

# The Importance of Technology in Banking during a Crisis <sup>\*</sup>

Nicola Pierri <sup>†</sup>

Yannick Timmer<sup>‡</sup>

## Abstract

We study the implications of information technology (IT) in banking for financial stability, using data on US banks' IT equipment and the tech-background of their executives. We find that one standard deviation higher pre-crisis IT adoption led to 10% fewer non-performing loans during the global financial crisis. We present several pieces of evidence that indicate a direct role of IT adoption in strengthening bank resilience; these include instrumental variable estimates exploiting the historical location of technical schools. Loan-level analysis reveals that high-IT adoption banks originated mortgages with better performance and did not offload low-quality loans.

**JEL Codes:** G21, G14, E44, D82, D83

**Keywords:** Technology, Financial Stability, IT Adoption, Non-Performing Loans

---

<sup>\*</sup>Most recent version [here](#). We are grateful to Tobias Adrian, Ufuk Akcigit, Tobias Berg, Barbara Biasi, Markus Brunnermeier, Giovanni Dell'Ariccia, Enrica Detragiache, Douglas Diamond, Ehsan Ebrahimi, Ann Harrison, Peter Hoffmann, Deniz Igan, Divya Kirti, Simone Lenzu, Davide Malacrino, Atif Mian, Sole Martinez Peria, Andrea Presbitero, Damien Puy, Lev Ratnovski, Gustavo Suarez, Suchanan Tambunlertchai, Gregor Weiss, Torsten Wezel, and seminar and conference participants at CAFR Workshop on Fintech, Cass Business School, ECB, FRB, IMF, and Villanova University for their insightful comments. Nicola Pierri is grateful to Nick Bloom for guidance and support during the early stages of the project. We thank Chenxu Fu, Hala Moussawi, and Huy Nguyen for excellent research assistance and Gladys Cheng and Sylvie Poirot for help with data sources. All remaining errors are our sole responsibility. This paper is a revised version of the IMF WP 20/14 with title "Tech in Fin before FinTech: Blessing or Curse for Financial Stability?". The views expressed in the paper are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or its Management.

<sup>†</sup>International Monetary Fund. Email: [npierri@imf.org](mailto:npierri@imf.org)

<sup>‡</sup>International Monetary Fund. Email: [ytimmer@imf.org](mailto:ytimmer@imf.org)

# 1 Introduction

The emergence of FinTech has triggered a debate on the effect of Information Technology (IT) on financial stability (FSB, 2019; Claessens et al., 2018; Nasiripour, 2019). The recent literature on FinTech has mostly focused on how the latest technological developments have been changing the way information is processed and the relative consequences for credit allocation and performance; for instance, see Berg et al. (2019b); Buchak et al. (2018); Di Maggio and Yao (2018); Fuster et al. (2019). However, predictive systems which are accurate in good times may fail to predict default in the event of an adverse systemic shock (Rajan et al., 2015, 2010). Since the era of FinTech has not been exposed yet to large financial shocks testing its resilience, the related empirical evidence cannot be conclusive about whether a more IT-driven financial system enhances or endangers financial stability in the medium- and long-run.<sup>1</sup>

To understand the potential impact of higher technology intensity in lending on financial stability, we study the non-performing loans on the balance sheet of US banks with a heterogeneous degree of IT adoption during the Global Financial Crisis (GFC) and the consequences for credit provision. The sign of the relationship between IT adoption and non-performing loans is a-priori ambiguous. Advances in technology can improve monitoring and screening thanks to the enhanced ability to collect, store, communicate, and process information (Liberti and Petersen, 2018). However, banks with more IT adoption might rely too much on “hard” information, which are easier to report and communicate, inducing them to neglect “soft” information (Rajan, 2006; Rajan et al., 2015).

We find that US commercial banks which were leaders in IT adoption before the GFC were significantly more resilient during the crisis. Figure 1 illustrates this striking pattern: high- and low-IT adoption banks had the same level of NPL over assets before the crisis, but as soon as the crisis hit, high-IT adoption banks experienced a significantly smaller increase of NPLs compared to their peers. Regression analysis reveals that a one standard deviation higher pre-GFC IT adoption is associated with 16 basis points lower NPL to assets ratio in the years between 2007 and 2010. This represents a 10% reduction with respect to the cross-sectional average and 14% of the cross-sectional standard deviation. In the panel dimension, we do not detect a significant correlation between pre-crisis IT adoption of banks and their non-performing loans outside the crisis. However, once the crisis hit, a one standard deviation increase in IT adoption could have lowered the surge in NPLs by 15% with respect to pre-crisis levels. The non-result in normal times reinforces the argument that it is important to study the effects of IT adoption when the economy faces a system-wide shock. We also find that high- and low-IT banks had the same share of loans on their portfolios before the GFC, but low-IT banks tighten lending significantly more during the GFC.

---

<sup>1</sup> Some papers explicitly recognize this limitation. For instance, Hughes et al. (2019) study the performance of personal loans made by a peer-to-peer lending platform (Lending Club) and write in their abstract: “caveat: we note that [...] the results may not hold under different economic conditions such as a downturn.”

Our main measure of IT adoption in banking is closely related to several seminal papers on IT adoption for non-financial firms, such as [Bloom et al. \(2012\)](#), [Beaudry et al. \(2010\)](#), [Bresnahan et al. \(2002\)](#), and [Brynjolfsson and Hitt \(2003\)](#). We access data on the number of personal computers (PCs) and the number of employees in a bank branch. Following this previous literature, we use the ratio of PCs per employee within a branch as the relevant measure of branch-level IT adoption.<sup>2</sup> We confirm that there is a strong correlation between the share of PCs per employee and other measures of IT adoption, such as IT budget or adoption of frontier technologies in 2016; these alternative measures are unfortunately unavailable in our data for the pre-2007 years.<sup>3</sup> We then map the bank branches to Bank Holding Companies (BHCs) and estimate a bank fixed effect after controlling for the geography of the branch (through county fixed effects) and other characteristics, such as the size of the branch. These bank fixed effects serve as our measure of bank-level IT adoption. To the best of our knowledge, this is the first paper to use this type of data to study financial firms.

Bank-level pre-crisis technology adoption may be correlated with other characteristics which impact non-performing loans during the crisis. Most importantly, demand for IT equipment and its productivity have been associated with firms' organizational forms and managerial quality ([Bresnahan et al., 2002](#); [Bloom et al., 2012](#); [McElheran and Forman, 2019](#)). Thus, the main concern for a causal interpretation of our results is that IT intense banks were simply "better run" and superior management practices shielded them from the impact of the crisis.

Our analysis uncovers several pieces of evidence in favor of a direct role of IT adoption, thus mitigating the concern that the correlation between IT and bank resilience is driven by unobservable factors, such as management quality. First, we find that IT adoption is not significantly correlated with banks' ex-ante exposure to the GFC in terms of their geographical footprint or business model as measured by funding sources, assets composition, and other balance sheet characteristics. IT adoption is also uncorrelated with employees' average wages or executives' compensation, which can be thought as measures of workforce human capital ([Becker, 2009](#)). The absence of a correlation between IT adoption and all these observable characteristics is a useful falsification test: it suggests that our measure is unlikely to be correlated with other unobservable predictors of exposure to the crisis. As management quality should also lead to better firm performance outside of a crisis ([Bloom et al., 2013, 2019](#)), but IT is uncorrelated with pre-crisis bank performance, our results suggest that IT adopters were not just "better managed" banks overall.

---

<sup>2</sup>We use the word "branch" as a general term for any bank establishment, including national or local headquarters and other offices.

<sup>3</sup>In fact, later waves of the same dataset provide additional information on IT-budget and adoption of Cloud Computing at the establishment level: the number of PCs per employee is a strong predictor of these other measures of IT adoption. For example, the correlation between the per capita share of PCs and the IT budget is 65% on the branch level. These correlations are likely understating the correlation in the pre-crisis period.

Second, we find that the estimated impact of IT on NPLs is unaffected by the inclusion of a rich set of variables as controls. Exploiting this coefficient's stability, we follow [Altonji et al. \(2005\)](#) and [Oster \(2019\)](#) to provide formal testing for the presence of bias from unobservable bank-level characteristics, finding no evidence of a sizeable bias.

Third, we complement our analysis with a set of instrumental variable (IV) specifications based on the distance between a bank's headquarter and land-grant colleges and universities. These institutions were established at the end of the nineteenth century in all US States to provide technical education. We show that their students are significantly more likely to major in technical subjects and less likely to major in business and management sciences, suggesting that these colleges are a shifter of the availability of technical knowledge rather than managerial capabilities. In addition, the location of land-grant colleges does not predict the presence of BHC headquarters in a county, indicating the distance between locations is independent with respect to the most relevant factors impacting banking business. We then show that banks whose headquarters are closer to these colleges have generally a higher level of IT adoption, supporting the idea that technical knowledge is an important factor in fostering technology adoption. The IV estimates confirm the main finding of the paper.

Fourth, to shed further light on the role of IT versus managerial quality, we study the biographies of the banks' top management. In fact, personal characteristics and experience of leaders matter for the outcomes of their organizations ([Benmelech and Frydman, 2015](#)). We apply a simple text-analysis algorithm to the biographies of top executives hired before 2007. We search for specific tech-related keywords and use them to measure the managers' predisposition toward IT. We find that banks led by more "tech-oriented" executives adopted IT more intensively and were also more resilient in the crisis. Interestingly, when we estimate the impact of tech-savviness of executives over time, we find a strikingly similar pattern compared to the one estimated with the baseline measure of IT adoption: banks run by tech-oriented executives had statistically indistinguishable levels of NPLs compared to their peers in any year before the crisis; however, once the crisis hit, their NPLs increased significantly less than banks led by executives with no tech-background. These findings support the hypothesis that IT adoption in banking, which can be partly caused by executives' personal experience and inclinations, led to more resilience during the crisis. Importantly, these results are completely unchanged if we control for managers' compensation or the variable share of their compensation (the latter can be thought as a measure of risk-taking incentives [Meiselman et al. \(2018\)](#)). As long as executive "overall" quality is, at least partially, priced in their compensation, these results suggest that it is "tech-orientation" that matters and not managerial talent or risk-taking incentives.

Adopting a "weight of evidence" approach, this collection of results points toward IT itself as the cause of lower NPLs and against a spurious correlation between the two variables created by unobserved

bank characteristics, such as managerial quality.

To understand the channels through which high-IT adopters succeeded in containing the surge of NPLs during the crisis, we analyze the performance of mortgages originated before 2007 and sold to Freddie Mac, one of the two large government-sponsored enterprises (GSEs). We find that mortgages sold by high-IT adoption banks were significantly less likely to be delinquent during the GFC than the ones sold by other banks. Therefore, the better performance of high-IT adopters during the crisis is driven—at least in part—by the screening of borrowers at origination.

This result has important implications for financial stability. If high-IT adopters were only better in offloading their bad loans to GSEs, such as Freddie Mac and Fannie Mae, then IT intensity would not enhance financial stability but instead lead to risk-shifting and exacerbate moral hazard. This is an important concern because securitization may reduce the incentives of banks to screen and monitor borrowers (Keys et al., 2009, 2010, 2012) and IT adoption may facilitate securitization. A related concern is that high-risk individuals, which were rejected by technology adopters, borrowed from banks with less IT operating in the same area. We test for these spillover effects and find no evidence, either. Both of these results suggest that IT adoption had positive aggregate effects and was not associated with a transfer of risk across parties.

The Freddie Mac data allows us to use granular loan-level characteristics in our analysis to investigate what type of information management drives the results. We control for the credit-score, the debt-to-income ratio, and the loan-to-value ratio of the borrower. These information are strong predictors of default and the most important “hard” information. If banks that adopted more IT had just better access to these data, the effect of IT adoption on delinquency would vanish. Our results are, instead, unchanged when we control for these characteristics in a linear regression or probit model. These results indicate that IT-adopters had either additional information available or they used these variables in a more sophisticated and effective manner.

Our results suggest that technology adoption in lending can enhance financial stability through better monitoring and screening. While the increase in NPLs may not be sufficient for a full assessment of financial stability, it has widely been considered an important indicator for banking sector distress (Demirgüç-Kunt and Detragiache, 2002) and has been shown to have severe adverse macroeconomic consequences (Peek and Rosengren, 2000; Caballero et al., 2008).<sup>4</sup> To investigate whether IT adoption indeed has an impact on the functioning of the financial system, we finally investigate the lending dynamics of the banks in our sample. We find that banks that adopted less IT before the crisis and banks which had higher NPLs in the crisis had significantly weaker loan growth in the crisis.

---

<sup>4</sup>The strong increase in NPLs during the GFC also impaired the functioning of the financial system (Berti et al., 2017; BIS, 2017).

Overall, our results indicate that technology adoption in lending can enhance the resilience of the financial system, and suggest that the recent rise in FinTech can be beneficial for financial stability through better screening abilities. An obvious limitation of extrapolating our results to the current FinTech era is that the type of technologies employed by commercial banks in the early 2000s might be different than today’s use of machine learning and big data. On the other hand, we can examine a period of severe and systemic financial turmoil while the growing FinTech literature cannot. Another strength of our approach is that it is likely to be more representative since our sample covers the vast majority of lending in the pre-crisis period, while FinTech is still a small fraction of credit markets.<sup>5</sup> Because of these important trade-offs the results presented in this paper should be seen as a relevant complement, rather than a substitute, to this literature.

The rest of the paper is structured as follows. In [section 2](#) we present a brief review of the relevant literature; in [section 3](#) we describe the several databases used; in [section 4](#) we present the main results on IT adoption and NPLs; in [section 5](#) we provide evidence on the roots of IT adoption and propose an instrumental variable strategy; in [section 6](#) we present additional results on mortgages performance to shed light on potential mechanisms; in [section 7](#) we provide evidence on lending dynamics; in [section 8](#) we conclude and discuss the relevance of our results for the ongoing policy debate.

## 2 Related Literature

This paper is related to the finance literature on technology adoption, which has been thriving in recent years thanks to the surge of FinTech. Because of its dynamism and breadth of scope, this literature is difficult to summarize and a surely non-exhaustive list includes [Frost et al. \(2019\)](#); [Gambacorta et al. \(2019\)](#); [Jagtiani and Lemieux \(2017\)](#); [Hughes et al. \(2019\)](#); [Fuster et al. \(2018, 2019\)](#); [Berg et al. \(2019b\)](#); [Di Maggio and Yao \(2018\)](#); [Buchak et al. \(2018\)](#); [Basten and Ongena \(2019\)](#); [Bartlett et al. \(2018\)](#); [Philippon \(2019\)](#); [Hau et al. \(2018, 2019\)](#); [Stulz \(2019\)](#); [Carlin et al. \(2019\)](#); [D’Acunto et al. \(2019\)](#); [Rossi \(2018\)](#); [Navaretti et al. \(2018\)](#). While most of these papers view FinTech as a positive development,<sup>6</sup> the FinTech era has not yet exposed to an adverse shock yet and therefore it is difficult to understand its impact on financial stability. We contribute by evaluating the impact of IT adoption in lending on financial stability and by studying both “normal times” and a severe systemic shock. Moreover, the data used in several papers of

---

<sup>5</sup> [Claessens et al. \(2018\)](#) use data from the Cambridge Centre for Alternative Finance and estimate FinTech credit to be 4% of the overall US market.

<sup>6</sup>For instance, [Berg et al. \(2019b\)](#) show that default prediction can be improved using borrowers’ digital footprint, [Bartlett et al. \(2018\)](#) find that algorithmic lending reduces racial disparities, [Jagtiani and Lemieux \(2017\)](#) show that Lending Club provides credit especially to areas that lose bank branches and in highly concentrated banking markets, and [Carlin et al. \(2019\)](#) argue that the introduction of a mobile application for a financial aggregation platform improves borrowers’ decision-making and lead them to take less high-interest unsecured debt and pay lower bank fees. Using data from a Chinese FinTech firm, [Gambacorta et al. \(2019\)](#) find that models based on non-traditional data and machine learning are better able to predict the consequences of an exogenous negative credit shock.

this literature are obtained from a single firm, e.g. [Berg et al. \(2019b\)](#), raising questions on the external validity of the results for financial stability. In contrast, our bank sample covers the vast majority of bank loans in the US.

A large literature has studied the demand for IT across different non-financial firms or geographical units and its effect on real outcomes, such as productivity and local wages. For instance, see [Akerman et al. \(2015\)](#); [Autor et al. \(2003\)](#); [Brynjolfsson and Hitt \(2003\)](#); [Bloom et al. \(2012\)](#); [Beaudry et al. \(2010\)](#); [Bresnahan et al. \(2002\)](#); [Bloom and Pierri \(2018\)](#); [Forman et al. \(2012\)](#); [McElheran and Forman \(2019\)](#); [Bessen and Righi \(2019\)](#). We contribute by using similar data and methodologies to study financial firms and financial stability.

Closer to us, in this respect, are a few papers that analyze certain features of IT adoption in banking before the GFC. [Beccalli \(2007\)](#) show that there are small productivity improvements from using IT in normal times. [Berger \(2003\)](#) argues that technology in banking led to improvements in cost and lending capacities. More recently, [Koetter and Noth \(2013\)](#) use IT data from Germany to re-estimate bank productivity with IT expenditure and they show that productivity is upward biased if IT expenditure is ignored. Compared to these papers we contribute by focusing on the effect of IT adoption across banks on their performance when a system-wide shock hits. Moreover, our paper provides a potential explanation for the “profitability paradox” ([Beccalli, 2007](#)): banks are heavy adopters of IT despite the small or negligible observed profitability gains in normal times because it may increase their resilience to large shocks. Relatedly, [Bostandzic and Weiss \(2019\)](#) study banks’ patent activities and their role as non-financial innovator while [Dante and Makridis \(2019\)](#) study causes and consequences of the diffusion of online banking. We differ because we study the overall adoption of information technology and because we focus on its impact on financial stability.

This paper is also related to the recent literature studying the increasing use of data in financial markets ([Bai et al., 2016](#); [Farboodi et al., 2018](#)). Our work differs because we focus on bank lending rather than asset pricing and trading.

We contribute to the literature on information in lending as advances in IT change the processing of information by helping firms to gather, store, distribute, and analyze information ([Liberti and Petersen, 2018](#); [Petersen and Rajan, 2002](#); [Degryse and Ongena, 2005](#); [Petersen and Rajan, 1994](#)). In the loan-level analysis we find that the impact of IT adoption on loans’ performance is robust to the inclusion of the most important predictors of default at origination in a linear regression or probit model. This indicates that either the high-IT lenders used additional information in the application decisions or they had a more sophisticated and effective way to use these data. While it would be interesting to distinguish between these two hypotheses, they have similar implications for financial stability.

New financial innovation can create moral hazard issues ([Rajan, 2006](#); [Gennaioli et al., 2012](#)). For



instance, [Keys et al. \(2010, 2012, 2009\)](#) show that securitization led to lax screening. We show that banks with more IT adoption did not offload low-quality loans to GSEs, mitigating the concern that IT adoption also had a destabilizing impact through securitization. In contrast, we show that IT adoption has a first-order beneficial effect on financial stability.

We contribute to the literature that focuses on the determinants of loan-performance in the crisis by highlighting the importance of lenders' technology. Besides, we find that larger banks and banks with more loans and wholesale funding had more NPLs in the crisis. Also, banks with more geographical exposure to the house price shock had more NPL than other banks. As in [Mian and Sufi \(2009\)](#) and [Mian and Sufi \(2011\)](#), we find that borrowers with lower credit scores and higher debt-to-income or loan-to-value ratios have higher probabilities of becoming delinquent.

We finally contribute to the literature highlighting the importance of top executives for firms' outcomes ([Benmelech and Frydman, 2015](#); [Bennedsen et al., 2006](#); [Bertrand and Schoar, 2003](#)). We document that the "tech-orientation" of the top executives of certain banks was an important factor in promoting IT adoption, which led to fewer NPLs during the GFC.

### 3 Data and Measurement

#### Regulatory Data on BHCs

We use bank balance sheet information from bank holding companies (BHCs) to assess the resilience of banks to the GFC. The data is collected by the Federal Reserve Bank of Chicago. We use the Financial Institution Reports which provides consolidated balance sheet information and income statements for domestic BHCs.

Our baseline NPLs are defined following [Hirtle et al. \(2018\)](#): Total loans, leasing financing receivables and debt securities and other assets - past due 90 days or more and still accruing (bhck5525) + Total loans, leasing financing receivables and debt securities and other assets - nonaccrual (bhck5526) - Debt securities and other assets - past due 90 days or more and still accruing (bhck3506) - Debt securities and other assets - nonaccrual (bhck3507). Our main dependent variable is the amount of NPLs scaled by total assets. We check the robustness of the main results of the paper to other definitions of NPLs (e.g. including loans with shorter delinquency periods) and alternative scaling choices (e.g. the use of loans as denominator), see [section 4](#).<sup>7</sup> [Figure A1](#) shows the distribution of the average NPLs ratio between 2007 and 2010 across banks. Most banks have an NPL ratio of around 1% in the crisis period, but there is a long right tail in the distribution. For some banks almost 5% of their balance sheet consists of NPLs.

---

<sup>7</sup>We rely on assets as a scaling variable for NPLs (rather than loans) since lending is endogenous to IT adoption and NPLs during the crisis, as we document in [section 7](#). Moreover, assets are commonly used to normalize the bank-level variables. The main raw patterns and results are robust to using loans instead of assets to normalize NPLs.



In addition to NPLs we construct the following variables as bank-level controls. The share of loans over total assets (*Loans*), the log of assets (in thousands of US Dollars) (*Size*), equity over assets (*Capital*), wholesale funding over assets (*Wholesale*), the return on assets (*ROA*), and the average log wage paid to employees (in thousands of US Dollars) (*LogWage*). All variables are averaged between the years 2001 and 2006 after winsorization. We winsorize all bank-level ratio at top 2.5 percent before taking averages, but results are robust to different treatment of outliers.

## IT Adoption

The IT data comes from an establishment survey on personal computers per employee by CiTBDs Aberdeen (previously known as “Harte Hanks”) for years 1999, 2003, 2004, 2006, and 2016. For the year 2016, we also have information on the IT budget and the usage of cloud computing of the establishment. The data also contains information about the type of establishment, i.e. whether it is the headquarter (HQ), a branch or a standalone establishment, the number of employees in the establishment as well as the location. The correlation between the IT budget of the establishment and the number of computers as a share of employees is very strong for later years, e.g. 65% in 2016. The R-squared of a cross-sectional regression of PCs per Employee on the per capital IT budget is 44%. There is also a positive correlation between PCs per Employee and the adoption of cloud computing. These correlations provide assurance that the number of personal computers per employee is a good measure of IT adoption, even more recently, but likely even more so in earlier years when other forms of IT adoption were less common.

We focus only on establishments in the banking sector (based on SIC2 classification) and drop savings institutions and credit unions (based on SIC 3). After these cleaning steps, we end up with 143,607 establishment-year observations.

We map bank branches from the Aberdeen dataset to the BHC data by using banks’ names and the BHC structure. We map 90% of the assets from the bank-level dataset to the IT data.

Our measure of IT adoption is based on a regression of the share of personal computers on a bank fixed effect controlling for the geography of the establishment and other characteristics. By doing so we can control for several characteristics that may be correlated with the number of personal computers per employee of the bank but are not informative about whether the bank has been at the technological frontier. This approach follows [Beaudry et al. \(2010\)](#), who measure IT adoption on the region-level controlling for establishments’ industry and size.

We estimate the following regression for years 1999, 2003, 2004, and 2006:

$$PCs/Emp_{i,t} = \widetilde{IT}_b + \theta_{type} + \theta_c + \theta_t + \gamma \cdot Emp + \epsilon_{i,t} \quad (1)$$

where  $PCs/Emp_{i,t}$  is the ratio of computers per employee in branch  $i$  survey wave  $t$  (capped at top 1%),  $\widetilde{IT}_b$  is a bank fixed effect,  $\theta_{type}$  is a establishment-type (HQ, standalone, branch) fixed effects,  $\theta_c$  is a county fixed effect,  $\theta_t$  is a year fixed effect and  $Emp$  is the log number of employees in the establishment.

The R-squared of the regression is 42%. The main part of the variation is explained by the bank fixed effect (60%). The year fixed effect explains 11%. The location of the establishment only explains 27% of the variation and the number of employees' and the bank types variables explanatory power is close to zero.

Our measure of IT adoption,  $IT_b$  hereafter, is a standardized version of the bank fixed effect. It is obtained by dividing  $\widetilde{IT}_b$  by its standard deviation after subtracting its mean. This adjustment is done considering the summary statistics for the sample of banks that we are able to match with BHC data only. The bottom panel of [Figure A1](#) plots the cross-sectional distribution of  $IT_b$ . We also check that our results are robust to an aggressive winsorization (5 % on both sides) of this variable and that they are not affected by the “generated regressor” problem ([Pagan, 1984](#)), see [subsection 4.1](#) and [Table A2](#).

We also compute a bank-level measure of the IT adoption of local competitors. For each bank, we first take the average  $IT_b$  of other banks in each county. Then we average for each bank the IT adoption of other banks across counties.

## House Price Data

As an additional control, we compute a variable that is capturing the exposure to the downturn in house prices, *HP Exposure*. We obtain county-level home value index from Zillow. For each county we construct the percentage change in the annual average house price between 2012 Q3 and 2007 Q4 (peak to through). We merge the county level decrease in the house price to the IT establishment data by county. We construct the exposure to the house price decrease by aggregating the decrease in the house price index across establishment for each bank.

## Data on Biographies of Executives

We obtain data on the biography of executives from S&P Global Market Intelligence. We have information on the Chief Executive Officer, the Chief Financial Officer, the Chief Operating Officer, and the President of the bank. We focus on the executives that have been hired before the GFC. We search the biography for the following words to characterize whether an executive is tech-prone: technology, engineering, math, computer, machine, system, analytic, technique, method, process, stem, efficiency, efficient, software, hardware, data, informatic. We count the number of occurrences of these words for each executive in the biography and scale the number by the total number of words in the biography. For each bank, we take the average across executives to construct a bank-level measure of the IT intensity of

their executives.

In addition to the biography, we also use data on the total compensation of the executives and the non-base share of the compensation, e.g. bonuses. This is in spirit of [Meiselman et al. \(2018\)](#) who construct a comprehensive dataset of CEO compensation complementing the Standard & Poor's Executive Compensation database. They show that higher payouts to CEOs are associated with significantly higher tail risk exposure.

### Land-Grant and County-level Data

We obtain the list of all the 70 Lang-Grant colleges (and universities) established in the mainland US during the nineteenth century (1862 and 1890) from the website of the US Department of Agriculture. We obtain data on enrolment by major and test scores from IPEDS (Integrated Postsecondary Education Data System) survey for 1996 and 2018 for several higher education institutions. We obtain county-level demographic characteristics from the 2000 US Census and from the American Community Survey from 2001 to 2006. County-level variables are averaged between 2000 and 2006.

### Freddie Mac Data

We use the Single Family Loan-Level Dataset from Freddie Mac.<sup>8</sup> The loan-level dataset covers the performance on mortgages that Freddie Mac bought starting in 1999. The data includes higher-quality loans which had to conform to agency guidelines ([Adelino et al., 2016](#)). We use the provided information on the postal code, credit score *FICO*, loan-to-value *LTV* and debt-to-income *DTI* ratio of the borrower as well as the origination year, the seller and the delinquency status of the loan. We define a loan as delinquent past due more than 90 days. The seller of the loan is only disclosed for sellers which have at least 1% of the total original mortgage balance of all loans in a quarter. We merge the seller of the loan with the IT dataset but due to the limited number of sellers reported we only have 22 banks with information on technology adoption.

## 4 IT adoption and NPLs

In this section, we investigate the relationship between banks' IT adoption before the GFC and their NPLs during and outside the crisis. As a preview, [Figure 1](#) shows the evolution of the ratio of NPLs to assets from 1996 to 2014 for banks in the bottom and top quartile of the IT adoption distribution. This raw data shows that the two series are virtually indistinguishable until 2007. However, in 2008 –as NPLs start to

---

<sup>8</sup>See [Goodman and Zhu \(2015\)](#) for a detailed description and summary statistics. The dataset is also used by [Adelino et al. \(2016\)](#) and [Bartlett et al. \(2018\)](#).

surge– the two lines diverge. The growth in NPLs is considerably more pronounced for banks with low IT adoption. The NPLs peak in 2010 and the two series start converging again from 2011.<sup>9</sup>

#### 4.1 Panel

In our sample, the sharp rise of NPLs over assets occurred in the years from 2007 to 2010. Therefore, we define these years as the “crisis” period. To investigate whether banks with different level of IT adoption experienced different levels of NPLs during this period, we rely on the following panel equation:

$$NPL_{b,t} = \alpha_b + \delta_t + \beta IT_b \cdot crisis_t + (X_b \cdot crisis_t)' \gamma + \epsilon_{b,t} \quad (2)$$

where  $NPL_{b,t}$  is the share of non-performing loans relative to assets for bank (BHC)  $b$  in year  $t$ .  $IT_b$  is our bank-level measure of IT adoption before the crisis as defined in [section 3](#),  $\alpha_b$  and  $\delta_t$  are bank and year fixed effects, respectively. The former capture bank time-invariant heterogeneity while the latter capture time-varying aggregate shocks, such as business cycle fluctuations.  $X_b$  is a vector of bank-level variables that may be associated with NPLs during the crisis. The vector contains several pre-crisis bank characteristics: the ratio of loans to assets, the size as measured by the log of total assets, the share of capital over assets, the share of wholesale funding over assets, ROA, and the log of average wages.<sup>10</sup> We also include two controls capturing the geographic distribution of banks’ branches. The first one measures the exposure to the house price drop, while the second one constructs the IT adoption of other banks operating in the same location. We include observations for years between 2001 and 2014 and keep only observation for which we have all the variables in  $X_b$ . We are left with 4,608 observations on 337 banks. Since we include bank fixed effects, the variables  $IT_b$  and  $X_b$  appear only interacted with the crisis period dummy.

[Table 1](#) presents the results from estimating different version of [Equation 2](#) via OLS, together with standard error double-clustered at the bank and year level. We first present a less saturated version of the above equation. Column (1) shows the result of [Table 1](#) without the inclusion of bank fixed and year fixed effects as well as without controls. The base effect of technology adoption on non-performing loans is negative but not statistically significant. We do not find that IT adoption significantly affects non-performing loans during normal times. However, the interaction between the crisis dummy and IT adoption is negative and statistically significant. In times of the crisis banks that adopted more IT before the crisis had a significantly lower share of non-performing loans than banks with less IT adoption. This

<sup>9</sup>The dynamics of NPLs for bank with intermediate adoption lies between the low- and high- adopters in most years, see [Figure A2](#). To check that this pattern is mainly driven by the numerator of the series (NPLs) rather than the denominator (Assets), we fix the value of assets to the bank-specific pre-crisis average and plot the adjusted ratio in [Figure A3](#), finding a very similar pattern.

<sup>10</sup>We take the simple averages between 2001 and 2006 to measure pre-crisis levels; however, results are unchanged if we use a different pre-GFC window, such as 2004-2006.

result is robust to the inclusion of bank and year fixed effects and various controls. In addition, the coefficient is stable across specifications, suggesting a low correlation between the controls included and the measure of IT adoption. A one standard deviation higher IT adoption is associated with a between 13 and 17 basis points lower NPL share increase during the crisis. The average share of NPLs was 1.5 percent in the crisis period, while its standard deviation was 1.13. Therefore, a one standard deviation higher IT adoption led to a reduction in NPLs between 9 and 11% with respect to the mean and between 12 and 15% with respect to the cross-sectional standard deviation. Moreover, the increase between the pre-crisis average and the crisis NPL share is 1.05 percentage points. Therefore, if we ignore potential heterogeneity in the effect of IT adoption, spillover between banks (which we test for, see below), and general equilibrium effects, we find that a one standard deviation uniform increase in IT adoption across all banks would have diminished the surge in NPLs between 12 and 16%.

Columns (5)-(12) successively introduce additional controls to the baseline specification with only bank and year fixed effects. We start by controlling for the share of loans relative to assets in the pre-crisis period as a control. Banks that had more loans as a share of assets had a stronger increase in NPLs, as expected. Next, we introduce the exposure to the drop in house prices. We compute this exposure by weighting the drop in county-level house prices by the number of branches a bank has in this county. Banks that had more branches in counties in which house price dropped more suffered a stronger increase in NPLs. In column (7), we include the size of the bank interacted with the crisis dummy as an additional control. Consistent with [Sullivan and Vickery \(2013\)](#) larger banks had a stronger increase in NPLs in the crisis than smaller banks. The pre-crisis capital position, the pre-crisis wholesale funding and the return on asset ratio, as well as the average wage of the banks' employees did not have a significant impact on NPLs in the crisis.

Lastly, we add the IT of local competitors as an additional control variable. This variable can shed light on whether there are negative spillover effects of IT adoption on other banks. Individuals who want to borrow but are rejected by a high-IT bank could apply for a loan at a low-IT bank in the same area. If the low-IT bank does not identify the borrower as risky, the bank may grant a loan, which defaults during the crisis and leads to an increase in NPLs for this bank. If this mechanism is at work, we would still find a significant difference between high- and low-IT banks in terms of their NPLs during the crisis, but the aggregate increase in NPLs would be the same if all banks adopted more IT with ambiguous implications for financial stability. Column (12) shows that banks, which are based in areas, where their competitors adopted more IT did not suffer a stronger increase in their NPLs relative to banks, where their local competitors did not adopt IT intensively. This evidence suggests that IT adoption does not have negative spillover effects to local competitors.

In [Table A1](#) we conduct several robustness test of [Equation 2](#). Column (1) repeats the baseline. Col-

umn (2) uses another measure of IT adoption. Instead of using the bank fixed effects from Equation 1, we take the average share of PCs per employee across branches for each bank. While the disadvantage of the approach is that we do not control for branch-specific characteristics that drive the share of PCs per employee of the branch, the results still hold with this simpler measure. In column (3) and column (4) we provide robustness with a stronger winsorization of the IT adoption measure and NPLs, respectively, to reassure our results are not driven by outliers. In column (5) and (7) we use a different denominator for our dependent variable. In column (5) we divide by the overall loans of the bank while in column (7) we divide by the pre-crisis assets to ensure the denominator is not driving our results. Column (6) uses a different definition of NPLs. In the baseline, we classify loans to be non-performing if the loan is past due 90 days or more. In column (6) we use a broader classification by including loans if the loan is past due 60 days or more. Finally, column (8) shows results when the standard errors are clustered at the bank-level instead of double clustered at the bank and year level.

Inference based on standard errors estimated jointly with the parameter of interest of Equation 2 could be invalidated by the so-called “generated regressor” problem (Pagan, 1984). In fact, our measure of IT adoption is a result of estimating Equation 1. Therefore, we follow common practice (see e.g. Ashraf and Galor (2013)) and use a bootstrap procedure to re-estimate, with each random sample, both Equation 1 and Equation 2. Sampling, with replacement, is performed at the BHC-level. Table A2 present t-statistics based on clustered standard errors (as in Table 1) and on the the distribution of the bootstrap estimates. The t-statistics are very similar in both cases and inference based on any of the two methods leads to the conclusion of rejecting the null of no impact of IT adoption on NPLs. Therefore, the “generated regressor” problem does not have a significant role in our setting.

Next, we allow the impact of pre-crisis IT adoption on NPLs to vary each year between 1996 and 2014 by year by estimating the following equation:

$$NPL_{b,t} = \alpha_b + \delta_t + \sum_{\tau \neq 2006} \beta_{\tau} \cdot IT_b \cdot 1[t = \tau] + \epsilon_{b,t} \quad (3)$$

The coefficient of 2006 is normalized to zero. Results are illustrated in Figure 2. The red dot shows the point estimates of the interaction between the IT adoption and the year dummies,  $\beta_{\tau}$ , with the black bars reflecting the 95%, and the grey shaded area the 99% confidence interval, according to standard errors double clustered at the bank and year level. The effect of IT adoption on NPLs is insignificant in the pre-crisis period between 1996 and 2007, except for a small negative effect which is statistically significant at the 5% level in 2002, likely due to the early 2000 recession. As shown in Table 1 banks which adopted more IT before the crisis had significantly lower NPLs than their counterparts in the crisis. In particular, between 2007 until 2010 the effect is negative and statistically significant at the 5% level and in 2009 and 2010 the effect is even statistically significant at the 1% level. The coefficient reaches its

maximum in 2010 with -0.3. In other words, a one standard deviation higher IT adoption was associated 30 basis points lower NPLs in 2010. The impact is still negative in 2011 and 2012 although not statistically significant anymore. We detect no impact in the two latest years of the sample, 2013 and 2014.

## 4.2 Cross-sectional analysis

In this section, we analyze the relationship between bank-level IT adoption and various other bank characteristics in the cross-section.

We apply OLS to the following equation:

$$Y_b = \alpha + \beta \cdot IT_b + \epsilon_b \quad (4)$$

where  $Y_b$  is either the share of NPLs over assets in the crisis period or one of the control variables in the set  $X_b$  described above and the independent variable is the pre-crisis IT adoption. Collapsing the data in this way has the advantage of avoiding the under-estimation of standard errors that can arise when estimating a diff-in-diff specification using panel data (Bertrand et al., 2004).

Table 2 present the relative results. Consistently with the panel (Table 1 and Figure 2) technology adoption is strongly negatively correlated with NPLs in the crisis period (column 1) with an R-squared of 2.6%. The magnitude of the coefficient (18 basis points) is slightly higher but very similar to the one estimated in the panel regressions.

Columns (2)-(8) test whether IT adoption on the bank-level is correlated with other bank-level variables that could be important in driving NPLs in the post-crisis period. We include several pre-crisis variables to capture banks' business model, profitability, and capital structure: ratio of loans, wholesale funding, and capital to assets, and ROA. We compute banks' exposure to the house price shock using the drop in house price for each county (peak to trough) and the pre-crisis geographical distribution of branches. We also use the log of pre-crisis assets to measure bank size, and the log of average wage as a proxy for employees' human capital. We find that IT adoption is not significantly related to any of these characteristics.<sup>11</sup> Moreover, the R-squared of column (1) is much larger (at least 4 times) than the ones of columns (2)-(8).

We, therefore, conclude that IT adoption is not correlated with any important bank-level characteristics that could predict their exposure to the GFC. This result is a comforting "falsification" test since it suggests IT is unlikely to be correlated to other unobservable characteristics that would also make them more exposed to the financial shocks and related recession.

---

<sup>11</sup>We compute additional variables, such as the share of residential or personal loans over the total amount of loans, and find no correlation of these variables with IT adoption either. Additionally, in section 6 we present direct evidence that the impact of IT on NPLs is not driven by location of lending activities.



We measure the IT adoption of banks' competitors by relying on the pre-crisis geographical distribution of branches. We find a positive correlation between the banks' own IT adoption and its competitors'.

In column (10) we include all the bank characteristics as controls in a regression of post-crisis NPLs on IT adoption. We find a very small change in the coefficient of IT adoption, despite the ten-fold increase in R-square. This suggests that the equation of column (1) does not suffer from an omitted variable bias and pointing towards a causal relationship between IT adoption and NPLs (Altonji et al., 2005; Oster, 2019). Oster (2019) provides formal statistical procedures to assess the stability of OLS coefficients to the inclusion of relevant control and test for the potential bias arising from the presence of other unobservable variables. We apply these procedures and find that the coefficient on IT adoption compatible with the inclusion of all unobservable variables is minus 14 basis points, which is similar to baseline and well below zero.<sup>12</sup> Therefore, our results seem to be robust to the presence of unobservable variables.

Some of the bank characteristics are important for the share of NPLs in the crisis period. Similarly to Table 1, banks with more loans and stronger geographic exposure to the house price shock suffer higher levels of NPLs. Besides, we find that banks that had more wholesale funding had larger increases in NPLs. Importantly, as in Table 1 the IT adoption of local competitors in the pre-crisis period did not have a significant impact on the increase in NPLs in the crisis period, suggesting that there are no negative spillover effects of IT adoption.

## 5 The Roots of IT Adoption

Why do some banks adopt less IT than others despite its beneficial effects?

In subsection 4.2 we show that several important pre-GFC characteristics, such as the share of loans in their balance sheet or their reliance on wholesale funding, do not predict the adoption of IT. In this section, we investigate two alternative sources of heterogeneity in IT adoption: the background and personal inclination of BHC top executives and the supply of technical skills generated by the establishment of Land-Grant colleges across the United States.

The focus on bank's top executives is grounded in the descriptive patterns documented in section 3: the explained variation in technology adoption at the branch-level is driven by bank characteristics (60%) relative to geographic characteristics (27%).

For the same reason, we investigate how the distance between a bank's headquarter and Land-Grant colleges, which were established at the end of nineteenth century, impacts IT adoption. We argue that, considered the findings of previous literature and complementary evidence presented in this paper, the

---

<sup>12</sup>This test has been used in other empirical studies in finance and economics, such as Mian and Sufi (2014). Following Oster (2019) jargon, we set the "hypothetical R-square" to 1, which is the most conservative choice. We find a treatment effect of -.14 and a relative degree of selection above 1. These results are valid under the assumption of "proportional selection of observables and unobservables" that is discussed in Altonji et al. (2005), Oster (2019), and Finkelstein et al. (2019).

distance of headquarters from these colleges can be used as a valid instrument for banks' pre-GFC adoption IT. IV estimates confirm the main findings of the paper.

## 5.1 Executives' Background

Economists have documented huge—and interrelated—dispersion in firm- and plant-level productivity (Syverson, 2011), management practices (Bloom et al., 2019), and IT adoption (see references in [section 2](#)). While a complete understanding of this dispersion is far from being accomplished, a robust literature has highlighted several relevant factors. In particular, it has been shown that frictions of different natures, such as lack of proper information (see Section VI.A in Bloom et al. (2013)), incentives (Atkin et al., 2017), or financial constraints (Duval et al., 2020; Manaresi and Pierri, 2019), have a prominent role in preventing or slowing down the adoption of superior practices and technologies.

A related strand of literature documents the importance of characteristics and background of executives for firm outcomes and performance (Benmelech and Frydman, 2015; Bennedsen et al., 2006; Bertrand and Schoar, 2003). Therefore, we conjecture that top executives that have a more tech-prone background and orientation may be important to overcome these frictions and promote a higher degree of IT adoption in the banks they lead.

To capture the tech-orientation of the top executives (CEOs, CFOs, COOs, and Presidents) hired before 2007, independently of whether they are still active, we search for tech-related keywords in their biographies, as described in [section 3](#). We compute an overall score for each bank regarding the “tech-intensity” of their executives. We can then match regulatory data, IT adoption, and executive biographies data for 249 banks.

To test whether these differences are statistically significant, we then estimate the following cross-sectional regression model:

$$Y_b = \alpha + \beta \cdot ExecIT_b + \epsilon_b \quad (5)$$

where  $b$  is a bank in our sample,  $ExecIT_b$  is the “tech-orientation” of  $b$ 's executives, and the dependent variable  $Y_b$  is either the pre-crisis IT adoption or the level of NPLs over assets during the crisis period. Both independent variables are standardized to have mean zero and variance one.

Results are presented in [Table 3](#). Column (1) repeats the baseline specification on this sub-sample of banks, finding similar results than in [Table 2](#) (although slightly smaller and noisier). Column (2) shows that banks led by more tech-oriented executives experienced lower NPLs during the crisis. Column (3), instead, shows that executives' background is a significant predictor of bank's IT adoption. Columns (2) and (3) are consistent with our hypothesis that tech-prone executives were instrumental in leading

banks to adopt IT more intensively and experience the related benefits afterward. The lower statistical significance of the coefficient in column (3) vs (2) may be due to differences in data sources: the left-hand side variable of column (2) is measured from regulatory data rather than a survey, and therefore it is likely to contain less measurement error.

We also reestimate [Equation 3](#), which sheds light on the time-varying impact on IT-adoption on NPLs, replacing our baseline IT measure with the tech-orientation of the executives. [Figure 3](#) shows strikingly similar results. Banks led by more tech-oriented executives had statistically indistinguishable NPLs in any year from the 90s to the GFC compared to banks with a less “tech-oriented” leadership. However, banks that were managed by tech-savvy executives experienced a significantly more limited increase in NPLs during—and right after—the GFC. The results are also similar in terms of economic magnitudes. While a one standard deviation higher IT-adoption using our baseline measure is associated with a 30 basis points lower NPLs during the peak of the crisis, a one standard deviation higher IT adoption in terms of the tech-savyness of the managers is associated with a 21 basis points lower NPL ratio. A similar pattern emerges from the raw data ([Figure A4](#)).

A potential concern with the results of [Table 3](#) and [Figure 3](#) is that the tech-orientation of executive may be capturing in general higher human capital of these executives. If tech-prone executives are just better managers, then the impact on NPLs may be due to better management practices unrelated to IT itself. We therefore re-estimate [Equation 5](#) including the (log of the) pre-crisis compensation of the executives as a control. Results, reported in [Table 4](#), are absolutely unaffected by the inclusion of such a control. Moreover, compensation itself does not have explanatory power neither for NPLs during the crisis nor for IT adoption before the crisis. Insofar as compensation can be used as a proxy for human capital ([Becker, 2009](#)), these results show that is specifically tech-orientation, and not general quality or skills, that matters for our variables of interests. Furthermore, the results are unaffected to the inclusion of the non-base share of total compensation, that can shape risk-taking incentives ([Meiselman et al., 2018](#)), as reported in [Table A3](#).

A further concern is that the list of words used to measure tech-orientation of executives from their biographies is somehow ad-hoc. We, therefore, test the robustness of our results to the choice of words. For each word, we compute an additional tech-orientation measure based on all the remaining words. We then re-estimate the cross-sectional regressions with each of the different tech-orientation measures. We plot the estimated coefficients in [Figure A5](#), where panels top to bottom refer to the columns (2) and (3) in [Table 3](#). We find that all our results are robust to the use of exclusion of any word in the list, as (a) the coefficients are all clustered around the estimates of [Table 3](#) (flagged by a dashed lined) and fairly close (b) all have the same sign (negative in top panel and positive in the bottom panel).

## 5.2 The Land-Grant Colleges

We have shown that branch-level IT adoption is predominately driven by bank-level characteristics rather than the location of the branches, with the technical background of the executives playing an important role. This suggest that the IT decisions at the headquarter of the bank are crucial for branch-level IT adoption and that these decisions are affected by the technical knowledge of decision makers. Therefore, factors affecting the supply of technical skills in the location of a BHC headquarters may affect the IT adoption in the whole group.

The Morrill Act of 1862 endowed federal land to states to found universities. The focus was to teach science, agriculture, and other technical subjects, due to a nationwide demand for more technical skills, and to contrast with liberal arts colleges. While some land-grant universities now offer degrees in both arts and science, their focus remains technical even until now. Indeed, in appendix A1 we show that students at land-grant colleges and universities are still much more likely to major in engineering and less likely to major in non-technical fields, such as education or business. We also find that SAT scores in math are higher for students of land-grant university students, but their writing scores are not.

The location of land-grant universities has been used as an instrument for the supply of skilled labor in a metropolitan area (Moretti, 2004) as their exact location is largely due to historical accidents.<sup>13</sup> Moreover, land-grant colleges are distributed evenly within a state and independent of Census regions; workers in areas close to a land-grant college are shown to be similar in terms of racial and demographic characteristics and have very close Armed Forces Qualification Test scores for a given level of education (Moretti, 2004; Shapiro, 2006). Land-grant universities were also not established in areas that were richer due to natural resources or other factors (Carstensen, 1963; Sawyer, 1981; Williams, 2010).

The location of many banks' headquarters<sup>14</sup> is also related to their historical heritage and usually predates the IT revolution. For instance, Bank of America's headquarter location in Charlotte (North Carolina) was established 1874 by the foundation of the "Commercial National Bank" (Blythe and Brockmann, 1961). More generally, in appendix A1 we show that the presence of a BHC headquarter in a US county is uncorrelated with proximity to the land-grant colleges. Therefore, distance from one of these colleges is plausibly exogenous with respect to the most important factors affecting banking industry and headquarter choice.

We conjecture that land-grant colleges may impact IT adoption at the bank level by altering the distribution of technical knowledge across the country. To test the plausibility of this inclusion restriction, we ask whether the distance between a bank's headquarter and a land-grant college is correlated to IT adoption. We estimate the set of linear regressions:

---

<sup>13</sup>As an example, the choice of Ithaca over Syracuse for the establishment of Cornell University was due to the fact that one of the two founders was robbed while visiting Syracuse (Andrews, 2019).

<sup>14</sup>For each BHC, we take the headquarter county from regulatory filings in 1995 or earliest available year.

$$IT_b = \delta + \gamma_j \cdot D_{b(c),j} + \epsilon_{b,j}$$

where  $j = 1, \dots, 70$  is one of the land-grant colleges and universities established in the nineteenth century in the mainland US,<sup>15</sup>  $D_{b(c),j}$  is the distance between the county where bank's  $b$  headquarter is located and the county where college  $j$  is located, expressed in log of miles (plus one).  $\gamma_j$  is statistically different than 0 in 33 cases out of 70 and it is negative 90% of these cases. Furthermore, if we order colleges by distance from the headquarter (so that  $j = 3$ , for instance, is the third closest college), then  $\gamma_j$  is negative every time it is statistically different than zero (57 out of 70 times). These results indicates that the closer a bank headquarter is to land-grant colleges the more intensely IT is adopted.

Under the assumption that the distance between land-grant colleges and headquarters impact NPLs only through the effect on technological adoption, we can use this variation as a valid instrument to recover the causal effect of IT adoption. A possible threat to this exclusion restriction is that banks lend in areas close to their headquarter and these areas are more resilient to a shock through the effect of land-grant colleges on local education. Since we focus on BHC, whose lending portfolio is usually geographically diversified, this is less of a concern than if we were to focus on smaller banks. Moreover, we show below that our IV estimates are robust to the inclusion of several characteristics of the county where the bank is headquartered, such as income and educational level. This also reassures that the land-grant instrument is capturing specifically the impact of technical knowledge rather than higher level of education in general. Finally, the finding that land-grant colleges have fewer students majoring in business and management science should eliminate the concern that their impact goes through better management practices rather than IT adoption itself.

We implement the IV empirical strategy by estimating the following 2SLS model:

$$IT_b = \delta + F(D_{b(c),j}, \gamma) + \eta_b \quad (6)$$

$$NPL_b = \alpha + \beta \cdot IT_b + \epsilon_b \quad (7)$$

where  $F(D_{b(c),j}, \gamma)$  is a function of the set of instruments  $D_{b(c),j}$ , known up to some parameter  $\gamma$ , and  $NPL_b$  is the level of NPLs over assets during the GFC. Results are reported by Table 5. Given the availability of many instruments the econometrician is left to find an appropriate function  $F(D_{b(c),j}, \cdot)$ . A first option would be to employ some simple statistics, such as the distance from the closest college, the average, and the median from each college. Unfortunately, we found that none of these statistics have enough power to estimate the impact of IT on NPLs during the financial crisis (results unreported). In column (2) we use a vector of 5 instruments, which are the distance between bank's headquarter and

---

<sup>15</sup>All the results of this section hold if we exclude Alaska and Hawaii from the analysis.

the closest 5 land-grant colleges. The estimated  $\beta$  is negative and statistically different from zero but it is not statistically different from OLS estimate. However, the F-statistics of the first stage is low, suggesting that a weak instrument problem might arise leading to poor small sample properties. We then use all the 70 instrument at our disposal, see column (3). The estimated  $\beta$  is negative, statistically different from zero, and close to OLS value. The F-stat of the first stage is very close to the “rule-of-thumb” value of 10. However, the weak-instruments diagnostic proposed by [Stock and Yogo \(2005\)](#) reveals a potential many-weak-instruments problem, as the Cragg-Donald Wald F statistic is low (1.081) compared to the reference value for 10% maximal IV size bias (10.99).

Hence, we use LASSO to optimally combined the available information.<sup>16</sup> In fact, [Belloni et al. \(2012\)](#) show that, under certain regularity and sparsity conditions, LASSO procedure delivers the optimal instrument. Columns (4) to (6) use the LASSO optimal instrument. Results are qualitatively consistent with the main findings of the paper: higher pre-crisis IT adoption leads to significantly lower NPLs during the crisis. The results are robust to the inclusion of several pre-GFC characteristics of the county where the bank is headquartered, such as the level of education and household income and several bank-level pre-GFC characteristics. The results are also unaffected by the inclusion state of the headquarter fixed effects. The F-statistics of the first stage is always above the value of ten. Moreover, diagnostic tests based of [Stock and Yogo \(2005\)](#) do not detect weak-instrument problems, except for the specification including state fixed effects: since we have only a few banks per each state the amount of variation available for estimation is limited.

The coefficients on IT adoption estimated by instrumental variables are larger in magnitude than the OLS one, reported by column (1). However, in 4 out of 5 columns, we cannot reject the null that the estimates are different than the baseline value of -.18 at 10% confidence level. Since we cannot reject the null of OLS’ consistency, these estimates should be preferred, as OLS is a more efficient estimator than the IV.

The results presented in this section are informative about the roots of the dispersion in IT adoption, as they highlight the importance of executives’ background and of the heterogeneity of technical knowledge across the country. Moreover, they are a strong support for the causal interpretation of the relationship between IT and NPLs. In fact, they point towards IT as the cause of the lower NPLs during the crisis, rather than other unobservable characteristics, such as the quality of management practices.

---

<sup>16</sup>LASSO (least absolute shrinkage and selection operator) is a machine learning technique used to predict an outcome variable using a parsimonious set of covariates selected from a large set of variables, possibly larger than the number of observations. Similarly to OLS, it works by minimizing the sum of square residuals but it is augmented with a penalty for any non-zero coefficient. We select the value of the penalty via cross-validation to avoid model overfitting.

## 6 Loan-Level Analysis

We use the Single Family Loan-Level Dataset to study the performance and characteristics of mortgages originated by banks with heterogeneous degrees of IT adoption and sold to Freddie Mac. This analysis is useful to investigate the channels through which high-IT adoption banks were able to limit the surge in NPLs. We estimate the following loan-level equation:

$$Delinquent_l = \alpha_{z(l)} + \delta_{o(l)} + \beta \cdot IT_{b(l)} + X_l' \gamma + \eta_l \quad (8)$$

where  $l$  is a mortgage held by Freddie Mac and originated before 2007 by a commercial bank in our IT sample;  $Delinquent_l$  is either the fraction of months, within the crisis period, during which the loan is delinquent or a dummy variable indicating whether the loan has ever been delinquent during the same period. For consistency with the bank-level regressions, we flag a loan as delinquent if it has a past due 90 days or above and we define the crisis period as the years between 2007 and 2010.<sup>17</sup>  $\delta_{o(l)}$  are fixed effects for the year of origination and  $\alpha_{z(l)}$  are fixed effects for the 3-digit postal code of the underlying property. While the origination-year fixed effects capture for example business cycle dynamics, the postal code fixed effects control for local heterogeneity that can arise, for instance, from the severity of the GFC and or from different house market dynamics.  $IT_b$  is bank-level technology adoption. Freddie Mac data does not report the name of all mortgage sellers, so we can only match 22 commercial banks.  $X_l$  is a vector of mortgage characteristics at origination: borrower's FICO score, Loan-to-Value (LTV) ratio, and Debt serving-to-Income (DTI) ratio.

Table 6 reports estimates of different versions of Equation 8, together with standard error clustered at the seller level. The equation is estimated with OLS except that for the last column where we rely on a probit model. Dependent variables are multiplied by 100 (except for the last column) while independent variables are normalized to have mean zero and unitary standard deviation so that coefficients are easily comparable. Column (1) shows without any controls that loans sold by technology adopters were delinquent fewer months in the crisis than other loans. Columns (2) and (3) include year-of-origination and location fixed effects and column (4) add the vector of mortgage characteristics at origination. All specifications illustrate that mortgage sold by banks with higher IT adoption spent a significantly smaller fraction of time in a delinquent status during the crisis. A loan that has been originated by a bank with a one standard deviation higher IT adoption has been delinquent 32 basis points fewer share of the time, i.e. around 10% less than the average loan.

The magnitude of this relationship is not trivial compared to loan-level characteristics that have been shown to be important predictors of default: a one standard deviation higher IT adoption has the same

---

<sup>17</sup>The results are robust to flagging loans to be delinquent if they have been past due for different periods.



predicted effect on delinquency that a third of a standard deviation lower LTV ratio, half of a standard deviation of lower DTI, or 13% of a standard deviation higher FICO score. In column (5) we allow the coefficient of IT adoption to be different for borrowers above and below the median credit score (735), finding that only mortgages given to relatively riskier borrowers are impacted by lenders' IT adoption. In column (6) and (7) we show that the results are qualitatively the same if we use a Linear Probability Model or a Probit Model to estimate the impact of IT adoption and mortgage characteristics on the probability that a mortgage has ever been delinquent during the crisis period.

The loan-level analysis serves multiple purposes. It allows us to dig deeper into the mechanisms behind the relationships between IT adoption and NPLs. It shows that at least part of the effect we document in [subsection 4.1](#) is due to the origination of more resilient loans before the crisis rather than other channels, such as a better ability to manage or dismiss NPLs. For instance, if the effect of IT on delinquencies for loans which are still on banks' balance sheets would solely come from banks better ability to collect mortgage payments, we would not expect the same effect for off-loaded mortgages.

It also shows that high-IT adoption banks were not offloading low-quality loans to GSEs. If technology-prone banks were simply better able to securitize and offload their bad loans, IT adoption would lead to lower on-balance sheet NPLs during the crisis, without reducing the amount of NPLs in aggregate ([Acharya et al., 2013](#)). If this was the case, technology adoption would only lead to risk shifting and increase moral hazard issues and not enhance financial stability.

The mortgage data allows us to control for additional characteristics of the loan, which also sheds more light on the channel through which IT adoption can affect NPLs, such as the postal code of the underlying property and the year of origination. The results confirm that the impact of IT adoption on NPLs is not fully driven by high-IT adopters lending to areas that were hit less by delinquency and foreclosures or originating a larger amount of loans in a particular year. Hence, the IT adoption of banks seems to give banks an informational advantage regarding the mortgage and not only about the location of the borrower or the business cycle. However, the estimated impact of IT on delinquency declines by almost a fourth once we control for location.

We can also control for a few of the most important characteristics that predict mortgage performance at origination: the borrower FICO score, the LTV, and the DTI ratios. We find that the impact of IT on delinquency does not disappear if we control for these characteristics in a linear fashion. This implies that either high-IT adopters had additional information available that affected their lending decisions or had more sophisticated ways to use these variables when deciding whether and how much to lend to a borrower.<sup>18</sup>

---

<sup>18</sup>For example, more IT might have allowed these banks to sustain more reliable internal rating systems. We refer to [Berg \(2015\)](#) and [Berg et al. \(2019a\)](#) for a description of internal rating systems.

In unreported results we find that IT adoption is uncorrelated with the default predictors mentioned above. This suggests that the high-IT adopters did not just choose to focus on a safer segment of the market before the crisis but actively selected better borrowers. Moreover, the availability of the FICO score allows us to investigate which types of mortgages were most affected by IT availability. We find that the delinquency rate is affected by IT adoption only for mortgage issued to borrowers with credit scores below the sample median.

The loan-level analysis has also some drawbacks. We can only match 22 banks with the dataset on IT adoption, limiting the statistical power of our analyses and our ability to further investigate the differences in lending practices between high- and low-IT adopters. Moreover, the loans in this dataset differ in some characteristics, such as the average credit score, compared to the portfolio of loans which were kept on the balance sheet of banks (Keys et al., 2010).

## 7 Bank Lending

High levels of NPLs weigh on banks' profitability and can, therefore, constrain their lending, depressing real economic activity. As IT adoption improves banks' resilience, it may also shield their ability to provide credit to customers during (and right after) financial turmoil.

Figure 4 reports the share of total loans (normalized by pre-GFC assets) for banks in the high- and low- IT adoption groups from 2001 to 2014. The two series are indistinguishable up to 2006, consistent with the lack of pre-crisis correlation between lending intensity and IT documented in Table 2. From 2007 on, the amount of loans provided by low-IT adopters is remarkably lower than the one provided by the more IT intense counterparts. The two series start converging from 2012 but the difference is still present in 2014. These patterns suggest that heterogeneity in IT adoption, perhaps through the impact on NPLs documented in section 4, has a role in explaining banks' different lending dynamics during and after the crisis.

In this section, we formally test whether banks with fewer NPLs during the crisis and with higher IT adoption were lending more during the crisis. We follow Peek and Rosengren (2000) in defining lending as the change in loans over total assets and, as in the rest of the paper, we take the average across the crisis period. We estimate the following cross-sectional specification:

$$\overline{\Delta Loans}_b^{GFC} = \alpha + \beta \cdot X_b + \epsilon_b \quad (9)$$

where  $X_b$  either the share of NPLs in the crisis period or the pre-crisis IT adoption. Results are presented in Table 7 and each regression is estimated with and without the set of controls discussed in subsection 4.2.

Similarly to [Peek and Rosengren \(2000\)](#), we find that a 100 basis points higher NPLs to assets ratio is associated with about 100 basis lower average loan growth. We also show a significant impact of IT adoption on lending: one standard deviation increase in IT adoption is associated with a 33 basis points higher loan growth during the crisis, which is about 20% of its mean.

These results show that IT adoption helped banks providing credit during (and after) the GFC. It is difficult to know whether IT adoption had an impact only because it mitigated the surge in delinquency rates ([section 6](#)) and NPLs ([section 4](#)) or it also improved banks' ability to function during the shock and expand afterward in other ways. In either case, our results indicate that IT intensity improved financial stability during the GFC.

## 8 Conclusion

As the financial industry becomes more and more reliant on Information Technology, as exemplified by the surge of FinTech players, it is extremely policy-relevant to understand the consequences for financial stability of a more intense use of IT in lending decisions.

In this paper, we measure the heterogeneous degree of IT adoption of US commercial banks before the GFC using a novel dataset. We show that high-IT-adopters experienced a significantly smaller increase in NPLs on their balance-sheets relative to other banks and provided more credit to the economy during the crisis. High- and low-IT-adopters were not differentially exposed to the GFC in terms of pre-crisis geographical footprint and business model. Moreover, loans originated by high-IT banks experienced lower delinquency rates during the crisis even when they were securitized and sold to Freddie Mac. Therefore, our results indicate that IT adoption helped banks to select better borrowers and produce more resilient loans. We also show that the roots of this heterogeneity seem to be partially related to the “tech-orientation” of their top-executives, which we capture with a simple text-analysis algorithm, and to the uneven diffusion of technical knowledge across the US. We finally propose an instrumental variable strategy based on the distance between the grant-land colleges and banks' headquarters. The IV estimates confirm our main findings.

The evidence presented in this paper suggests that the “FinTech era” is likely to be beneficial to financial stability. The main caveat of using our results to inform the debate on FinTech and financial stability is that the technologies adopted by commercial banks before the GFC might be significantly different than the ones that banks, FinTech firms, and financial arms of BigTech companies are implementing nowadays.

We offer three main considerations to support the relevance of our results despite this obvious limitation. First, our measure of IT adoption, while based on the simple counting of computers divided

by the number of employees within a branch, is still informative about technological intensity more broadly defined in very recent years. A simple regression of this measure against the overall IT-budget of an establishment in 2016 delivers an R-square of 44% and a correlation coefficient of 65%. Moreover, the adoption of frontier technologies, such as Cloud Computing, is also positively correlated with our simple measure.

Second, many of the IT-driven changes in the financial industry are so recent that they have not experienced yet—at the time of the writing— a large financial systemic shock testing their resilience. In addition, the share of lending provided by “FinTech” firms is still small in most countries. Therefore, it is important to collect the best possible empirical evidence from past systemic shocks to inform the current debate. Analogously, despite the ever-changing features of economies and financial systems, the lessons learned during the Great Depression were useful in shaping the policy-response to the Global Financial Crisis (Bernanke, 2015). Conversely, ignoring evidence regarding past crises because of the observed differences with the present scenario might lead to highly undesirable outcomes (Reinhart and Rogoff, 2009).

Third, if one focuses on the lending business, there are several commonalities between the IT-intensive methods used before GFC and the most recent advancements. Statistical models to predict defaults were widely used during the decade preceding the GFC (Rajan et al., 2015). The up-to-date machine learning techniques that are used to predict borrowers’ behavior are more powerful versions of the previously available statistical tools, rather than radically different systems. In fact, introductory university courses on machine learning often list linear regressions, probit, or logit models as simple examples.<sup>19</sup> The collection and use of new data to inform application decisions, such as the digital footprint (Berg et al., 2019b), is not conceptually different than the use of credit scores. The main difference lies in the requirements—in terms of infrastructure and know-how—to acquire, store, manage, and employ these data. The decline in the cost of computing power, that allows this progress to happen, occurs at a constant rate (Moore’s law).

We conclude by underlining that, since we study IT adoption of traditional banks, we are silent on many institutional features associated to FinTech, such as the connection with shadow banking and the room for regulatory arbitrage (Buchak et al., 2018). These features may be relevant for financial stability.

---

<sup>19</sup>For instance, the Linear Regression is one of the two topics covered in lecture 2 of “CS229: *Machine Learning*” at Stanford University, according to the 2019 syllabus (available at <http://cs229.stanford.edu/syllabus.html>). Weighted Least Squares and Logistic Regression are covered in lecture 3.

## References

- Acharya, Viral V, Philipp Schnabl, and Gustavo Suarez, "Securitization without risk transfer," *Journal of Financial Economics*, 2013, 107 (3), 515–536.
- Adelino, Manuel, Antoinette Schoar, and Felipe Severino, "Loan originations and defaults in the mortgage crisis: The role of the middle class," *The Review of Financial Studies*, 2016, 29 (7), 1635–1670.
- Akerman, Anders, Ingvil Gaarder, and Magne Mogstad, "The skill complementarity of broadband internet," *The Quarterly Journal of Economics*, 2015, 130 (4), 1781–1824.
- Altonji, Joseph G, Todd E Elder, and Christopher R Taber, "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 2005, 113 (1), 151–184.
- Andrews, Michael J, "Local Effects of Land Grant Colleges on Agricultural Innovation and Output," *NBER working paper*, 2019.
- Ashraf, Quamrul and Oded Galor, "The 'Out of Africa' hypothesis, human genetic diversity, and comparative economic development," *American Economic Review*, 2013, 103 (1), 1–46.
- Atkin, David, Azam Chaudhry, Shamyra Chaudry, Amit K Khandelwal, and Eric Verhoogen, "Organizational barriers to technology adoption: Evidence from soccer-ball producers in Pakistan," *The Quarterly Journal of Economics*, 2017, 132 (3), 1101–1164.
- Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly Journal of Economics*, 2003, 118 (4), 1279–1333.
- Bai, Jennie, Thomas Philippon, and Alexi Savov, "Have financial markets become more informative?," *Journal of Financial Economics*, 2016, 122 (3), 625–654.
- Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-Lending Discrimination in the Era of FinTech," *Unpublished working paper. University of California, Berkeley*, 2018.
- Basten, Christoph and Steven Ongena, "The Geography of Mortgage Lending in Times of FinTech," *Swiss Finance Institute Research Paper*, 2019.
- Beaudry, Paul, Mark Doms, and Ethan Lewis, "Should the personal computer be considered a technological revolution? Evidence from US metropolitan areas," *Journal of Political Economy*, 2010, 118 (5), 988–1036.
- Beccalli, Elena, "Does IT investment improve bank performance? Evidence from Europe," *Journal of Banking & Finance*, 2007, 31 (7), 2205–2230.
- Becker, Gary S, *Human capital: A theoretical and empirical analysis, with special reference to education*, University of Chicago press, 2009.
- Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen, "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 2012, 80 (6), 2369–2429.
- Benmelech, Efraim and Carola Frydman, "Military ceos," *Journal of Financial Economics*, 2015, 117 (1), 43–59.

- Bennedsen, Morten, Francisco Perez-Gonzalez, and Daniel Wolfenzon, “Do CEOs Matter?,” *NYU Working paper no. FIN-06-032*, 2006.
- Berg, Tobias, “Playing the devil’s advocate: The causal effect of risk management on loan quality,” *The Review of Financial Studies*, 2015, 28 (12), 3367–3406.
- , Manju Puri, and Jorg Rocholl, “Loan Officer Incentives, Internal Rating Models and Default Rates,” *Review of Finance (forthcoming)*, 2019.
- , Valentin Burg, Ana Gombović, and Manju Puri, “On the rise of FinTechs—Credit scoring using digital footprints,” *The Review of Financial Studies (forthcoming)*, 2019.
- Berger, Allen N, “The economic effects of technological progress: Evidence from the banking industry,” *Journal of Money, Credit and Banking*, 2003, pp. 141–176.
- Bernanke, Ben S, *The courage to act: A memoir of a crisis and its aftermath*, WW Norton & Company, 2015.
- Berti, Katia, Christian Engelen, Borek Vasicek et al., “A macroeconomic perspective on non-performing loans (NPLs),” *Quarterly Report on the Euro Area (QREA)*, 2017, 16 (1), 7–21.
- Bertrand, M and A Schoar, “The Managing with Style: The Effect of Managers on Corporate Policy,” *Quarterly Journal of Economics*, 2003, 118 (4), 1169.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, “How much should we trust differences-in-differences estimates?,” *The Quarterly journal of economics*, 2004, 119 (1), 249–275.
- Bessen, James E and Cesare Righi, “Shocking Technology: What Happens When Firms Make Large IT Investments?,” *Boston Univ. School of Law, Law and Economics Research Paper*, 2019, (19-6).
- BIS, “Resolution of non-performing loans—policy options,” *BIS FSI Insights on policy implementation*, 2017, 3.
- Bloom, Nicholas and Nicola Pierri, “Cloud Computing Is Helping Smaller, Newer Firms Compete,” *Harvard Business Review*, 2018.
- , Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, “Does management matter? Evidence from India,” *The Quarterly Journal of Economics*, 2013, 128 (1), 1–51.
- , Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen, “What drives differences in management practices?,” *American Economic Review*, 2019, 109 (5), 1648–83.
- , Raffaella Sadun, and John Van Reenen, “Americans do IT better: US multinationals and the productivity miracle,” *American Economic Review*, 2012, 102 (1), 167–201.
- Blythe, LeGette and Charles Raven Brockmann, *Hornets’ Nest: The Story of Charlotte and Mecklenburg County*, Public Library of Charlotte and Mecklenburg County, 1961.
- Bostandzic, Deneafa and Gregor NF Weiss, “Innovating Banks and Local Lending,” *Available at SSRN 3421123*, 2019.
- Bresnahan, Timothy F, Erik Brynjolfsson, and Lorin M Hitt, “Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence,” *The Quarterly Journal of Economics*, 2002, 117 (1), 339–376.

- Brynjolfsson, Erik and Lorin M Hitt, "Computing productivity: Firm-level evidence," *Review of Economics and Statistics*, 2003, 85 (4), 793–808.
- Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru, "Fintech, regulatory arbitrage, and the rise of shadow banks," *Journal of Financial Economics*, 2018, 130 (3), 453–483.
- Caballero, Ricardo J, Takeo Hoshi, and Anil K Kashyap, "Zombie lending and depressed restructuring in Japan," *American Economic Review*, 2008, 98 (5), 1943–77.
- Carlin, Bruce I, Arna Olafsson, and Michaela Pagel, "FinTech and Consumer Financial Well-Being in the Information Age," 2019.
- Carstensen, Vernon, "The State Universities and Democracy. By Allan Nevins.(Urbana: University of Illinois Press, 1962. x+ 171 pp. Notes, appendix, and index.)," 1963.
- Claessens, Stijn, Jon Frost, Grant Turner, and Feng Zhu, "Fintech credit markets around the world: size, drivers and policy issues," *BIS Quarterly Review September*, 2018.
- D'Acunto, Francesco, Alberto Rossi, and Michael Weber, "Crowdsourcing financial information to change spending behavior," *Chicago Booth Research Paper*, 2019, (19-09).
- Dante, Hugo and Christos A. Makridis, "The Rise of Mobile Banking and the Displacement of Brick-and-Mortar Branches, 2014-2019," *mimeo*, 2019.
- Degryse, Hans and Steven Ongena, "Distance, lending relationships, and competition," *The Journal of Finance*, 2005, 60 (1), 231–266.
- Demirgüç-Kunt, Asli and Enrica Detragiache, "Does deposit insurance increase banking system stability? An empirical investigation," *Journal of Monetary Economics*, 2002, 49 (7), 1373–1406.
- Duval, Romain, Gee Hee Hong, and Yannick Timmer, "Financial frictions and the great productivity slowdown," *The Review of Financial Studies*, 2020, 33 (2), 475–503.
- Farboodi, Maryam, Adrien Matray, and Laura Veldkamp, "Where has all the big data gone?," *working paper*, 2018.
- Finkelstein, Amy, Matthew Gentzkow, and Heidi L Williams, "Place-based drivers of mortality: Evidence from migration," *NBER working paper series*, 2019.
- Forman, Chris, Avi Goldfarb, and Shane Greenstein, "The Internet and local wages: A puzzle," *American Economic Review*, 2012, 102 (1), 556–75.
- Frost, Jon, Leonardo Gambacorta, Yi Huang, Hyun Song Shin, and Pablo Zbinden, "BigTech and the changing structure of financial intermediation," 2019.
- FSB, "FinTech and market structure in financial services: Market developments and potential financial stability implications," *Basel: Financial Stability Board*, 2019.
- Fuster, Andreas, Matthew Plosser, Philipp Schnabl, and James Vickery, "The role of technology in mortgage lending," *The Review of Financial Studies*, 2019, 32 (5), 1854–1899.
- , Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther, "Predictably unequal? the effects of machine learning on credit markets," *The Effects of Machine Learning on Credit Markets (November 6, 2018)*, 2018.



- Gambacorta, Leonardo, Yiping Huang, Han Qiu, and Jingyi Wang, “How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm,” 2019.
- Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, “Neglected risks, financial innovation, and financial fragility,” *Journal of Financial Economics*, 2012, 104 (3), 452–468.
- Goodman, Laurie S and Jun Zhu, “Loss severity on residential mortgages: evidence from Freddie Mac’s newest data,” *The Journal of Fixed Income*, 2015, 25 (2), 48–57.
- Hau, Harald, Yi Huang, Hongzhe Shan, and Zixia Sheng, “FinTech Credit and Entrepreneurial Growth,” Technical Report, Mimeo, University of Geneva 2018.
- , —, —, and —, “How FinTech Enters China’s Credit Market,” in “AEA Papers and Proceedings,” Vol. 109 2019, pp. 60–64.
- Hirtle, Beverly, Anna Kovner, and Matthew C Plosser, “The impact of supervision on bank performance,” *FRB of NY Staff Report*, 2018.
- Hughes, Joseph P, Julapa Jagtiani, and Choon-Geol Moon, “Consumer lending efficiency: Commercial banks versus a fintech lender,” *Departmental Working Papers, Rutgers University, Department of Economics*, 2019.
- Jagtiani, Julapa and Catharine Lemieux, “Fintech lending: Financial inclusion, risk pricing, and alternative information,” 2017.
- Keys, Benjamin J, Amit Seru, and Vikrant Vig, “Lender screening and the role of securitization: evidence from prime and subprime mortgage markets,” *The Review of Financial Studies*, 2012, 25 (7), 2071–2108.
- , Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, “Financial regulation and securitization: Evidence from subprime loans,” *Journal of Monetary Economics*, 2009, 56 (5), 700–720.
- , —, —, and —, “Did securitization lead to lax screening? Evidence from subprime loans,” *The Quarterly Journal of Economics*, 2010, 125 (1), 307–362.
- Koetter, Michael and Felix Noth, “IT use, productivity, and market power in banking,” *Journal of Financial Stability*, 2013, 9 (4), 695–704.
- Liberti, José María and Mitchell A Petersen, “Information: Hard and soft,” *Review of Corporate Finance Studies*, 2018, 8 (1), 1–41.
- Maggio, Marco Di and Vincent W Yao, “FinTech Borrowers: Lax-Screening or Cream-Skimming?,” *Working Paper*, 2018.
- Manaresi, Francesco and Nicola Pierri, “Credit supply and productivity growth,” *IMF Working Paper*, 2019.
- McElheran, Kristina and Chris Forman, “Firm Organization in the Digital Age: IT Use and Vertical Transactions in US Manufacturing,” *Available at SSRN 3396116*, 2019.
- Meiselman, Ben S, Stefan Nagel, and Amiyatosh Purnanandam, “Judging Banks’ Risk by the Profits They Report,” *Working Paper*, 2018.
- Mian, Atif and Amir Sufi, “The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis,” *The Quarterly Journal of Economics*, 2009, 124 (4), 1449–1496.

- and — , “House prices, home equity-based borrowing, and the US household leverage crisis,” *American Economic Review*, 2011, 101 (5), 2132–56.
- and — , “What explains the 2007–2009 drop in employment?,” *Econometrica*, 2014, 82 (6), 2197–2223.
- Moretti, Enrico, “Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data,” *Journal of econometrics*, 2004, 121 (1-2), 175–212.
- Nasiripour, Shahien, “Fintech Lenders Tighten Standards as They Become More Like Banks,” *Bloomberg*, Dec 16, 2019.
- Navaretti, Giorgio Barba, Giacomo Calzolari, José Manuel Mansilla-Fernandez, and Alberto F Pozzolo, “Fintech and Banking. Friends or Foes?,” *Friends or Foes*, 2018.
- Oster, Emily, “Unobservable selection and coefficient stability: Theory and evidence,” *Journal of Business & Economic Statistics*, 2019, 37 (2), 187–204.
- Pagan, Adrian, “Econometric issues in the analysis of regressions with generated regressors,” *International Economic Review*, 1984, pp. 221–247.
- Peek, Joe and Eric S Rosengren, “Collateral damage: Effects of the Japanese bank crisis on real activity in the United States,” *American Economic Review*, 2000, 90 (1), 30–45.
- Petersen, Mitchell A and Raghuram G Rajan, “The benefits of lending relationships: Evidence from small business data,” *The Journal of Finance*, 1994, 49 (1), 3–37.
- and — , “Does distance still matter? The information revolution in small business lending,” *The Journal of Finance*, 2002, 57 (6), 2533–2570.
- Philippon, Thomas, “On Fintech and Financial Inclusion,” *NBER Working Paper*, 2019.
- Rajan, Raghuram G, “Has finance made the world riskier?,” *European Financial Management*, 2006, 12 (4), 499–533.
- Rajan, Uday, Amit Seru, and Vikrant Vig, “Statistical default models and incentives,” *American Economic Review*, 2010, 100 (2), 506–10.
- , — , and — , “The failure of models that predict failure: Distance, incentives, and defaults,” *Journal of Financial Economics*, 2015, 115 (2), 237–260.
- Reinhart, Carmen M and Kenneth S Rogoff, *This time is different: Eight centuries of financial folly*, princeton university press, 2009.
- Rossi, Alberto G, “Predicting stock market returns with machine learning,” 2018.
- Sawyer, William E, “The Magnificent Charter: The Origin and Role of the Morrill Land-Grant Colleges and Universities,” 1981.
- Shapiro, Jesse M, “Smart cities: quality of life, productivity, and the growth effects of human capital,” *The review of economics and statistics*, 2006, 88 (2), 324–335.
- Stock, James and Motohiro Yogo, “Asymptotic distributions of instrumental variables statistics with many instruments,” *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, 2005, pp. 109–120.

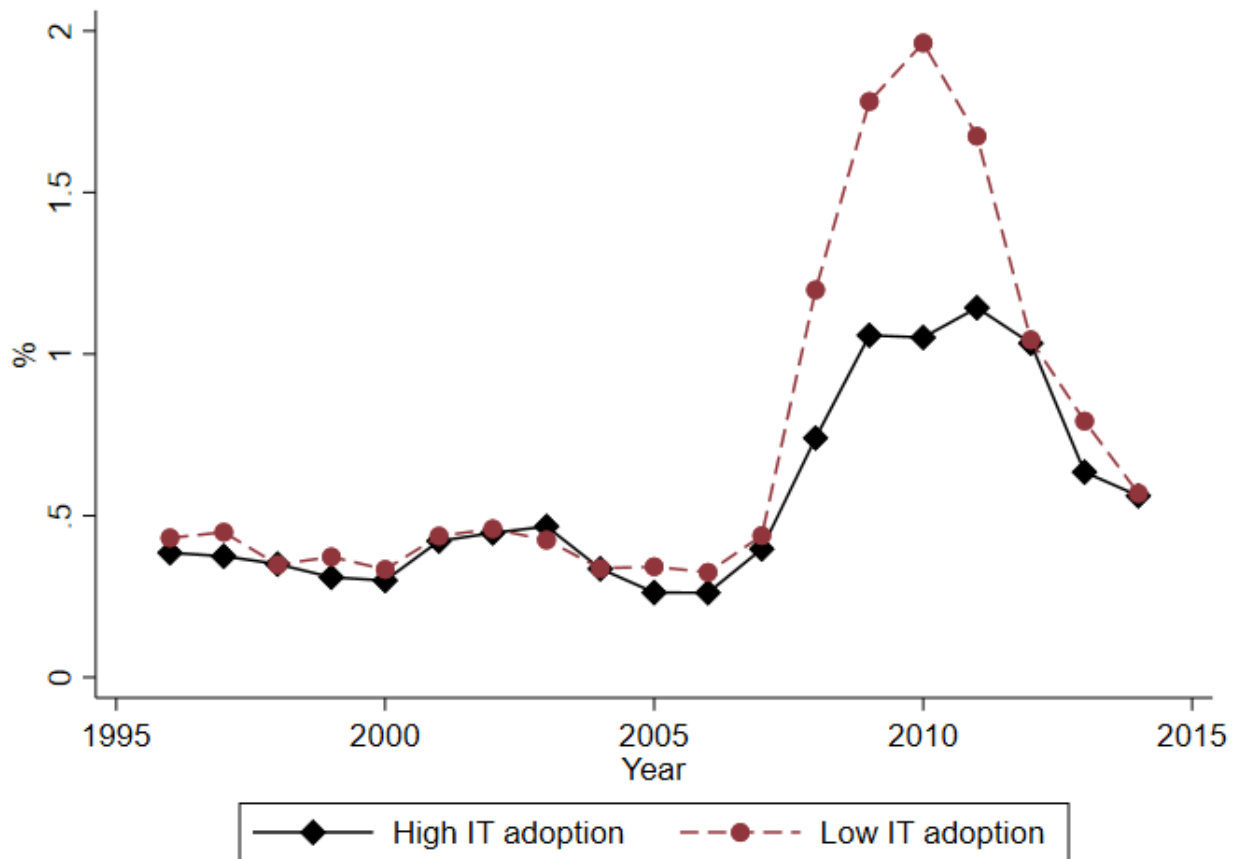
Stulz, René M, “FinTech, BigTech, and the Future of Banks,” *NBER Working Paper*, 2019.

Sullivan, Tara and James Vickery, “A Look at Bank Loan Performance,” *Federal Reserve Bank of New York Liberty Street Economics*, 2013.

Syversen, Chad, “What determines productivity?,” *Journal of Economic literature*, 2011, 49 (2), 326–65.

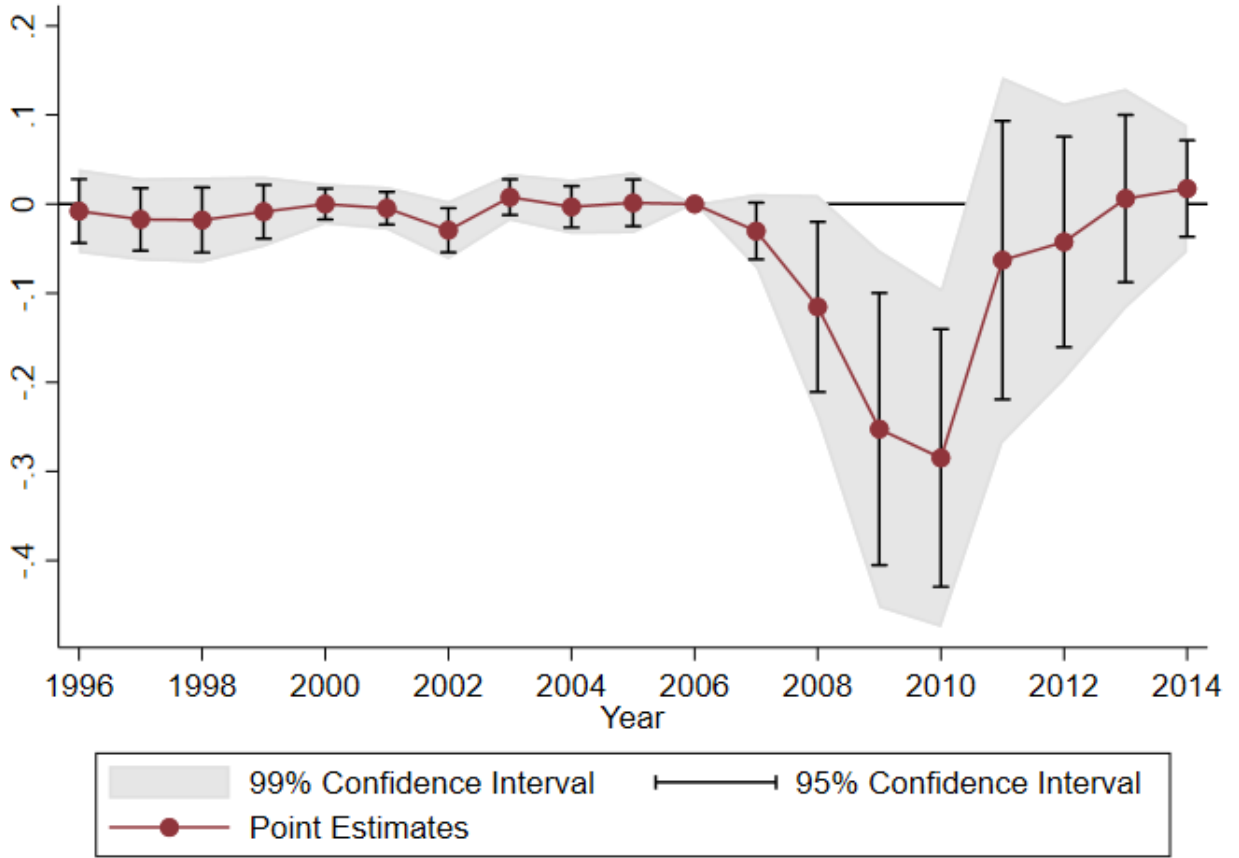
Williams, Roger L, *Origins of federal support for higher education: George W. Atherton and the land-grant college movement*, Penn State Press, 2010.

Figure 1: NPLs over Assets by pre-GFC IT adoption



This Figure plots the median share of NPLs over assets for high- and low-IT adopters. “High IT adoption” is the median share of NPLs over assets for banks with  $IT_b$  above the 75th percentile. “Low-IT adoption” is the median share of NPLs over assets for banks with  $IT_b$  below the 25th percentile. We include only banks for which we have regulatory data for at least 14 years. See [subsection 4.1](#) and [section 3](#) for more details.

Figure 2: Time-varying Effect of IT adoption on NPLs

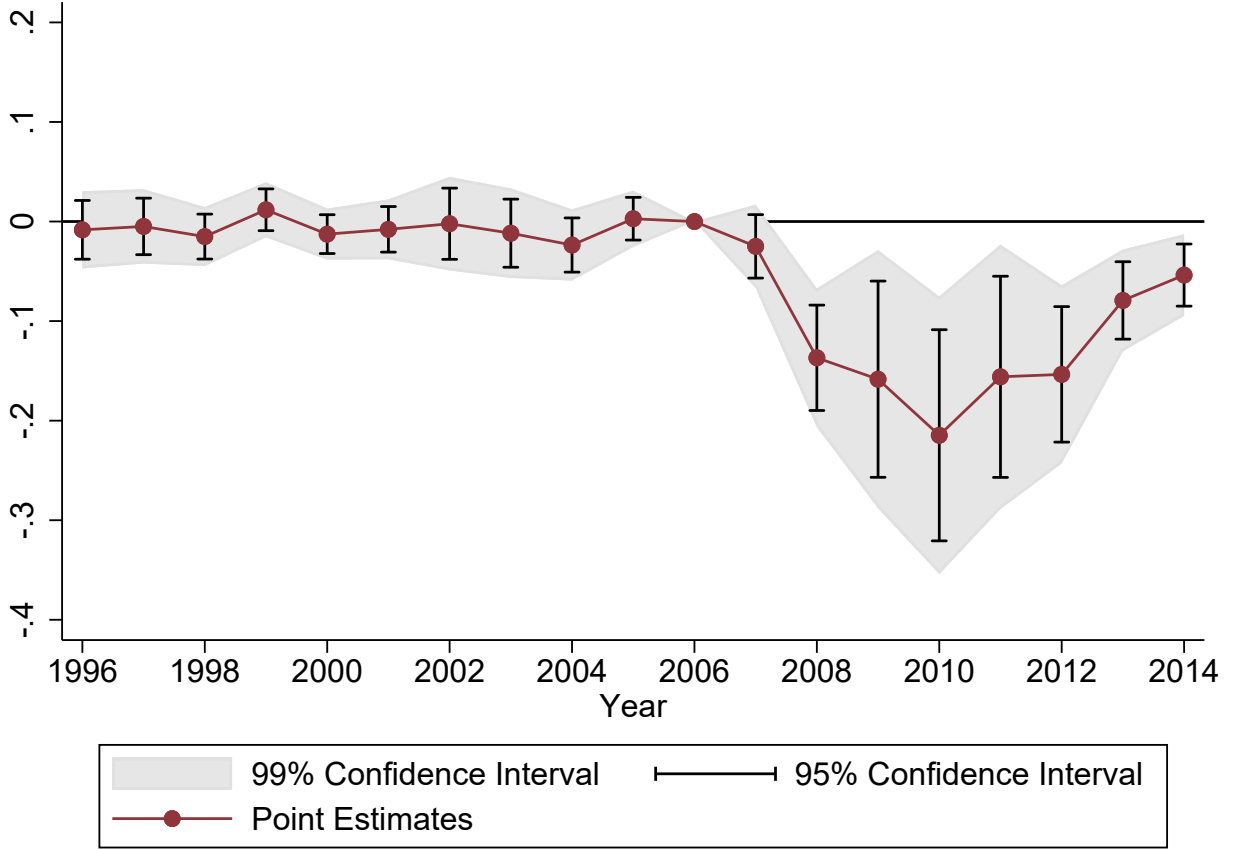


This Figure plots the coefficient and the 95% and 99% confidence intervals of  $\beta_\tau$  from the following estimated equation:

$$NPL_{b,t} = \alpha_b + \delta_t + \sum_{\tau \neq 2006} \beta_\tau IT_b \cdot 1[t = \tau] + \epsilon_{b,t}$$

where  $b$  is a bank (BHC),  $t$  one year between 1996 and 2014,  $\alpha_b$  are bank fixed effects, and  $\delta_t$  are year fixed effects. The dependent variable  $NPL_{b,t}$  is the share of NPLs over assets in  $b$ 's regulatory filing for year  $t$ .  $IT_b$  is the pre-crisis IT adoption of bank  $b$  estimated as described in [section 3](#). The coefficient of 2006 is normalized to zero. Confidence intervals are based on double-clustered standard errors at the bank and year level. See [subsection 4.1](#) and [section 3](#) for more details.

Figure 3: Time-varying Effect of tech-background of executives on NPLs

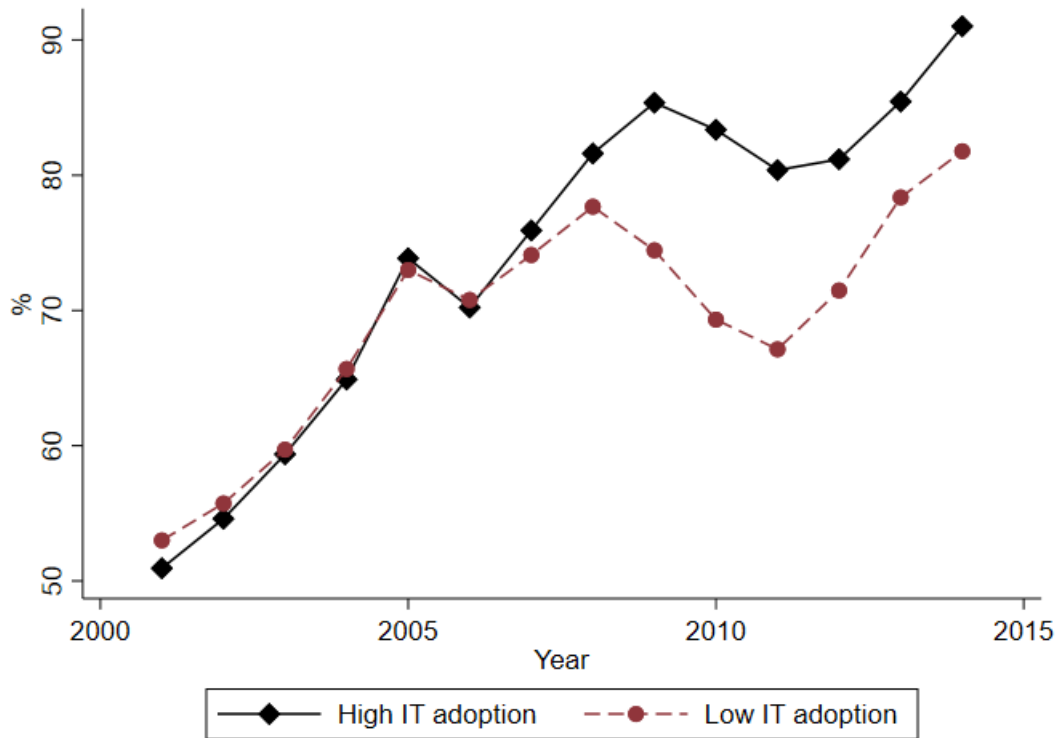


This Figure plots the coefficient and the 95% and 99% confidence intervals of  $\beta_\tau$  from the following estimated equation:

$$NPL_{b,t} = \alpha_b + \delta_t + \sum_{\tau \neq 2006} \beta_\tau ExecIT_b \cdot 1[t = \tau] + \epsilon_{b,t}$$

where  $b$  is a bank (BHC),  $t$  one year between 1996 and 2014,  $\alpha_b$  are bank fixed effects, and  $\delta_t$  are year fixed effects. The dependent variable  $NPL_{b,t}$  is the share of NPLs over assets in  $b$ 's regulatory filing for year  $t$ .  $ExecIT_b$  is the average “tech-orientation” of bank's  $b$  top executives (CEOs, CFOs, and Presidents). The “tech-orientation” of a banks' executives is computed by dividing the total amount of “tech-related” keywords over the total amount of words in their biographies, see [subsection 5.1](#) and [section 3](#) for more details. The coefficient of 2006 is normalized to zero. Confidence intervals are based on double-clustered standard errors at the bank and year level. See [subsection 4.1](#) and [section 3](#) for more details.

Figure 4: Loans over pre-crisis Assets by pre-GFC IT adoption



This Figure plots the median share of total loans scaled by average pre-crisis (2001-2006) assets for high- and low-IT adopters. “High IT adoption” is the median share of Loan over pre-crisis assets for banks with  $IT_b$  above the 75th percentile. “Low IT adoption” is the median share of Loan over pre-crisis assets for banks with  $IT_b$  below the 25th percentile. We include only banks for which we have regulatory data for at least 14 years. See [section 7](#) and [section 3](#) for more details.



Table 1: Panel Regressions

	Dependent Variable: NPLs during GFC											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
IT adoption	-0.0230 (0.017)		-0.0280 (0.018)									
IT adoption × crisis	-0.160** (0.063)	-0.168** (0.065)	-0.157** (0.066)	-0.170** (0.068)	-0.158** (0.063)	-0.139** (0.057)	-0.131** (0.055)	-0.131** (0.056)	-0.131** (0.056)	-0.131** (0.056)	-0.131** (0.056)	-0.143** (0.063)
Loans × crisis					0.0201*** (0.006)	0.0199*** (0.006)	0.0177*** (0.006)	0.0175*** (0.005)	0.0186*** (0.006)	0.0189*** (0.006)	0.0189*** (0.006)	0.0192*** (0.006)
HP Exposure × crisis						0.0191** (0.009)	0.0172* (0.008)	0.0172* (0.008)	0.0172* (0.008)	0.0171* (0.008)	0.0171* (0.009)	0.0174* (0.009)
Size × crisis							0.128** (0.052)	0.126** (0.050)	0.115** (0.050)	0.123** (0.055)	0.124* (0.062)	0.124* (0.062)
Capital × crisis								-0.000672 (0.005)	0.000698 (0.005)	0.000933 (0.005)	0.000805 (0.005)	0.00113 (0.005)
Wholesale × crisis									0.00831 (0.007)	0.00779 (0.007)	0.00779 (0.007)	0.00762 (0.007)
ROA × crisis										-0.0204 (0.048)	-0.0206 (0.048)	-0.0211 (0.049)
Log Wage × crisis											-0.0108 (0.125)	-0.00388 (0.124)
IT of local competitors × crisis												0.0475 (0.046)
(Within) R-squared	0.112	0.140	0.0111	0.00997	0.0363	0.0553	0.0613	0.0611	0.0621	0.0620	0.0617	0.0623
N	4,608	4,608	4,608	4,608	4,608	4,608	4,608	4,608	4,608	4,608	4,608	4,608
Bank FE	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Results of estimating the following equation:

$$NPL_{b,t} = \alpha_b + \delta_t + \beta IT_b \cdot crisis + (X_b \cdot crisis_t)' \gamma + \epsilon_{b,t}$$

where  $b$  is a bank (BHC),  $t$  one year between 2001 and 2014,  $crisis_t$  a dummy variable indicating years 2007 to 2010,  $\alpha_b$  are bank fixed effects, and  $\delta_t$  are year fixed effects. The dependent variable  $NPL_{b,t}$  is the share of NPLs over assets in  $b$ 's regulatory filing for year  $t$ .  $IT_b$  is the pre-crisis IT adoption of bank  $b$  estimated as described in [section 3](#). The bank-level set of controls  $X_b$  includes the pre-crisis (2001-2006) average of: the loans to assets ratio, the capital to assets ratio, the wholesale funding ratio, ROA, the (log of) average wages in thousands of USD, and the (log of) assets size in thousands of USD.  $X_b$  also includes the average IT adoption of local competitors and a measure of exposure to the house price shocks (HP Exposure) based on the combination of the observed drop in prices (peak to trough) in each county and the location of banks' branches. Columns (1) and (3) exclude bank fixed effect, while column (1) and (2) exclude year fixed effects. See [subsection 4.1](#) and [section 3](#) for more details. Sample size is kept constant by dropping observations with missing values for any variable. Standard errors (in parentheses) are double-clustered on bank and year level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 2: Cross-Sectional Regressions

Dependent Variable:	NPLs during GFC	Loans pre-GFC	HP Exposure	Size pre-GFC	Capital pre-GFC	Wholesale pre-GFC	ROA pre-GFC	Log Wage pre-GFC	IT of local competitors	NPLs during GFC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IT adoption	-0.183*** (0.061)	-0.648 (0.700)	-0.896 (0.664)	-0.0931 (0.057)	-0.195 (0.420)	-0.0459 (0.372)	-0.0282 (0.049)	-0.0227 (0.018)	0.275*** (0.083)	-0.157*** (0.058)
Loans										0.0305*** (0.004)
HP Exposure										0.0242*** (0.005)
Size										0.0857 (0.077)
Capital										0.00605 (0.007)
Wholesale										0.0251*** (0.008)
ROA										-0.0452 (0.070)
Log Wage										-0.145 (0.210)
IT of local competitors										0.0773 (0.047)
R-squared	0.0262	0.00220	0.00550	0.00712	0.000427	0.0000383	0.00107	0.00414	0.0750	0.243
N	337	337	337	337	337	337	337	337	337	337
Mean	1.54	62.69	15.83	13.9	13.02	15.92	2.55	4.84	0	1.54
Std.Dev.	1.13	13.8	12.06	1.1	9.43	7.41	.86	.35	1	1.13

Results of estimating the following equation:

$$Y_b = \alpha + \beta IT_b + \epsilon_b$$

where  $b$  is a bank (BHC) and  $IT_b$  is the pre-crisis IT adoption of  $b$ , estimated as described in [section 3](#). The dependent variable  $Y_b$  is either the share of NPLs over assets in bank  $b$  regulatory filing (averaged over 2007 to 2010) or one of the variables of the set  $X_b$ , defined as follow.  $X_b$  includes the pre-crisis (2001-2006) average of: the loans to assets ratio, the capital to assets ratio, the wholesale funding ratio, ROA, the (log of) average wages in thousands of USD, and the (log of) assets size in thousands of USD.  $X_b$  also includes the average IT adoption of local competitors and a measure of exposure to the house price shocks (HP Exposure) based on the combination of the observed drop in prices (peak to trough) in each county and the location of banks' branches. In column (10) the dependent variable is the share of NPLs over assets and the set of covariates  $X_b$  are included as controls. See [subsection 4.2](#) and [section 3](#) for more details. Sample size is kept constant by dropping observations with missing values for any variable. Robust standard errors are reported in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 3: NPLs, IT adoption, and Executives' "tech-orientation"

Dependent Variable:	NPLs during GFC (1)	NPLs during GFC (2)	IT adoption (3)
IT adoption	-0.138* (0.076)		
Executives' "tech-orientation"		-0.155*** (0.047)	0.0900* (0.051)
R-squared	0.0141	0.0210	0.00967
N	249	249	249

Results of estimating the following equation:

$$Y_b = \alpha + \beta X_b + \epsilon_b$$

where  $b$  is a bank (BHC). The dependent variable  $Y_b$  is either the ratio of NPLs to assets averaged between 2007 and 2010 (columns 1 and 2), or the pre-crisis IT adoption (column 3), estimated as described in [section 3](#). The independent variable  $X_b$  is either the pre-crisis IT adoption (column 1) or the average "tech-orientation" of bank's  $b$  top executives (CEOs, CFOs, and Presidents). The "tech-orientation" of a banks' executives is computed by dividing the total amount of "tech-related" keywords over the total amount of words in their biographies, see [subsection 5.1](#) and [section 3](#) for more details. Sample size is kept constant by dropping observations with missing values for any variable. Robust standard errors are reported in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 4: Executives' "tech-orientation" and Compensation

	NPLs during GFC	NPLs during GFC	IT adoption	IT adoption
	(1)	(2)	(3)	(4)
Executives' "tech-orientation"	-0.173*** (0.062)	-0.168*** (0.062)	0.104* (0.057)	0.104* (0.057)
Log Compensation		-0.0375 (0.060)		-0.00208 (0.053)
R-squared	0.0226	0.0244	0.0136	0.0136
N	237	237	149	149

Results of estimating the following equation:

$$Y_b = \alpha + \beta \text{ExecutiveIT}_b + \gamma \text{Comp}_b + \epsilon_b$$

where  $b$  is a bank (BHC). The dependent variable  $Y_b$  is either the ratio of NPLs to assets averaged between 2007 and 2010 (columns 1 and 2) or the pre-crisis IT adoption (columns 3 and 4), estimated as described in [section 3](#). The independent variable  $\text{ExecutiveIT}_b$  is the average "tech-orientation" of bank's  $b$  top executives (CEOs, CFOs, and Presidents). The "tech-orientation" of a banks' executives is computed by dividing the total amount of "tech-related" keywords over the total amount of words in their biographies, see [subsection 5.1](#) and [section 3](#) for more details. The controls  $\text{Comp}_b$  is the log of the average total compensation earned by top executives in 2007.  $\text{Comp}_b$  is excluded in columns (1) and (3). Sample size is kept constant by dropping observations with missing values for any variable. Robust standard errors are reported in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 5: Land-Grant Colleges Instruments

Instrument(s)	Dependent Variable: NPLs during GFC					
	OLS	IV	IV	IV	IV	IV
		5 closest	All	LASSO	LASSO	LASSO
	(1)	(2)	(3)	(4)	(5)	(6)
IT adoption	-0.183*** (0.055)	-0.949* (0.489)	-0.301** (0.127)	-0.837** (0.350)	-0.541** (0.230)	-0.546** (0.241)
N	337	337	337	337	337	337
P-value: IV = OLS		0.117	0.353	0.0619*	0.118	0.132
Controls	No	No	No	No	Yes	Yes
State FEs	No	No	No	No	No	Yes
F-stat of First Stage		2.192	9.948	14.06	12.42	10.76
Cragg-Donald Wald F		1.258	1.081	22.959	17.509	5.817
Stock and Yogo (2005)'s value		10.83	10.99	16.38	16.38	16.38

Results of estimating the following 2SLS equation:

$$IT_b = \delta + F(D_{b(c),j}, \gamma) + \eta_b$$

$$NPL_b = \alpha + \beta \cdot IT_b + \epsilon_b$$

where  $b$  is a bank (BHC).  $NPL_b$  is the ratio of NPLs to assets averaged between 2007 and 2010.  $IT_b$  is pre-crisis IT adoption, estimated as described in [section 3](#).  $F(D_{b(c),j}, \gamma)$  is a function of the set of instruments  $D_{b(c),j}$ , know up to some parameter  $\gamma$ . In column (2) the instruments are the distances of the 5 closest land-grant colleges, in column (3) the instruments are the distances to all land-grant colleges, in column (4)-(6) the instrument is the LASSO prediction of  $IT_b$ . Controls include bank headquarter county-level and bank-level controls. F-stat of First Stage refers to the null hypothesis that all instruments are jointly zero. [Stock and Yogo \(2005\)](#)'s value are the values corresponding to the 10% maximal IV relative bias and should be compared to the Cragg-Donald Wald F-statistics. Controls include both bank-level and county-level variables. Bank level controls are the pre-crisis (2001-2006) average of: the loans to assets ratio, the capital to assets ratio, the wholesale funding ratio, ROA, the (log of) average wages in thousands of USD, and the (log of) assets size in thousands of USD. County-level controls are: education (share of adults with bachelor degree or higher), (log) median household income, log population, and log population density of the county where the BHC is headquartered. Headquarters' location are taken from regulatory filings in 1995 or earliest year available. See [subsection 5.2](#) for details. Standard errors (in parentheses) are clustered at the state level, except in column (3) as the number of clusters would be insufficient to calculate robust covariance matrix singleton dummy variable. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 6: Loan-Level Regressions

Dependent Variable:	Delinquency during GFC						
	Share of months with past due>90 days				Ever past due>90 days		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IT adoption	-0.471** (0.191)	-0.459** (0.169)	-0.348** (0.145)	-0.323** (0.118)		-0.106** (0.041)	-0.0377** (0.016)
FICO score				-2.578*** (0.284)	-1.125*** (0.181)	-1.042*** (0.088)	-0.258*** (0.036)
DTI				0.565*** (0.052)	0.248*** (0.022)	0.246*** (0.019)	0.0994*** (0.012)
LTV				1.075*** (0.129)	0.543*** (0.056)	0.541*** (0.058)	0.185*** (0.006)
IT adoption × Low FICO					-0.198*** (0.064)		
IT adoption × High FICO					-0.00732 (0.029)		
Estimation Method	OLS	OLS	OLS	OLS	OLS	OLS	Probit
Org. Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Postal Code FE	No	No	Yes	Yes	Yes	Yes	No
R-squared	0.000942	0.00643	0.0210	0.0659	0.0184	0.0185	
N	3,451,671	3,451,671	3,451,671	3,451,671	3,451,671	3,451,671	3,451,671
Mean	3.44	3.44	3.44	3.44	3.44	1.5	.015
Std.Dev. of dept. var.	14.32	14.32	14.32	14.32	14.32	12.15	.1215

Results of estimating the following equation:

$$Delinquent_l = \alpha_{z(l)} + \delta_{o(l)} + \beta IT_{b(l)} + X_l' \gamma + \eta_l$$

where  $l$  is a mortgage held by Freddie Mac and originated before 2007,  $\alpha_{z(l)}$  are fixed effects for the 3-digit postal code of the underlying property, and  $\delta_{o(l)}$  are fixed effects for the year of origination.  $IT_{b(l)}$  is the pre-crisis IT adoption of the bank which sold the mortgage to Freddie Mac, estimated as described in [section 3](#) (22 banks). The dependent variable  $Delinquent_l$  is either the share of months between 2007 and 2010 during which the loan was in a delinquent status (> 90 days past due) or a dummy variable indicating whether loan was ever delinquent between 2007 and 2010. Both are multiplied by 100, except than in column (7). The vector of controls  $X_l$  includes the FICO score, the debt servicing to Income (DTI), and the Loan-to-Value (LTV) ratios at origination. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Column (1) excludes all fixed effects and controls, column (2) excludes  $X_l$  and  $\alpha_{z(l)}$ , while column (3) excludes  $X_l$ . Column (5) interacts banks' IT adoption with a dummy variable indicating whether the borrower has a FICO score above or below the median. We also include this dummy in the regression. See [section 6](#) and [section 3](#) for more details. Sample size is kept constant by dropping observations with missing values for any variable. Standard errors (in parentheses) are cluster at the bank-level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 7: Lending Regressions

Dependent Variable :	Loan Growth (crisis)			
	(1)	(2)	(3)	(4)
NPLs during the GFC	-0.926*** (0.159)	-1.030*** (0.187)		
IT adoption			0.378** (0.182)	0.331* (0.196)
R-squared	0.0127	0.0928	0.0961	0.175
N	343	336	343	336
Controls	No	Yes	No	Yes

Results of estimating the following equation:

$$\overline{\Delta Loans_b}^{GFC} = \alpha + \beta X_b + \epsilon_b$$

where  $b$  is a bank (BHC). The dependent variable  $\overline{\Delta Loans_b}^{GFC}$  is the loan growth over assets in bank  $b$  regulatory filing (averaged over 2007 to 2010).  $X_b$  is either  $IT_b$  is the pre-crisis IT adoption of  $b$ , estimated as described in [section 3](#) or the share of NPLs over assets in bank  $b$  regulatory filing (averaged over 2007 to 2010). See [subsection 5.1](#) and [section 3](#) for more details. Robust standard errors are reported in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## Internet Appendix

### A1 Land-Grant Colleges Characteristics and Headquarter Selection

In this section we study the characteristics of land-grant colleges' students and how the location of land-grant colleges impact banks' headquarters location. These results are complementary to the analysis in [subsection 5.2](#). The IPEDS survey provides data on the major and SAT scores for students enrolled in more than 1,400 higher education institutions in the US during fall 2018. We then estimate the following regressions:

$$Share_{u,M} = \alpha_M + \beta_M LandGrant_u + \epsilon_{u,M}$$

and

$$SAT\_Score_u = \alpha + \beta LandGrant_u + \epsilon_u$$

where  $Share_{u,M}$  is the share of students in each of 6 fields of study  $M$  we have information on (we drop dentistry school) in institution  $u$ ,  $SAT\_Score_u$  is the 25th or 75th percentile of SAT score in reading or math of the enrolled students, and  $LandGrant_u$  is a dummy flagging whether  $u$  is a land-grant college or university. Results are reported in [Table A4](#) and [Table A5](#). Land-grant institutions have a much larger share of students enrolled in engineering and slightly more students in other scientific disciplines, such as biology and physics. Conversely, they have much fewer students in business and management science and also less students in education. Moreover, students at land-grant colleges have significantly higher math scores (whether we look at the 25th or 75th percentile) but similar reading scores. These results indicate that land-grant colleges are mainly technical schools. We repeat the analysis using data from fall 1996 (since we take bank headquarter location in 1995 when possible) and find very similar results.

We then move to analyse whether the distance from land-grant colleges predicts headquarters' location. We estimate the following linear probability model:

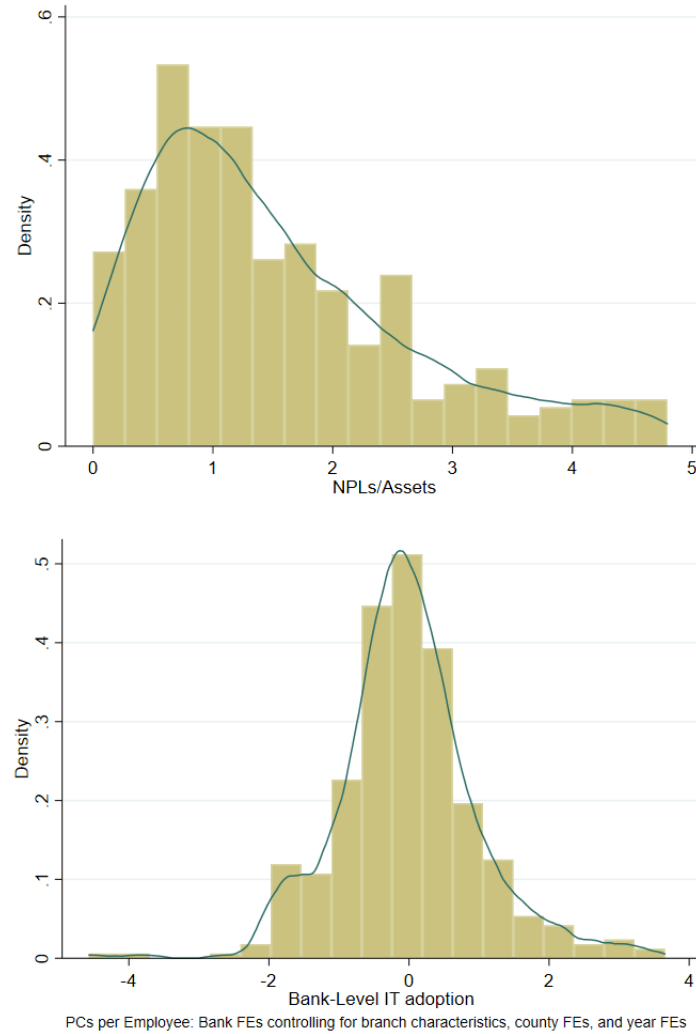
$$BankHQ_c = \alpha + \beta Distance\_Landgrant_c + \gamma X_c + \epsilon_c$$

where  $BankHQ_c$  is a dummy variable indicating whether the county  $c$  host the headquarter of one of the 337 BHC of our main sample, and  $X_c$  is set of controls including pre-GFC education, income, and state fixed effects.  $Distance\_Landgrant_c$  is one of four measure of distance (in log of miles plus one) of county  $c$  to land-grant colleges: closest college, median across all colleges, mean across all colleges, and the LASSO instrument used in [subsection 5.2](#). Results are presented in [Table A6](#). No measure has statistically significant predictive power. The results are robust to estimating a probit model rather than a linear probability model.



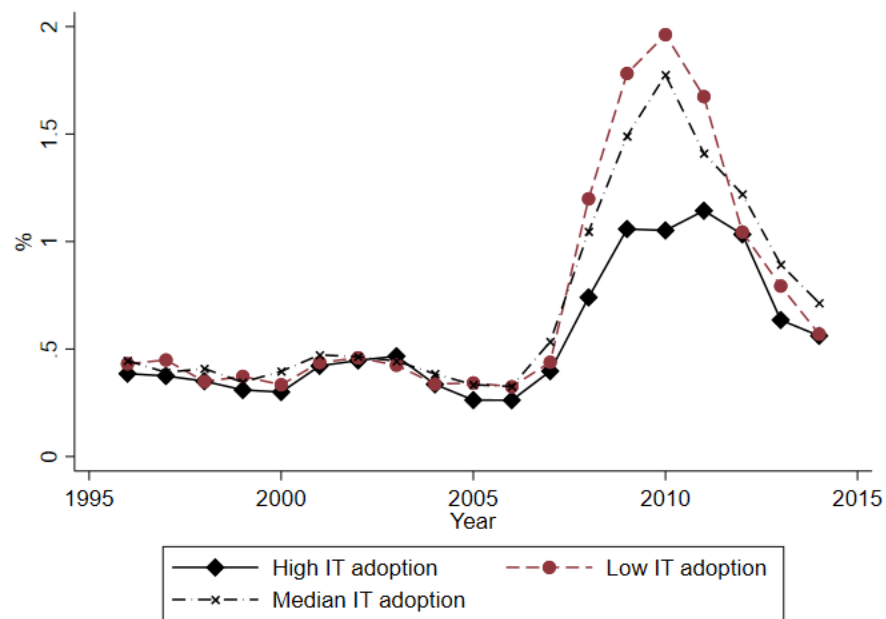
## Appendix: Figures and Tables

Figure A1: Cross-sectional distribution of NPLs over Assets (crisis) and IT adoption (pre-crisis)



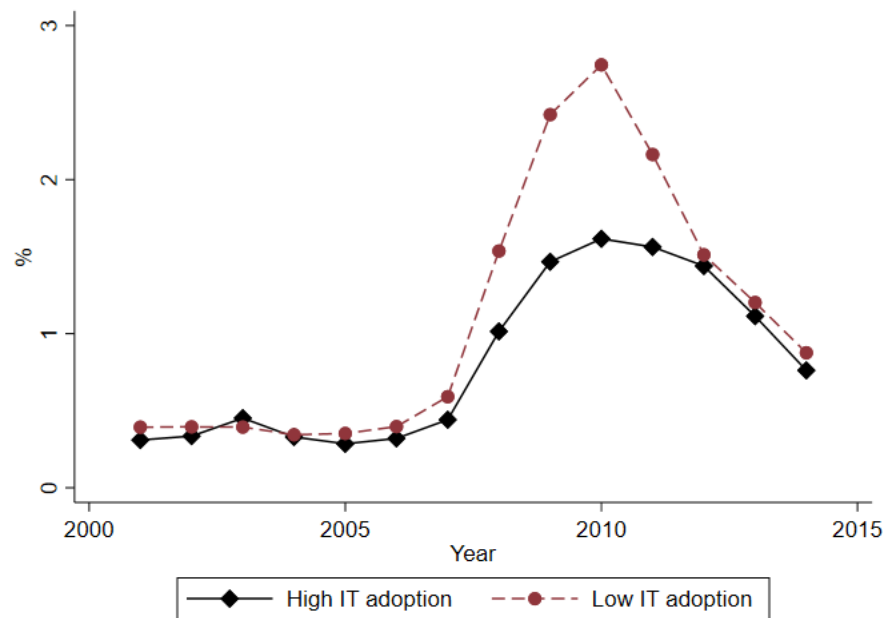
This Figure plots the cross-sectional distribution of the ratio of NPLs to assets averaged between 2007 and 2010 (top panel) and of the pre-crisis IT adoption  $IT_b$ . See [section 3](#) for more details.

Figure A2: NPLs over Assets by pre-GFC IT adoption



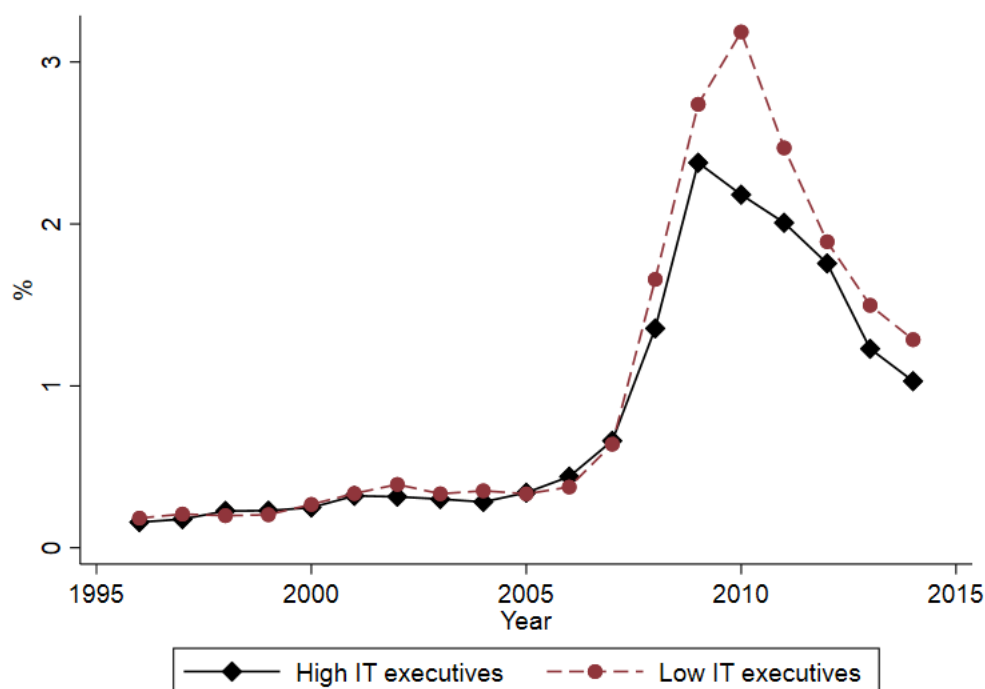
This Figure plots the median share of NPLs over assets for high, medium, and low-IT adopters. “High IT adoption” is the median share of NPLs over assets for banks with  $IT_b$  above the 75th percentile. “Low IT adoption” is the median share of NPLs over assets for banks with  $IT_b$  below the 25th percentile. “Median IT adoption” is the median share of NPLs over assets for banks with  $IT_b$  between the 25th percentile and the 75th percentile. We include only banks for which we have regulatory data for at least 14 years. See [subsection 4.1](#) and [section 3](#) for more details.

Figure A3: NPLs over pre-GFC Assets by pre-GFC IT adoption



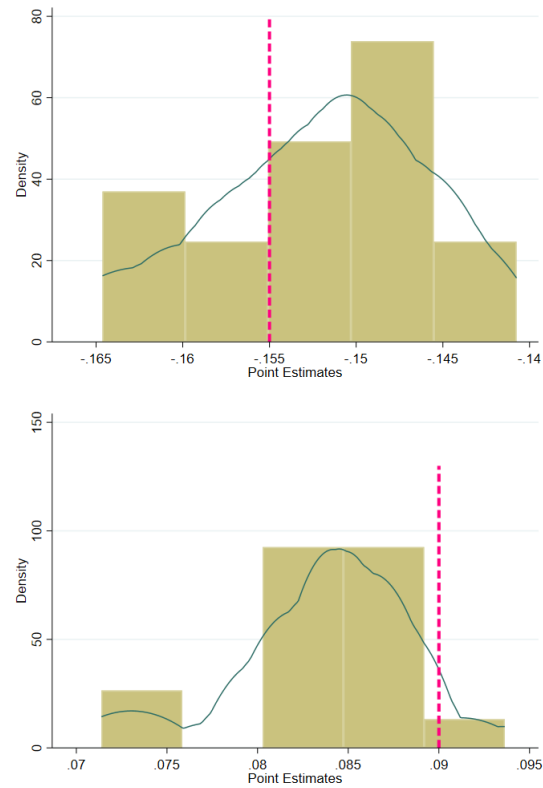
This Figure plots the median share of NPLs scaled by average pre-crisis (2001-2006) assets for high- and low-IT adopters. “High IT adoption” is the median share of NPLs over pre-crisis assets for banks with  $IT_b$  above the 75th percentile. “Low-IT adoption” is the median share of NPLs over pre-crisis assets for banks with  $IT_b$  below the 25th percentile. We include only banks for which we have regulatory data for at least 14 years. See [subsection 4.1](#) and [section 3](#) for more details.

Figure A4: NPLs over pre-GFC Assets by bank top executives' technology orientation



This Figure plots the median share of NPLs scaled by average pre-crisis (2001-2006) assets for banks with high and low executives' "tech-orientation". "High-IT executive" is the median share of NPLs over pre-crisis assets for banks with executives "tech-orientation" above the 75th percentile. "Low-IT executive" is the median share of NPLs over pre-crisis assets for banks with executives "tech-orientation" at or below the 25th percentile. We include only banks for which we have regulatory data for at least 14 years. The "tech-orientation" of banks' executives is computed by dividing the total amount of "tech-related" keywords over the total amount of words in their biographies. We then compute a bank-level measure by averaging over the top executives (CEOs, CFOs, COOs, and Presidents) hired before 2007. See [subsection 5.1](#) and [section 3](#) for more details.

Figure A5: Robustness of the Executives' results to changes in the keywords list



This Figure plots the coefficient of columns (2)-(3) of [Table 3](#) for different measures of bank top executives' technology orientation. For each word used in defining the technology orientation of executives, we create a new measure in which we leave out this particular word and build the measure based on all remaining words. The dashed line reflect the estimates of columns (2) and (3) of [Table 3](#). See [subsection 5.1](#) and [section 3](#) for more details.

Table A1: Robustness of Main Panel Regression

Dependent Variable: NPLs								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IT adoption $\times$ crisis	-0.165** (0.068)	-0.243* (0.120)	-0.158** (0.069)	-0.161** (0.063)	-0.242** (0.095)	-0.214** (0.080)	-0.380* (0.183)	-0.165*** (0.051)
Exercise	Baseline	PCs per Emp	HW IT	HW NPLs	Loans	Broad def.	As of 2006	Bank Clustering
R-squared	0.00944	0.00376	0.00794	0.0108	0.00867	0.00993	0.00530	0.00944
N	4692	5035	4692	4692	4692	4692	4655	4692
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Results of estimating the following equation:

$$NPL_{b,t} = \alpha_b + \delta_t + \beta IT_b \cdot crisis + \epsilon_{b,t}$$

where  $b$  is a bank (BHC),  $t$  one year between 2001 and 2014,  $crisis_t$  a dummy variable indicating years 2007 to 2010,  $\alpha_b$  are bank fixed effects, and  $\delta_t$  are year fixed effects. The dependent variable  $NPL_{b,t}$  is the share of NPLs over assets in  $b$ 's regulatory filing for year  $t$ .  $IT_b$  is the pre-crisis IT adoption of bank  $b$  estimated as described in [section 3](#). In column (2) the IT adoption is measured by the average PCs per employee in bank  $b$ 's branches. In column (3) the IT adoption measure is winsorized after estimation at 5 percent on each side. In column (4) the NPLs are winsorized at 5 percent on each side. In column (5) NPLs are normalized by the amount of loans rather than assets. In column (6) NPLs are defined according to a broader definition, which includes loans with shorter delinquency period. In column (7) we normalized NPLs by the average amount of assets that each bank had in the pre-crisis period (2001 to 2006) rather than contemporaneous assets. In column (8) we cluster standard errors only on the bank-level. Standard errors (in parentheses) are double-clustered on bank and year level for columns (1)-(7). See [subsection 4.1](#) and [section 3](#) for more details. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A2: Two-stages bootstrapped standard errors: robustness to the “generated regressor” problem

Dependent Variable: NPLs during GFC					
	(1)	(2)	(3)	(4)	(5)
IT adoption×crisis	-0.160	-0.170	-0.143		
IT adoption				-0.183	-0.157
baseline t-stat	-2.555	-2.491	-2.283	-3.009	-2.705
bootstrap t-stat	-2.819	-3.090	-2.732	-2.804	-2.601
Specification	Panel	Panel	Panel	Cross-Section	
	No controls	No controls	All controls	No controls	All controls
	No FEs	with FEs	with FEs		

Results of estimating equations:

$$NPL_{b,t} = \alpha_b + \delta_t + \beta IT_b \cdot crisis + (X_b \cdot crisis_t)' \gamma + \epsilon_{b,t}$$

and

$$NPL_b = \alpha + \beta IT_b + \epsilon_b$$

Columns (1)-(3) report the same results as Table 1, columns (1), (4), and (12); baseline t-statistics are based on double clustered (at bank and year level) standard errors. Columns (4) and (5) report the same results as Table 2, columns (1) and (10); baseline t-statistics are based on robust standard error. Bootstrap t-statistics are based on 500 simulations. With each random sample, we first re-estimate the first stage (Equation 1) to obtain a new estimate of bank-level IT adoption, and then we estimate the equation of interest. Standard errors are then calculated as the standard deviation of the bootstrap coefficients of interests.

Table A3: Robustness of Executives' "tech-orientation" to the inclusion of Non-Base Compensation

	Dependent Variable: NPLs during GFC			
	(1)	(2)	(3)	(4)
Executives' "tech-orientation"	-0.477*** (0.151)	-0.475*** (0.149)	0.212* (0.114)	0.213* (0.115)
Non-Base Compensation		-0.281 (0.520)		-0.191 (0.665)
R-squared	0.0844	0.0875	0.0241	0.0262
N	80	80	79	79

Results of estimating the following equation:

$$Y_b = \alpha + \beta ExecutiveIT_b + \gamma NonBaseComp_b + \epsilon_b$$

where  $b$  is a bank (BHC). The dependent variable  $Y_b$  is either the ratio of NPLs to assets averaged between 2007 and 2010 (columns 1 and 2) or the pre-crisis IT adoption (columns 3 and 4), estimated as described in [section 3](#). The independent variable  $ExecutiveIT_b$  is the average "tech-orientation" of bank's  $b$  top executives (CEOs, CFOs, COOs, and Presidents). The "tech-orientation" of a banks' executives is computed by dividing the total amount of "tech-related" keywords over the total amount of words in their biographies, see [subsection 5.1](#) and [section 3](#) for more details. The controls  $NonBaseComp_b$  is the average share of non-base compensation over total compensation for top executives in 2007.  $NonBaseComp_b$  is excluded in columns (1) and (3). Sample size is kept constant by dropping observations with missing values for any variable. Robust standard errors are reported in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$



Table A4: Enrollment by Major

	Share of Enrollment by Major					
	Biology	Business	Education	Engineering	Medicine	Physics
	(1)	(2)	(3)	(4)	(5)	(6)
Landgrant	0.0167*	-0.124***	-0.0892***	0.191***	-0.00226	0.00730**
	(0.009)	(0.012)	(0.013)	(0.017)	(0.002)	(0.003)
R-squared	0.000726	0.0154	0.0100	0.0660	0.000103	0.000720
N	1,468	1,468	1,468	1,468	1,468	1,468

Results of estimating the following equation:

$$Share_{u,M} = \alpha_M + \beta_M LandGrant_u + \epsilon_{u,M}$$

where  $u$  is a higher-education institution, and  $M$  is a major of study. The dependent variable  $Share_{u,M}$  is the ratio of enrollment in major  $M$  to enrollment in all degrees in Fall 2018. The independent variable  $LandGrant_u$  is a dummy variable that takes the value one if the institution is a land-grant college and zero otherwise. Robust standard errors are reported in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A5: SAT Score

	SAT Score Reading		SAT Score Math	
	25th percentile	75th percentile score	25th percentile	75th percentile
	(1)	(2)	(3)	(4)
Landgrant	10.19	10.31	21.61**	25.35**
	(8.652)	(9.922)	(10.922)	(11.650)
R-squared	0.00116	0.00112	0.00461	0.00644
N	1,144	1,144	1,144	1,144

Results of estimating the following equation:

$$SAT\_Score_u = \alpha + \beta LandGrant_u + \epsilon_u$$

where  $u$  is a higher-education institution. The dependent variable  $SAT\_Score_u$  is entry SAT score for either math or reading for the 75th or 25th percentile. The independent variable  $LandGrant_u$  is a dummy variable that takes the value one if the university is a land-grant college and zero otherwise. Robust standard errors are reported in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A6: Bank Headquarter Location

	(1)	(2)	(3)	(4)
Dependent Variable: HQ of Bank				
Average distance from land-grant colleges	-0.0291 (0.066)			
Median distance from land-grant colleges		0.0178 (0.063)		
Distance from closest land-grant college			-0.00691 (0.006)	
LASSO land-grant college IV				0.159 (0.115)
Education	0.326*** (0.075)	0.323*** (0.075)	0.306*** (0.077)	0.328*** (0.075)
Income	0.0627*** (0.005)	0.0632*** (0.005)	0.0625*** (0.005)	0.0623*** (0.005)
R-squared	0.146	0.146	0.147	0.147
N	3144	3144	3144	3144

Results of estimating the following equation:

$$BankHQ_c = \alpha + \beta Distance\_Landgrant_c + \gamma X_c + \epsilon_c$$

where  $c$  is a county. The dependent variable  $BankHQ_c$  is a dummy that equals one if one of the BHC of our main sample has its headquarter in the county and zero otherwise.  $Distance\_Landgrant_c$  is either the average distance to all land-grant colleges (column 1), the median distance to all land-grant colleges (column 2), the distance to the closest land-grant colleges (column 3), or the LASSO land-grant college IV as described in [subsection 5.2](#) and [Table 5](#) in (column 4). County level controls  $X_c$  include the share of people with bachelor degrees, the log average household income, and state fixed effects. Robust standard errors are reported in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$