1659–1701.

Croux, C., Jagtiani, J., Korivi, T., and Vulanovic, M. (2020). Important factors determining fintech loan default: Evidence from a lendingclub consumer platform. *Journal of Economic Behavior & Organization*, 173:270–296.

Di Maggio, M., Ratnadiwakara, D., and Carmichael, D. (2022). Invisible primes: Fintech lending with alternative data. Working Paper 29840, National Bureau of Economic Research.

Di Maggio, M. and Yao, V. (2020). Fintech Borrowers: Lax Screening or Cream-Skimming? *The Review of Financial Studies*, 34(10):4565–4618.

D'Andrea, A., Pelosi, M., and Sette, E. (2021). Broadband and bank intermediation. *Working Paper*.

Edelberg, W. (2006). Risk-based pricing of interest rates for consumer loans. *Journal of Monetary Economics*, 53(8):2283–2298.

Einav, L., Jenkins, M., and Levin, J. (2013). The impact of credit scoring on consumer lending. *The RAND Journal of Economics*, 44(2):249–274.

Ellison, G. and Pathak, P. A. (2021). The efficiency of race-neutral alternatives to race-based affirmative action: Evidence from chicago's exam schools. *American Economic Review*, 111(3):943–75.

Frost, J. (2020). The economic forces driving fintech adoption across countries. *BIS Working Papers No 838*.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., and Walther, A. (2022). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, 77(1):5–47.

Fuster, A., Plosser, M., Schnabl, P., and Vickery, J. (2019). The Role of Technology in Mortgage Lending. *The Review of Financial Studies*, 32(5):1854–1899.

Glode, V. and Ordoñez, G. (2022). Technological Progress and Rent Seeking. *Working Paper*.

Haendler, C. (2022). Keeping up in the digital era: How mobile technology is reshaping the banking sector. *Working Paper*.

Hannan, T. H. and McDowell, J. M. (1984). The determinants of technology adoption: The case of the banking firm. *The RAND Journal of Economics*, 15(3):328–335.

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. *Journal of Financial Economics, Forthcoming*.

He, Z., Jiang, S., Xu, D., and Yin, X. (2021). Investing in lending technology: It spending in banking. *Working Paper*.

Hertzberg, A., Liberman, A., and Paravisini, D. (2018). Screening on Loan Terms: Evidence from Maturity Choice in Consumer Credit. *The Review of Financial Studies*, 31(9):3532–3567.

Jagtiani, J. and Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending: Evidence from the lendingclub consumer platform. *Financial Management*, 48(4):1009–1029.

Jiang, E. X., Lee, Y., and Liu, W. S. (2022a). The role of loan officers in the fintech era: Implications for disparities in consumer credit. *Working Paper*.

Jiang, E. X., Yu, G. Y., and Zhang, J. (2022b). Bank competition amid digital disruption: Implications for financial inclusion. *Working Paper*.

Jørring, A. (2020). Financial Sophistication and Consumer Spending. *Working Paper*.

Keys, B. J., Mukherjee, T., Seru, A., and Vig, V. (2009). Financial regulation and securitization: Evidence from subprime loans. *Journal of Monetary Economics*, 56(5):700–720. Carnegie-Rochester Conference Series on Public Policy: Distress in Credit Markets: Theory, Empirics, and Policy November 14-15, 2008.

Keys, B. J., Mukherjee, T., Seru, A., and Vig, V. (2010). Did securitization lead to lax screening? evidence from subprime loans. *The Quarterly Journal of Economics*, 125(1):307–362.

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

Lin, C., Ma, C., Sun, Y., and Xu, Y. (2021). The telegraph and modern banking development, 1881–1936. *Journal of Financial Economics*, 141(2):730–749.

Maggio, M. D., Ratnadiwakara, D., and Carmichael, D. (2021). Invisible primes: Fintech lending with alternative data. *Working Paper*.

Manove, M., Padilla, A. J., and Pagano, M. (2001). Collateral versus project screening: A model of lazy banks. *The RAND Journal of Economics*, 32(4):726–744.

Marquez, R. (2002). Competition, adverse selection, and information dispersion in the banking industry. *The Review of Financial Studies*, 15(3):901–926.

Ridder, M. D. (2021). Market Power and Innovation in the Intangible Economy. *Working Paper*.

Ruckes, M. (2004). Bank Competition and Credit Standards. *The Review of Financial Studies*, 17(4):1073–1102.

Scharfstein, D. and Sunderam, A. (2016). Market power in mortgage lending and the transmission of monetary policy. *Unpublished working paper.*, Harvard University.

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.

Storeygard, A. (2016). Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa. *The Review of Economic Studies*, 83(3):1263–1295.

Straka, J. W. (2000). A shift in the mortgage landscape: The 1990s move to automated credit evaluations. *Journal of Housing Research*, 11(2):207–232.

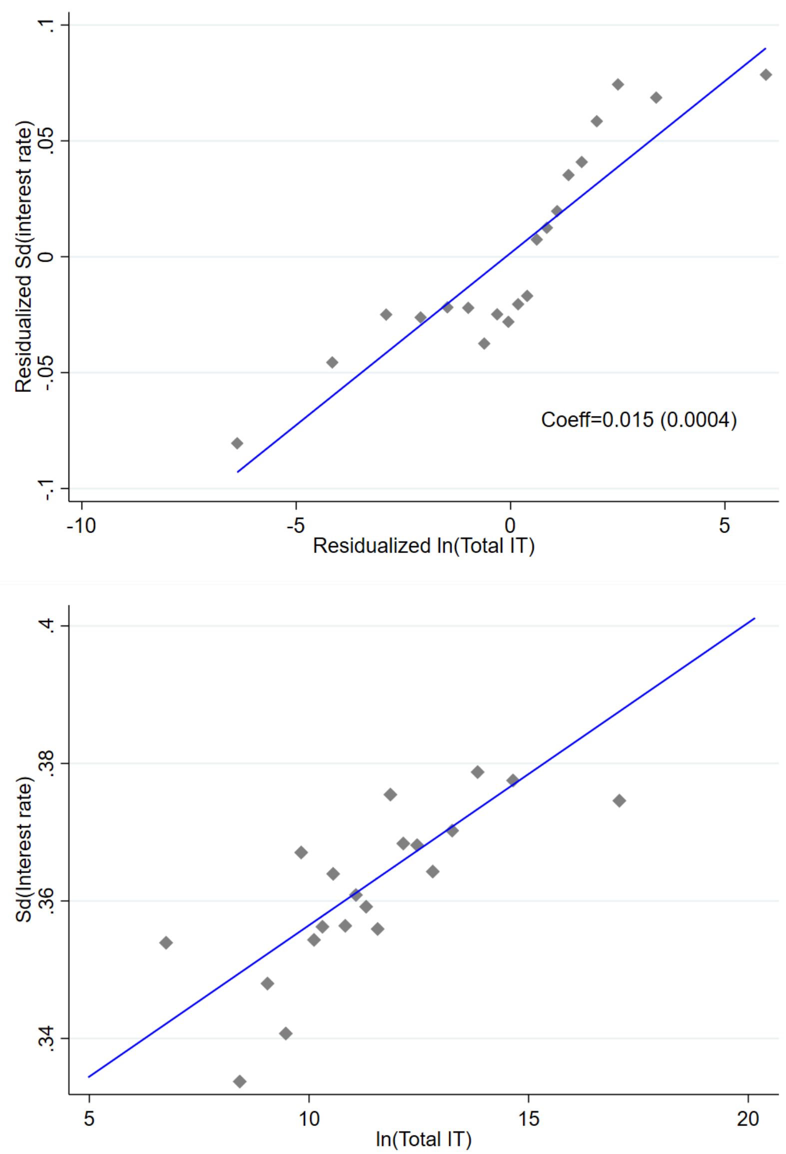
Thakor, A. V. (1996). Capital requirements, monetary policy, and aggregate bank lending: Theory and empirical evidence. *The Journal of Finance*, 51(1):279–324.

Vanasco, V. (2017). The downside of asset screening for market liquidity. *The Journal of Finance*, 72(5):1937–1982.

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

Yannelis, C. and Zhang, A. L. (2021). Competition and selection in credit markets. *Working Paper*.

Figure 2. Residualized Interest Rate Dispersion and Residualized IT Spending

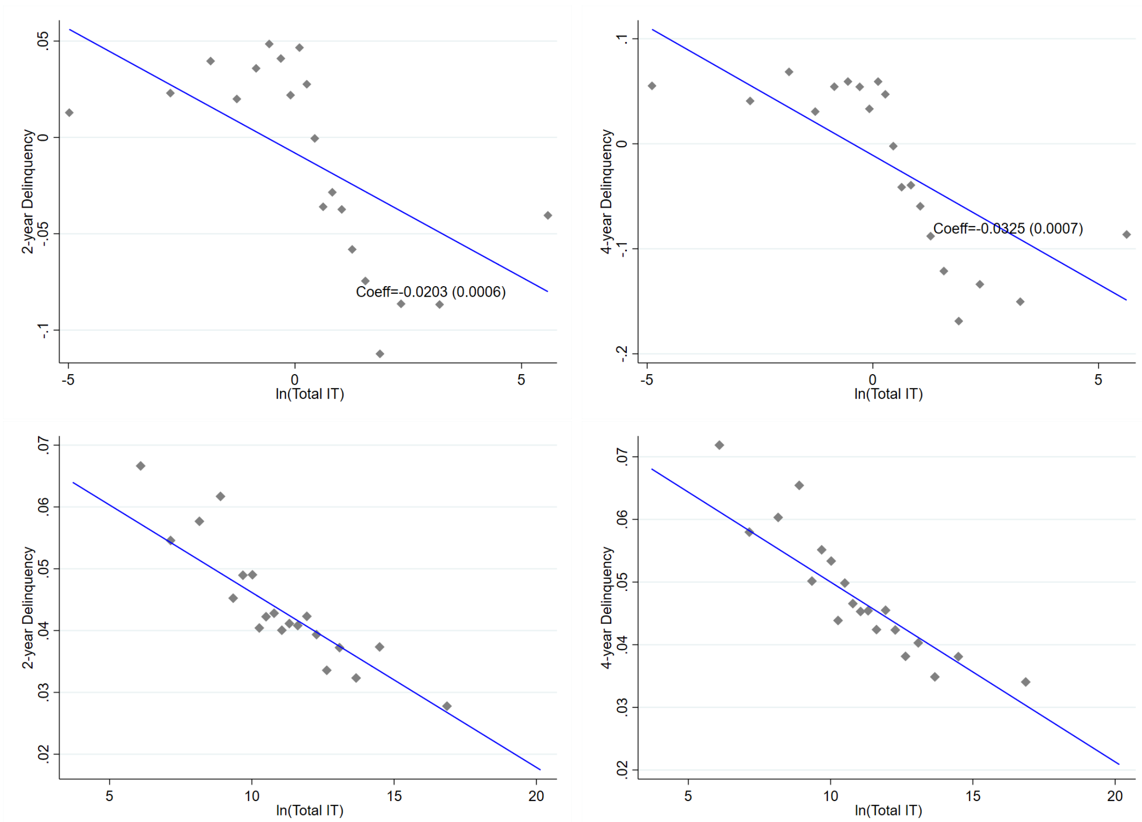


Notes: The below shows the binned scatter plot between the residualized bank IT spending and the residualized interest rate dispersionas well as the binned scatter plot between IT spending and interest rate dispersion following Cattaneo et al. (2022). Specifically, we estimate the residuals of interest rate dispersion $\nu_{i,c,t}$ and residuals of bank IT spending $\phi_{i,c,t}$ with the following regression specification:

$$\begin{aligned}\text{Int-rate disp}_{i,c,t} &= \alpha_i + \mu_c + \eta_t + \beta_1\text{FICO disp} + \beta_2\text{LTV disp} + \beta_4\mathbf{X}_{i,c,t} + \nu_{i,c,t}\\ \ln(\text{Total IT})_{i,c,t} &= \alpha_i + \mu_c + \eta_t + \beta_1\text{FICO disp} + \beta_2\text{LTV disp} + \beta_4\mathbf{X}_{i,c,t} + \phi_{i,c,t}\end{aligned}$$

"FICO disp" and "LTV disp" are the standard deviation of FICO scores and LTV ratios of all the loans issued by bank i in county c in year t. In the control variables, we include the standard deviation of logarithmic of loan amounts of all the mortgages issued by bank i in county c in year t, we control for the net revenue scaled by total assets and total deposits scaled by total assets of bank i in year t. Time-varying county-level controls include population growth rate, real GDP growth rate, changes in unemployment rate, house price index growth, and local banks deposit HHI.

Figure 3. Residualized Delinquency and Residualized IT Spending



Notes: The below shows the binned scatter plot between the residualized bank IT spending and the residualized delinquency rates as well as the binned scatter plot between IT spending and delinquency rates following Cattaneo et al. (2022). Specifically, we estimate the residuals of the delinquency rate $\psi_{i,c,t}$ and residuals of bank IT spending $\phi_{i,c,t}$ with the following regression specification:

$$\begin{aligned}\text{Delinq rate}_{i,c,t} &= \alpha_i + \mu_c + \eta_t + \beta_1 \text{Ave FICO}_{i,c,t} + \beta_2 \text{Ave LTV}_{i,c,t} + \beta_4 \mathbf{X}_{i,c,t} + \psi_{i,c,t} \\ \log(\text{Total IT})_{i,c,t} &= \alpha_i + \mu_c + \eta_t + \beta_1 \text{Ave FICO}_{i,c,t} + \beta_2 \text{Ave LTV}_{i,c,t} + \beta_4 \mathbf{X}_{i,c,t} + \phi_{i,c,t}\end{aligned}$$

"Deliq rate$_{i,c,t}$" is the proportion of loans issued by bank i in county c in year t that were deliquent after 2 years/4 years after origination. "Ave FICO" and "Ave LTV" are the average of FICO scores and LTV ratios of all the loans issued by bank i in county c in year t. In the control variables, we include average logarithmic of loan amounts of all the mortgages issued by bank i in county c in year t, we control for the net revenue scaled by total assets and total deposits scaled by total assets of bank i in year t. Time-varying county-level controls include population growth rate, real GDP growth rate, changes in unemployment rate, house price index growth, and local banks deposit HHI.

Table 1. Summary Statistics

Panel A: Loan Characteristics	N	Mean	S.d.	p25	p50	p75
Loan amount	10430621	455527.2	7140893	125000	213000	345000
Interest rate	1592892	4.475	0.527	4.000	4.500	4.875
Dispersion of Interest rate	192145	0.481	0.205	0.191	0.401	0.523
Debt-to-income (bin)	4212828	3.515	1.309	2.000	4.000	4.000
ln(Applicant income)	7756177	11.117	0.748	10.645	11.112	11.599
FICO score	1848515	753.475	43.362	723.000	762.000	790.000
LTV ratio	4259597	81.164	15.534	74.588	82.692	95.000
Application rejected	8019505	0.553	0.497	0.000	1.000	1.000
Approved but denied by applicant	8019505	0.150	0.358	0.000	0.000	0.000
Loan originated	8019505	0.297	0.457	0.000	0.000	1.000
New Purchase	8019505	0.778	0.415	1.000	1.000	1.000
Female applicant	7351722	0.376	0.484	0.000	0.000	1.000
Delinquency in 2 years	1850612	0.051	0.221	0.000	0.000	0.000
Delinquency in 2 years	1850755	0.060	0.238	0.000	0.000	0.000
Delinquency in 6 years	1850859	0.062	0.241	0.000	0.000	0.000
Panel B: Bank IT (Bank-year level)	N	Mean	S.d.	p25	p50	p75
ln(Assets)	21560	12.977	1.425	12.025	12.734	13.625
Total IT/Assets	21560	0.003	0.005	0.000	0.001	0.003
Total IT/Revenue	21527	0.268	0.463	0.015	0.059	0.309
Total IT/Expenses	21527	0.305	0.521	0.017	0.069	0.356
ROA	21527	-0.001	0.000	-0.001	-0.001	-0.001
Equity/Assets	21547	0.108	0.029	0.091	0.103	0.119
Deposits/Assets	21547	0.833	0.062	0.801	0.846	0.879
Agricultural loan/Total loan	21545	0.036	0.072	0.000	0.003	0.034
C&I loan/Total loan	21205	0.123	0.084	0.064	0.109	0.167
Real estate loan/Total loan	21205	0.771	0.139	0.694	0.789	0.873
Panel C: Bank IT (Bank-County-year level)	N	Mean	S.d.	p25	p50	p75
ln(Total IT)	935555	11.314	2.174	9.847	11.248	12.521
ln(Software IT)	935555	10.223	2.277	8.833	10.275	11.517
ln(Communication IT)	935555	8.650	2.320	7.377	8.649	9.913
ln(Employee)	937168	2.665	1.388	1.792	2.398	3.401
ln(Revenue)	716120	15.438	1.563	14.509	14.914	16.213
Number of branches	2261397	1.213	1.921	1.000	1.000	1.000

Notes: Table below provides the summary statistics of the trajectory of mortgage loan origination and banks' IT spending. Panel A show the summary statistics of loan decision, mortgage rates, loan characteristics, and loan performances of the 30% randomly generated matched HMDA-GSE sample. Panel B shows the banks' IT spending and balance sheet conditions at bank-year level. Panel C shows banks' IT spending, revene, and employees at bank-county-year level.

Table 2. Determinants of IT Spending: I

Panel A: Local Tech Infrastructure	ln(Total IT)				
	(1)	(2)	(3)	(4)	(5)
% Pop with access	0.0429***				
	(0.0043)				
Average speed (population weighted)		0.0544***			
		(0.0045)			
Max downstream speed			0.0913***		
			(0.0074)		
Max upstream speed				0.0667***	
				(0.0056)	
Ave dist					-0.0221***
					(0.0061)
Year FE	✓	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓	✓
N	434,303	434,303	434,303	434,303	846,159
R^2	0.45	0.45	0.59	0.59	0.58
Panel B: Local Credit Market Profile	ln(Total IT)				
	(1)	(2)	(3)	(4)	(5)
L.FICO	-0.0069				
	(0.0102)				
L.DTI		0.0088***			
		(0.0030)			
L.Securitization			-0.4125***		
			(0.0124)		
L.Conforming				-0.4935***	
				(0.0128)	
L.Delinquency					0.0712**
					(0.0286)
Year FE	✓	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓	✓
N	33,256	51,872	60,181	60,181	59,537
R^2	0.53	0.56	0.51	0.52	0.48

Notes: This table presents the how banks' IT spending is influenced by local technology infrastructure and local credit market profile. The regression specification is as follows:

$$\log(\text{Total IT})_{i,c,t} = \alpha_i + \mu_t + \beta \text{Local Characteristics}_{i,c,t} + \pi \text{X}_{c,t} + \epsilon_{i,c,t}$$

The dependent variable $\log(\text{Total IT})_{i,c,t}$ is the natural logarithmic of total IT spending of bank i in county c in year t. In Panel A, %Pop with access is the percentage of population that has at one supplier offering fixed broadband services. %Average speed is the population-weighted average downstream speed of fixed broadband services. Max downstream speed and Max upstream speed are the maximum downstream and upstream speed offered by local broadband service providers to local businesses. "Ave dist" is the average transportation distance from the county to the computing hardware and data storage device providers' warehouse. In Panel B, "L.FICO" is the lagged average fico score of loans issued by bank i in county c, "L.DTI" is the lagged debt-to-income ratio of loans issued by bank i in county c, "L.Securitization proportion" is the lagged proportion of mortgaged loans that was securitized by bank i in county c, "L.Conforming proportion" is the lagged proportion of mortgage issued by bank i satisfying conforming loan limit in county c, and "L.Delinquency" is the lagged proportion of mortgage loans issued by bank i in county c that were delinquent within 2 years. Control variables include real GDP per capita and population growth. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets.

Table 3. Determinants of IT Spending: II

Panel A: Local Economic Condition	ln(Total IT)				
	(1)	(2)	(3)	(4)	(5)
HPI	0.0645***				
	(0.0061)				
% College graduates		0.0663***			
		(0.0037)			
Unemployment rate			-0.0168***		
			(0.0046)		
Δ Establishments				0.0171***	
				(0.0021)	
% Pop 65 and older					-0.0414***
					(0.0035)
Year FE	✓	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓	✓
N	282,939	297,469	297,469	297,469	290,735
R^2	0.53	0.53	0.53	0.53	0.53
Panel B: Bank Local Presence	ln(Total IT)				
and Balance Sheet Condition	(1)	(2)	(3)	(4)	(5)
Deposit share	0.1242***				
	(0.0064)				
Deposit HHI		-0.0725***			
		(0.0099)			
Deposits/Loan			0.0485***		
			(0.0047)		
Profitability				0.1027***	
				(0.0334)	
HMDA share					0.0365***
					(0.0019)
Year FE	✓	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓	✓
N	100,297	100,297	297,770	297,469	618,492
R^2	0.45	0.44	0.53	0.53	0.59

Notes: This table presents the how banks' IT spending is determined by local economic condition and banks' balance sheet condition. The regression sepcification is as follows:

$$\log(\text{Total IT})_{i,c,t} \quad = \quad \alpha_i + \mu_t + \beta \text{Local Characteristics}_{i,c,t} + \pi \text{X} + \epsilon_{i,c,t}$$

The dependent variable $\log(\text{Total IT})_{i,c,t}$ is the natural logarithmic of total IT spending of bank i in county c in year t. In Panel A, "HPI" is the FHFA house price index, "% Adult with College and Higher" is the percentage of adult with college degree or higher, "Unemployment rate" is the unemployment rate of a county in a given year, "Δ Establishment is the change in change in logarithmic of total business establishments from last year, "%Pop 65 and older" is the percentage of population that is older than 65 years old. In Panel B, "Deposit HHI" is the deposit concentration measured by HHI in county c, "Deposit share" is bank i's deposit market share in county c, "Deposits/Assets" is bank i's deposit as a proportion of total asset in year t, "Profitability" is bank i's total revenue scaled by total number of employees in county c and year t, "HMDA share" is bank i's market share of mortgage lending in county c and year t. Control variables include real GDP per capita and population growth. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets.

Table 4. IT Spending and Loan Decision Making

Panel A: OLS	$\mathbb{1}_{\text{Rejection}}$	$\mathbb{1}_{\text{Approve but denial}}$	$\mathbb{1}_{\text{Originated}}$
	(1)	(2)	(3)
ln(Total IT)	-0.0186***	-0.0096***	0.0218***
	(0.0019)	(0.0018)	(0.0020)
Controls	✓	✓	✓
Loan feature FE	✓	✓	✓
Lender FE	✓	✓	✓
Year FE	✓	✓	✓
County FE	✓	✓	✓
N	1,992,869	1,992,869	1,992,869
AdjR^2	0.09	0.04	0.09

Panel B: 2SLS	1-st stage	1[Rejection]	1[Approval but denial]	1[Originated]
	(1)	(2)	(3)	(4)
HP exposure	0.0733***			
	(0.0134)			
ln(Total IT)		-0.1115***	-0.0936**	0.1508***
		(0.0392)	(0.0430)	(0.0490)
Controls	✓	✓	✓	✓
Loan feature FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓
County FE	✓	✓	✓	✓
F-stat	450			
N	1,993,062	1,993,062	1,993,062	1,993,062
AdjR^2	0.63	0.01	-0.01	0.00

Notes:
This tables show how does banks' IT spending impact banks' loan decision making in OLS and 2SLS regression specifications. The OLS regression equation is as follows:

$$\text{Loan decision}_{l,i,c,t} = \alpha_i + \mu_c + \eta_t + \beta \times \log(\text{Total IT})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \textbf{FE's} + \epsilon_{l,i,c,t}$$

The 2SLS regreesion specification is as follows:

$$\begin{aligned} \log(\text{Total IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{HP exposure}_{i,t} + \pi_1 \mathbf{X} + \textbf{FE's} + \epsilon_{i,c,t} \\ \text{Loan decision}_{l,i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Total IT}})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \textbf{FE's} + \epsilon_{l,i,c,t} \end{aligned}$$

The dependent variables in the above OLS regression are the dummy variable indicating three possible decision status of a loan l received by bank i in county c in year t. The definitions of the dependent variables are provided in Section 5.2. "HP exposure$_{i,t}$" is weighted sum of house price index of a county, weighted by the banks' deposit in the county among the bank's total deposit. The detailed definition of "HP exposure$_{i,t}$" is provided in Section 5.2.2. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include logarithmic of loan amount and logarithmic of borrower income. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table 5. Decision and IT by Observable Default Risk

Panel A: OLS	$\mathbb{1}_{\text{Rejection}}$		$\mathbb{1}_{\text{Approve but denial}}$	
	Low def risk	High def risk	Low def risk	High def risk
ln(Total IT)	-0.0027	-0.0110***	-0.0103***	-0.0013
	(0.0051)	(0.0041)	(0.0021)	(0.0032)
Controls	✓	✓	✓	✓
Loan feature FE	✓	✓	✓	✓
Lender FE	✓	✓	✓	✓
County FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
N	300,359	298,322	300,359	298,322
AdjR^2	0.09	0.09	0.04	0.04

Panel B: 2SLS	$\mathbb{1}_{\text{Rejection}}$		$\mathbb{1}_{\text{Approve but denial}}$	
	Low def risk	High def risk	Low def risk	High def risk
$\ln(\widehat{\text{Total IT}})$	-0.1081**	-0.2026***	-0.0391***	0.0115
	(0.0511)	(0.0547)	(0.0090)	(0.0458)
Controls	✓	✓	✓	✓
Loan feature FE	✓	✓	✓	✓
Lender FE	✓	✓	✓	✓
County FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
N	125,644	104,038	126,821	104,038
AdjR^2	-0.14	-0.17	-0.18	-0.13

Notes: This table shows how is banks' loan making decisions differently associated with banks' IT spending for *high-risk borrowers* and *low-risk borrowers*. The regression specifications are identical to those in Panel B of Table 4. The dependent variables are the dummy variable indicating possible decision status of a loan l received by bank i in county c in year t. The definitions of the dependent variables are provided in Section 5.2. The regression specification is ran for "High def risk" borrowers and "Low def risk" borrowers respectively. "High def risk" borrowers are borrowers whose estimated 2-year delinquency rates based on observables characteristics are above median in a county and a year. Explanation of constructions of borrowers' risk profile is provided in Section 5.2.3. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include logarithmic of loan amount and logarithmic of borrower income. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table 6. Pricing Precision: Interest Rate Dispersion

	Sd(Interest rate)		
	(1)	(2)	(3)
HP Exposure	0.0323***		
	(0.0041)		
ln(Total IT)		0.7567***	0.0438***
		(0.2540)	(0.0166)
Controls	✓	✓	✓
Loan feature FE	✓	✓	✓
Lender FE	✓	✓	✓
Year FE	✓	✓	✓
County FE	✓	✓	✓
F-stat	32		
N	812,756	18,458	20,354
AdjR^2	0.47	-0.29	0.12

Notes: This table shows how does banks' IT spending affect the loan pricing dispersion. The 2SLS regression is specified as follows:

$$\begin{aligned}
\log(\text{Total IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{HP exposure}_{i,t} + \pi_1 \mathbf{X} + \mathbf{FE's} + \epsilon_{i,c,t} \\
\text{Interest rate dispersion}_{i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Total IT}})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE's} + \epsilon_{i,c,t}
\end{aligned}$$

The OLS regression specification is as follows

$$\text{Interest rate dispersion}_{i,c,t} = \alpha_i + \mu_c + \eta_t + \beta \times \log(\text{Total IT})_{i,c,t} + \beta_1 \text{X} + \epsilon_{i,c,t}$$

The dependent variable is the standard deviation of interest rates of all mortgage originated by bank i in county c in year t. "HP exposure$_{i,t}$" is weighted sum of house price index of a county, weighted by the banks' deposit in the county among the bank's total deposit. The detailed definition of "HP exposure$_{i,t}$" is provided in Section 5.2.2. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include average logarithmic of loan amount, average logarithmic of borrower income, the average FICO scores and average loan-to-value ratios of all the loans made by a bank in a county in a specific year. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table 7. Pricing Precision: Interest Rate Level by Observable Default Risk

	Interest rate			
	(1)	(2)	(3)	(4)
1[High Def risk]	0.0998***	0.0961***		0.0689***
	(0.0050)	(0.0053)		(0.0052)
1[High spender]×1[High def risk]		0.0327**		0.0323**
		(0.0152)		(0.0151)
1[High spender]		-0.0238***	-0.0110*	-0.0211***
		(0.0069)	(0.0065)	(0.0071)
1[High spender]×1[Low def risk]			-0.0170**	-0.0070
			(0.0076)	(0.0059)
1[Low def risk]			-0.1038***	-0.0885***
			(0.0031)	(0.0028)
Controls	✓	✓	✓	✓
Loan feature FE	✓	✓	✓	✓
Lender FE	✓	✓	✓	✓
County FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
N	392,229	392,211	392,211	392,211
AdjR^2	0.40	0.40	0.40	0.41

Notes: This table shows how mortgage interest rates offered to borrowers with different observable delinquency risks (delinquency risk estimated based on observable characteristics) are differently correlated with the banks' IT spending. The regression specification is as follows

$$\begin{aligned} \text{Interest rate}_{l,i,c,t} &= \beta \times 1[\text{High spender}]_{i,c,t} \times 1[\text{Low/High Def risk}]_{l,c,t} + \beta_1 1[\text{High spender}]_{l,c,t} \\ &+ \beta_2 1[\text{Low/High Def risk}]_{l,c,t} + \alpha_i + \mu_c + \eta_t + \beta_3 \mathbf{X} + \mathbf{FE}s + \epsilon_{l,i,c,t} \end{aligned}$$

The dependent variable is the interest rate of loan l received by bank i in county c and in year t. "High def risk" borrowers are borrowers with estimated default probabilities above 75-th percentile among all borrowers in a year. "Low def risk" borrowers are borrowers with estimated default probabilities below 25-th percentile among all borrowers in a year. A bank is defined as a "high spender" if the bank's IT spending scaled by total revenue is above median among all banks in the same county in a specific year. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include the logarithmic of loan amount and the logarithmic of borrower income. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table 8. Pricing Precision: Interest Rate Level by Observable Default Risk (2SLS)

	ln(Total IT) (1-st stage)	Interest rate (2nd stage)	
	(1)	(2)	(3)
	1-st stage	High def risk	Low def risk
HP exposure	0.5292***		
	(0.1580)		
ln(Total IT)		2.9923***	-0.3173**
		(0.5815)	(0.1491)
Controls	✓	✓	✓
Loan feature FE	✓	✓	✓
Lender FE	✓	✓	✓
County FE	✓	✓	✓
Year FE	✓	✓	✓
N	414,251	89,035	119,714
AdjR^2	0.70	-0.05	-0.06

Notes: This table shows how mortgage interest rates offered to borrowers with different observable delinquency risks (delinquency risk estimated based on observable characteristics) are differently correlated with the banks' IT spending in 2SLS specification. The regression specification is as follows

$$\begin{aligned} \log(\text{Total IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{HP exposure}_{i,t} + \pi_1 \mathbf{X} + \mathbf{FE's} + \epsilon_{i,c,t} \\ \text{Interest rate}^{\text{high def risk}}_{l,i,c,t} / \text{Interest rate}^{\text{low def risk}}_{l,i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Total IT}})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE's} + \epsilon_{i,c,t} \end{aligned}$$

The dependent variables in the second stage are the interest rate of loan l received by bank i in county c and in year t, where the borrower is either the a "high def risk" borrower or a "low def risk borrower." "High def risk" borrowers are borrowers with estimated default probabilities above 75-th percentile among all borrowers in a year. "Low def risk" borrowers are borrowers with estimated default probabilities below 25-th percentile among all borrowers in a year. "HP exposure$_{i,t}$" is weighted sum of house price index of a county, weighted by the banks' deposit in the county among the bank's total deposit. The detailed definition of "HP exposure$_{i,t}$" is provided in Section 5.2.2. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include the logarithmic of loan amount and the logarithmic of borrower income. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table 9. Loan Outcome: Delinquency Status After Origination

Panel A: OLS		$\mathbb{1}_{\text{Delinq 2 years}}$	$\mathbb{1}_{\text{Delinq 4 years}}$	$\mathbb{1}_{\text{Delinq 6 years}}$
	(1)	(2)	(3)	(4)
ln(Total IT)		-0.0132***	-0.0121***	-0.0115***
		(0.0046)	(0.0045)	(0.0044)
Controls	✓		✓	✓
Loan feature FE	✓		✓	✓
Lender FE, County FE, Year FE		✓	✓	✓
N	136,195		136,195	136,195
AdjR^2	0.03		0.03	0.03
Panel B: 2SLS	1-st stage	$\mathbb{1}_{\text{Delinq 2 years}}$	$\mathbb{1}_{\text{Delinq 4 years}}$	$\mathbb{1}_{\text{Delinq 6 years}}$
	(1)	(2)	(3)	(4)
HP exposure	0.0733***			
	(0.0134)			
ln(Total IT)		-0.5157***	-0.5302***	-0.5154***
		(0.1260)	(0.1275)	(0.1280)
Controls	✓	✓	✓	✓
Loan feature FE	✓	✓	✓	✓
Lender FE, County FE, Year FE	✓	✓	✓	✓
F-stat	450			
N	1,993,062	136,306	136,306	136,306
AdjR^2	0.63	-0.03	-0.04	-0.04

Notes: This table shows how does banks' IT spending impact the loan portfolio performance as measured by loan delinquency rates. The regression specification is as follows:

$$\mathbb{1}_{\text{Delinq x years},l,i,c,t} = \alpha_i + \mu_c + \eta_t + \beta \times \log(\text{Total IT})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE's} + \epsilon_{l,i,c,t}$$

The dependent variables in the above OLS regressions are the dummy variable indicating whether loan l originated by bank i in county c and in year t went into delinquency after x years of origination. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include the logarithmic of loan amount, the logarithmic of borrower income, FICO, and LTV ratio. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table 10. Loan Outcome: Delinquency Status for High-risk Borrowers

	$\mathbb{1}_{\text{Delinq 2 years}}$	$\mathbb{1}_{\text{Delinq 4 years}}$	$\mathbb{1}_{\text{Delinq 6 years}}$
	(1)	(2)	(3)
1[High spender]	0.0072	0.0084	0.0084
	(0.0065)	(0.0069)	(0.0069)
1[High spender]×1[High def risk]	-0.0132*	-0.0180**	-0.0180**
	(0.0073)	(0.0078)	(0.0078)
1[High def risk]	0.0180***	0.0216***	0.0216***
	(0.0018)	(0.0019)	(0.0019)
Controls	✓	✓	✓
Loan feature FE	✓	✓	✓
Lender FE	✓	✓	✓
County FE	✓	✓	✓
Year FE	✓	✓	✓
N	111,899	111,899	111,899
AdjR^2	0.02	0.03	0.03

Notes: This table shows how loan performances are correlated with the lending banks' IT spending for borrowers that were perceived to be high-risk borrowers based on observables at origination. The regression specification is as follows:

$$\begin{aligned} 1[\text{Delinquent in x years}]_{l,i,c,t} \quad &= \alpha_i + \mu_c + \eta_t + \beta \times 1[\text{High spender}] \times 1[\text{High def risk}] \\ &\beta_1 1[\text{High def risk}] + \beta_2 1[\text{High spender}] + \beta_3 \mathbf{X} + \mathbf{FE}s + \epsilon_{l,i,c,t} \end{aligned}$$

The dependent variables are the dummy variables indicating whether loan l originated by bank i in county c and in year t went into delinquency after x years of origination. "High def risk" borrowers are borrowers with estimated default probabilities above 50-th percentile among all borrowers in the sample. Bank fixed effects, year fixed effects and county fixed effects are included. We include the borrowers' FICO, LTV ratio and logarithmic of loan amount in the loan controls. Other control variables include population growth rate, real GDP per capita, unemployment rate, and house price index growth that capture time-varying county-level economic characteristics. We further control for time-varying bank characteristics including the logarithmic of bank i's total assets in year t, the net revenue scaled by total assets of bank i in year t and the total deposits scaled by total loans of bank i in year t. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table 11. Interest Rate Conditional on Delinquency

	Interest rate	
	(1)	(2)
$\mathbb{1}_{\text{Delinq 2 years}}$	0.1302***	
	(0.0049)	
1[High spender]	-0.0221***	-0.0120*
	(0.0072)	(0.0069)
1[High spender]$\times\mathbb{1}_{\text{Delinq 2 years}}$	0.0284*	
	(0.0165)	
$\mathbb{1}_{\text{Delinq 4 years}}$		0.1101***
		(0.0044)
1[High spender]$\times\mathbb{1}_{\text{Delinq 4 years}}$		0.0467***
		(0.0158)
Controls	✓	✓
Loan feature FE	✓	✓
Lender FE, County FE, Year FE	✓	✓
N	111,194	111,194
AdjR^2	0.25	0.31

Notes: This table shows how IT is banks' IT spending correlated with the mortgage interest rates it offers to borrowers who actually defaulted after the origination. The regression specification is as follows

$$\text{Interest rate}_{l,i,c,t} = \alpha_i + \mu_c + \eta_t + \beta \times \mathbb{1}_{\text{Delinq x years},i,c,t} + \beta_1\mathbf{X} + \mathbf{FE}s + \epsilon_{l,i,c,t}$$

The dependent variable is the interest rate of loan l received by bank i in county c and in year t. The main dependent variable is the indicator variable $\mathbb{1}_{\text{Delinq x years},i,c,t}$, which equals to 1 if the loan was delinquent within x-year time window after the origination. Column (1), (3) and (5) are for the sub-sample of loans issued by high IT spenders, and column (2), (4) and (6) are for the sub-sample of loans issued by low IT spenders. A bank is defined as a "high spender" if the bank's IT spending scaled by total revenue is above median among all banks in the same county in a specific year. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include the logarithmic of loan amount and the logarithmic of borrower income. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Appendix: Additional Tables and Figures

Table A1. Summary Statistics of Local Characteristics

	N	Mean	S.d.	25-th	Median	75-th
Local Technology Infrastructure						
Property damage exposure	297166	1.030	1.316	0.040	0.388	1.672
Duration	281647	1.610	2.573	1.000	1.000	1.333
Max downstream speed	970409	1118.335	1305.594	1000.000	1000.000	1000.000
Max upstream speed	970409	904.793	1378.880	115.000	1000.000	1000.000
Ave Dist (miles)	2188986	1148.149	248.697	1014.886	1076.952	1169.768
Local Credit Market Profile						
L.FICO	73264	750.307	24.063	736.900	752.250	766.150
L.Delinquency	67129	0.001	0.009	0.000	0.000	0.000
L.DTI	190579	39.861	8.742	34.481	39.667	45.000
L.Securitization	67129	0.197	0.236	0.026	0.118	0.269
L.Conforming	67129	0.333	0.281	0.125	0.250	0.500
Local Economic Condition						
HPI	31321	145.688	31.223	125.150	139.610	159.480
% College graduates	29595	0.218	0.095	0.152	0.195	0.261
Unemployment rate	41782	0.066	0.033	0.042	0.059	0.083
Δ Establishments	26633	-0.003	0.038	-0.017	-0.000	0.014
% Pop 65 and older	39558	0.177	0.044	0.148	0.173	0.200
Bank Balance Sheet Condition						
Deposit share	297166	0.119	0.160	0.014	0.059	0.158
Deposit HHI	297171	0.220	0.140	0.130	0.179	0.264
Deposits/Loan	849680	0.783	0.112	0.735	0.805	0.859
Profitability	849312	-0.001	0.044	-0.001	-0.001	-0.001
HMDA share	922929	0.034	0.079	0.002	0.008	0.030

Notes: The table below shows the summary statistics of key determinant of local banks' IT spending. The detailed definition and the analysis are provided in Section 4.

Table A2. Summary of Biggest Technology Hardware Producers

Ticker	Company Name	Sales	Total Assets	Warehouse location
EMC	EMC CORP/MA	21863.1	40252.69	25017
CNHI	CNH INDUSTRIAL NV	27570.67	48606.67	
DE	DEERE & CO	32722.7	60774.82	
JCI	JOHNSON CONTROLS INTL PLC	35032.82	38848.55	
CAJ	CANON INC	35873.69	43892.97	34023
HPE	HEWLETT PACKARD ENTERPRISE	41320.5	64689	48225
CSCO	CISCO SYSTEMS INC	47456	102971.8	6085
CAT	CATERPILLAR INC	50974.64	79077.09	
DELL	DELL TECHNOLOGIES INC	74725.75	90656.88	48015
HPQ	HP INC	88627.55	76658.64	6085

Notes: The table below summarizes the biggest 10 commercial and industrial machinery producers and their warehouse locations based in U.S. The shaded companies are producers of information technology hardware, storage devices and server devices utilized in constructing the instrumental variables in our sample. Companies in the table are ranked by their average sales volumes during 2010-2019.

Table A3. Summary on the Number of Applications Taken by Lenders

	N	Mean	S.d.	25-th	Median	75-th
All						
Num applications	12725	207.264	214.094	119.000	151.000	222.000
Num applications/employee	11540	21.371	36.251	5.829	12.781	26.000
High IT spender						
Num applications	2642	196.181	148.737	120.000	151.000	216.000
Num applications/employee	2635	5.327	10.397	1.464	3.360	5.833
Other						
Num applications	10083	210.168	228.061	119.000	151.000	224.000
Num applications/employee	8905	26.118	39.652	9.667	16.704	30.500

Notes: The table below shows the summary statistics of the total number of applications received by high IT spending lenders and low IT spending lenders at lender-year level. "Num applications" is the total number of applications of a lender in a county in a year, "Num applications/employees" is the number of applications received per employee of a lender in a county per year.

Table A4. Bank IT Spending and Estimated Default Probability

	Estimated default risk	Sd (Estimated default risk)
	(1)	(2)
ln(Total IT)	0.0193***	0.0815***
	(0.0046)	(0.0109)
N	921,393	918,132
Lender FE	Y	Y
County FE	Y	Y
Year FE	Y	Y
AdR^2	0.19	0.31

Notes: The table below the correlation between loan applicants' estimated default probability based on observable characteristics. "Estimated default risk" is the is estimated using the LASSO estimation with borrower income, loan amount, borrower ethnicity, borrower gender, borrower age, borrower race, loan-to-value ratio (LTV), and whether there is delinquency within 2 years after the loan was originated. "Sd(Estimated default risk)" is the standard deviation of the estimated default risk of all loan applicants a bank i receives in county c in year t. Control variable include population growth rate, real GDP growth rate, changes in unemployment rate, house price index growth, and local banks deposit HHI that capture time-varying local economic and banking sector characteristics. We further control for time-varying bank characteristics including the logarithmic of a bank's total revenue in county c in year t, the logarithmic of bank i's total number of employees in county c and in year t, the net revenue scaled by total assets of bank i in year t and the total deposits scaled by total assets of bank i in year t. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table A5. Loan Decision and IT Spending: Low Income Group

	Desk reject	Approval but reject	Loan originated
	(1)	(2)	(3)
1[Low income]	0.1125***	-0.0414***	-0.0711***
	(0.0031)	(0.0020)	(0.0028)
1[High spender]×1[Low income]	-0.0254***	0.0004	0.0248***
	(0.0038)	(0.0028)	(0.0038)
1[High spender]	-0.0174***	0.0040	0.0134***
	(0.0037)	(0.0035)	(0.0032)
Loan controls	✓	✓	✓
Bank controls	✓	✓	✓
Loan feature FE	✓	✓	✓
Lender FE	✓	✓	✓
County controls	✓	✓	✓
Year FE	✓	✓	✓
N	740,673	740,673	740,673
AdR-squared	0.34	0.14	0.46

Notes: This table presents how are loan loan decisions correlated with banks' IT spending and applicants from different income groups. The regression specification is as follows

$$\begin{aligned}\text{Loan decision}_{l,i,c,t} &= \alpha_i + \mu_c + \eta_t + \beta \times \mathbb{1}_{\text{Low income}} \times \mathbb{1}_{\text{High spender}} + \beta_1 \mathbb{1}_{\text{Low income}} \\ &+ \beta_2 \mathbb{1}_{\text{High spender}} + \beta_3 \mathbf{X}_{i,c,t} + \textbf{FE's} + \epsilon_{l,i,c,t}\end{aligned}$$

The dependent variables are the dummy variable indicating whether loan application l in received by bank i in county c in year t was rejected by the lender and whether the lender made an offer to the borrower but got rejected by the borrower. Bank fixed effects, year fixed effects and county fixed effects are included. The detailed definition of the dependent variable is provided in Section 5.2. A loan applicant is defined as from "low-income" group if his/her income is below 25-th percentile of applicants from the same county and year. Control variable include population growth rate, real GDP growth rate, changes in unemployment rate, house price index growth, and local banks deposit HHI that capture time-varying local economic and banking sector characteristics. We further control for time-varying bank characteristics including the logarithmic of bank i's total revenue in county c in year t, the logarithmic of bank i's total number of employees in county c and in year t, the net revenue scaled by total assets of bank i in year t and the total deposits scaled by total assets of bank i in year t. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table A6. Loan Decision and IT Spending: Minority Group

	Desk reject	Approval but reject	Loan originated
	(1)	(2)	(3)
1[Minority]	0.0977***	-0.0245***	-0.0732***
	(0.0085)	(0.0034)	(0.0075)
1[High spender]×1[Minority]	-0.0269***	0.0006	0.0253***
	(0.0060)	(0.0037)	(0.0047)
1[High spender]	-0.0053***	0.0008	0.0045***
	(0.0014)	(0.0006)	(0.0015)
Loan controls	✓	✓	✓
Bank controls	✓	✓	✓
Loan feature FE	✓	✓	✓
Lender FE	✓	✓	✓
County controls	✓	✓	✓
Year FE	✓	✓	✓
N	680,877	680,877	680,877
AdR-squared	0.33	0.14	0.45

Notes: This table presents how are loan decisions correlated with banks' IT spending and applicants from different ethnic groups. The regression specification is as follows

$$\begin{aligned}\text{Loan decision}_{l,i,c,t} &= \alpha_i + \mu_c + \eta_t + \beta \times \mathbb{1}_{\text{Minority}} \times \mathbb{1}_{\text{High spender}} + \beta_1 \mathbb{1}_{\text{Minority}} \\ &+ \beta_2 \mathbb{1}_{\text{High spender}} + \beta_3 \mathbf{X}_{i,c,t} + \textbf{FE's} + \epsilon_{l,i,c,t}\end{aligned}$$

The dependent variables are the dummy variable indicating whether loan application l in received by bank i in county c in year t was rejected by the lender, whether the lender made an offer to the borrower but got rejected by the borrower, and whether the loan is originated. Bank fixed effects, year fixed effects and county fixed effects are included. The detailed definition of the dependent variable is provided in Section 5.2. A loan applicant is defined as from "Minority" group if his/her ethnicity is Hispanic or Lantino. Control variable include population growth rate, real GDP growth rate, changes in unemployment rate, house price index growth, and local banks deposit HHI that capture time-varying local economic and banking sector characteristics. We further control for time-varying bank characteristics including the logarithmic of bank i's total revenue in county c in year t, the logarithmic of bank i's total number of employees in county c and in year t, the net revenue scaled by total assets of bank i in year t and the total deposits scaled by total assets of bank i in year t. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table A7. Loan Decision and IT Spending: Minority Group

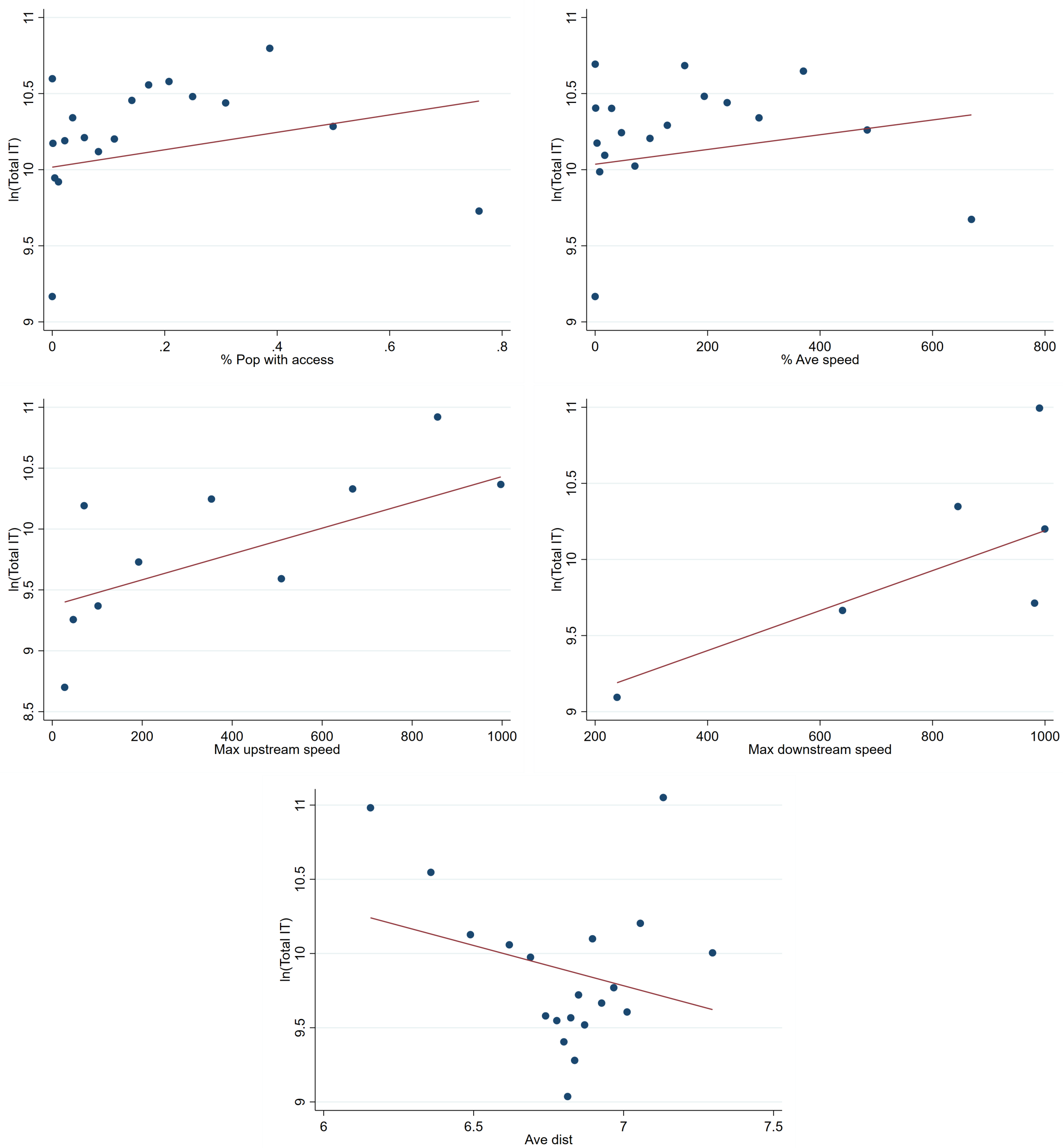
	Loan interest rate	
	(1)	(2)
1[Low income]	-0.1065***	
	(0.0047)	
1[High spender]×1[Low income]	0.0194***	
	(0.0055)	
1[High spender]	-0.0134***	-0.0068
	(0.0048)	(0.0045)
1[Minority]		0.0267***
		(0.0078)
1[High spender]×1[Minority]		0.0049
		(0.0098)
Loan controls	✓	✓
Bank controls	✓	✓
Loan feature FE	✓	✓
Lender FE	✓	✓
County controls	✓	✓
Year FE	✓	✓
N	173,351	163,196
AdR-squared	0.32	0.32

Notes: This table shows how interest rates are correlated with banks' IT spending for borrowers from different income groups and ethnicity groups. The regression specification is as follows

$$\begin{aligned}\text{Interest rate}_{l,i,c,t} &= \beta \times 1[\text{High spender}]_{i,c,t} \times 1[\text{Low/High Def risk}]_{l,c,t} + \beta_1 1[\text{High spender}]_{l,c,t} \\ &+ \beta_2 1[\text{Low/High Def risk}]_{l,c,t} + \alpha_i + \mu_c + \eta_t + \beta_3 \mathbf{X} + \mathbf{FE}s + \epsilon_{l,i,c,t}\end{aligned}$$

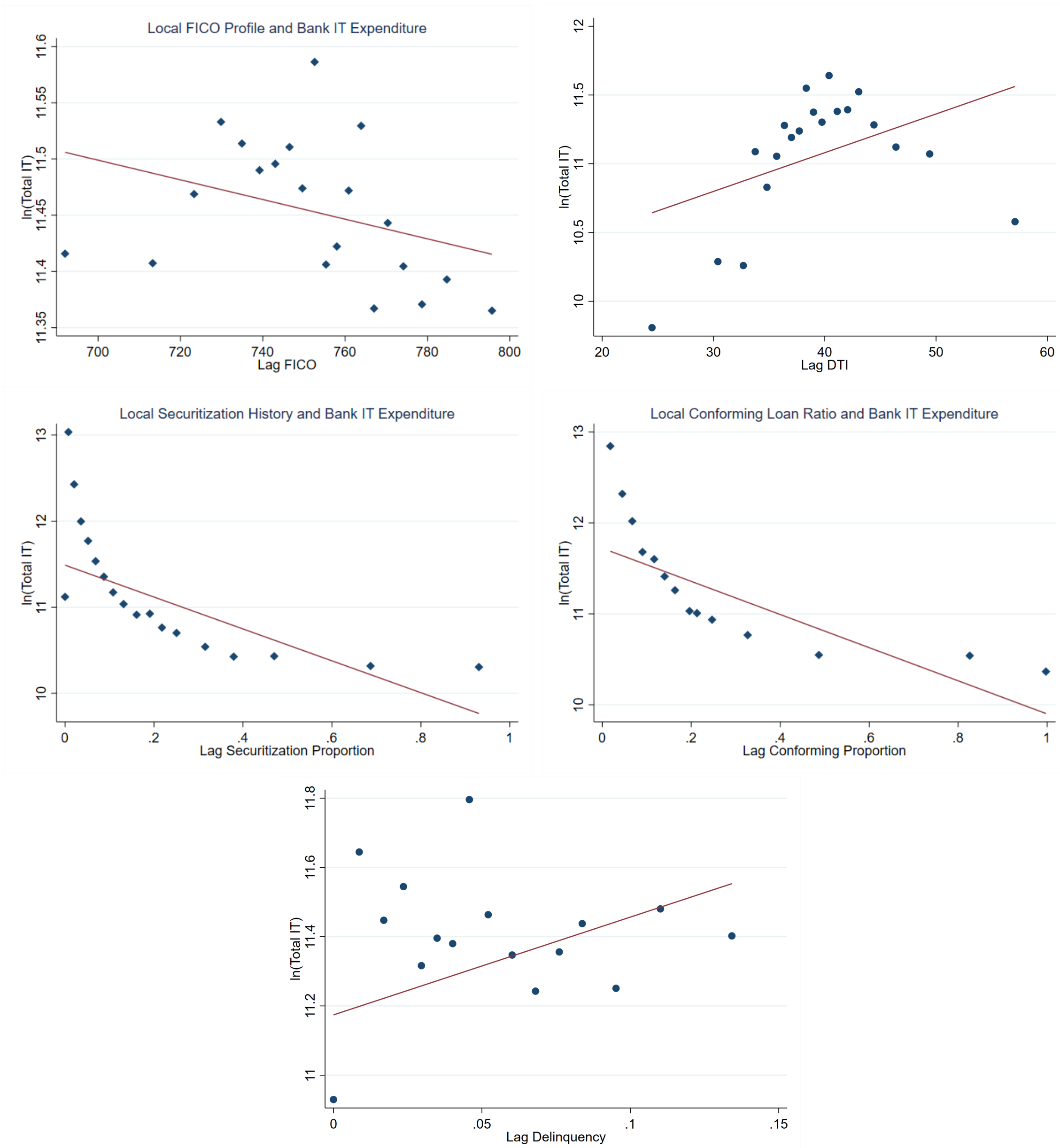
The dependent variable is the interest rate of loan l received by bank i in county c and in year t. A loan applicant is defined as from "Minority" group if his/her ethnicity is Hispanic or Lantino. A loan applicant is defined as from "low-income" group if his/her income is below 25-th percentile of applicants from the same county and year. A bank is defined as a "high spender" if the bank's IT spending scaled by total revenue is above median among all banks in the same county in a specific year. Bank fixed effects, year fixed effects and county fixed effects are included. Control variable include population growth rate, real GDP growth rate, changes in unemployment rate, house price index growth, and local banks deposit HHI that capture time-varying local economic and banking sector characteristics. We further control for time-varying bank characteristics including the logarithmic of bank i's total revenue in county c in year t, the logarithmic of bank i's total number of employees in county c and in year t, the net revenue scaled by total assets of bank i in year t and the total deposits scaled by total assets of bank i in year t. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Figure A1. Determinants of IT Spending: Local Technology Infrastructure



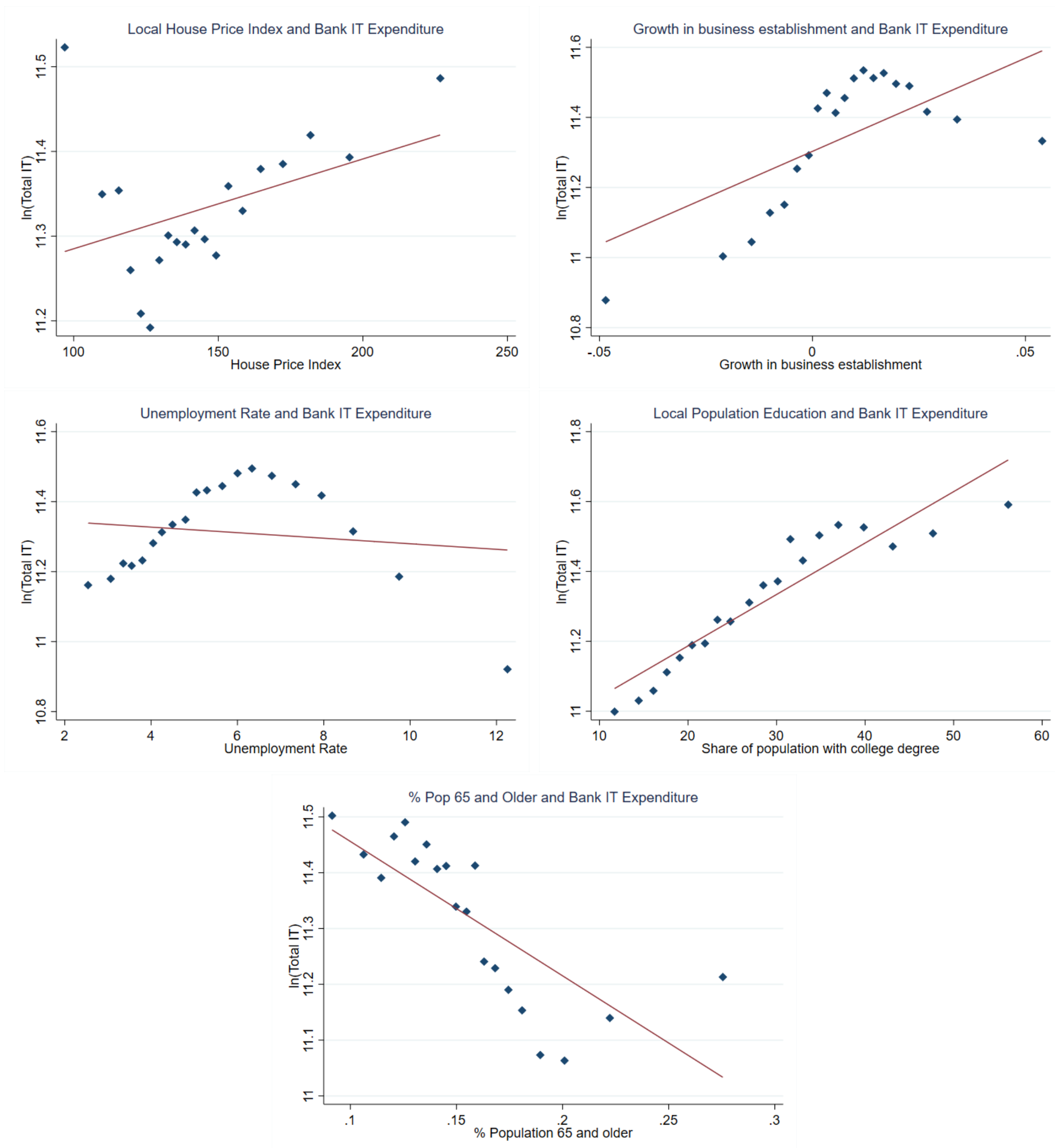
Notes: The figures are binned scatter plot demonstrating the correlation between banks' average technology spending in a county and the technology infrastructure of a county. The y-axis of each figure is the average annual logarithmic of total IT spending of a bank in a county during 2010-2020. The x-axis are the explanatory variables split into 20 equalized bins. "%Pop with access" is the percentage of population that has at one supplier offering fixed broadband services. "%Average speed" is the population-weighted average downstream speed of fixed broadband services. "Max downstream speed" and "Max upstream speed" are the maximum downstream and upstream speed offered by local broadband service providers to local businesses. "Ave dist" is the average transportation distance from the county the computing hardware and data storage device providers' warehouse.

Figure A2. Determinants of IT Spending: Local Credit Market Profile



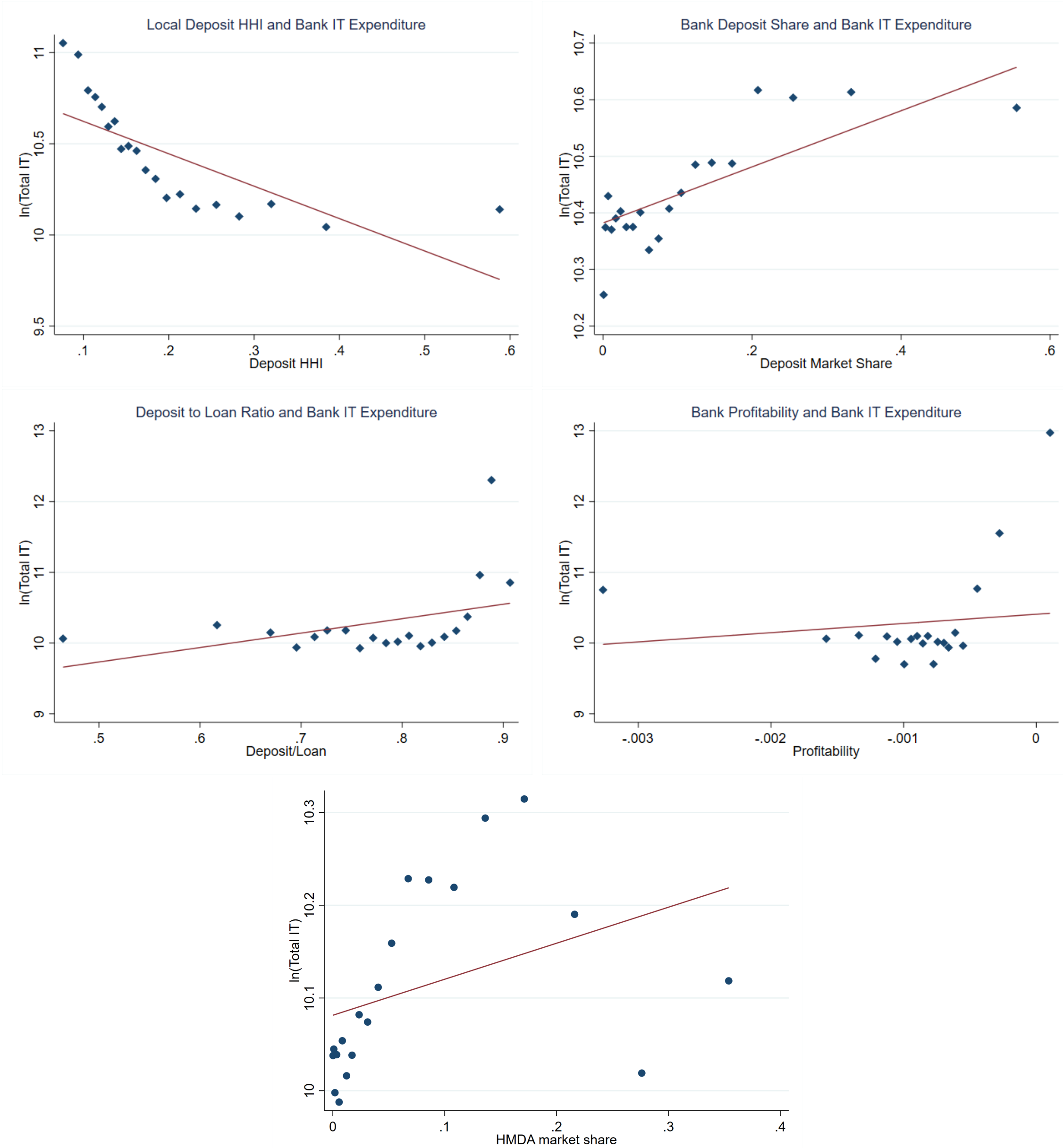
Notes: The figures are binned scatter plot demonstrating the correlation between banks' average technology spending in a county and the local credit market profile. The y-axis of each figure is the average annual logarithmic of total IT spending of a bank in a county during 2010-2020. The x-axises are the explanatory variables split into 20 equalized bins. "Lag FICO" is the lagged average fico score of loans issued by a bank in county c, "Lag DTI" is the lagged debt-to-income ratio of loans issued by a bank in county c, "Lag Securitization proportion" is the lagged proportion of mortgaged loans that was securitized by a bank in county c, "Lag Conforming proportion" is the lagged proportion of mortgage issued by a bank satisfying conforming loan limit in county c, and "Lag Delinquency" is the lagged proportion of mortgage loans issued a bank in county c that were delinquent within 2 years.

Figure A3. Determinants of IT Spending: Local Economic Condition



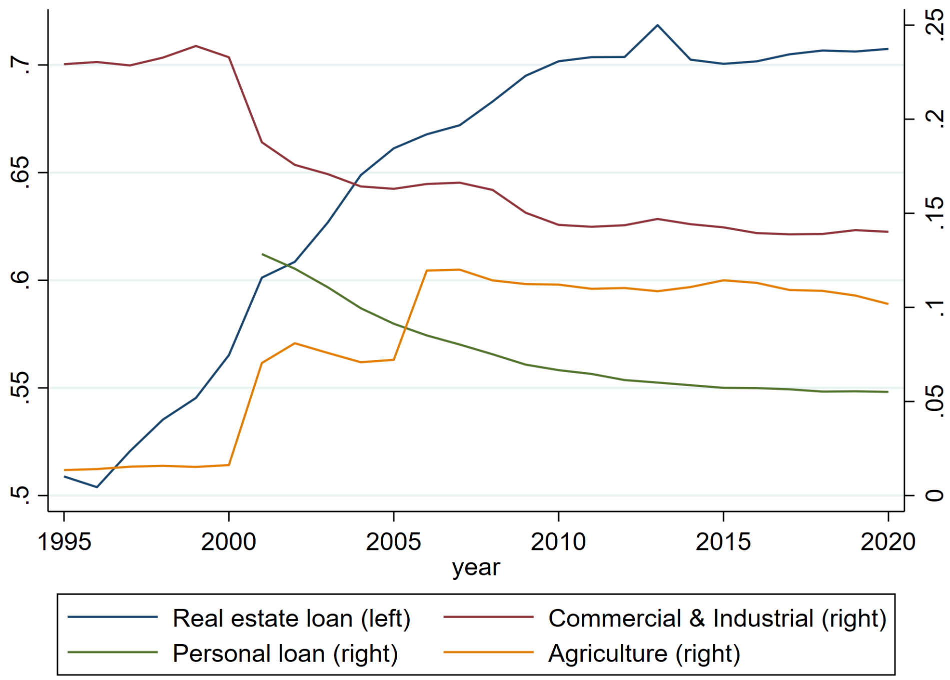
Notes: The figures are binned scatter plot demonstrating the correlation between banks' average technology spending in a county and the local economic condition. The y-axis of each figure is the average annual logarithmic of total IT spending of a bank in a county during 2010-2020. The x-axises are the explanatory variables split into 20 equalized bins. "House Price Index" is the county level house price index with 2000 base year provided by FHFA, "Growth in business establishments" is the county-level growth in business establishments, "Unemployment rate" is the county level unemployment rate, "Local Population Education" is the proportion of adults with college degree or high in a county, and "% Population 65 and older" is the proportion of population of age 65 and older in a county.

Figure A4. Determinants of IT spending: Bank Balance Sheet



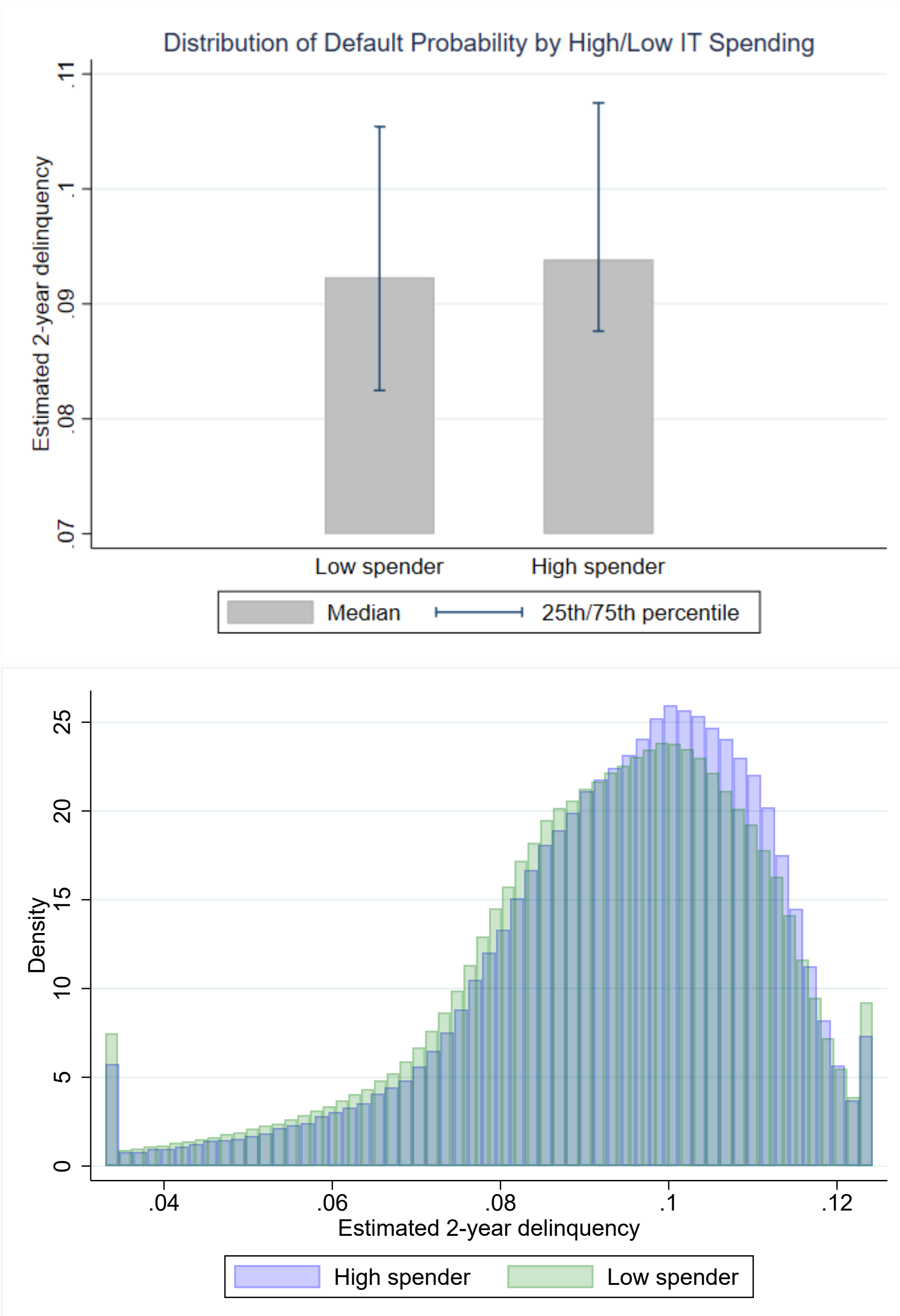
Notes: The figures are binned scatter plot demonstrating the correlation between banks' average technology spending in a county and banks' balance sheet condition. The y-axis of each figure is the average annual logarithmic of total IT spending of a bank in a county during 2010-2020. The x-axises are the explanatory variables split into 20 equalized bins. "Deposit HHI" is the deposit concentration measured by HHI of banks' deposit shares in a county and year, "Deposit share" is bank i's deposit market share in a county and year, "Deposits/Assets" is bank i's deposit as a proportion of total asset in a year, "Profitability" is bank i's total revenue scaled by total assets in a year, "Mortgage market share" is bank i's market share of mortgage lending in a county and year.

Figure A5. Evolution of Loan Portfolio of U.S. Banks



Notes: The figure below shows the evolution of banks' loan portfolio over time. The loan profile is constructed using information in "Call Report". Each line shows the average (weighted by total assets) of a specific type of loans scaled by total loans among all banks in "Call Report".

Figure A6. IT spending and Borrower Type Distribution



Notes: The figures below show the distribution of observable default risk taken by high-spending banks and low-spending banks. The observable default risk is estimated using the LASSO estimation with borrower income, loan amount, borrower ethnicity, borrower gender, borrower age, and borrower race and whether there is delinquency within 2 years after the loan was originated. The top figure shows the median, the 25-th percentile and 75-th percentile of of the estimated default risk. The bottom figure shows the overall distribution of the estimated default risk for borrowers taken by high-spending banks and low-spending banks.

Table A8. Software Spending, Communication Spending and Loan Decision

	1-st stage	$\mathbb{1}_{\text{Rejection}}$	$\mathbb{1}_{\text{Approve but denial}}$	$\mathbb{1}_{\text{Originated}}$	$\mathbb{1}_{\text{Rejection}}$	$\mathbb{1}_{\text{Approve but denial}}$	$\mathbb{1}_{\text{Originated}}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HP growth exposure	0.0391***						
	(0.0057)						
ln(Software IT)		-0.0149**	-0.0341***	0.0530***	-0.0027***	-0.0009***	0.0036***
		(0.0074)	(0.0083)	(0.0156)	(0.0003)	(0.0002)	(0.0004)
Year FE	✓	✓	✓	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓
Loan feature FE	✓	✓	✓	✓	✓	✓	✓
F-stat	318						
N	760,215	674,840	674,840	674,840	1,992,869	1,992,869	1,992,869
AdjR^2	0.69	0.02	-0.07	-0.03	0.09	0.05	0.09
	1-st stage	$\mathbb{1}_{\text{Rejection}}$	$\mathbb{1}_{\text{Approve but denial}}$	$\mathbb{1}_{\text{Originated}}$	$\mathbb{1}_{\text{Rejection}}$	$\mathbb{1}_{\text{Approve but denial}}$	$\mathbb{1}_{\text{Originated}}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance	-0.2450***						
	(0.0541)						
ln(Communication IT)		-0.0382***	-0.0062***	0.0444***	-0.0029***	-0.0010***	0.0039***
		(0.0030)	(0.0017)	(0.0033)	(0.0003)	(0.0002)	(0.0004)
Year FE	✓	✓	✓	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓	✓	✓	✓
County FE	N	N	N	N	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓
Loan feature FE	✓	✓	✓	✓	✓	✓	✓
F-stat	354						
N	2,270,169	1,994,598	1,994,598	1,994,598	1,994,597	1,994,597	1,994,597
AdjR^2	0.57	-0.01	-0.00	-0.01	0.09	0.05	0.09

Notes: This table shows how software IT spending and communication IT spending affect banks' decision making respectively. The 2SLS regression equations are as follows:

$$\begin{aligned}
\log(\text{Software IT})_{i,c,t} / \log(\text{Communication IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{HP exposure}_{i,t} / \text{Ave dist}_c + \pi_1 \mathbf{X} + \textbf{FE's} + \epsilon_{i,c,t} \\
\text{Loan decision}_{l,i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Software IT}})_{i,c,t} / \log(\widehat{\text{Communication IT}})_{i,c,t} \\
&\quad + \beta_1 \mathbf{X}_{i,c,t} + \textbf{FE's} + \epsilon_{i,c,t}
\end{aligned}$$

The OLS regression specification is as follows

$$\text{Loan decision}_{l,i,c,t} = \alpha_i + \mu_c + \eta_t + \beta \times \log((\text{Software IT}))_{i,c,t} / \log((\text{Communication IT}))_{i,c,t} + \beta_1 \text{X} + \epsilon_{i,c,t}$$

The dependent variables above 2SLS specification are the dummy variable indicating three possible decision status of a loan l received by bank i in county c in year t. The definitions of the dependent variables are provided in Section 5.2. "HP exposure$_{i,t}$" is weighted sum of house price index of a county, weighted by the banks' deposit in the county among the bank's total deposit. The detailed definition of "HP exposure$_{i,t}$" is provided in Section 5.2.2. "Ave dist" is the logarithmic average transportation distances between a county where a bank is located and major tech product suppliers' warehouses. The location of the major tech suppliers' warehouses are provided in Table A2. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include logarithmic of loan amount and logarithmic of borrower income. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table A9. Software Spending, Communication Spending and Loan Pricing Dispersion

	1-st stage	Sd(Interest rate)	
	(1)	(2)	(3)
HP growth exposure	0.0751***		
	(0.0228)		
ln(Software IT)		0.0686***	0.0051***
		(0.0246)	(0.0017)
Controls	✓	✓	✓
Lender FE	✓	✓	✓
Year FE	✓	✓	✓
County FE	✓	✓	✓
Loan feature FE	✓	✓	✓
F-stat	18		
N	20,354	18,458	20,354
AdjR^2	0.50	-0.20	0.13
	1-st stage	Sd(Interest rate)	
	(1)	(2)	(3)
Distance	-0.0431***		
	(0.0066)		
ln(Communication IT)		0.0235**	0.0045**
		(0.0082)	(0.0018)
Controls	✓	✓	✓
Lender FE	✓	✓	✓
Year FE	✓	✓	✓
County FE	N	N	✓
Loan feature FE	✓	✓	✓
F-stat	16		
N	20,354	18,461	20,354
AdjR^2	0.53	-0.01	0.13

Notes: This table presents how software IT spending and communication IT spending affect banks' pricing menu dispersion. The 2SLS regression equations are as follows:

$$\begin{aligned}\log(\text{Software IT})_{i,c,t} / \log(\text{Communication IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{HP exposure}_{i,t} / \text{Ave dist}_c + \pi_1 \mathbf{X} + \mathbf{FE's} + \epsilon_{i,c,t} \\ \text{Interest rate dispersion}_{i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Software IT}})_{i,c,t} / \log(\widehat{\text{Communication IT}})_{i,c,t} \\ &\quad + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE's} + \epsilon_{i,c,t}\end{aligned}$$

The dependent variable is the standard deviation of interest rates of all mortgage originated by bank i in county c in year t. "HP exposure$_{i,t}$" is weighted sum of house price index of a county, weighted by the banks' deposit in the county among the bank's total deposit. The detailed definition of "HP exposure$_{i,t}$" is provided in Section 5.2.2. "Ave dist" is the logarithmic average transportation distances between a county where a bank is located and major tech product suppliers' warehouses. The location of the major tech suppliers' warehouses are provided in Table A2. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include average logarithmic of loan amount, average logarithmic of borrower income, the average FICO scores and average loan-to-value ratios of all the loans made by a bank in a county in a specific year. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table A10. Software Spending, Communicaition Spending and Loan Performance

	1-st stage	$\mathbb{1}_{\text{Delinq 2 years}}$	$\mathbb{1}_{\text{Delinq 4 years}}$	$\mathbb{1}_{\text{Delinq 2 years}}$	$\mathbb{1}_{\text{Delinq 4 years}}$
	(1)	(2)	(3)	(4)	(5)
HP growth expension	0.0391***				
	(0.0057)				
ln(Software IT)		-0.0732***	-0.0865***	-0.0016***	-0.0017***
		(0.0242)	(0.0253)	(0.0006)	(0.0006)
Year FE	✓	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓
F-stat	13				
N	760,215	56,274	59,063	136,261	136,261
R^2	0.69	-0.19	-0.24	0.05	0.05
	1-st stage	$\mathbb{1}_{\text{Delinq 2 years}}$	$\mathbb{1}_{\text{Delinq 4 years}}$	$\mathbb{1}_{\text{Delinq 2 years}}$	$\mathbb{1}_{\text{Delinq 4 years}}$
	(1)	(2)	(3)	(4)	(5)
Distance	-0.9692***				
	(0.3259)				
ln(Communication IT)		-0.0107**	-0.0112**	-0.0016***	-0.0015**
		(0.0054)	(0.0091)	(0.0075)	(0.0006)
Year FE	✓	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓	✓
County FE	N	N	N	✓	✓
Controls	✓	✓	✓	✓	✓
F-stat	20				
N	136,364	136,364	59,227	136,169	136,169
R^2	0.57	-0.00	0.00	0.05	0.05

Notes: This table shows how does banks' software and communication spending impact the loan portfolio performance as measured by loan delinquency rates. The regression specification is as follows:

$$
\begin{aligned}
\log(\text{Software IT})_{i,c,t}/\log(\text{Communication IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{HP exposure}_{i,t}/\text{Ave dist}_c + \pi_1 \mathbf{X} + \mathbf{FE's} + \epsilon_{i,c,t} \\
\mathbb{1}_{\text{Delinq x years},l,i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Software}}\text{ IT})_{i,c,t}/\log(\widehat{\text{Communication}}\text{ IT})_{i,c,t} \\
&\quad + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE's} + \epsilon_{l,i,c,t}
\end{aligned}
$$

The dependent variables in the OLS regression and in the second-stage of 2SLS are the dummy variable indicating whether loan l originated by bank i in county c and in year t went into delinquency after x years of origination. "HP exposure$_{i,t}$" is weighted sum of house price index of a county, weighted by the banks' deposit in the county among the bank's total deposit. The detailed definition of "HP exposure$_{i,t}$" is provided in Section 5.2.2. "Ave dist" is the logarithmic average transportation distances between a county where a bank is located and major tech product suppliers' warehouses. The location of the major tech suppliers' warehouses are provided in Table A2. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include the logarithmic of loan amount, the logarithmic of borrower income, FICO, and LTV ratio. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table A11. IV Analysis: IT Spending and Loan Decision

	1-st stage	1[Rejection]	1[Approval but denial]	1[Originated]
	(1)	(2)	(3)	(4)
Distance	-0.2409***			
	(0.0534)			
ln(Total IT)		-0.0387***	-0.0063***	0.0450***
		(0.0030)	(0.0018)	(0.0034)
Controls	✓	✓	✓	✓
Loan feature FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓
County FE	✓	✓	✓	✓
F-stat	354			
N	2,268,302	1,994,598	1,994,598	1,994,598
AdjR^2	0.58	-0.01	-0.00	-0.01

Notes: This table shows how does banks' IT spending impact banks' loan decision making in 2SLS specification. The regression specification is as follows:

$$\begin{aligned}\log(\text{Total IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{Distance}_c + \pi_1 \mathbf{X} + \mathbf{FE's} + \epsilon_{i,c,t} \\ \text{Loan decision}_{l,i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Total IT}})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE's} + \epsilon_{l,i,c,t}\end{aligned}$$

The dependent variables above 2SLS specification are the dummy variable indicating three possible decision status of a loan l received by bank i in county c in year t. "Distance" is the logarithmic average transportation distances between a county where a bank is located and major tech product suppliers' warehouses. The location of the major tech suppliers' warehouses are provided in Table A2. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include logarithmic of loan amount and logarithmic of borrower income. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table A12. IV Analysis: IT Spending and Loan Pricing Dispersion

	Sd(Interest rate)	
	(1)	(2)
Distance	-0.0271***	
	(0.0101)	
ln(Total IT)		0.0248***
		(0.0557)
Controls	✓	✓
Loan feature FE	✓	✓
Lender FE	✓	✓
Year FE	✓	✓
County FE	✓	✓
N	18,322	18,322
AdjR^2	0.65	-0.09

Notes: This table shows how does banks' IT spending affect the loan pricing dispersion. The 2SLS regression is specified as follows:

$$\begin{aligned}
\log(\text{Total IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi\text{Distance}_c + \pi_1\mathbf{X} + \mathbf{FE's} + \epsilon_{i,c,t} \\
\text{Interest rate dispersion}_{i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Total IT}})_{i,c,t} + \beta_1\mathbf{X}_{i,c,t} + \mathbf{FE's} + \epsilon_{i,c,t}
\end{aligned}$$

The dependent variable is the standard deviation of interest rates of all mortgage originated by bank i in county c in year t. "Distance" is the logarithmic average transportation distances between a county where a bank is located and major tech product suppliers' warehouses. The location of the major tech suppliers' warehouses are provided in Table A2. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include average logarithmic of loan amount, average logarithmic of borrower income, the average FICO scores and average loan-to-value ratios of all the loans made by a bank in a county in a specific year. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table A13. IV Analysis: IT Spending and Loan Performance

	1-st stage	$\mathbb{1}_{\text{Delinq 2 years}}$	$\mathbb{1}_{\text{Delinq 4 years}}$	$\mathbb{1}_{\text{Delinq 6 years}}$
	(1)	(2)	(3)	(4)
Distance	-0.2409***			
	(0.0534)			
ln(Total IT)		-0.3385***	-0.3770***	-0.3977***
		(0.0852)	(0.0946)	(0.0992)
Controls	✓	✓	✓	✓
Loan feature FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓
County FE	✓	✓	✓	✓
F-stat	20			
N	136,456	136,456	136,456	136,456
AdjR^2	0.57	-0.01	-0.00	-0.00

Notes: This table shows how does banks' IT spending impact the loan portfolio performance as measured by loan delinquency rates. The regression specification is as follows:

$$\begin{aligned} \log(\text{Total IT})_{i,c,t} &= \tilde{\alpha}_i + \tilde{\eta}_t + \pi \text{Distance}_c + \pi_1 \mathbf{X} + \mathbf{FE's} + \epsilon_{i,c,t} \\ \mathbb{1}_{\text{Delinq x years},l,i,c,t} &= \alpha_i + \eta_t + \beta \times \log(\widehat{\text{Total IT}})_{i,c,t} + \beta_1 \mathbf{X}_{i,c,t} + \mathbf{FE's} + \epsilon_{l,i,c,t} \end{aligned}$$

The dependent variables in the OLS regression and in the second-stage of 2SLS are the dummy variable indicating whether loan l originated by bank i in county c and in year t went into delinquency after x years of origination. "Distance" is the logarithmic average transportation distances between a county where a bank is located and major tech product suppliers' warehouses. The location of the major tech suppliers' warehouses are provided in Table A2. We control for bank fixed effects, county fixed effects, and year fixed effects. "Loan features" fixed effects include borrower ethnicity, borrower gender, and borrower race. Loan level controls include the logarithmic of loan amount, the logarithmic of borrower income, FICO, and LTV ratio. Other control variables include county real GDP per capita, unemployment rate, house price index growth, logarithmic of lenders' revenue at county-year level, logarithmic of lenders' employee at county-year level, logarithmic of lenders' total assets at lender-year level, net revenue scaled by total assets at bank-year level, and total deposits-to-loan ratio at bank-year level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at county level.

Table A14. Potential Bank Expansion, Other Expanses and HP Exposure

	ln(branches)	ln(employee)	ln(printing exp)	ln(advertising)	ln(Total IT)
	(1)	(2)	(3)	(4)	(5)
HP exposure	0.0000	-0.0130	-0.0050*	-0.0002	0.0089***
	(0.0013)	(0.0201)	(0.0026)	(0.0027)	(0.0014)
ln(assets)	0.0202	0.1648	0.8233***	0.7585***	0.2894***
	(0.0200)	(0.3413)	(0.0455)	(0.0740)	(0.0259)
Deposit/assets	-0.0670	2.3880*	0.2336	-0.0247	2.4729***
	(0.0912)	(1.2751)	(0.2572)	(0.2528)	(0.0916)
Income/assets	0.0593	13.7697	-122.3034***	-129.2614***	-129.1010***
	(0.2290)	(45.8013)	(41.6651)	(39.2333)	(7.2473)
Bank FE	✓	✓	✓	✓	✓
Bank controls	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
N	794,177	366,160	299,158	617,668	366,108
AdjR^2	0.18	0.26	0.98	0.99	0.54

Notes: This table presents how does housing price exposure ("HP exposure") correlate with banks' potential expansion and potential increases in other types of expenses:

$$\log(\text{Total IT})_{i,t} \quad = \quad \alpha_i + \mu_t + \beta \text{HP exposure}_{i,t} + \pi_1 \mathbf{X} + \textbf{FE's} + \epsilon_{i,t}$$

"ln(branches)" are the logarithmic of total number of branches of a bank in a county in a given year. "ln(employee" is the logarithmic of total number of employees of a bank in a county in a given year. "ln(printing exp)" is the logarithmic of total printing expenses of a bank in a given year. "ln(advertising)" is the logarithmic of total advertising and marketing expenses of a bank in a year. "ln(Total IT)" is the logarithmic of total IT expenses of a bank in a county and in a given year. We control for bank bank asset size, deposit/asset ratio, and income/assets. The dependent variables ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Robust standard errors are reported in the brackets. Standard errors are clustered at bank level.

Appendix: Data Set Construction

We create our panel data set by merging the Home Mortgage Disclosure Act (HMDA) panel with three different supplementary data sets. We first merge HMDA with the Fannie Mae and Freddie Mac Data Dynamics, Government-Sponsored Enterprise (GSE) panel data sets to augment information on loan evolution. We next merge HMDA with the Harte Hanks Market Intelligence Computer Intelligence Technology Database (CiTDB) panel data set which provides information on technology investment and capital for various mortgage-offering establishments. Finally, we build loan portfolios for borrowers in HMDA where we group by various loan/applicant characteristics. The process by which we merge HMDA with the three separate data sets and cluster borrowers is described in the subsequent sections.

1 HMDA-GSE Merge

The HMDA and GSE data sets share several loan characteristics which allow us to match loans. Data richness varies across time periods in both the HMDA and GSE data, so the process by which we merge varies correspondingly. Our merge occurs sequentially across two time periods, namely 2010-2017 and 2018-2020, and two iterations.

1.1 Merging Variables

In time period 2010-2017, overlapping loan observables in HMDA and GSE are restricted to the year of origination, the state in which the property lies, the loan amount, and the Metropolitan Statistical Area (MSA). The overlapping loan observables are richer in HMDA and GSE for time period 2018-2020. Therefore, our merging variables are expanded to include the interest rate of the loan, the 3-digit zip code of the property, the loan-to-value (LTV) ratio, and the debt-to-income (DTI) ratio. We also exclude the property's MSA for these years since the data are sparse by nature (i.e. not all properties necessarily belong to an MSA). We now describe the merging iterations.

1.2 Merging Iterations

We first restrict the loans to have our project conforming characteristics and, similarly, to satisfy the GSE one-unit, conforming loan amount limit.

1. **Iteration One:** In the first iteration of our merge, we match loans on the available, year-dependent loan characteristics described above. In order to uniquely identify loans, we drop loans that have identical merging variables prior to matching. In other words, to correctly match loans, we cannot pair a HMDA loan with more than one GSE loan, and similarly, we cannot pair a GSE loan with more than one HMDA loan.

2. **Iteration Two:** In the second iteration, we include the loan seller as a merging variable for both time periods and drop the matches from the previous iteration. Since the originator and the seller of the loan are not necessarily the same, we exclude this variable in our merge for iteration one to prevent matching false positives. We similarly identify unique loans by dropping loans with identical merging variables.

1.3 Merging Rates

We now present the merging rates for each time period-iteration pair. In total, we match approximately 3.7 million loans giving us a merge rate of approximately 0.464. To obtain the match rates in the table on the year margin, we avoid double-counting uniquely identified loans in the first iteration that were not matched, and would hence, also be uniquely identified in the second iteration. For example, suppose there were y_1 uniquely identified loans in iteration one of the merge and x_1 loans were matched with $x_1 \leq y_1$. Then, suppose in the second iteration there were y_2 uniquely identified loans and x_2 loans were matched. Then, on the year margin, the match rate for that year would be:

$$\text{merge rate} = \frac{x_1 + x_2}{x_1 + y_2}$$

	Merge Rates by Year and Iteration		
	Iteration 1	**Iteration 2**	**Total**
2010-2017	(224,968 / 475,919) ≈ 0.473	(497,778 / 2,715,745) ≈ 0.183	(722,746 / 2,940,713) ≈ 0.246
2018-2020	(2,461,540 / 4,172,618) ≈ 0.590	(518,538 / 2,582,573) ≈ 0.201	(2,980,078 / 5,044,113) ≈ 0.591
Total	(2,686,508 / 4,648,537) ≈ 0.578	(1,016,316 / 5,298,318) ≈ 0.192	(3,702,824 / 7,984,826) ≈ 0.464

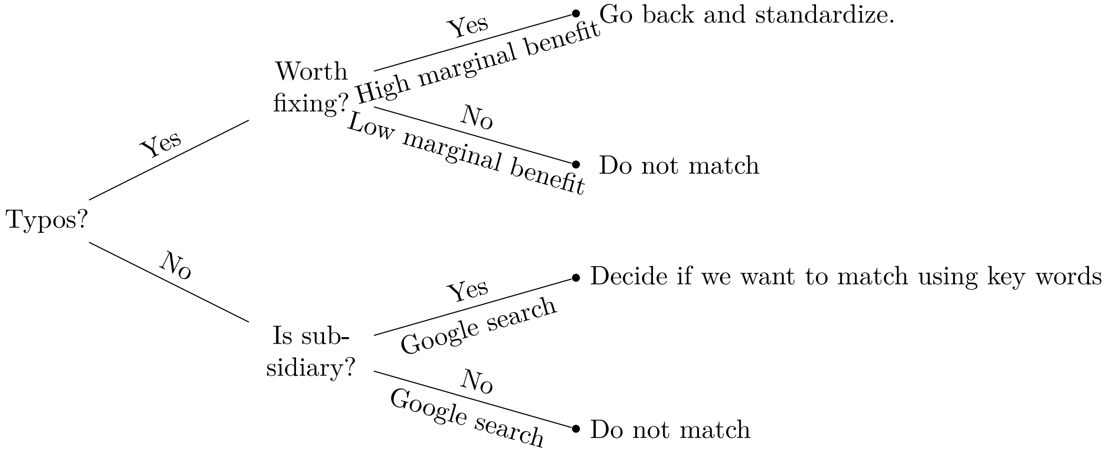
2 HMDA-CiTDB Merge

We now outline how we match HMDA loans to branches in the CiTDB data set. Since there is no indication of which specific branch at which each borrower applied for their mortgage in HMDA, the process of matching HMDA loans to IT sites is rather involved. The matching process involves four steps which we now outline. Due to bank name inconsistencies in both data sets, we first homogenize names so that we allow for as many possible loan-branch matches in each year. With rich location data, we next impute tract and county level geographies for each site. The CiTDB data has in-depth site-level categorizations describing what sort of services each site engages in. To minimize mismatches, we next rank sites by how likely their site description implies they are mortgage-offering institutions. Finally, we match loans to sites based on their common variables. We describe each step in detail below:

2.1 Bank Name Matching

Naming conventions in HMDA and the CiTDB data set are both inconsistent across year. As a result, we perform a relatively extensive standardizing procedure. Some of the biggest issues that prevent exact name matches are inconsistencies in entity type ("Federal Credit Union", "Federal Savings Bank", "Credit Union", "National Association", etc.) reporting. The stnd_compname command written by Nada Wasi and Aaron Flaaen often normalizes/deletes general entity types that are used across most names ("Corporation", "Company", "Incorporated", etc.), abbreviates numbers, gets rid of unnecessary words, etc. It fails, however, to generalize entity types that are specific to the banking industry. Thus, we spend a lot of time normalizing the typographical errors and abbreviating bank-specific entity types. There are also commonly-used words ("Bank", "Community", "Financial", "Employees", "Mortgage", etc.) found in the company names which we put a lot of effort into standardizing.

We also expend a lot of our effort in matching names that have the highest marginal benefit. Therefore, we focus on matching banks across HMDA and IT that have a large market share in HMDA (i.e. banks that originate many loans). To do so, we first run an exact merge on name. Due to the extensive effort allotted to standardization, most of our exact matches come from this merge. Then, looking at the banks with the highest market share in HMDA, we look at all IT names which share key words with the HMDA banks. We use the following decision tree to manually match HMDA and IT banks:



To be more concrete, we first check if there are typographical errors in the subset of names. If there are any, and the marginal benefit of fixing the error is high, we standardize this error. We then check to see if any of the sites in the subset are subsidiaries of banks by doing a Google search. For example, "Wells Fargo Funding" is a subsidiary of "Wells Fargo, National Association." If subsidiaries are present, we check to see if the subsidiary/branch could be mortgage-offering using key words/phrases. We list the key words and phrases below.

The following words indicate a subsidiary likely offers mortgages: [Mortgage, Finance/Financial, Home Loan, Lending, Financial Services, Credit, Funding]

The following words indicate a subsidiary unlikely offers mortgages: [Capital, Trust, Securities, Retail, Equipment, Auto, Holding, Leasing, Financial Advisors, Investment, Ventures, Commercial, Equity, Insurance, Operations, Shares, ATM, Enterprise, Asset(s), Audit, Brokerage, Pension, International]

We currently match approximately 82% of HMDA names to some IT name and 25% of IT names to some HMDA name.

2.2 Imputing Census Tracts

To match a single branch to each loan, we include a distance criterion to select from the subset of branches that match by name. More specifically, when matching branches to loans, we iterate over different geographies for a given HMDA loan and keep branches within the given geography that also match on name. The geography data in HMDA are restricted to census-level geographies, and therefore, to find branches in IT that match at these census-level geographies, we impute these geographies in ArcGIS Pro using the address data in CiTDB. We describe how to use this software to geocode the addresses and impute census tract geographies:

Geocoding IT data

1. We begin by opening up ArcGIS Pro on a Windows computer. We create and save a new ArcGIS project file.
2. We import the raw IT data, by choosing the "Map" tab, "Add Data", then "Data". In the dialogue box we choose the CiTDB data. The raw IT data becomes visible in the Contents Pane.
3. From the Contents Pane, we right click on the raw IT data and select "Geocode Table". We follow the prompts to geocode the address of each bank. We use the "USA_StreetAddress" geocoder available from Stanford.
4. The geocoding process then completes. It may take a long time.

Mapping census districts

1. Download Census Bureau .shp files here: https://www.census.gov/cgi-bin/geo/shapefiles/index.php
2. Import the .shp to ArcGIS by selecting "Add Data", "Data", then choose the .shp file you downloaded.
3. Finally, we need to calculate the centroids of each census tract. To do so, we first right click the census data from the Contents pane, and choose "Attribute Table".

4. We need to create two empty variables for the X and Y coordinates of the census tract centroid. Choose “Add” from the Attribute Table dialogue, then add the variables X and Y.
5. We right click on the new X variable and choose “Calculate Geometry”. In the “Field” section, we select X. In the “Property” section, we choose “Centroid x-coordinate”, and we do the same for Y (but instead choose “Centroid y-coordinate”. Press “OK” and ArcGIS will add the centroids as new columns in your Census data file.

Spatial Join

1. Finally, we spatially join the geocoded banks to their census tract and their centroid. We open the “Spatial Join” tool (we search for it in the search box in the top right corner of ArcGIS).
2. We set our IT addresses to be the “Target Features” and census tracts to the “Join Features”. We use the “Join one to one” option. See this url for an in-depth explanation of this process.
3. Running this command will generate the final dataset, and we now export the IT data which will now contain each banks’ census tract and census tract centroid.

2.3 Creating Match Ranks

Many large banks have a variety of subsidiaries that are unlikely to offer mortgages. To prevent matching loans to non-mortgage offering subsidiaries, we utilize the SIC descriptions provided by the CiTDB data set. In particular, the “SIC4” site description relays the services each site provides. We list the descriptions that are indicative of a site having mortgage-offering services for our subset of sites below:

1. National Commercial Banks
2. State Commercial Banks
3. Regional Commercial Banks
4. Federal Savings Institutions
5. Mortgage Bankers & Correspondents
6. Commercial Banks
7. Federal Credit Unions
8. National Credit Institutions
9. Commercial Banks NEC
10. National Savings Institutions
11. Savings Institutions
12. State Credit Unions
13. Regional Credit Institutions
14. Personal Credit Institutions
15. Savings Institutions Except National
16. Federal Credit Unions
17. Business Credit Institutions
18. Misc Business Credit Institutions

If a site has any of the above SIC 4 site descriptions, then we designate these sites with the highest match rank. This means that, when we match loans to sites in the next step of the merging process, these sites are given priority in a way we will describe. For all sites within the same bank-year as any site with the highest match rank, we drop so we do not erroneously match loans to sites that are unlikely to have mortgage-offering services. There are site descriptions that we deem as mortgage-related services and we give sites with these site descriptions a match rank of two. We list these below:

1. Loan Brokers
2. Insur-agents & Brokers
3. Agents, Brokers, Adjusters Claims Services

Similarly, to not incorrectly match loans to sites that are within the same bank-year as sites with any of the site descriptions above, we drop sites in the same bank-year as these sites as well. Finally, we give sites a match rank of three if they have any of the following site descriptions:

1. Real Estate-Agents
2. Real Estate Agents, Managers, Appraisers
3. Invest-Real Estate
4. Real Estate-Apartment
5. Real Estate-Developers
6. Real Estate Investment Trusts

Once again, we drop sites within the same bank-year as a site with any of the above site descriptions. Since some bank-years only have a single SIC 4 code across all sites within a given year, we also give sites a match rank of one if this criterion holds. Otherwise, all other sites are dropped.

2.4 Collapsing CiTDB Data and Matching Loans to Sites

Before matching sites to loans, we collapse the data by unique bank-tract-year-match rank quadruples to prevent the possibility of multiple branches being the same distance away from a tract centroid. When collapsing, we aggregate capital measures like total computers, total servers, total printers, etc. and investment measures like total budget, software budget, hardware budget, etc. Although tracts are not necessarily convex, we take the average of the latitudes and the average of the longitudes for all sites within the quadruple, and we take the first ID, state, city, and county, since tracts are completely within counties and states.

With the data structure now ready for matching, we describe how we match loans in HMDA to sites in CiTDB. The goal of this part of the matching process is to use the HMDA-IT loan-branch pairs found in the first subsection in tandem with the geographies and ranks we imputed in the following two subsections and use a distance criterion to find the "closest" site to the property. Keeping unique tract-bank-year triples, we loop through all loans in HMDA parallelized by state (i.e. run a loop for each state separately) and find all IT sites that match on name and year. Then, we subset all of these matches into four different geography levels; the sites that match the tract of the property, the sites that match the county of the property, the state, and finally all sites within the country. With the goal of finding the "closest" site, we take the subset of branches from the most granular geography that is also non-empty and calculate the distance between each site and the tract centroid of the loan. We then utilize the match ranks we created in the previous subsection and use the following steps to find the best match:

1. Find the site that is closest to the tract centroid of the given loan using Haversine's distance formula.
2. Check the geography level and match rank of the site
   - i If the site has a match rank of one and was found at the any geography level, match the given loan with the given site.
   - ii If the site was found at the tract level and the match rank is equal to two or equal to three, check for any sites at the tract level that have a match rank of one.
   - iii If the site was found at the tract level and the match rank is equal to two or equal to three and there are no sites with a match rank of one at the tract level, check for any sites at the county level that have a match rank of one. If there exists such a site, then find the closest one.
   - iv If the site was found at the tract level and the match rank is equal to two or equal to three and there are no sites at either the tract or county level with a match rank of one, return the site matched in the tract.
   - v If the site was found at the county level and the match rank is equal to two or equal to three, check for any sites at the county level that have a match rank of one. If there exists such a site, then find the closest one.
   - vi If the site was found at the county level and the match rank is equal to two or equal to three and there are no sites at the county level with a match rank of one.
   - vii Return the closest site regardless of match rank if it was found at the state or country level.
3. Match the unique site attributes with the loan and return this output and repeat the steps for the next loan.

Once all the best matches are found, we save all of these to be merged later on in the pipeline when we complete the panel.

2.5 Merging Rates

We now present the loan to site merge rates by year and the distribution of matches by geography. We subset the mortgages in HMDA to those that meet our project criteria which amounts to 30,432,445 loans. We were able to match some IT site to 24,021,772 of these loans, corresponding to a match rate of approximately 78.9%. We now partition these matches by geography

Geography Level	Merge Counts and Rates	
	Total Matches	Proportion of Matches
Tract	1,047,354	(1,047,354 / 27,592,338) $\approx$ 0.03
County	9,445,698	(9,445,698 / 27,592,338) $\approx$ 0.342
Bank-State	6,621,730	(6,621,730 / 27,592,338) $\approx$ 0.24
Bank	6,906,990	(6,906,990 / 27,592,338) $\approx$ 0.25
No Data Available for Bank-Year	3,570,566	(3,570,566 / 27,592,338) $\approx$ 0.129

3 Building HMDA Loan Portfolios

It could certainly be that a potential borrower applies for more than one loan, and thus, appears multiple times in HMDA. With the goal of tracking these borrowers, we build loan portfolios for the applications in HMDA using an iterative procedure that groups loans based on loan/borrower characteristics. In particular, we group loans by borrower/loan characteristics that are unvarying and then refine these groups using loan characteristics that are susceptible to distortion using several grouping rounds. We now describe how we perform this procedure:

3.1 Application Grouping

In the first round, we cluster loans by loan characteristics that are likely to be exact. HMDA reports applicant observables that make it easy to identify the same borrower across different applications. We group applications by year, state, tract, ethnicity, race, sex, co-applicant ethnicity, co-applicant race, and co-applicant sex for both 2010-2017 data and 2018-2020 data. Since 2018-2020 data also includes the applicant age, we also use this as a grouping variable. These groups form potential portfolios for applicants, so we create a group ID that defines a cluster for a given borrower. With the HMDA data being the universe of mortgages, there could certainly be more than one borrower with the same characteristics (i.e. two applicants match on all characteristics). This would mean that, for example, a portfolio made based on borrower observables should actually be two portfolios for two different applicants. To better define the portfolios, we separate them if there are big discrepancies in income for two separate observations within the same portfolio. After sorting based on income, we loop through each application within each portfolio and check if the income reported in row N is within a prespecified interval of the income reported in row $N-1$. If so, we separate these loans into two separate portfolios. If not, we keep them within the same portfolio. We then perform a second round where instead of refining portfolios using income, we use the loan amount. To choose these intervals, we first group applications based on the discrete applicant covariates. We then calculate the difference in loan amount/income between each application within each group and examine this distribution.

3.2 Visual Representation

To better understand the clustering process, we demonstrate with a visual representation. Suppose we have five observations with only three portfolios that were made based on applicant characteristics.
Since rows 1 and 2 matched on applicant characteristics and the reported income is the same, they remain in the same portfolio from Table 1 to Table 2. However, in rows 3-5, although all of these rows had the same applicant characteristics, they differ in income. In particular, it seems unlikely for application three and application four to belong to the same borrower while it may be likely that application three and application four belong to the same borrower since the discrepancy in income is small. Thus, we separate application three and application four into different portfolios while we keep applications four and five in the same portfolio.

Row	Temp ID	Income
1	1	50,000
2	1	50,000
3	2	100,000
4	2	90,000
5	2	89,000

Table I 1. Portfolios Prior to Refinement

Row	Temp ID	Income	Output ID
1	1	50,000	1
2	1	50,000	1
3	2	100,000	2
4	2	90,000	3
5	2	89,000	3

Table I 2. Portfolios Post Refinement

4 Project Conforming Flag

We restrict loans in HMDA such that they do not have any unconventional features. We create a flag based on the loan features that are in HMDA and describe the project conforming characteristics below:

1. Conventional loans not insured or guaranteed by FHA, VA, RHS, or FSA
2. Home purchases only
3. Secured by first lien
4. Principle residence
5. Single family homes
6. No reverse mortgage
7. No open-end line of credit
8. Not for business or commercial purposes
9. Must have no prepayment penalty term
10. No introductory rate period
11. No interest only payments
12. No balloon payments
13. Loan must be for one housing unit
14. The loan term must be 3 years

5 Miscellaneous Data Issues

Some of the data sets have some miscellaneous inconsistencies across years that need to be standardized, and we describe all of them here.

5.1 Home Mortgage Disclosure Act

To maintain the anonymity of borrowers, HMDA rounds loan amounts. However, this occurs inconsistently across different years. Prior to 2018, amounts are rounded to the nearest thousand, with $500 rounded up to the next thousand, and loan amounts were reported in thousands of dollars. For example, if the loan was for $152,500, institutions were instructed to enter 153; if the loan was for $152,300, institutions were instructed to enter 152. Starting in 2018, the CFPB modifies loan amount in Public HMDA so that amounts are disclosed as the midpoint of the nearest $10,000 interval for which the reported value falls. For example, if the reported loan amount is $117,823, in Public HMDA, this would be disclosed as $115,000 as the midpoint between values $110,000 and $120,000.

5.2 Harte Hanks Market Intelligence Computer Intelligence Technology Database

The CiTDB data for the time period 2010-2019 have a different schema relative to the data in 2020. In particular, the 2010-2019 data are made under version 3.14 and the 2020 data are made under version 4.1. A minor difference is that revenue is reported in millions of USD in version 3.14, so revenue has to be normalized by one million in version 4.1.

The newer schema updates the site IDs, and the first column, "aod_id", of the "IDs_LooLuP" file will be the site IDs in the newer version of 2020 data, and the third column "aod_alt_id" will be the site IDs in the previous version of data.

Appendix: Algebraic Proof

In this section, we provide detailed the algebraic details for our calculations that prove the propositions proposed in Section 3. We start with the proof for Proposition 3, as our analysis characterizing the loan origination process is conducted in a backward manner.

Proof of Proposition 3

Suppose a lender with screening accuracy θ chooses to screen a borrower with observable risk profile p. First consider the case if a g signal is generated. In this case, the expected payoff to the lender from approving the application and offering a net interest rate r that is taken by the borrower is

$$\begin{aligned}\Pi_g(r;p,\theta) &\equiv \pi_g(p;\theta)r + [1-\pi_g(p;\theta)]\,[\lambda r + (1-\lambda)(-l)] \\ &= \frac{p\theta r + (1-p)(1-\theta)(-l)}{p\theta + (1-p)(1-\theta)}\end{aligned}$$

Therefore we have

$$\begin{aligned}V_g(p;\theta) &\equiv \max_r (1-F(r))\Pi_g(r;p,\theta) + F(r)\cdot 0 \\ &= \frac{(1-r)\,[p\theta r - (1-p)(1-\theta)l]}{p\theta + (1-p)(1-\theta)}\end{aligned}$$

Taking the first order condition, the optimal choice of interest rate offering r_g is thus determined as

$$r_g = \frac{1}{2} + \frac{(1-p)(1-\theta)l}{2p\theta}$$

and the optimized payoff is

$$\begin{aligned}V_g(p;\theta) &= \frac{(1-r_g)\,[p\theta r_g - (1-p)(1-\theta)l]}{p\theta + (1-p)(1-\theta)} \\ &= \frac{\left[\frac{1}{2} - \frac{(1-p)(1-\theta)l}{2p\theta}\right]\left[\frac{p\theta}{2} - \frac{(1-p)(1-\theta)l}{2}\right]}{p\theta + (1-p)(1-\theta)} \\ &= \frac{\left[1 - \frac{(1-p)(1-\theta)l}{p\theta}\right][p\theta - (1-p)(1-\theta)l]}{4\,[p\theta + (1-p)(1-\theta)]}\end{aligned}$$

Similarly, in the case where a b signal is generated, the expected payoff to the lender from approving the application and offering a net interest rate r that is taken by the borrower is

$$\begin{aligned}\Pi_b(r;p,\theta) &\equiv \pi_b(p;\theta)r + [1-\pi_b(p;\theta)]\,(-l) \\ &= \frac{p(1-\theta)r + (1-p)\theta(-l)}{p(1-\theta) + (1-p)\theta}\end{aligned}$$

Therefore we have

$$\begin{aligned}V_b(p;\theta) &\equiv \max_r (1-F(r))\Pi_b(r;p,\theta) + F(r)\cdot 0 \\ &= \frac{(1-r)\,[p(1-\theta)r - (1-p)\theta l]}{p(1-\theta) + (1-p)\theta}\end{aligned}$$

Taking the first order condition, the optimal choice of interest rate offering r_b is thus determined as

$$r_b = \frac{1}{2} + \frac{(1-p)\theta l}{2p(1-\theta)}$$

and the optimized payoff is

$$\begin{aligned} V_b(p;\theta) &= \frac{(1-r_b)\left[p(1-\theta)r_b-(1-p)\theta l\right]}{p(1-\theta)+(1-p)\theta} \\ &= \frac{\left[1-\frac{(1-p)\theta l}{p(1-\theta)}\right]\left[p(1-\theta)-(1-p)\theta l\right]}{4\left[p(1-\theta)+(1-p)\theta\right]} \end{aligned}$$

It is thus easy to see that with the prior probability (observable risk profile) p held fixed, the optimal interest rate r_g (r_b) offered by the lender when getting a g (b) signal is decreasing (increasing) in θ. ■

Proof of Proposition 2

We now move on to showing that in equilibrium each lender adopts a cutoff strategy in deciding whether to direct reject an applicant or to incur the flow cost to further screen the application.

When a lender with screening accuracy θ receives an application from a borrower with observable risk profile p, the net payoff from conducting screen on the borrower is

$$V_A(p;\theta) \equiv \sum_{s=g,b} \Pr(s;p,\theta)V_s(p;\theta) - \tau$$

where $\Pr(s;p,\theta)$ are the probabilities of receiving signal $s \in \{g,b\}$ conditional on screening, given by

$$\Pr(g;p,\theta) \equiv p\theta + (1-p)(1-\theta);\ \Pr(b;p,\theta) \equiv p(1-\theta) + (1-p)\theta.$$

Plug in the optimized payoff $V_s(p;\theta)$ as calculated in our above proof for Proposition 3, we get

$$\begin{aligned} V_A(p;\theta) &= \frac{1}{4}\left[1-\frac{(1-p)(1-\theta)l}{p\theta}\right]\left[p\theta-(1-p)(1-\theta)l\right] + \frac{1}{4}\left[1-\frac{(1-p)\theta l}{p(1-\theta)}\right]\left[p(1-\theta)-(1-p)\theta l\right] - \tau \\ &= \frac{1}{4}\left(p\theta - 2(1-p)(1-\theta)l + \frac{\left[(1-p)(1-\theta)l\right]^2}{p\theta}\right) \\ &\quad + \frac{1}{4}\left(p(1-\theta) - 2(1-p)\theta l + \frac{\left[(1-p)\theta l\right]^2}{p(1-\theta)}\right) - \tau \\ &= p - 2(1-p)l + \frac{(1-p)^2}{p}\left[\frac{(1-\theta)^2}{\theta} + \frac{\theta^2}{1-\theta}\right] l^2 - \tau \end{aligned}$$

To the monotonicity of $V_A(p;\theta)$ as a function of p with fixed θ, we have

$$\begin{aligned} \frac{\partial V_A(p;\theta)}{\partial p} &= 1 + 2l - \left[\frac{(1-p)^2}{p^2} + \frac{2(1-p)}{p}\right]\left[\frac{(1-\theta)^2}{\theta} + \frac{\theta^2}{1-\theta}\right] l^2 \\ &\geq 1 + 2l - \left[\frac{(1-\underline{p})^2}{\underline{p}^2} + \frac{2(1-\underline{p})}{\underline{p}}\right]\left[\frac{(1-\theta)^2}{\theta} + \frac{\theta^2}{1-\theta}\right] l^2 \\ &\geq 1 + 2l - \left[\frac{(1-\underline{p})^2}{\underline{p}^2} + \frac{2(1-\underline{p})}{\underline{p}}\right]\left[\frac{(1-\bar{\theta})^2}{\bar{\theta}} + \frac{\bar{\theta}^2}{1-\bar{\theta}}\right] l^2 \end{aligned}$$

where the last line utilizes the fact that function $\frac{(1-\theta)^2}{\theta} + \frac{\theta^2}{1-\theta}$ is increasing in θ on $\theta \in [\frac{1}{2}, \bar{\theta}]$. Now by Assumption 1, we have

$$\frac{(1-\underline{p})l}{\underline{p}} \leq \frac{1-\bar{\theta}}{\bar{\theta}},$$

and therefore

$$\begin{aligned}\frac{\partial V_A(p;\theta)}{\partial p} &\geq 1+2l-\left[\frac{(1-\bar{\theta})^2}{\bar{\theta}^2}+\frac{2(1-\bar{\theta})l}{\bar{\theta}}\right]\left[\frac{(1-\bar{\theta})^2}{\bar{\theta}}+\frac{\bar{\theta}^2}{1-\bar{\theta}}\right]\\ &= 1+2l-\left[\frac{(1-\bar{\theta})^4}{\bar{\theta}^3}+\frac{2(1-\bar{\theta})^3 l}{\bar{\theta}^2}+(1-\bar{\theta})+2\bar{\theta}l\right]\\ &\geq 1+2l-[\bar{\theta}+2(1-\bar{\theta})l+(1-\bar{\theta})+2\bar{\theta}l]\\ &= 0\end{aligned}$$

Hence for any fixed $\theta \in [\frac{1}{2}, \bar{\theta}]$, $V_A(p;\theta)$ is an increasing function of p on $[\underline{p}, 1]$. As such, since the lender chooses to directly reject the application if and only if

$$V_A(p;\theta) \leq 0 \equiv V_R(p;\theta),$$

it thus follows that the lender with screening accuracy θ adopts a cutoff strategy by rejecting applicants with observable risk profile below $\hat{p}(\theta)$, where $\hat{p}(\theta)$ solves equation

$$V_A(p;\theta) = 0$$

Finally, to further show that this threshold $\hat{p}(\theta)$ is decreasing as a function of screening accuracy θ, apply the implicit function theorem we can get

$$\hat{p}'(\theta) = -\frac{\partial V_A(p;\theta)}{\partial p}\left(\frac{\partial V_A(p;\theta)}{\partial \theta}\right)^{-1}.$$

Since $\frac{\partial V_A(p;\theta)}{\partial p} \geq 0$ and

$$\begin{aligned}\frac{\partial V_A(p;\theta)}{\partial \theta} &= \frac{(1-p)^2}{p}\left[\frac{2\theta}{1-\theta}+\frac{\theta^2}{(1-\theta)^2}-\frac{2(1-\theta)}{\theta}-\frac{(1-\theta)^2}{\theta^2}\right]l^2\\ &\geq 0\end{aligned}$$

for any $\theta \in [\frac{1}{2}, \bar{\theta}]$. Therefore, we have $\hat{p}'(\theta) \geq 0$ for any for any $\theta \in [\frac{1}{2}, \bar{\theta}]$. ■

Proof of Proposition 1

The last proposition concerns banks' ex-ante incentive in making fixed cost investment to improve their screening technology. In equilibrium, lenders' optimal choice of screening accuracy θ_E satisfies

$$\frac{\partial E_p[V(p;\theta)]}{\partial \theta} - c'(\theta) = 0,$$

such that the marginal return to screening accuracy meets the marginal cost of making further improvement.

Note that

$$\frac{\partial V_A(p;\theta)}{\partial \theta} = \frac{(1-p)^2}{p}\left[\frac{2\theta}{1-\theta}+\frac{\theta^2}{(1-\theta)^2}-\frac{2(1-\theta)}{\theta}-\frac{(1-\theta)^2}{\theta^2}\right]l^2$$

As such, we have

$$\frac{\partial V_A^2(p;\theta)}{\partial \theta \partial l} \geq 0$$

Therefore, it follows that a larger loss l associated with borrower default will make lenders' payoff more sensitive to their screening accuracy, i.e., increasing the value of marginal benefit $\frac{\partial E_p[V(p;\theta)]}{\partial \theta}$ from more accurate screening. Under the assumption that the investment cost function $c(\theta)$ is sufficiently convex and

$\lim_{\theta \to \frac{1}{2}} c'(\theta) = 0$, it can be guaranteed that at the equilibrium level θ_E, marginal cost curve $c'(\theta)$ crosses the marginal benefit curve $\frac{\partial E_p[V(p;\theta)]}{\partial \theta}$ from below. Therefore, it follows that an increasing l pushes up the marginal benefit curve and leads to a higher equilibrium choice of θ_E. ■