# Data Breach Announcement Effect on Bank Operations and Performance

Isarin Durongkadej[*], Heng (Emily) Wang[†‡]

August 7, 2024

**Abstract**

A downside of the digital economy is that banks are prone to experiencing data breaches resulting from, for example, cyberattacks, system glitches, and employee negligence. We investigate how a data breach announcement affects bank operations and stock performance. Our findings show that banks increase their lending after a data breach announcement, potentially, to increase bank's income and preserve stock value. We also provide evidence showing that the CEOs of banks with data breach have a stronger financial incentive to maintain stock value than the CEOs of non-breached banks. In terms of deposits, banks have an outflow of insured and brokered deposits after a data breach announcement. Furthermore, we find that deposits transfer from banks with a data breach to nearby banks with no data breach. However, we do not find a systematic long-term impact of the data breach announcement on bank operations. Data breaches negatively affect stock returns in both the short term and long term which could be explained by significantly lower operating cash flow after the announcement.

*JEL classification*: G14, G21

*Keywords*: bank operations and performance, cybersecurity, data breach, bank stability.

[*] Lucas College and Graduate School of Business, Department of Accounting and Finance, San Jose State University, San Jose, CA, 95192; E-mail: isarin.durongkadej@sjsu.edu [Corresponding author]

[†] Martha and Spencer Love School of Business, Department of Finance, Elon University, Elon, NC 27244; Email: ewang3@elon.edu

[‡] We are grateful to Sara Yasar, Samuel Rosen, Chris Clark, Leng Ling, Brent Evans, and seminar participants at 2020 COBRA Conference, 2020 Southern Finance Association Annual Conference, 2021 Financial Management Association Annual Meeting, and 2022 Eastern Finance Association Annual Meeting for their helpful comments and suggestions. We thank Jonah Paul and Donavan Lingerfelt for research assistance. All errors are our own. Isarin Durongkadej received a research support grant from Georgia College and State University.

*"In 2019, Capital One data breach compromised data of over 100 million people. The bank agreed to pay $190 million to settle claims"*

The New York Times on December 23, 2021

# 1. Introduction

In the past few years, data breach incidents have increased sharply. Equifax, one of the three largest credit reporting agencies in the US, announced a data breach in 2017 which affected more than 150 million Americans. Equifax paid $650 million to settle claims from customers and investigations from federal and state governments. In 2019, Capital One revealed a data breach of 140,000 Social Security numbers and 80,000 bank accounts. The breach cost $190 million for Capital One to settle claims. More recent examples include the ransomware attacks on JBS, one of the America's largest beef producers, and Colonial Pipeline, a company that delivers most of the gasoline in the East Coast. These cyberattacks cause firms' extra expenses such as expenses of settlements and inventing new preventative measures. However, an important indirect cost of a data breach is losing the trust of customers. Customers may lose trust in firms with a data breach and decide to switch to another firm that offers similar products or services. If the same situation happens to a bank whose majority of depositors lose confidence and decide to withdraw their money from the data breach bank, will this affect banks' operations such as deposit and loan activities? The increasing frequency and size of data breaches indicates that the problem of data breaches is a matter of continuing concern. Data breaches not only influence banks' operations, but also affect banks' stock prices after a data breach is announced to the public. A decrease in stock price could further decrease people's confidence in banks.

Given a potential impact on the economy through changes in banking activities, we believe that the data breach impact on banks' operation and stock performance is an important question to investigate. On the one hand, we expect that, after a data breach, people lose trust in banks, withdraw their deposits, and sell banks' stocks owing to expected higher expenses and lower revenues. On the other hand, people might trust their banks regardless of data breach risks and decide to take no action. People may believe that banks are able to handle the situation, or that banks will eventually compensate them for any damage resulted from a data breach. Besides, customers have a switching cost that prevents them from conveniently moving their accounts away. For example, banks may tie up customers with other services such as mortgages, payroll, car loans, and investment portfolios. It would be inconvenient for customers to move their accounts away. In this case, there should not be much effect on bank operations or stock performance. Since there are two possible effects from a data breach on bank operations and performance, our paper strives to understand which effect dominates. The results in our paper can provide banks and regulators with more information and implication about the consequences of a data breach.

To address our research questions, we gather information about bank operations from Call Reports, data-breach announcements from Privacy Rights Clearing House[1], and stock data from CRSP. In terms of methodologies, we use difference-in-difference (DiD) to analyze the impact on banks' operations before and after a data breach announcement. Specifically, we examine the effect of a data breach announcement on several key variables of bank operations such as deposits and loans. In addition to evaluating only on the quarter of a data breach announcement, we evaluate the impact on the subsequent quarters and a quarter prior to the data breach announcement. Since there could be a delay between the time that banks announce their data breach and the time that customers receive the information, the impact might appear later in a subsequent period. We also analyze a data breach impact a quarter before the announcement as news about the breach may leak to the public before banks announce their data breach. To reduce a confounding factor, we apply propensity score matching to find a set of control banks. We did an event study by using Cumulative Abnormal Returns (CARs) to find the impact of the data breach on banks' performance

[1] https://privacyrights.org/data-breaches?terms=&f%5B0%5D=years%3A2019

surrounding the date that a data breach was announced to the public. We find that a data breach announcement has significant impacts on banks' operations and stock performance. However, we do not see evidence of a long-term impact on bank operations.

The existing literature shows that banks with data breach tend to be riskier.[2] We find that banks with data breach tend to have higher unused commitment, larger wholesale funding, more non-performing loans (NPLs), lower capital, larger size, and smaller portion of loans from real estate. Our analysis of stock performance shows that CARs are significantly negative within different windows surrounding the data breach announcement date. The data breach announcement affects banks' stock returns not only in the short term but also in the long term. We examine the long-term effect of a data breach announcement and find consistent results that CARs are continuously decreasing on average through the period from the announcement date to the third quarter after the announcement. To understand how data breach announcements affect bank operations, we examine the channel of impact and find that banks with data breach have significantly lower operating cash flow after data breach announcements. Our analysis shows that banks increase their lending after a data breach announcement, potentially, to increase bank's income and preserve stock value. We also provide a supplementary analysis showing that the CEOs of banks with data breach have a stronger financial incentive to maintain stock value than the CEOs of non-breached banks.

In terms of deposits, banks experience a decrease in insured deposits a quarter after a data breach announcement. Higher risk aversion of insured depositors could be the reason. For instance, insured depositors may try to reduce the future risk of their breached information being exploited by withdrawing money from the breached banks and depositing it in a bank with no data breach. We find evidence supporting this argument that depositors withdraw money from a breached bank and deposit it to another bank with no history

[2] For example, banks with more loan commitments have higher risk (Avery and Berger, 1991). If banks rely on wholesale funding more than retail's, they assume larger risk (Shin, 2009; Pérignon, Thesmar, and Vuillemey, 2018)

of data breach. Banks may have multiple breaches. We find that depositors are more sensitive to the first few breaches and, again, when the number of breach reaches double digit. The effect of data breach announcements are more pronounced for breaches from hacking, portable devices, and stationary devices.[3] Conditional on the number of records lost, the effect of data breach announcements on bank deposits are on time deposit accounts two quarters after the announcement quarter.

The contribution of our paper is, first, related to the literature on the liquidity of financial intermediaries as sufficient liquidity is important for banks to maintain the economic role of capital providers (e.g., Diamond and Rajan, 2001; Gatev, Schuermann, and Strahan, 2009; Imbierowicz and Rauch, 2014; Chen, Chen, and Huang, 2021). Direct or indirect costs of data breaches may negatively reduce economic activities. Our study gives implications to banks about preparation for the new digital age on data breaches.

Second, our paper contributes to the literature of data breach impact in finance. Data breaches have different characteristics than other corporate events and are worth exploring in financial research. The existing literature related to cyberattacks mainly focuses on the impact of data breach incidents on stock markets (Lending, Minnick, and Schorno, 2018; Wang, Wang, and Wu, 2022), bond markets (Iyer, Simkins, and Wang, 2020), options markets (Piccotti and Wang, 2022), firm policies (Kamiya, Kang, Kim, Milidonis, and Stulz, 2021), and credit markets (Mikhed and Vogan, 2018). To our knowledge, Lending et al. (2018) is the only article that generally tested the impact of data breach on sales of non-bank firms and deposits of banks. We extended their work in several ways. First, our sample focuses on U.S. banks only as opposed to all U.S. companies. This allows us to delve more into details about how data breach could impact bank operations. For example, we use Call Reports which can break down the effect of data breach on different types of deposits and find that the impact is mostly on the smaller accounts that have the amount lower than $250,000 (insured deposits). Second, beyond the impact on deposits, we examine the impact

[3] Breaches from portable devices are, for example, lost, discarded, or stolen laptop, smartphone, etc. Breaches from stationary devices refer to lost or stolen computer or server not designed for mobility.

of data breach on bank operations such as lending activities, stock returns, deposit spillover, and deposits with multiple breaches. Third, our analyses are more granular in terms of the data frequency. We use quarterly data which allows us to examine the effect of data breach on bank operations dynamically (i.e., a quarter before, at the quarter, 1-4 quarters after the data breach announcement). In addition, we follow Huang and Wang (2021) to conduct a cleaner staggered DiD with different effective dates of notification law in each state.

Third, our study contributes to the literature of emerging operational risks (e.g., Chernobai, Ozdagli, and Wang, 2021; Berger, Curti, Mihov, and Sedunov, 2022). The financial losses of a data breach could be one of the most hard-hitting consequences for a company both directly and indirectly (e.g., Lending et al., 2018; Huang and Wang, 2021; Foerderer and Schuetz, 2022). The financial losses span from bank revenues to capital markets such as stock markets, bond markets, and options markets. In addition, a data breach induces other types of losses such as reputational damage, operational downtime, legal action, and loss of sensitive data. All of them would deteriorate the existing situation and lead to additional financial losses. According to 2020 McAfee's report, the annual worldwide cost associated with cyberattacks was estimated to be $1 trillion which is about one percent of the global GDP. Cyber risk is worth to analyze separately from other operational risks because of its unique characteristics and its growing threat to financial stability (Curti, Gerlach, Kazinnik, Lee, and Mihov, 2006).

Lastly, our paper contributes to the bank financial crisis and contagion (e.g., Allen and Gale, 2000; Brusco and Castiglionesi, 2007; Baur, 2012; Cont and Schaanning, 2019). The scope of the damage from a data breach could go beyond the stakeholders of the banks to the entire economy. If our society can get benefits from moving financial transactions to a digital platform, banks should be proactive to prevent the violation of customer privacy due to a data breach. This is to preserve the bank's confidence and the spillover effect on the financial system. Though, currently, we do not see a bank failing after a data breach, our paper shows that data breach incidents have an impact on bank operations and stock performance. The financial system has progressed rapidly to an online platform. The data breach issues are expected to be more, not less. To keep the data breach problem in check, banks, financial markets, and regulators should be aware of the impact of a data breach that could occur.

The rest of the paper is outlined as follows. Section 2 provides literature of the data breach in financial markets and hypothesis development. Data and variable construction are in Section 3. Section 4 describes methodologies. Section 5 provides results and discussions. Section 6 concludes.

## 2. Literature and Hypothesis Development

The global economy has been evolving into a digital platform. With much higher performance of smartphones and computers today, we can perform financial transactions more conveniently. However, the good and the bad usually come hand in hand. Concerning data breaches, companies have experienced system hacking, computer hardware stealing, and system glitches. In July 2019, Capital One revealed a data breach of 98 million US consumers. Capital One needs to pay the settlement of 190 million dollars.[4] In 2019, the Financial Crimes Enforcement Network (FINCEN) reported more than 12,500 cases of cyberattacks on banks.[5] Equifax was downgraded by Moody's in 2017 because of the cyberattack. *"We are treating this with more significance because it is the first time that cyberattack has been a named factor in an outlook change."* Joe Mielenhausen, a spokesperson for Moody's, told CNBC. [6]

Data breaches can be very costly. Equifax, a credit bureau company, announced in September 2017 that customers' information such as Social Security and driver's license numbers were compromised. Equifax paid around $650 million to settle claims from cus-

[4] https://www.capitalonesettlement.com/en

[5] https://www.fincen.gov/reports/sar-stats

[6] https://www.cnbc.com/2019/05/22/moodys-downgrades-equifax-outlook-to-negative-cites-cybersecurity.html

tomers and investigations from federal and state governments. The incident exposed the personal information of more than 145 million people. After this incident, the stock price of Equifax tumbled, and the CEO was forced to resign.[7] In Europe, policymakers adopted a new law, the General Data Protection Regulation, known as the GDPR. The GDPR allows regulators in each EU country to charge fines of up to 4 percent of revenue for a breach.[8] This regulation scopes the limit of fines that banks have to pay.

## *2.1. Data breach banks' characteristics*

The first step before we examine the effect of data breach announcements on bank operations is to understand the characteristics of banks with data breach relative to banks without data breach. Lending et al. (2018) show that breached firms are larger in size with lower stock return standard deviation. Huang and Wang (2021) find that large firms with high return on assets are likely to be breached. However, our samples are different from the samples in the previous studies. We focus solely on banking samples while Lending et al. (2018) include non-banking samples and Huang and Wang (2021) examine non-financial firms only. Therefore, the characteristics of banks with data breach could be different by analyzing banks only. The cost and benefit trade-off may result in more breaches for large and more profitable firms for the banking sample as well. In other words, larger and more profitable banks may have more valuable data that increase the reward if a breach is successful. The characteristics of banks with data breach are further discussed in Section 3 and Section 5.1.

## *2.2. Effects on bank loans*

Bank managers and shareholders may have a conflicting goal (Allen and Saunders, 1992). Bank managers may focus on a short-term performance, but shareholders focus more on a long-term goal. In this context, bank managers may try to improve bank's profits after a data

[7] https://www.nytimes.com/2019/07/19/business/equifax-data-breach-settlement.html?searchResultPosition=1

[8] https://www.nytimes.com/2019/07/08/business/british-airways-data-breach-fine.html?searchResultPosition=9

breach announcement to compensate for a potential drop in bank stock values. One method to improve financial performance is to increase lending. Our conjecture can be supported by the literature on bank window dressing. A study by Bank for International Settlements by Garcia, Lewrick, and Sečnik (2021) shows that Banks in Europe suppress their year-end balance sheet to avoid being designated as Global Systemically Important Banks (G-SIBs), because G-SIBs will be subject to more scrutiny and higher capital requirement. Ho, Huang, Lin, and Yen (2016) find that over confident CEOs tend to lower lending standards and raise leverage before a financial crisis. Banks with over confident CEOs may increase lending after a data breach announcement if banks do not experience a severe liquidity outflow after a data breach announcement. On the other hand, we may not find any effect on loan activities after a data breach announcement, since loan transactions have high switching costs. Sharpe (1990) finds that borrowers and banks have accumulated relationships and reduced the monitoring costs. The banks that have long-term relationships with borrowers know their customers better and provide an appropriate lending rate matching the customer's risk. If another bank offers loans to the borrowers whom the bank has never had a relationship with, the borrowers might be charged a higher rate. The reasons could be that the bank has not developed enough relationships with the new customers to fully understand customers' risk profiles and the bank compensates for their risk by charging a higher lending rate. Therefore, our first hypothesis tests whether data breach announcements have an impact on banks' loan activities.

## *2.3. Effects on bank deposits*

The consequences after a data breach in the banking industry could go far beyond the settlement that banks need to pay to their customers or federal regulators. Data breaches may reduce the deposit level of banks due to depositors losing confidence in bank security. Bank liquidity problems may arise because of the lower level of confidence (Diamond and Dybvig, 1983) or the level of uncertainty (Arifovic and Jiang, 2019). When there are more uncertainties, depositors may withdraw money from banks with a data breach and deposit it into another bank without a data breach; therefore, the deposit outflow after a data breach announcement may occur.[9]

Many articles have examined how financial crises affect banks' operations. For example, Peria, Soledad, and Schmukler (2001) find that depositors in Argentina, Chile, and Mexico discipline banks by withdrawing deposits during banking crises in the 1980s-1990s. Acharya and Mora (2015) examine how the financial crisis during 2007-2009 affects deposits and loans. Data breaches can affect a sense of trust within the financial system. People may move their accounts from a bank that experienced a data breach to another bank without a data breach. From a macro perspective, customers may lose trust in a bank security system when their private data is compromised.[10] The worst-case scenario is that bank runs may occur, and further cause economy-wide damage.[11] The short-term stock performance of these banks should be negatively affected for the same reason of lower confidence in the security system of the bank after a data breach. Other than the impact of data breaches on stock prices, researchers also explore how a data breach impacts on firm values (Iyer et al., 2020). To our knowledge, however, the real economic cost of a data breach on bank operations and performance has not been examined.

Even so, customers may not move their accounts to other banks for several reasons, such as a trust that banks can handle the situation or have higher switching costs (Sharpe, 1990; Kim, Kliger, and Vale, 2003; Vesala, 2007). For example, a customer may have many financial transactions with a bank such as mortgages, payroll, financial investments, etc. Therefore, it could be inconvenient to move their accounts to another bank. Concerning switching costs, it is possible that data breach has no impact on bank deposits.[12] We test the second

[9] This is also consistent with Diamond and Rajan (2000). They show in their model that a bank run could occur when the level of uncertainty increases

[10] Campbell, Gordon, Loeb, and Zhou (2003) find that the stock market reacted negatively when the publicly traded US corporations reported their information security breaches in newspapers.

[11] Diamond and Dybvig (1983) and Bryant (1980) show in their models that when depositors lose confidence in their banks and withdraw their money at the same time, bank runs would occur and damage the economy

[12] Sharpe (1997) finds that banks have higher monopoly power when the switching cost is high. Kim et al.

hypothesis whether banks experience deposit outflows after a data breach announcement and whether the effect is stronger for any type of deposits. This test extends the analysis of Lending et al. (2018) that analyzes total deposits only. We break down deposits into total, core, insured, brokered, and time deposits.

### *2.3.1. Where deposits go after a data breach announcement?*

If people are risk-averse and weigh the risk of their private information being compromised higher than the hassle of switching their bank account to another bank, they would withdraw their money from a bank with data breach and deposit to another bank without data breach. If bank customers' still value a face-to-face service, they would switch their banking services to a close-by bank. We conduct a test to examine if deposits flow from a bank with data breach to a nearby bank without a data breach.

### *2.3.2. Effects of banks with multiple breaches*

Iyer et al. (2020) find a negative return for companies that were attacked four times by examining the bond valuation. Peng, Zhang, Mao, and Xu (2023) find that second-time data breaches hurt the firm value more than the respective first breaches. Bank customers may or may not be sensitive to multiple data breach announcements. We examine if customers who might not be sensitive to the first data breach announcement because of high switching cost decide to switch their bank accounts after banks have multiple breaches. The opposite could also be true that customers may be more sensitive to the first data breach announcement, but they are less sensitive after multiple breaches, especially when they do not see any impact on their bank accounts or believe that it will have no impact for the next breach.

(2003) find that the switching costs also exist in banks' lending.

### *2.3.3. Effects on deposits conditional on breach type and number of records lost*

The effect from a data breach announcement on bank operations could be stronger depending on the size of records lost and type of breach. Lending et al. (2018) find that the stock return of firms with data breach is lower for larger number of records lost. Huang and Wang (2021) examine the effect of data breach announcements on loan terms and find that the effect is stronger for criminal-type breaches. Following the literature, we examine the deposit flows based on different types of breach and levels of records lost.

## *2.4. Channel of impact*

Furthermore, we follow Huang and Wang (2021) to examine the channel of impact. For example, if customers expect that banks may have less profits or higher chance of bankruptcy, customers may switch their banking to another bank with stronger financial prospect. In addition, for publicly traded banks, investors may sell the bank shares when they expect higher risk and lower return from their investment. We examine bank reputation , financial performance, default risk, and information risk after a data breach announcement in Section 5.4.

## *2.5. Effects on bank stock returns*

Additionally, the impact of data breach announcements could affect banks' stock returns as well. Spanos and Angelis (2016) conduct a survey analysis showing that more than 37 papers study the effect of data breaches on stock prices from 2003 to 2015. The stock market could negatively react to data-breach announcements because a data breach has a negative effect on the firm's profit such as potential financial losses from settlement claims or a bad reputation resulting in losing customers to another bank with no data breach. Some investors may choose to sell their stock holding when they feel uncertain about the future performance of banks with a data breach. In this case, stock investors perceive the data breach event as negative news that could negatively affect the bank's future revenues or costs. We employ CARs to estimate the stock market performance of banks surrounding the data breach announcement date.

In addition, banks with a larger number of breach records may experience a larger amount of negative CARs. For example, a bank that lost a large number of records of customers' information should have a larger impact on their stock price than a bank that lost a smaller number of records. Fang and Peress (2009) demonstrate that media coverage has an impact on stock performance. A larger number of breach records affect more customers, and they could potentially convey the message to others. After investors receive the information, they may sell the stock of a bank with a data breach to avoid future capital losses. Therefore, banks with a higher number of data breach records may experience a larger drop in CARs than the banks with a lower number of breach records.

# 3. Data and Variable Construction

We obtain the data breach announcement dataset from Privacy Right Clearing House (PRC) for the period of 2005 to 2018. Other than the banking sector, the types of companies in the PRC database include other sectors such as retail, health care, and non-profit organizations. The data contain many interesting pieces of information: date made public, company name, city, state, type of breach, type of organization, total records, description of the incident, and information source. Appendix B provides examples of data breaches. The full sample of the data is from the PRC website.[13] The types of data breaches include CARD (Payment Card Fraud), HACK (Hacking or Malware), INSD (Insider), PHYS (Physical Loss), PORT (Portable Device), STAT (Stationary Device), DISC (Unintended Disclosure) and UNKN (Unknown).[14]

[13] https://privacyrights.org/data-breaches.

[14] CARD involves debit and credit cards that are not accomplished via hacking, such as skimming devices at point-of-service terminals. HACK refers to being hacked by an outside party or infected by malware. INSD is caused by insiders with legitimate access who intentionally breach information, such as an employee,

Our banking variables are from Bank Regulatory (Call Reports). The identifiers (i.e. GVKEY, PERMNO, CIK, and CUSIP) of the firms in the data breach dataset from PRC are hand collected. Table 1 demonstrates sample description. The number of data breach announcements by financial institutions for the period of 2005 to 2018 is 209, with 124 unique institutions. After merging the PRC data with the Call Reports, the number of data breach announcements by banks is 87 with 39 unique banks with available PERMCO-RSSDID link. To merge the datasets, first, the bank Call Reports is merged with the PERMCO-RSSDID linking table to obtain PERMNO/PERMCO ID. The Federal Reserve Bank of New York provides the PERMCO-RSSDID linking table starting from June 30, 1986.[15] PERMCO and RSSDID are the unique ID for CRSP companies and banks, respectively. The number of observations of treatment banks after merging data breach events with the Call Reports is 4,491. We filter bank-quarter observations when the quarterly asset growth is more than 10 percent to remove the merger effect following Acharya and Mora (2015). The number of observations for treatment banks is 4,249 after the filter. We use propensity score matching (PSM) to find control firms. For control banks, the number of observations before matching and after the 10 percent asset growth filter is 312,244. After matching, the number of observations is 4,641 with 87 unique banks. Our final dataset including both treatment and control banks has 8,890 bank-quarter observations. We explain in more detail in the methodology section about our propensity score matching procedure.

**[Insert Table 1 near here]**

We create bank variables in the same spirit as Acharya and Mora (2015). The description of variables is provided in Appendix A. Our sample period is quarterly from 2005 to 2018. We started the sample in 2005 which is the beginning year of PRC data. Appendix C provides

contractor or customer. PHYS includes paper documents that are lost, discarded or stolen (non-electronic). PORT includes lost, discarded or stolen laptop, PDA, smartphone, memory stick, CDs, hard drive, data tape, etc. STAT refers to stationary computer loss (lost, inappropriately accessed, discarded or stolen computer or server not designed for mobility). DISC is unintended disclosure not involving hacking, intentional breach or physical loss, such as sensitive information posted publicly, mishandled or sent to the wrong party via publishing online, sending in an email, sending in a mailing, or sending via fax.

[15] https://www.newyorkfed.org/research/banking_research/datasets.html

specific details on how we construct banking variables.[16] Bank-level variables are from the bank's quarterly Call Reports. We merge the banking data at the bank and the bank holding level. For the bank holding, we aggregate banks under the same holding company to the top holder and we treat it as a single banking organization. In this paper, "banks" refer to banking organizations and individual banks. The standard errors in our analysis are clustered at the banking organization and individual bank levels. To remove the merger effect in the banking industry, we exclude samples with quarterly growth of total assets larger than 10 percent. This filter follows Acharya and Mora (2015). All the growth rates are computed from the Call Reports and winsorized at the 1 percent tails. Our regression specification includes fixed effects for banks, bank district, and time.

Our main banking operation variables are the growths of deposits and loans. For the deposits, we create five different deposit variables: total, core, insured, brokered, and time deposits. For loans, there are three different types of loans: total, commercial and industrial (CI), and credit (loan and unused commitment). The control variables include other liquidity demand and bank solvency. A bank's exposure to liquidity demand is proxied by a bank's unused commitments ratio. The unused commitments ratio is computed as the ratio of unused loan commitments to the sum of loans and unused commitments. The parts of the credit lines that have not been drawn down are unused commitments. We need to control for other liquidity demand because it could affect the level of deposit at banks when the demand is transferred from an off-balance sheet to an on-balance sheet. Other control variables related to a bank's liquidity and solvency are net wholesale funding, nonperforming loans (NPL), capital, real estate exposure, and size. Net wholesale funding is the liabilities net core deposit and liquid assets. The wholesale borrowing includes gross federal fund bought net gross federal fund sold and repos net reverse repos. Non-performing loans are loans that are past due for 90 days and nonaccruing. The capital ratio is the ratio of book capital to assets. Real estate exposure is controlled by loans backed by real estate to total loan

[16] We use a modified code provided by Professor Schnabl http://pages.stern.nyu.edu/ pschnabl/data/data_callreport.htm

outstanding. Table 2 shows summary statistics for the variables used in our analysis both for controls and dependent variables. Banks in our sample are healthy based on the bank capital ratio of 12 percent. A 0.8 percent mean of the net wholesale funding ratio indicates that the larger part of bank operations is backed by wholesale funds. The proportion of NPL to loans is 1.7 percent whereas the proportion of real estate loans to loans is 47.1 percent. On average, core and insured deposits have a higher growth quarter to quarter than other types of deposits such as brokered deposits. Overall, loan growth has a similar growth rate as deposits based on the total growth of both deposits and loans. Relative to the sample of banks in Acharya and Mora (2015), our banks are larger with higher real estate loan and unused commitment.

[**Insert Table 2 near here**]

Table 3 shows summary statistics of the data breach announcement. Panel A shows the number of data breach announcements for each year. From 2005 to 2018, there are 87 breach announcements in total. The highest number of data breach announcements is in 2010 which has 12 data breach announcements. Panel B shows summary statistics of data breach types. The top two reasons for a data breach in banks are Portable Devices and Insider. Portable Devices reason is lost, discarded or stolen laptop, USB, hard drive, or any devices designed to be movable. The Insider reason is when an employee, contractor, or customer intentionally takes the data out or sells the data to a third party. Payment Card Fraud, Unintended Disclosure, and Hacking or Malware are tied for the third reason of data breach.[17] The top three average number of records lost by type is data breach events from insider, portable devices, hacking or malware. Panel C presents the summary statistics of data breach records lost. Of 87 data breach events, 44 events are known lost records while 43 events are unknown

[17] For more explanation of other reasons, Payment Card Fraud is fraud relating to debit and credit cards such as skimming devices at point-of-service terminals. Unintended Disclosure is when the data were disclosed unintentionally. For example, sending private information to a wrong party falls into this category. Hacking or malware reason is when banks are hacked by an outside party or infected by malware. Unknown is when banks do not know the reason. Physical Loss is paper documents that are lost, discarded or stolen. Stationary Computer Loss is when the data breach is from losing, inappropriately accessing, discarding, stealing computer or server not designed for moving

lost records. Known-record events mean banks know how many customer records were lost, such as Social Security numbers, addresses, etc. On average, the records lost are 898,489 with the highest number of records of 17,000,000. Based on the median records lost of 6,000, the average of records lost is highly right skewed. The largest loss of 17,000,000 records was from Countrywide Financial Corp on August 2, 2008. A former employee of the company stole and sold sensitive personal information to outside parties.

**[Insert Table 3 near here]**

# 4. Testing Methods

Our first analysis is to understand different characteristics between banks with and without data breach. We run a probit regression with important bank characteristics that determine bank operations and performance.

Next, we examine the impact of data breach announcements on bank operations using difference-in-difference (DiD) approach. To test our hypotheses, we adapt the DiD model to handle different event dates. We create interaction effects to handle the event date in Equation (1) below.

$$Y_{i,t} = \alpha_t + c_i + k_d + \sum_{q=-1}^{4} \beta_q D_{i,t+q} + \theta X_{i,t} + \epsilon_{i,t} \tag{1}$$

where $Y_{i,t}$ is a dependent variable based on our hypotheses such as deposit and loan growths for bank $i$ at quarter $t$, $\alpha_t$ is a time fixed effect, $c_i$ is a bank fixed effect, $k_d$ is a bank district fixed effect, $D_{i,t}$ are dummies of quarters before, at, and after the data breach announcement. We consider six dummies to capture the timing effect of data breach announcement (before the announcement quarter, the announcement quarter, and four quarters after). After banks announce a data breach, customers may not know right away after the data breach has been announced. For example, if banks announce the breach via email and customers miss it, customers will not know about the breach for some time. Therefore, analyzing the subsequent quarters after the announcement quarter could capture the delayed effect. On the other hand, news about the data breach could be revealed before the official announcement from banks. Hence, we also analyze the effect of data breach a quarter before the announcement. $X$ is the vector of all controls.

To test the stock market reaction of data breach announcements, we compute Cumulative Abnormal Returns (CARs) over the window (0,+1d), (0,+5d), (-1,+1d), and (-2,+1d) based on the CRSP value-weighted return. The number in the front (back) indicates the number of days before (after) a data breach announcement. For example, (-1,+1d) is the window between a day before and a day after the date of data breach announcement. In this case, zero means the event date when a data breach was announced. The longer term is also tested under the window from ten days prior to the announcement date and three quarters after the announcement date to examine if there is any long-term effect of a data breach announcement. Apart from the CAR analysis for different windows surrounding the date of a data breach announcement, we also examine CAR for each type of breach and different levels of breach records lost. We also test the multivariate CARs following Lending et al. (2018) to examine the relationship between bank characteristics and its returns.

## *4.1. Propensity Score Matching*

To reduce a confounding effect and lessen the randomness selection of a data breach bank, we match banks with and without a data breach announcement on five control variables using propensity score matching (PSM). The five control variables are unused commitment ratio, net wholesale funding, NPL to loans, capital ratio, and real estate loan share. These five control variables follow Acharya and Mora (2015) to examine bank operations (e.g. deposit and loan growth).[18] We use probit one-to-one matching without replacement, because this method satisfies the parallel trend assumption the most among all the methods. It is noted

[18] We also include indicator for large bank (top five largest back in each quarter by assets), but it was dropped during the first stage due to no top five banks in the control group.

that, for a difference-in-difference model, the goal is to match sample and control banks so that dependent variables have the same trend prior to a data breach announcement, though their levels are not required to be the same. The final sample consists of 8,890 bank quarter observations.

Table 4 shows the mean of dependent variables and controls between banks with (Treatment) and without (Control) a data breach.[19] Banks with data breach have lower capital ratio, use more wholesale funding, are larger, have larger NPL and unused commitment ratio. In terms of deposit and loan growth, banks with data breach tend to have larger deposit growth but smaller loan growth. Overall, banks with data breach tend to be riskier based on lower capital, higher NPL, more wholesale funding (e.g. Shin, 2009; Pérignon et al., 2018) and larger unused commitment (Avery and Berger, 1991).

**[Insert Table 4 near here]**

## *4.2. Parallel Trend Test*

A key assumption of the difference-in-difference (DiD) approach is, without the treatment, the average change in the dependent variables would have been the same for both treatment and control groups (parallel trend assumption). In our case, if there is no data breach announcement, the change in the dependent variables should be the same for both banks with and without the announcement. We test the parallel trend hypothesis following

[19] We are aware that one variable (real estate loan share) has indifferent mean between treatment and control banks whereas other four variables have significant mean differences. We have tried different methods of PSM including logit 1:1 with no replacement, Local Linear Regression (LLR) with Epanechnikov Kernel, radius with 0.25 standard deviation of pscore as caliper following Stuart and Rubin (2008), all Kernel with different weights: normal, biweight, and uniform. The results of mean differences for each method are in Table D.1 of Appendix D. First, compared with the no matching, matching significantly improves the mean differences based on the t-value of each variable. Logit matching provides very similar results of the mean-difference test while LLR, radius, and kernel matching methods (Continued on the next page for Table D.1) provide the same set of control variables. This indicates that the set of controls is not varied much based on the types of matching method. Similarly, it also means that our results are not sensitive to the matching methodology. For the dependent variables, t-values of the mean-difference test show less significance than the matching variables. This supports the parallel trend assumption. We choose probit over logit as it gives better parallel trend test results. We explain in details about the parallel trend test in the next section.

the Equation (2) on the time period before the data breach announcement.

$$Y_{i,t} = \alpha + \beta_1 Time_t + \beta_2 Treat_i + \beta_3 Time_t \times Treat_i + \epsilon_{i,t} \tag{2}$$

$Y_{i,t}$ is a dependent variable. $Time_t$ is a time fixed effect. $Treat_i$ is a dummy equal to one if a bank announced its data breach. The parallel trend assumption is satisfied if $\beta_3$ is not statistically different from zero. If $\beta_3$ is not different from zero, it means that the changes in dependent variables prior to a data breach announcement are not different between banks with and without a data breach announcement. Since we test the hypothesis for each data breach announcement date, the results in Table 5 show average t-statistic values for each coefficient. The most concerning variable that may not satisfy the parallel trend test is the quarterly growth of total deposit. However, our results mostly find the effect of data breach announcements on insured deposit which satisfies the parallel trend test. We finally chose probit method over logit as its average t-value for all variables is lower. Nonetheless, we also run all the results with sample created from logit matching. The results are qualitatively the same. We find that $\beta_3$ of quarterly growth of core deposit, credit, and time deposit are significant at 5 or 10 percent level. Though we dropped these variables, our paper main findings and contributions still entail by and large the same story. Specifically, on the deposit side, most of the significant effects and tests are on insured deposit. On the loan side, the only variable that may not satisfy the parallel trend test is the quarterly growth of credit (loan+commitment). We find the effect of data breach announcements on the quarterly growth of C&I loan and total loan.[20]

[20] One reason that we could not render the matching results as in Lending et al. (2018) and Huang and Wang (2021) is that our sample only includes one industry which is banking. Lending et al. (2018) and Huang and Wang (2021) include a number of industries. Including multiple industries in the sample allows more flexibility in matching variables. Banking industry is clustered and regulated mostly by asset size. The top 5 largest banks are very different from the rest of the banks in terms of size and operations. For example, our sample includes JP Morgan Chase which is the largest bank in the US by a great margin. Its size is larger than the second largest (Bank of America) by 37%. In the same vein, the second largest bank is larger than the third largest bank by 46% (Source: https://www.federalreserve.gov/releases/lbr/current/). In terms of operations, for instance, pertaining to capital ratio which is one of the matching variables, Berger, DeYoung, Flannery, Lee, and Öztekin (2008) shows that large banks can manage to have their capital level amply higher than the required minimum requirement as they can raise capital easier than smaller banks. There is

**[Insert Table 5 near here]**

# 5. Empirical Results

## *5.1. Data Breach Banks' Characteristics*

To understand the likelihood of data breach, we examine the first-stage propensity score matching in Table 6 through a Probit regression of a data breach event on bank characteristics. The data breach event is a dummy variable equal to one if banks have a data breach announcement and zero otherwise. In Table 6, we find that banks with higher unused commitment, wholesale funding, non-performing loans (NPLs), capital, and non-real estate loans are more likely to have a data breach. Based on the banking literature, except higher capital and non-real estate loans, banks with data breach seem to have higher operational risk. Avery and Berger (1991) shows that banks with more loan commitments expose to larger risk. Unlike a traditional loan, commitments allow borrowers to take out loans in the future conditional on the credit risk today. After commitments are given, borrowers' credit risk may deteriorate or borrowers may assume more risk than they would have assumed under traditional loans. Wholesale funding is another source of risk for banks if they rely on it more than retail funding (e.g. Shin, 2009; Pérignon et al., 2018). Banks with more NPLs have higher credit risk as banks may not be able to collect the full amount of their bad loans. In addition, their future financial prospect is dimmer as banks need to reserve more capital for probable losses from the NPLs. With lower lending and higher reserve, banks are poised to worsen financial performance.

**[Insert Table 6 near here]**

also interplay among bank size, capital, and lending (Kishan and Opiela, 2000). This complex relationship among matching variables and concentration in banking industry exacerbates the matching procedure in terms of finding a comparable control. Therefore, finding a comparable match within the banking industry is challenging though necessary. We have attempted to reduce the confounding effect through multi-step process of methodological selection (PSM and parallel tests) and robustness tests.

## *5.2. Effects on bank loans*

Table 7 shows that the credit activities increase in a quarter after a data breach announcement. Credits are the loans and unused commitments, such as home equity and credit card lines. Increases in Credit imply that banks tend to issue more loans after a breach announcement. We also find a sign of total loan growth increase a quarter prior to the data breach announcement quarter albeit weaker significance. According to our hypotheses, there are two possible effects of data breach announcements on loan growths: no effect (Sharpe, 1990) or more lending (Allen and Saunders, 1992). Our results are consistent with Allen and Saunders (1992) that banks try to offset the bad reputation perceived by the public after a data breach announcement by issuing more loans in order to have better financial performance.

**[Insert Table 7 near here]**

Consistent with stronger incentive to take risk (DeYoung, Peng, and Yan, 2013), large banks may lend more after a data breach announcement than small banks. Therefore, we also investigate further for large bank samples. To capture the effect of a data breach announcement on large banks, we use difference-in-difference-in-differences (DDD) as provided in Equation (3).

$$Y_{i,t} = \alpha_t + c_i + k_d + l_{i,t} + \sum_{q=-1}^{4} \alpha_q D_{i,t+q} + \sum_{q=-1}^{4} \beta_q D_{i,t+q} l_{i,t} + \theta X_{i,t} + \epsilon_{i,t} \tag{3}$$

$l_{i,t}$ is a dummy equal to one when the banks' asset size is in the top five largest banks in a given quarter, and other variables have the same definition as Equation (1). For all specifications, we control for the quarter, bank, and bank district fixed effects. The standard errors are clustered at the individual bank and bank holding levels.

Table 8 presents the impact of data breach announcement on large bank loans. It shows that the loan activities increase two quarters after the data breach announcement. Total loan increases 1.5 percent while commercial and industrial (CI) loan increases 0.3 percent two quarters after the data breach announcement quarter. The results are consistent with the findings in Table 7 that banks with a data breach potentially try to offset negative effects from a data breach with better future financial performance.

**[Insert Table 8 near here]**

To further understand the catalyst of higher lending after a data breach announcement. We examine the conflict of values between bank managers and shareholders stated in Allen and Saunders (1992). If bank managers' compensations tie to the stock value, bank managers might increase lending to improve company's financial performance which in turn affects the stock value. We apply a CEO's compensation proxy used in Ho et al. (2016) to examine whether CEOs of data breach banks have any financial incentive to increase lending after a data breach announcement. Ho et al. (2016) use CEO's option moneyness to gauge CEO over confidence.[21] They find that banks with over confident CEO tend to lower lending standard and increase leverage before financial crises. We expect to see higher option moneyness for CEOs of banks with data breach, because moneyness proxies for deep in-the-money options which indicate the amount of financial incentive tied to the stock value. To examine if CEOs of banks with data breach have higher moneyness, we create a dummy variable, *Breached Bank Dummy*, equal to one when banks have a data breach announcement. Dependent variable is moneyness with all the compensation controls along with bank, district, and time fixed effects. In Table 9, we find that CEOs of banks with data breach have higher deep in-the-money options than banks without data breach. This finding is consistent with the higher lending after a data breach announcement, potentially, to preserve banks' stock value by increasing interest income from lending.

**[Insert Table 9 near here]**

[21] The option moneyness is based on the estimated strike price and per option realizable value following Core and Guay (2002) and Campbell, Gallmeyer, Johnson, Rutherford, and Stanley (2011). To find the estimated strike price, first, find total realizable value per share for exercisable option using data from ExecuComp (OPT_UNEX_EXER_EST_VAL/OPT_UNEX_EXER_NUM). Then, subtract the total realizable value per share for exercisable option from the stock price at the fiscal year end (PRCCF). The option moneyness is the per option realizable value divided by the estimated strike price.

## *5.3. Banks deposit and data-breach announcements*

### *5.3.1. Effects on deposit flows*

As we described in the hypothesis development section, a data breach might or might not affect the deposit growth of a bank. On the one hand, if the switching cost is high or customers trust that banks can handle the situation well, we would not see much effect of the data breach announcement on the deposit flows. On the other hand, the effect of a data breach announcement could affect the overall reputation and operational losses of a bank and result in deposit outflow after a breach. Columns (1) to (5) of Table 10 show the regression results of the data breach announcement effect on different types of deposit accounts.

Insured deposits show a significant outflow of 0.5 percent at the data breach announcement quarter. Insured deposits are deposit amounts no more than $100,000 before 2009Q3. After 2009Q3, the insured amount increased to $250,000.[22] Since insured deposits are deposit accounts with a known limited amount of no more than $250,000 ($100,000 before 2009Q3), the decrease in insured deposits implies that depositors who have a deposit amount that does not exceed this cap are more sensitive to the data breach announcement. In addition, the result of core deposit growth shows that the deposit accounts with an amount less than $100,000 are not sensitive to the data breach announcement. The results of both unchanged core deposit growth and an decrease in insured deposit growth imply that the deposit accounts with the amount between $100,000 to $250,000 are sensitive to a data breach announcement. Deposit insurance literature provides evidence that depositors are less sensitive to bank risk when deposits are insured by Federal Deposit Insurance Corporation (FDIC). Park and Peristiani (1998) show a positive relation between the probability of bank failure and the subsequent outflow of uninsured deposits. Moreover, Karas, Pyle, and Schoors (2013) find that depositors are less sensitive to bank risk when their deposits are insured. This reduces market discipline by depositors. However, it is noted that the FDIC insures deposits only when a bank fails. If depositors lose their money owing to a

[22] The insured amount increased to $250,000 in 2006Q2 for retirement account.

data breach, and the bank does not fail, the FDIC will not be responsible for the loss. In addition, the FDIC does not provide deposit insurance for deposit loss from identity theft. This could be a reason why depositors who are potentially more sensitive to banks' risk withdraw their funds after a data breach announcement. The FDIC states that unauthorized access to deposits can be covered by the Electronic Funds Transfer Act (EFT Act) and other consumer protections.[23] However, the EFT Act requires depositors to report an unauthorized electronic fund transfer that appears on a periodic statement within 60 days; otherwise, the depositors could be liable to the losses.[24] Though deposits could be covered by the EFT Act, depositors need to monitor their accounts themselves. Unlike the EFT Act, FDIC deposit insurance will automatically cover deposits when a bank fails without additional effort from the depositors to monitor their own accounts. As a result, it is possible that insured depositors try to avoid the inconvenience of the EFT Act reporting requirement by moving their deposits to another bank with no data breach history.

Therefore, our findings are potentially not related to banks' likelihood of failure since their deposits are insured in that case. Our finding is more in line with the precautionary action to prevent losses from their information being exploited and depositors losing trust in the banks in handling their private information. Based on the restuls in Table 10, we do not see a long-term effect after a data breach announcement. A sign of negative insured deposit growth only appears at the data breach announcement quarter without any subsequent impacts in later quarters.

We also find a 0.3 percent decrease in the brokered deposit growth almost a year after a data breach announcement. The delayed outflow of the brokered deposits could be explained by the types of inflow that created brokered deposits. Brokered deposits can be divided into Primary Purpose (PP) and Primary Purpose Exception (PPE). Not all third-party deposits are considered brokered deposits.[25] The Primary Purpose indicates that if the goal of the

[23] https://www.fdic.gov/deposit/covered/notinsured.html

[24] FDIC Law, Regulations, Related Acts- 6500-ConsumerFinancial Protection Bureau-Part 1005 Electronic Fund Transfers (Regulation E) §1005.6 https://www.fdic.gov/regulations/laws/rules/

[25] https://www.fdic.gov/news/board-matters/2020/2020-12-15-notice-dis-a-fr.pdf

third-party deposits is to facilitate the deposit flows for the purpose of deposit, the deposit is considered a brokered deposit. A large proportion of Primary Purpose deposits are in the form of Certificate Deposits (CDs) which require customers to deposit their money for a certain time period. Typical CDs last between 12 to 36 months. In the case of Primary Purpose, some depositors may not want to lose the interest income from an early withdrawal though they may like to move to another bank after a data breach announcement. For the Primary Purpose Exception, some deposits from the third-party may also qualify to be brokered deposits though their sole purpose is not to facilitate the deposit flow for customers. However, they need to satisfy either the rule of 1) No more than 25 percent of the assets are deposited 2) 100 percent of assets are in transactional accounts that have no interest or fees paid to depositors. For the Primary Purpose Exception, customers deposit their money through a third-party and may not even know which bank has their money. Unless the third party itself has a data breach, customers may not withdraw the money and, consequently, there would be no immediate brokered deposit decrease in banks with a data breach.

[**Insert Table 10 near here**]

### *5.3.2. Where deposits go after a data breach announcement*

Our main results show that banks with a data breach experience insured deposit outflows at the data breach announcement quarter. An interesting follow-up question is whether those outflows become the inflows of banks with no data breach. We hypothesize that depositors would withdraw money from banks with a data breach and deposit it to a nearby bank without a data breach if depositors are sensitive to the data breach and still value the face-to-face service.[26] To test this conjecture, we follow Equation (4) below.

[26] We examine only the insured deposit, since the outflows occur during the data breach event quarter. The outflows of brokered deposit occur much later and are subject to potential confounding effects.

$$Y_{i,t} = \alpha_t + c_i + k_d + \beta_b DepositGrowth_{b,t} + \sum_{q=-1}^{4} \alpha_q D_{i,t+q} + \sum_{q=-1}^{4} \beta_q D_{i,t+q} DepositGrowth_{b,t} + \theta X_{i,t} + \epsilon_{i,t} \tag{4}$$

where $Y_{i,t}$ is insured deposit growth of no data-breach bank $i$ at quarter $t$, $\alpha_t$ is a time fixed effect, $c_i$ is a bank fixed effect, $k_d$ is a bank district fixed effect. $DepositGrowth_{b,t}$ is the insured deposit growth of banks $b$ with a data breach at quarter $t$. Since we assume an immediate withdrawal from a data breach bank and depositing into a no data breach bank, we match the same quarter of deposit growths between banks with and without a data breach announcement. $X$ is a vector of all controls. $D_{i,t}$ are dummies of quarters prior to, at, and four quarters after the data breach announcement. A negative $\beta_q$ is consistent with depositors withdrawing money from banks with a data breach and depositing it to banks without a data breach.

To test Equation (4), we first match banks with and without a data breach based on their zip code.[27] Then, we calculate the distance between the two banks in miles. We have three distance matching categories: less than 5 miles, less than 10 miles, and between 10 to 30 miles. customers may withdraw deposits from breached banks and redeposit in non-breached banks that are farther than 30 miles away from the breached banks. In this case, we should not find any evidence of deposit flowing to a non-breached bank within 30 miles of a breached bank. For example, under a technological advancement, depositors may open an account online and transfer via an online channel without visiting the physical branch. In this case, the zip code of the newly opened bank account can be the zip code of the bank headquarters, which could be far away from the bank with a data breach. If this argument is true, we should not find a significant relationship between the deposit growth of banks with and without a data breach that are located near each other. We also evaluate the hypothesis

[27] It is noted that we use all available non-breached banks in this step which is different from our main results that only use matched non-breached banks from the propensity score matching. We use all available non-breached banks in this matching step because we assume that customers can move their deposits to any bank. The zip code (RSSD9220) and state (RSSD9200) information are retrieved from the Call Reports.

for banks within the same state. The motivation for the same-state analysis is that banks are regulated at both the federal and state levels.[28]

Table 11 shows the results of deposits outflow analysis between deposit growth of banks with and without a data breach. We find evidence that deposits outflow from the banks with a data breach to the banks without a data breach. At the event quarter (Qevent), we find negative coefficients for banks with and without data breach located less than 10 miles, 10 to 30 miles, and within the same state. The interpretation of the coefficients is, for example, a one percent drop in insured deposit growth in the data breach announcement quarter results in 0.173 percent increase in the insured deposit growth of banks without a data breach announcement that is located less than 10 miles away from breached banks. For banks located less than 5 miles from each other, we find negative coefficient a quarter before banks announce their data breach. An explanation could be that since our event quarter is when banks announced their data breach, not when banks had the breach, there could be some delay in their announcement. As a result, there could be information about the breach leaking out to the public before banks decided to announce the breach.

[**Insert Table 11 near here**]

### *5.3.3. Effects of banks with multiple breaches*

Banks can experience more than one data breach. In this section, we analyze if responses from bank customers are different when banks have multiple breach announcements. On the support of stronger effect, bank customers may perceive a bank with multiple data breaches as having an unsafe security system. Consequently, customers could move their accounts away from a bank with multiple breaches. On the other hand, customers may learn from the previous breaches that their accounts are still safe and may decide to maintain the service with their bank. For the latter case, we may find less impact on the deposit flows for

[28] https://www.frbsf.org/education/publications/doctor-econ/2006/november/commercial-banks-regulation/

repeated data breaches. Our approach is to first create a dummy variable for each data breach announcement. For example, on the first data breach announcement, we assign the first-breach dummy equal to one and zero otherwise. On the second data breach announcement, the second-breach dummy is assigned a value of one and zero otherwise, and so on. We also interact the number of data breach announcement dummies with the announcement quarter and four quarters after. We perform the analysis on the insured deposit account because we find significant insured deposit outflows at the announcement quarter in our main results. In the results presented in Table 12, each column represents each announcement order. For example, column (1) shows the results for the first announcement dummy. *BreachNumber* is equal to one when the announcement is the first time a data breach occurred. Table 12 shows no significant movement in deposit flows after the fifth announcement and then the movement starts again on the ninth announcement. The two largest and most significant deposit outflows are when banks announced their fourth and tenth breaches. Specifically, banks have 4.1 percent deposit outflows three quarters after the fourth announcement. Then, on the tenth announcement, banks experience 4.8 percent deposit outflow a quarter after the announcement. The results demonstrate that customers may be less sensitive with a data breach after a few announcements and they start to worry again when the number of breach reaches double digit.

[**Insert Table 12 near here**]

### *5.3.4. Effects on deposits conditional on breach type and the number of records lost*

In this section, we test the deposit outflow based on the types of breach and the number of records lost. Table 13 shows the results of data breach effect on deposits by types of breach. Huang and Wang (2021) find that criminal-type data breach has more effect on the loan spreads than other types. We find that hacking, portable device, and stationary device are the three types that generate the most outflow on bank deposits. For data breach from hacking (HACK), we find evidence of total deposit outflow even a quarter prior to the data breach announcement quarter. Consequently, some evidence of insured deposit outflow shows up during the announcement quarter and a quarter after. For a data breach causing from portable device (PORT), the effect of data breach announcements is on core and insured deposits mostly at the announcement quarter. Insured deposit also shows signs of long-term effects as the same magnitude of outflows also shows up a quarter and three quarters after the announcement. Data breach from stationary device (STAT) causes banks total and brokered deposit outflows though the effect is a little slower, a quarter after the announcement quarter compared with the HACK and PORT. The results of deposit outflows from hacking are in line with Huang and Wang (2021) that groups the criminal-type breach by including payment card fraud and hacking.

For the number of records lost analysis, we create a dummy variable equal to one, *High-records*, if the number of records lost is greater than the median number of records lost. We interact with the data breach event and treatment variable to create the analysis of three interactions. Table 14 provides the results of high-record data breach. We find an outflow of time deposit two quarters after the data breach announcement quarter. The results show that the effect of data breach announcements on bank operations conditional on the number of records lost maybe less responsive as there are no significant effects at the quarter or a quarter after a data breach announcement. Alternatively, depositors may have to wait until they can withdraw the time deposit due to the required amount of time to stay in the account; hence, the effect is delayed.

**[Insert Table 13 and 14 near here]**

## *5.4. Channels of impact*

From the previous sections, we find that data breach announcements impact banks' operations. In this section, we investigate in which channels data breach could impact banks' operations. We follow Huang and Wang (2021) to test the channels of impact. First, we construct proxy variables for reputation (loss of major customers and market share growth), financial performance (ROA and cash flow from operations), default risk (Z-Score), and information risk (stock illiquidity and standard deviation of stock returns). The full descriptions of variables are in Appendix A and Appendix C. Then, we apply the staggered DiD methodology utilizing the effective date of data breach notification laws by states from Table 7 of Huang and Wang (2021). Specifically, our specifications follow Equation (5) below.

$$Y_{i,t} = \alpha_t + c_i + k_d + \beta Treat * AfterLaw_{i,t} + \theta X_{i,t} + \epsilon_{i,t} \tag{5}$$

where $Y_{i,t}$ is a dependent variables represented reputation, financial performance, default risk, and information risk for bank $i$ at quarter $t$, $\alpha_t$ is a time fixed effect, $c_i$ is a bank fixed effect, $k_d$ is a bank district fixed effect. $Treat$ is a dummy variable equal to one if banks have a data breach announcement. $AfterLaw$ is a dummy variable equal to one when a data breach announcement quarter is after the effective date of data breach notification laws by states. We identify banks' location based on their zip code (RSSD9220) and state (RSSD9200) information from Call Reports. We incorporate bank and time fixed effects. Therefore, the single covariates of $Treat$ and $AfterLaw$ are already controlled for.

Based on the results in Table 15, they show no evidence of data breach announcement effect on reputation and default risk. In terms of the results on default risk, they are consistent with our findings that deposit outflows are mostly short-lived with no evidence of long-term and sustaining impacts. Since the outflows of deposits are not long-term, the likelihood of default (or run) should be minimal. However, we find that bank significantly have lower cash flow from operations and lower standard deviation of return. The result on the standard deviation of return is consistent with Lending et al. (2018). They find that banks with data breach tend to be larger in size and have lower standard deviation of returns.

**[Insert Table 15 near here]**

### *5.5. Effects on banks' stock returns*

In this section, we discussed the effects of data breach announcements on banks' stock returns. Panel A of Table 16 shows that data breach announcements have a negative impact on banks' stock returns, which is consistent with our hypothesis. CARs are significantly negative within different windows, (0,+1d), (0,+5d), (-1d,+1d), (-2d,+1d). Among them, the CAR of (0,+5d) has the lowest value on average, -0.83%. The CAR within (-1d, +1d) is -0.4%, which is smaller than -0.8% from Kamiya et al. (2021), possibly because Kamiya et al. (2021) examine attacks involving the loss of personal financial information only. However, we find a larger loss within (-1d, +1d) than the CAR, -0.31%, of the stocks taken advantage of by short sellers (Wang et al., 2022). Comparing with the CARs within (-1d, +1d), -0.1%, of all breaches in Lending et al. (2018), we find much larger losses, -0.4%, in the financial industry. Among several types of data breaches, Hacking or malware (HACK) has a more severe impact on the stock market, lower than -0.9% of CARs over the window (0,+1d) and window (-2d,+1d). This is consistent with the findings of Lending et al. (2018) that HACK induces larger losses in the stock market than other types of breaches. Large negative CARs also occur when the cause of a data breach is from losing physical documents (PHYS). The results on (0,+1d) are also consistent with our finding in multivariate analysis of data breach announcement effects on deposit growth by breach type. We find that breaches from hacking, portable devices and stationary devices, are the top three types that produce the most outflow on bank deposits.

Panel B of Table 16 provides the CARs for different quantiles of total records. We divide financial firms into four groups from Q1 (highest number of breach records) to Q4 (lowest number of breach records). Unknowns are the banks with an unknown number of breach records. The results show that Q1 has the significantly largest loss of stock returns, which are, -1.76% over (0,+1d), -4.55% over (0,+5d) and -1.66% over (-1d,+1d). Overall, the losses are decreasing from Quantile 1 to Quantile 4 although most CARs are not significant in groups Q3 to Q4. The CARs of Unknown firms range from -0.3% to -0.46% yet they are not statistically significant. The results confirm our expectation that banks with a higher number of lost records experience a larger drop in CARs than banks with a lower number of lost records.

To eliminate the possibility that the data limitation may bias the overall results, we present CARs of the banks with available Call Reports and stock data in Panel C of Table 16. The results of banks with available Call Reports and stock data are qualitatively the same. Compared to all banks with data breach, banks in our sample after merging all the data have slightly smaller losses (-0.33% vs. -0.49%) over (0,+1d) and slightly larger losses (-1.26% vs. -0.83%) over (0,+5d). The reasons are that the banks with missing call reports and stock data are either relatively small banks or the data breach incidents have unknown records. Panel C reports the CARs of large banks only (top five largest banks in a given quarter). Large banks experienced a loss of -0.39% over (0,+1), but did not retain significant losses 5 days after the data announcements. The insignificant losses over (0,+5) could be from investors' confidence in the large banks and large banks' quick and responsible reactions to data breaches. For example, Capital One disclosed a data breach on July 19, 2019. They immediately fixed the issue and promptly began working with federal law enforcement. The following clear reports are publicly available about who is responsible for the data breach incident, how the incident impacts customers, what Capital One did to protect customers, etc.[29]

[**Insert Table 16 near here**]

We also investigate the long-term effect of data breach announcements as shown in Figure 1. Consistent with the findings in the short-term effect, the CAR is continuously decreasing to around -5% through the period from the announcement date to three quarters post announcement. We acknowledge that the results can be noisy when we test for a longer period.

[29] https://www.capitalone.com/digital/facts2019/faq/

**[Insert Figure 1 near here]**

We follow Lending et al. (2018) and conduct a multivariate analysis on CARs.[30] The results are in Table 17. We perform six different CARs' windows. The first four columns are CARs' windows from our CARs' univariate analysis. The last two columns are longer windows following Lending et al. (2018). For the choice of explanatory variables, since we try to explain what causes CARs to change based on different banks' characteristics, we follow Huang and Wang (2021) by adding more banks' characteristics that can be created from Call Reports: firm size, ROA, leverage, and Z-Score. These variables are on top of the control variables in the main analyses following Acharya and Mora (2015). The results show that if banks are one of the largest five banks in each quarter, they tend to have larger negative impact on their return during the window between one day before and one day after the data breach announcement date (-1 to 1d). Banks with larger NPL to loans tend to have better CARs during the window between two days before and one day after the data breach announcement date (-2 to 1d). Larger banks with more real estate loan and NPL to loans tend to perform better during the return window between 30 days before and 2 days before the data breach announcement date (-30 to -2d). The opposite pattern applies to banks that use more wholesale funding during the same window (-30 to -2d).

**[Insert Table 17 near here]**

Overall, based on the CARs' univariate analysis, we find that a data breach announcement produces negative CARs surrounding the announcement date. Compared to other banks, large banks have less impact 5 days after the data breach announcement date. Hacking, portable and stationary devices, and physical document loss are the types of breaches with the most severe losses. In addition, data breaches are likely to have a long-term destructive impact on stock value. In case of multivariate analysis, top five largest banks have larger negative impact on their returns on the announcement date, but the effect recedes during

[30] It is noted that, following Lending et al. (2018), we only use banks with data breach announcement to conduct the CARs' multivariate analysis.

the one month leading to the announcement date. Banks with more real estate loan and larger NPLs tend to fare better while banks that use more wholesale funding tend to perform worse.

## 5.6. *Robustness*

### 5.6.1. *Breach disclosure law*

Romanosky, Telang, and Acquisti (2011) examine how disclosure laws on identity theft during 2002 to 2007 affect the number of identity thefts. They find that the law had marginal effect on the number of identity thefts. We use the breach law adoption data in Table 7 of Huang and Wang (2021) regarding the notification laws' effective date for each state to run a cleaner DiD model. We provide more details about the model specifications in Section 5.4. From Table 18, with the DiD incorporating notification laws, we find a negative outflow of 0.4 percent in the core deposits when the announcement is after the breach law adoption date compared with a negative outflow of 0.5 percent in the insured deposits after data breach announcements in the main results. Overall, the results in Table 18 show a similar deposit outflow in smaller bank accounts (Core deposit less than $100,000 VS Insured deposit less than $250,000).

[**Insert Table 18 near here**]

### 5.6.2. *Spillover effect*

An earlier data breach announcement could create a psychological impact on customers overall confidence in their bank, even though their banks are not the one which announces the data breach incident (spillover effect). For example, when customers hear about a data breach in another bank and are afraid that their banks may also have a data breach in the future, they may decide to hold cash or move their deposits somewhere. Our results partly could be from the spillover effect of the earlier data breach announcement from another bank. To address the omitted variable concern, we control the earlier breach announcement effect by creating a cumulative number of data breach announcements as a control variable. Specifically, we create *Cumulative Breach* variable which is equal to one for the first breach announcement in our data. Then, in chronological order, the next announcement is assigned number two and so on. We add the *Cumulative Breach* variable as part of the control variables and rerun the regression. Table 19 shows that the results are qualitatively robust to the main results.

[**Insert Table 19 near here**]

### *5.6.3. Non-random shock*

One could argue that banks may be targeted by criminals in a non-random fashion. For example, large banks could be targeted by hackers more than small banks because the benefit from breaching large banks is higher. However, hacking as a cause of a data breach in our sample is about 15 percent of the total. For other causes such as insider, payment card fraud, and unintended disclosure, a counter argument would be that large banks may have a better protocol to prevent a data breach from these types of breach than small banks could have. For instance, it could be easier for employees in smaller banks to take customers' information outside and sell it to a third party. The argument that some banks are targeted more than other banks could be true when the cause is hacking. However, for other causes which are the majority of the data breach types in our sample, the argument that some banks are prone to have more data breach is debatable. Besides, in the main results, we control for the large bank effect with large bank indicator as our control to lessen the issue of non-random target based on the hacking as the cause of breach. Large bank indicator is the dummy equal to one if a bank's asset size is the top five largest asset size in each quarter.

# 6. Conclusion

Data breaches could negatively affect banks' operations and performance when they announce that their customers' information such as their Social Security and account numbers are compromised. Customers may move their accounts (e.g., deposits) to another bank with no previous history of a data breach because they lose confidence in the security system of the bank or to prevent their accounts from being exploited. Meanwhile, customers may not move their accounts elsewhere because of a high switching cost or trust that banks can handle the issue well. A consequence after a data breach is an important question for the banking industry and the economy since the banking industry plays a key role in our economy as evidenced by the subprime financial crisis. To evaluate the effect of a data breach announcement on banks' operations and stock performance, we employ a difference-in-difference approach and an event study.

Relative to banks with no data breach, banks with data breach have lower capital ratio, use more wholesale funding, are larger, have larger NPL and unused commitment ratio. The characteristics of banks with data breach indicate that they tend to be riskier. After a data breach announcement, banks may try to offset bad news from a data breach with better financial performance in a later period by reducing costs and increasing revenue. We find evidence that banks increase their lending after a data breach announcement, supporting our argument that banks try to increase their revenue after a data breach announcement. To further support our argument, we show that CEOs of banks with data breach have significantly higher deep in-the-money options which give them incentives to maintain or improve financial performance.

We find a sign of insured deposit outflows a quarter after a data breach announcement and brokered deposit outflows a year after. A typical US bank borrows short-term and lends long-term. Without stable deposits, banks may have a liquidity issue. Bank customers are less sensitive to a data breach announcement when banks have multiple data breaches. customers may learn from an earlier breach that their accounts are still handled well by banks and have no desire to move their money away. Based on our results, the outflows of deposits show no sign of long-term impact. The results are consistent with the real world situation that, so far, we have not seen any bank fails after a data breach and banks usually compensate for unauthorized transactions. However, banks' customers are concerned about the data breach again after a number of data breach reach double digits. We find evidence that the deposits after a data breach announcement outflow to near-by banks with no data breach. Conditional on the number of records lost, a larger number of records affect more on time deposit accounts. Data breach caused by hacking, portable device, and stationary device have larger impacts on banks' operations. Our main findings show that data breach announcements affect banks' operations, and the effects are short-term.

In terms of the effect of data breach announcement on stock returns, we conduct analyses on Cumulative Abnormal Returns (CARs) surrounding the event date. We find both short-term and long-term effect on CARs. We examine the channels in which data breach announcement could impact bank operations and returns. We find that, after a data breach announcement, banks have significantly lower operating cash flows.

Our paper gives the first step to understand how a data breach affects bank operations and performance. The contribution of our paper is to understand more about the new bank risk factors in the technology era such as data breaches and cybersecurity. In the future, if banks do not keep up with the new technology and security system, a data breach may trigger a liquidity problem and affect economic welfare. Areas of future research could be, first, how changes in banking activities such as decreases in insured deposits affect economic growth or financial decisions of different economic units. Second, more studies can further examine bank strategic actions in response to different alarming situations, such as various types of a data breach.

Fig. 1. Cumulative Abnormal Return: Mean & 95% Confidence Limits
This Figure provides CARs from pre-10 day to post-270 days relative to the data-breach announcement date. The estimation model is based on the CRSP value-weighted return.

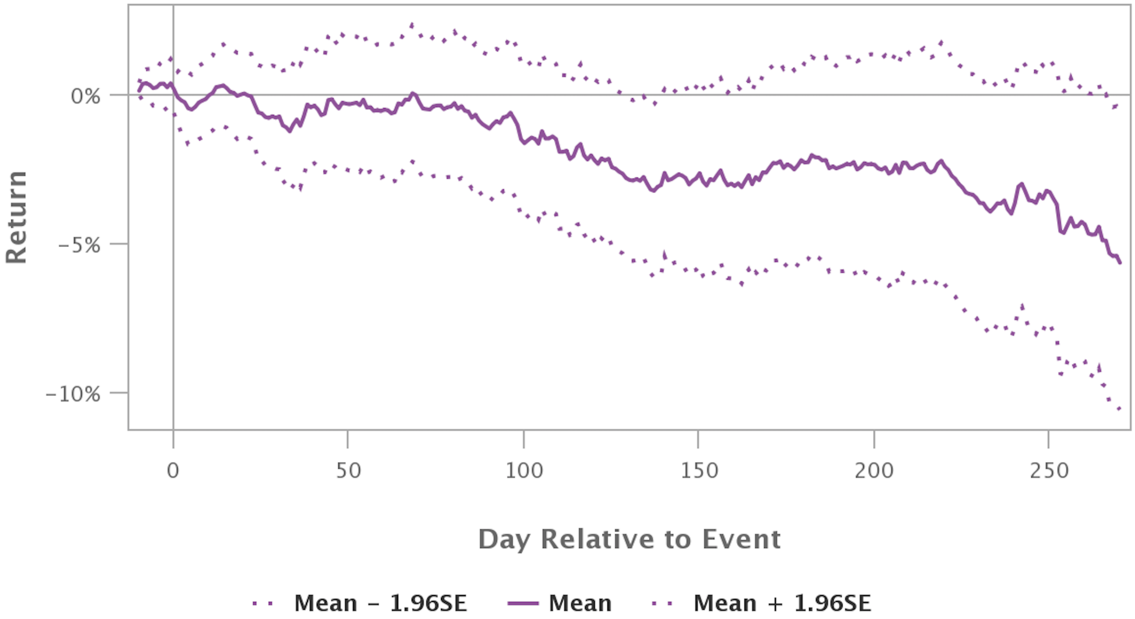

Table 1: Sample Description

This table shows the development of sample construction. The data are from 2005 to 2018. The data breach announcements are from Privacy Right Clearinghouse. The banking data are from Call Reports. For banks with BHCID level, we aggregate them at the bank holding level. Number of banks includes both bank holding (with BHCID level) and banks (with no BHCID level). Privacy Right Clearinghouse and Call Reports are merged through RSSDID-PERMCO link provided by Federal Reserve Bank of New York (https://www.newyorkfed.org/research/banking_research/crsp-frb). We remove bank quarters with asset growth more than 10% following Acharya and Mora (2015) to reduce M&A effect. We use propensity score matching (PSM) probit 1:1 without replacement to find control banks. For control banks, we also require control banks to have at least 30 quarters available during 2005 to 2018.

| | |
|---|---:|
| Number of data breach events from 2005 to 2018 (financial firms) | 209 |
| Number of data breach events with available merging ID (RSSDID-PERMCO Link) | 87 |
| Number of unique treatment banks | 39 |
| Number of observations of treatment banks after merging data breach events with the Call Reports | 4,491 |
| Number of observations of treatment banks after removing bank quarter with asset growth more than 10% | 4,249 |
| Number of control banks with available PSM matching variables | 5,102 |
| Number of observations of control banks | 333,282 |
| Number of observations of control banks after removing bank quarter with asset growth more than 10% | 312,244 |
| Number of observations of control banks after PSM matching | 4,641 |
| Number of unique control banks after PSM matching | 87 |

Table 2: Summary Statistics of Variables.

This table provides the mean, standard deviation, 25 percentile, median, 75 percentile and the number of available observations (N) of control variables and dependent variables. The bank-level variables are quarterly data from 2005 to 2018. The banking data is obtained from Bank Regulatory (Call Reports). Stock returns data are from CRSP. Customer information is from Compustat segment. The full description of each variable can be found in Appendix A.

| Variables | Mean | StdDev | P25 | Median | P75 | N |
|---|---|---|---|---|---|---|
| Quarterly growth of deposits | 0.011 | 0.035 | -0.007 | 0.009 | 0.027 | 8,718 |
| Qaurterly growth of core deposits | 0.009 | 0.031 | -0.006 | 0.006 | 0.022 | 8,718 |
| Quarterly growth of insured deposits | 0.007 | 0.030 | -0.004 | 0.002 | 0.011 | 8,718 |
| Qaurterly growth of brokered deposits | 0.001 | 0.013 | -0.001 | 0.000 | 0.001 | 8,718 |
| Quarterly growth of time deposits | 0.000 | 0.023 | -0.004 | 0.000 | 0.004 | 8,718 |
| Qaurterly growth of loans | 0.010 | 0.030 | -0.005 | 0.006 | 0.022 | 8,718 |
| Quarterly growth of C&I loans | 0.002 | 0.011 | -0.001 | 0.001 | 0.005 | 7,327 |
| Quarterly growth of credit (loans+commitments) | 0.008 | 0.038 | -0.004 | 0.001 | 0.023 | 8,718 |
| Qaurterly growth of net interest margin | 0.001 | 0.014 | -0.002 | 0.007 | 0.009 | 8,718 |
| Asset growth | 0.009 | 0.036 | -0.009 | 0.009 | 0.029 | 8,718 |
| Capital ratio (book capital to assets) | 0.120 | 0.041 | 0.097 | 0.111 | 0.130 | 8,890 |
| Firm size (log of total assets) | 14.633 | 2.599 | 12.035 | 15.445 | 17.114 | 8,890 |
| Indicator for large banks | 0.230 | 0.421 | 0.000 | 0.000 | 0.000 | 8,890 |
| Market share growth | -0.004 | 0.046 | -0.024 | -0.003 | 0.018 | 8,718 |
| Net wholesale funding ratio (wholesale -liquid) | 0.008 | 0.192 | -0.100 | 0.001 | 0.143 | 8,890 |
| Nonperforming loans to loans | 0.017 | 0.020 | 0.003 | 0.010 | 0.021 | 8,890 |
| ROA | 0.017 | 0.021 | 0.004 | 0.009 | 0.038 | 8,890 |
| Real estate loan share | 0.471 | 0.251 | 0.298 | 0.472 | 0.659 | 8,890 |
| Unused commitment ratio | 0.196 | 0.150 | 0.028 | 0.207 | 0.363 | 8,890 |
| Z-Score | 2.021 | 1.470 | 0.771 | 1.863 | 2.680 | 8,890 |
| Leverage | 0.874 | 0.137 | 0.832 | 0.894 | 1.000 | 6,282 |
| Tangibility | 0.004 | 0.008 | 0.000 | 0.001 | 0.004 | 4,681 |
| Return Std | 0.069 | 0.072 | 0.030 | 0.055 | 0.084 | 4,194 |
| Operating Cash Flow (CFO) | 0.011 | 0.020 | 0.002 | 0.009 | 0.019 | 4,531 |
| Illiquidity | -1.482 | 1.384 | -2.328 | -1.864 | -1.204 | 4,196 |
| Loss Customer Dummy | 0.297 | 0.457 | 0.000 | 0.000 | 1.000 | 1,118 |

Table 3: Summary Statistics of Data Breach Announcements

Panel A shows the number of data breach announcements for each year from 2005 to 2018. Panel B shows the number of data breach types. Panel C shows summary statistics of data breach records lost. Data breach data are retrieved from Privacy Rights Clearinghouse, https://privacyrights.org/data-breaches

Panel A: Data-Breach-Announcement Year

| Year | Freq. | Percent |
|---|---|---|
| 2005 | 9 | 10.34 |
| 2006 | 11 | 12.64 |
| 2007 | 5 | 5.75 |
| 2008 | 6 | 6.90 |
| 2009 | 3 | 3.45 |
| 2010 | 12 | 13.79 |
| 2011 | 9 | 10.34 |
| 2012 | 8 | 9.20 |
| 2013 | 7 | 8.05 |
| 2014 | 6 | 6.90 |
| 2015 | 2 | 2.30 |
| 2017 | 2 | 2.30 |
| 2018 | 7 | 8.05 |
| Total | 87 | 100 |

Panel B: Type of Data Breach

| Type of Breach | Freq. | Percent | Avg. Record Lost |
|---|---|---|---|
| Portable Devices | 17 | 19.54 | 1,638,329 |
| Insider | 16 | 18.39 | 2,378,791 |
| Payment Card Fraud | 13 | 14.94 | 7,961 |
| Unintended Dsiclosure | 13 | 14.94 | 132,789 |
| Hacking or Malware | 13 | 14.94 | 171,607 |
| Unknown | 9 | 10.34 | 13,764 |
| Physical Loss | 4 | 4.60 | All unknown |
| Stationary Computer Loss | 2 | 2.30 | 45,500 |
| Total | 87 | 100 | 898,489 |

Panel C: Summary Statistics of Data Breach Records Lost

| Mean | Std | P25 | Median | P75 | Min | Max | N |
|---|---|---|---|---|---|---|---|
| 898,489 | 3,164,006 | 332 | 6,000 | 118,000 | 0 | 17,000,000 | 44 |

Table 4: Summary Statistics of Banks with and without Data Breach

This table shows the mean of bank variables with (Treatment) and without (Control) a data breach. First, banks with a data breach from Privacy Right Clearing House (PRC) are merged with the PERMCO-RSSDID linking table. PERMCO-RSSDID linking table links the bank entity and CRSP-PERMCO. Then, we match banks with a data breach with banks without a data breach using propensity score matching (PSM). We use probit one-to-one matching on all control variables without replacement. Banking data is from the Federal Reserve Call Reports. Please see Appendix C for instructions on how to construct each variable.

| Variables | Treatment | Control | Difference | t-value |
|---|---|---|---|---|
| Capital ratio (book capital to assets) | 0.115 | 0.124 | -0.009 | 10.17 |
| Net wholesale funding ratio (wholesale -liquid) | 0.070 | -0.049 | 0.119 | -30.77 |
| Nonperforming loans to loans | 0.021 | 0.012 | 0.009 | -22.31 |
| Unused commitment ratio | 0.263 | 0.134 | 0.129 | -44.75 |
| Real estate loan share | 0.470 | 0.473 | -0.003 | 0.49 |
| Indicator for large bank | 0.482 | 0.000 | 0.482 | -65.68 |
| Quarterly growth of deposits | 0.012 | 0.011 | 0.001 | -1.15 |
| Qaurterly growth of core deposits | 0.011 | 0.008 | 0.002 | -3.15 |
| Quarterly growth of insured deposits | 0.007 | 0.007 | 0.000 | -0.10 |
| Qaurterly growth of brokered deposits | 0.001 | 0.001 | 0.000 | -1.37 |
| Quarterly growth of time deposits | 0.001 | 0.000 | 0.001 | -1.19 |
| Qaurterly growth of loans | 0.009 | 0.011 | -0.002 | 2.92 |
| Quarterly growth of C&I loans | 0.002 | 0.003 | -0.001 | 3.18 |
| Quarterly growth of credit (loans+commitments) | 0.006 | 0.009 | -0.003 | 3.37 |
| Qaurterly growth of net interest margin | 0.001 | 0.001 | 0.000 | 0.23 |
| Number of observations | 4,249 | 4,641 | | |

Table 5: Parallel Trend Test

$$Y_{i,t} = \alpha + \beta_1 Time_t + \beta_2 Treat_i + \beta_3 Time_t \times Treat_i + \epsilon_{i,t}$$

This table shows the parallel trend test. $Y_{i,t}$ is dependent variable (Quarterly growth of each variable). $Time_t$ is a time fixed effect. $Treat_i$ is a dummy equal to one if a bank announced its data breach. The parallel trend assumption is satisfied if $\beta_3$ is not statistically different from zero. We test the assumption for each data breach announcement date. The results in Table 5 show average t-statistic values for each coefficient. The banking data is from Call Reports from 2005 to 2018. Please see the full definition for each variable in Appendix A. ***,**,* are significant at 1, 5, 10 percent, respectively.

| Dependent Variables | $\alpha$ | $\beta 1$ | $\beta 2$ | $\beta 3$ |
|---|---|---|---|---|
| Total deposit | 8.47 | 2.29 | 3.41 | 2.94*** |
| Core deposit | 4.28 | 1.58 | 3.08 | 2.06** |
| Insured deposit | 4.21 | 3.04 | 1.26 | 1.35 |
| Brokered deposit | 4.68 | 2.93 | 1.14 | 1.49 |
| Time deposit | 9.22 | 8.97 | 1.68 | 2.26** |
| Total loan | 14.56 | 7.06 | 0.79 | 0.51 |
| C&I loan | 8.10 | 3.14 | 1.00 | 0.60 |
| Credit (loan+commitment) | 13.97 | 8.65 | 1.45 | 1.83* |
| Net interest margin | 4.32 | 3.16 | 0.78 | 0.67 |

Table 6: Probit Regression

Dependent variable is a dummy variable equal to one for banks with data breach. This table shows the first-stage propensity score matching. Please see the full definition for control variables in Appendix A. ***,**,* are significant at 1, 5, 10 percent, respectively.

| | (1) |
|---|---|
| VARIABLES | Treated |
| Unused Commitment Ratio | 0.629*** |
| | (198.98) |
| Net Wholesale Funding | 0.067*** |
| | (58.77) |
| NPL to Loans | 0.403*** |
| | (39.69) |
| Capital Ratio | 0.125*** |
| | (22.68) |
| Real Estate Loan Share | -0.131*** |
| | (-116.30) |
| Constant | 0.017*** |
| | (7.80) |
| Observations | 316,493 |
| R-squared | 0.1908 |
| Bank FE | YES |
| District FE | YES |
| Qtr FE | YES |

Table 7: Data Breach Announcements and Loans

The results show the data breach announcement's effect on loan growths. Columns (1) to (3) show results for different types of loan growth as dependent variables. *Loan* is quarterly growth of all types of loans. *CI* is the quarterly growth of commercial and industrial loans. *Credit* is quarterly growth of loans and commitments. Columns (4) is the result of the regression for Net Interest Margin (NIM) as dependent variables. Please see the full definition for each type of loan and other control variables in Appendix A. Q1before, Qevent, and Q1-4after, are a dummy variable equal to one if the interaction terms between a data breach bank dummy and a time dummy of data breach announcement prior quarter, quarter, and 1-4 quarter after the announcement, respectively, are equal to one. ***,**,* are significant at 1, 5, 10 percent, respectively.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VARIABLES | Loans | CI | Credit | NIM |
| Q1before | 0.004* | 0.001 | 0.001 | -0.000 |
| | (1.73) | (1.07) | (0.36) | (-0.25) |
| Qevent | -0.003 | -0.001 | 0.000 | -0.001 |
| | (-1.15) | (-0.65) | (0.13) | (-0.85) |
| Q1after | 0.002 | -0.000 | 0.008** | -0.000 |
| | (0.89) | (-0.22) | (2.22) | (-0.85) |
| Q2after | 0.003 | 0.001 | 0.005 | 0.001 |
| | (0.91) | (1.19) | (1.39) | (1.63) |
| Q3after | -0.001 | -0.000 | -0.000 | 0.000 |
| | (-0.55) | (-0.34) | (-0.03) | (0.45) |
| Q4after | 0.004 | -0.001 | 0.004 | -0.001* |
| | (1.07) | (-1.29) | (0.87) | (-1.71) |
| Unused Commitment Ratio | 0.004 | -0.003 | 0.010 | -0.001 |
| | (0.49) | (-0.95) | (0.82) | (-1.40) |
| Net Wholesale Funding | 0.053*** | 0.013*** | 0.042*** | 0.001 |
| | (8.01) | (3.35) | (5.92) | (1.40) |
| NPL to Loans | -0.315*** | -0.070*** | -0.369*** | -0.011*** |
| | (-7.01) | (-4.40) | (-6.66) | (-3.01) |
| Capital Ratio | 0.054 | 0.026 | 0.080 | 0.002 |
| | (1.17) | (1.10) | (1.39) | (0.26) |
| Real Estate Loan Share | -0.003 | -0.015*** | 0.009 | 0.001 |
| | (-0.28) | (-4.16) | (0.95) | (0.83) |
| Large Bank Indicator | 0.006 | 0.004*** | 0.009 | 0.001 |
| | (1.10) | (2.68) | (1.20) | (0.79) |
| Constant | 0.018* | 0.012*** | 0.004 | 0.007** |
| | (1.67) | (2.83) | (0.38) | (2.50) |
| Observations | 8,718 | 7,327 | 8,718 | 8,718 |
| R-squared | 0.2094 | 0.1483 | 0.2806 | 0.8862 |
| Bank FE | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES |

Table 8: Data Breach Announcements and Loans for Large Banks

The results show the large bank data breach announcement effect on loan growths. Columns (1) to (3) show results for different types of loan growth as dependent variables. *Loan* is quarterly growth of all types of loans. *CI* is the quarterly growth of commercial and industrial loans. *Credit* is quarterly growth of loans and commitments. Columns (4) is the result of the regression for Net Interest Margin (NIM) as dependent variables. Please see the full definition for each type of loan and other control variables in Appendix A. Q1before, Qevent, and Q1-4after, are a dummy variable equal to one if the interaction terms between a data breach bank dummy and a time dummy of a data breach announcement prior quarter, quarter, and 1-4 quarter after the announcement, respectively, are equal to one. *Largebank* is a dummy equal to one if banks' asset size is in the top five largest in a given quarter. ***,**,* are significant at 1, 5, 10 percent, respectively.

|  | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VARIABLES | Loans | CI | Credit | NIM |
| Q1before*Largebank | 0.004 | 0.001 | 0.001 | -0.000 |
|  | (1.19) | (0.67) | (0.22) | (-0.08) |
| Qevent*Largebank | 0.003 | 0.001 | 0.002 | -0.000 |
|  | (0.49) | (0.64) | (0.40) | (-0.07) |
| Q1after*Largebank | -0.005 | -0.000 | -0.002 | 0.001 |
|  | (-1.05) | (-0.14) | (-0.24) | (0.61) |
| Q2after*Largebank | 0.015*** | 0.003** | 0.010 | 0.000 |
|  | (3.00) | (2.23) | (1.40) | (0.03) |
| Q3after*Largebank | -0.004 | -0.000 | -0.001 | -0.001 |
|  | (-1.08) | (-0.43) | (-0.29) | (-0.88) |
| Q4after*Largebank | -0.009 | 0.001 | -0.008 | 0.001 |
|  | (-1.52) | (0.68) | (-1.02) | (1.29) |
| Observations | 8,718 | 7,327 | 8,718 | 8,718 |
| R-squared | 0.2101 | 0.1486 | 0.2809 | 0.8863 |
| Bank FE | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES |
| Controls | YES | YES | YES | YES |

Table 9: Data Breach Announcements and Moneyness

Dependent variable is the option moneyness, *Moneyness*. The option moneyness is based on the estimated strike price and per option realizable value following Core and Guay (2002) and Campbell et al. (2011). To find the estimated strike price, first, find total realizable value per share for exercisable option using data from ExecuComp (OPT_UNEX_EXER_EST_VAL/OPT_UNEX_EXER_NUM). Then, subtract the total realizable value per share for exercisable option from the stock price at the fiscal year end (PRCCF). The option moneyness is the per option realizable value divided by the estimated strike price. *Breached Bank Dummy* is equal to one if banks have a data breach announcement and zero otherwise. *Delta* is the change in option value when the underlying stock price changes by 1 percent. *Vega* is the change in option value when the stock volatility changes by 1 percent. *Insidedebt* is the sum of present value of accumulated pension benefits and deferred compensation. *Salary* is the annual salary. *Bonus* is the dollar value of a bonus earned during the fiscal year. *Stock* is value of stock-related awards that do not have option like features. ***,**,* are significant at 1, 5, 10 percent, respectively.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | Moneyness | Moneyness | Moneyness | Moneyness | Moneyness |
| Breached Bank Dummy | 123.7138*** | 120.9423*** | 120.9579*** | 120.9474*** | 104.4373*** |
| | (15.15) | (14.93) | (14.90) | (14.93) | (7.31) |
| Delta | | 0.0177** | | 0.0409 | 0.0390 |
| | | (2.17) | | (1.21) | (1.16) |
| Vega | | | 0.0551** | -0.1296 | -0.1584 |
| | | | (2.38) | (-0.87) | (-1.08) |
| Insidedebt | | | | | 0.0084 |
| | | | | | (0.51) |
| Salary | | | | | -0.0082 |
| | | | | | (-1.62) |
| Bonus | | | | | -0.0012 |
| | | | | | (-1.08) |
| Stock | | | | | 0.0014* |
| | | | | | (1.66) |
| Constant | 93.1021*** | 96.2244*** | 96.4683*** | 96.0527*** | 35.8404** |
| | (7.20) | (7.74) | (7.66) | (7.75) | (2.01) |
| Observations | 6,220 | 6,034 | 6,034 | 6,034 | 5,323 |
| R-squared | 0.5702 | 0.5048 | 0.5037 | 0.5056 | 0.5313 |
| Bank FE | YES | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES | YES |

Table 10: Data Breach Announcements and Deposits

The results show the data breach announcement effect on deposit growths. Dependent variables from column (1) to (5) are different types of quarterly deposit growth. *Deposits* quarterly growth of all types of deposits. *Core* is quarterly growth of transaction, saving, and time deposits less than $100,000. *Insured* is quarterly growth of insured deposits. *Brokered* is the quarterly growth of deposits from brokers. *Time* is the time deposit account. Please see the full definition for each type of deposit and other control variables in Appendix A. Q1before, Qevent, and Q1-4after, are a dummy variable equal to one if the interaction terms between a data breach bank dummy and a time dummy of a data breach announcement prior quarter, quarter, and 1-4 quarter after the announcement, respectively, are equal to one. ***,**,* are significant at 1, 5, 10 percent, respectively.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | Deposits | Core | Insured | Brokered | Time |
| Q1before | 0.001 | 0.001 | -0.001 | 0.001 | 0.001 |
| | (0.28) | (0.44) | (-0.58) | (1.24) | (0.37) |
| Qevent | -0.002 | -0.001 | -0.005** | 0.001 | -0.000 |
| | (-0.77) | (-0.22) | (-2.23) | (0.63) | (-0.18) |
| Q1after | -0.000 | 0.002 | -0.000 | -0.000 | -0.002 |
| | (-0.11) | (0.74) | (-0.19) | (-0.08) | (-1.01) |
| Q2after | 0.001 | 0.001 | 0.003 | 0.001 | -0.001 |
| | (0.33) | (0.52) | (0.79) | (0.75) | (-0.43) |
| Q3after | -0.004 | -0.002 | -0.003 | -0.001 | 0.000 |
| | (-1.35) | (-0.80) | (-1.07) | (-0.91) | (0.12) |
| Q4after | 0.002 | 0.001 | -0.002 | -0.003** | -0.001 |
| | (0.60) | (0.50) | (-0.81) | (-2.39) | (-0.37) |
| Unused Commitment Ratio | -0.008 | -0.010 | -0.008 | -0.005 | 0.012 |
| | (-0.58) | (-0.92) | (-0.78) | (-0.69) | (1.39) |
| Net Wholesale Funding | -0.024** | -0.033*** | 0.000 | 0.001 | 0.013** |
| | (-2.40) | (-3.85) | (0.09) | (0.28) | (2.04) |
| NPL to Loans | -0.214*** | -0.103* | -0.170*** | -0.049*** | -0.090*** |
| | (-4.31) | (-1.98) | (-4.53) | (-2.66) | (-2.84) |
| Capital Ratio | -0.147*** | -0.083*** | -0.012 | 0.000 | 0.031 |
| | (-3.79) | (-2.71) | (-0.33) | (0.03) | (1.01) |
| Large Bank Indicator | 0.012*** | 0.009*** | 0.007*** | 0.001 | 0.009*** |
| | (4.35) | (2.87) | (3.45) | (0.97) | (4.57) |
| Real Estate Loan Share | -0.019* | -0.003 | -0.007 | -0.006 | 0.001 |
| | (-1.87) | (-0.37) | (-1.39) | (-1.59) | (0.18) |
| Constant | 0.059*** | 0.033*** | 0.017** | 0.000 | -0.009* |
| | (5.08) | (3.93) | (2.15) | (0.01) | (-1.69) |
| Observations | 8,718 | 8,718 | 8,718 | 8,718 | 8,718 |
| R-squared | 0.1488 | 0.1434 | 0.2518 | 0.0843 | 0.3402 |
| Bank FE | YES | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES | YES |

Table 11: Data Breach and Deposit Flows

This table shows the results of deposit flows from breached banks to non-breached banks. The dependent variable is the insured deposit growth of non-data breached banks. *Breached Bank Insured Gr* is insured deposit growth of data breached banks. We first match the data breached banks with non breached banks by zip code. Then, calculate the distance in miles between breached and non-breached banks. Column (1) is for the distance of less than 5 miles between breached and non-breached banks. Column (2) is a distance of less than 10 miles. Column (3) is the distance between 10 to 30 miles. Column (4) is for the results if breached and non-breached banks are in the same state. The zip code (RSSD9220) and state (RSSD9200) information are retrieved from the Call Reports. Please see the full definition for the control variables in Appendix A. Q1before, Qevent, and Q1-4after, are a dummy variable equal to one if the interaction terms between a data breach bank dummy and a time dummy of a data breach announcement prior quarter, quarter, and 1-4 quarter after the announcement, respectively, are equal to one.$^{***,**,*}$ are significant at 1, 5, 10 percent, respectively.

|  | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VARIABLES | < 5 miles | < 10 miles | 10 to 30 | Same State |
| Q1before*Breached Bank Insured Gr | -0.175** | -0.042 | -0.017 | -0.022 |
|  | (-2.09) | (-1.28) | (-0.50) | (-1.48) |
| Qevent*Breached Bank Insured Gr | 0.251 | -0.173*** | -0.277* | -0.180*** |
|  | (1.04) | (-3.14) | (-1.91) | (-5.05) |
| Q1after*Breached Bank Insured Gr | -0.141 | -0.099** | 0.058 | -0.040* |
|  | (-1.68) | (-2.40) | (0.89) | (-1.92) |
| Q2after*Breached Bank Insured Gr | 0.014 | 0.053 | 0.002 | -0.005 |
|  | (0.13) | (0.92) | (0.05) | (-0.21) |
| Q3after*Breached Bank Insured Gr | -0.542** | -0.089 | -0.053 | 0.078** |
|  | (-2.38) | (-0.65) | (-0.66) | (2.13) |
| Q4after*Breached Bank Insured Gr | 0.079 | 0.206** | -0.056* | 0.016 |
|  | (0.83) | (2.54) | (-1.77) | (0.48) |
| Observations | 10,381 | 22,541 | 36,421 | 112,331 |
| R-squared | 0.3141 | 0.2901 | 0.3494 | 0.3167 |
| Bank FE | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES |
| Controls | YES | YES | YES | YES |

Table 12: Deposit Effects for Banks with Multiple Breach

The results show the effect of multiple breach announcements on insured deposit growth. Each column represents the order of the data breach announcement. 1st is the first announcement, 2nd is the second announcement, and so on. *BreachNumber* is a dummy variable equal to one when the data breach incident is announced based on the number of breach announcement. For instance, BreachNumber for column (1) is the dummy variable equal to one when it is the first data breach announcement. The maximum number of data breach announcements is 10 times from the same bank. Please see the full definition for each type of deposit and other control variables in Appendix A. Qevent (Q1,2,3,4after) is a dummy variable equal to one if the interaction terms between a data breach bank dummy and a time dummy of a data breach announcement quarter (1,2,3,4 quarters after a data breach announcement quarter) dummy is equal to one. ***,**,* are significant at 1, 5, 10 percent, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | 1st | 2nd | 3rd | 4th | 5th | 6th | 7th | 8th | 9th | 10th |
| Qevent*BreachNumber | -0.001 | -0.011* | 0.005 | 0.008 | -0.009 | 0.001 | -0.002 | -0.010 | 0.055*** | 0.028*** |
| | (-0.13) | (-1.97) | (0.65) | (0.84) | (-0.92) | (0.21) | (-0.51) | (-0.54) | (12.85) | (6.27) |
| Q1after*BreachNumber | 0.003 | 0.006 | -0.011* | -0.006 | -0.002 | -0.002 | 0.002 | 0.008 | 0.029*** | -0.048*** |
| | (0.49) | (0.82) | (-1.79) | (-0.77) | (-0.24) | (-0.53) | (0.29) | (0.93) | (7.38) | (-7.95) |
| Q2after*BreachNumber | -0.002 | -0.002 | 0.001 | 0.016* | -0.000 | 0.003 | 0.001 | 0.006 | -0.006** | 0.004 |
| | (-0.33) | (-0.40) | (0.12) | (1.75) | (-0.02) | (0.48) | (0.27) | (0.99) | (-2.31) | (0.88) |
| Q3after*BreachNumber | -0.007 | 0.014* | 0.008 | -0.041*** | -0.009* | 0.006 | 0.000 | -0.006 | -0.003 | 0.011** |
| | (-1.03) | (1.83) | (1.50) | (-3.02) | (-1.87) | (1.08) | (0.04) | (-0.54) | (-0.75) | (2.52) |
| Q4after*BreachNumber | 0.008 | -0.001 | -0.012** | 0.014* | 0.001 | -0.002 | -0.000 | -0.001 | 0.021*** | 0.004 |
| | (1.36) | (-0.10) | (-2.12) | (1.95) | (0.19) | (-0.31) | (-0.03) | (-0.20) | (4.61) | (0.82) |
| Observations | 8,718 | 8,718 | 8,718 | 8,718 | 8,718 | 8,718 | 8,718 | 8,718 | 8,718 | 8,718 |
| R-squared | 0.1490 | 0.1493 | 0.1493 | 0.1492 | 0.1489 | 0.1495 | 0.1491 | 0.1499 | 0.1503 | 0.1491 |
| Bank FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |

Table 13: Data Breach Announcements and Deposits by Types

The results show the effect of data breach announcements on different types of deposits by data breach types. HACK is a breach from hacking or malware. PORT is a breach from portable devices. STAT is a breach from stationary computer loss. *Deposits* quarterly growth of all types of deposits. *Core* is quarterly growth of transaction, saving, and time deposits less than $100,000. *Insured* is quarterly growth of insured deposits. *Brokered* is the quarterly growth of deposits from brokers. Please see the full definition for each type of deposit and other control variables in Appendix A. Q1before, Qevent, and Q1-4after, are a dummy variable equal to one if the interaction terms between a data breach bank dummy and a time dummy of a data breach announcement prior quarter, quarter, and 1-4 quarter after the announcement, respectively, are equal to one. ***,**,* are significant at 1, 5, 10 percent, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | HACK | HACK | HACK | HACK | PORT | PORT | PORT | PORT | STAT | STAT | STAT | STAT |
| VARIABLES | Deposits | Core | Insured | Brokered | Deposits | Core | Insured | Brokered | Deposits | Core | Insured | Brokered |
| Q1before | -0.014** | -0.005 | -0.010 | 0.000 | -0.001 | -0.003 | -0.003 | -0.001 | 0.024 | -0.002 | -0.006 | -0.017 |
| | (-2.37) | (-0.53) | (-0.92) | (0.10) | (-0.17) | (-0.69) | (-1.02) | (-0.38) | (0.82) | (-0.22) | (-1.27) | (-1.43) |
| Qevent | -0.008 | -0.005 | -0.015* | -0.001 | -0.002 | -0.008* | -0.004** | -0.001 | -0.018 | -0.003 | 0.001 | 0.012 |
| | (-1.19) | (-0.63) | (-1.73) | (-0.17) | (-0.34) | (-1.78) | (-2.05) | (-0.34) | (-0.52) | (-0.30) | (0.17) | (1.09) |
| Q1after | -0.003 | 0.000 | -0.011** | -0.002 | -0.010 | -0.009 | -0.004** | -0.004 | -0.030*** | 0.004 | -0.002 | -0.039*** |
| | (-0.39) | (0.03) | (-2.39) | (-0.27) | (-1.29) | (-1.58) | (-2.28) | (-1.35) | (-5.30) | (0.76) | (-0.27) | (-21.98) |
| Q2after | 0.010 | 0.007 | 0.040* | 0.008 | 0.005 | 0.002 | -0.002 | 0.000 | -0.031* | 0.013 | -0.004 | -0.023 |
| | (0.84) | (0.63) | (1.83) | (1.43) | (1.01) | (0.49) | (-0.81) | (0.10) | (-1.91) | (0.55) | (-0.54) | (-1.66) |
| Q3after | -0.020** | -0.018*** | -0.022* | -0.002 | 0.000 | -0.003 | -0.009** | -0.001 | -0.037 | -0.031 | 0.016*** | -0.009 |
| | (-2.52) | (-2.64) | (-1.96) | (-1.03) | (0.05) | (-0.35) | (-2.41) | (-0.98) | (-1.57) | (-0.99) | (3.41) | (-1.09) |
| Q4after | -0.001 | -0.000 | -0.008 | -0.014** | -0.012 | -0.007 | 0.000 | -0.003*** | 0.015*** | 0.012*** | 0.008*** | 0.000 |
| | (-0.11) | (-0.07) | (-0.68) | (-2.32) | (-1.41) | (-0.94) | (0.16) | (-2.92) | (2.84) | (2.66) | (2.69) | (0.18) |
| Observations | 5,115 | 5,115 | 5,115 | 5,115 | 5,452 | 5,452 | 5,452 | 5,452 | 4,670 | 4,670 | 4,670 | 4,670 |
| R-squared | 0.1480 | 0.1550 | 0.2763 | 0.0823 | 0.1529 | 0.1614 | 0.2806 | 0.0820 | 0.1597 | 0.1682 | 0.2853 | 0.0891 |
| Bank FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |

Table 14: Data Breach Announcements and Deposits by the Number of Records

The results show the data breach announcement effects on deposit growths for a data breach with high records (*Highrecords*). *Highrecords* is a dummy variable equal to one if the number of records is greater than the median of the number of records. Dependent variables from columns (1) to (5) are different types of quarterly deposit growth. *Deposits* quarterly growth of all types of deposits. *Core* is quarterly growth of transaction, saving, and time deposits less than $100,000. *Insured* is quarterly growth of insured deposits. *Brokered* is the quarterly growth of deposits from brokers. *Time* is the time deposit account. Please see the full definition for each type of deposit and other control variables in Appendix A. Q1before, Qevent, and Q1-4after, are a dummy variable equal to one if the interaction terms between a data breach bank dummy and a time dummy of a data breach announcement prior quarter, quarter, and 1-4 quarter after the announcement, respectively, are equal to one. ***,**,* are significant at 1, 5, 10 percent, respectively.

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | Deposits | Core | Insured | Brokered | Time |
| Q1before*Highrecords | 0.007 | -0.000 | -0.000 | -0.001 | -0.000 |
|  | (0.78) | (-0.07) | (-0.02) | (-0.47) | (-0.05) |
| Qevent*Highrecords | -0.002 | -0.005 | 0.004 | 0.002 | 0.002 |
|  | (-0.27) | (-1.25) | (1.04) | (0.86) | (0.47) |
| Q1after*Highrecords | 0.004 | 0.003 | -0.003 | -0.001 | -0.001 |
|  | (0.67) | (0.59) | (-0.83) | (-0.49) | (-0.36) |
| Q2after*Highrecords | -0.002 | 0.001 | 0.010 | -0.002 | -0.009** |
|  | (-0.40) | (0.20) | (1.05) | (-0.86) | (-2.50) |
| Q3after*Highrecords | -0.009 | -0.010 | -0.011* | -0.004* | -0.004 |
|  | (-1.26) | (-1.48) | (-1.76) | (-1.85) | (-1.27) |
| Q4after*Highrecords | -0.011 | -0.008 | 0.007 | -0.002 | -0.006 |
|  | (-1.56) | (-1.28) | (1.24) | (-0.67) | (-0.98) |
| Observations | 8,718 | 8,718 | 8,718 | 8,718 | 8,718 |
| R-squared | 0.1492 | 0.1437 | 0.2523 | 0.0845 | 0.3406 |
| Bank FE | YES | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES | YES |
| Controls | YES | YES | YES | YES | YES |

Table 15: Data Breach Announcements and Channel of Effects

*Treat* is a dummy variable equal to one for firms with a data breach announcement and zero otherwise. *AfterLaw* is a dummy variable equal to one when data breach announcements are after the disclosure law adoption date. We identify banks' location based on their zip code (RSSD9220) and state (RSSD9200) information from Call Reports. *Z-Score* indicates a bankruptcy probability. *MKTSHARE* is a firm's market share. *ROA* is return on asset. *ASSETGR* is asset growth. *LOGILL* logarithmic of illiquidity. *STDRET* is standard deviation of return. *CFO* is cash flow from operations. *LOSSCUS* is loss of major customers. Please see the full definition for each type of control variables in Appendix A. ***,**,* are significant at 1, 5, 10 percent, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| VARIABLES | Z-Score | MKTSHARE | ROA | ASSETGR | LOGILL | STDRET | CFO | LOSSCUS |
| Treat*AfterLaw | -0.001 | 0.001 | -0.000 | 0.002 | 0.028 | -0.007** | -0.005*** | -0.092 |
| | (-0.33) | (0.58) | (-0.26) | (1.06) | (1.45) | (-2.13) | (-4.73) | (-1.03) |
| Unused Commitment Ratio | 0.376*** | 0.003 | -0.008*** | 0.003 | 0.454*** | -0.068*** | 0.001 | -0.198 |
| | (40.92) | (0.56) | (-4.02) | (0.59) | (6.01) | (-5.06) | (0.43) | (-0.93) |
| Net Wholesale Funding | -0.018** | 0.023*** | 0.007*** | 0.023*** | -0.115 | -0.047*** | 0.016*** | -1.160*** |
| | (-2.49) | (5.66) | (4.46) | (5.73) | (-1.59) | (-3.68) | (4.63) | (-3.99) |
| NPL to Loans | -0.118** | -0.248*** | -0.091*** | -0.257*** | -2.674*** | 0.455*** | 0.179*** | 9.305*** |
| | (-2.23) | (-8.63) | (-8.26) | (-9.04) | (-6.01) | (5.78) | (8.42) | (5.87) |
| Capital Ratio | 1.028*** | -0.181*** | 0.002 | -0.178*** | -0.684* | -0.239*** | 0.080*** | -8.141*** |
| | (31.59) | (-9.64) | (0.26) | (-9.60) | (-1.86) | (-3.68) | (5.07) | (-7.66) |
| Large Bank Indicator | 0.073*** | 0.000 | 0.016*** | -0.000 | 0.116*** | -0.001 | -0.001 | -0.716** |
| | (7.57) | (0.06) | (8.04) | (-0.02) | (2.65) | (-0.14) | (-0.54) | (-2.45) |
| Real Estate Loan Share | -0.194*** | -0.021*** | 0.009*** | -0.018*** | 0.159* | -0.016 | 0.006 | -2.085*** |
| | (-19.79) | (-3.90) | (4.62) | (-3.34) | (1.65) | (-0.94) | (1.36) | (-5.51) |
| Constant | 0.061*** | 0.044*** | -0.012*** | 0.048*** | -0.745*** | 0.102*** | -0.058*** | 3.747*** |
| | (3.93) | (5.14) | (-3.84) | (5.63) | (-3.09) | (3.40) | (-6.98) | (6.74) |
| Observations | 8,890 | 8,718 | 8,890 | 8,718 | 4,161 | 4,159 | 4,753 | 1,122 |
| R-squared | 0.8193 | 0.4563 | 0.6066 | 0.1537 | 0.9602 | 0.5424 | 0.4283 | 0.6693 |
| Bank FE | YES | YES | YES | YES | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES | YES | YES | YES | YES |

Table 16: Cumulative Abnormal Returns (CARs) Results

Cumulative Abnormal Returns (CARs) following a data breach under the windows, (0,+1d), (0,+5d), (-1,+1d), and (-2,+1d). Panel A reports the CARs based on the CRSP value-weighted average returns for each data breach category. The categories of data breaches include CARD (Payment Card Fraud), HACK (Hacking or Malware), INSD (Insider), PHYS (Physical Loss), PORT (Portable Device), STAT (Stationary Device), DISC (Unintended Disclosure) and UNKN (Unknown). Panel B provides the Cumulative Abnormal Returns (CARs) based on quantiles of total records. The samples are divided into 4 groups based on the total records of data breaches. Q1 is the group with the highest breach records and Q4 is the group with the lowest breach records. Unknowns are the firms with an unknown number of breach records. Panel C shows the results for all available bank with Call Reports relative to the large bank samples. ***,**,* are significant at 1, 5, 10 percent, respectively.

Panel A: Full Sample

| | (0,+1d) | | | (0,+5d) | | | (-1,+1d) | | | (-2,+1d) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Median | N | Mean | Median | N | Mean | Median | N | Mean | Median | N |
| Overall | -0.49%*** | -0.28% | 171 | -0.83%* | -0.20% | 171 | -0.40%* | -0.16% | 172 | -0.52%** | -0.37% | 172 |
| | (-2.68) | | | (-1.95) | | | (-1.76) | | | (-2.06) | | |
| CARD | -1.35% | -0.10% | 14 | -0.45% | 0.42% | 14 | -1.32% | -0.12% | 14 | -1.64% | -0.41% | 14 |
| | (-1.26) | | | (-0.33) | | | (-1.14) | | | (-1.33) | | |
| DISC | -0.19% | 0.00% | 27 | -0.31% | -0.69% | 27 | -0.43% | -0.40% | 27 | 0.04% | -0.06% | 27 |
| | (-0.47) | | | (-0.50) | | | (-1.01) | | | (0.09) | | |
| HACK | -0.98%* | -0.37% | 36 | -2.35% | -0.18% | 36 | -0.97% | -0.70% | 36 | -1.33%* | -0.84% | 36 |
| | (-1.93) | | | (-1.40) | | | (-1.41) | | | (-1.90) | | |
| INSD | 0.35% | 0.44% | 23 | 0.21% | 0.36% | 23 | 0.40% | 0.80% | 23 | 0.40% | 0.65% | 23 |
| | (1.08) | | | (0.28) | | | (0.85) | | | (1.01) | | |
| PHYS | -0.89% | -0.18% | 9 | -1.09% | 0.56% | 9 | -1.42%** | -0.26% | 9 | -1.51%** | -0.76% | 9 |
| | (-1.14) | | | (-1.09) | | | (-2.07) | | | (-2.16) | | |
| PORT | -0.48%* | -0.39% | 44 | -0.72% | -0.35% | 44 | -0.28% | -0.36% | 44 | -0.48% | -0.67% | 44 |
| | (-1.89) | | | (-1.34) | | | (-1.07) | | | (-1.14) | | |
| STAT | -0.67%** | -0.79% | 4 | -0.58% | -0.56% | 4 | 1.19% | -0.06% | 5 | 1.62% | 0.39% | 5 |
| | (-2.05) | | | (-1.17) | | | (0.74) | | | (1.20) | | |
| UNKN | -0.06% | -0.57% | 14 | -0.27% | -0.07% | 14 | 0.42% | 0.05% | 14 | -0.13% | -0.07% | 14 |
| | (-0.09) | | | (-0.48) | | | (0.42) | | | (-0.13) | | |

Panel B: Quantiles of Total Records

| | (0,+1d) | | | (0,+5d) | | | (-1,+1d) | | | (-2,+1d) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Median | N | Mean | Median | N | Mean | Median | N | Mean | Median | N |
| Overall | -0.49%*** | -0.28% | 171 | -0.83%* | -0.20% | 171 | -0.40%* | -0.16% | 172 | -0.52%** | -0.37% | 172 |
| | (-2.68) | | | (-1.95) | | | (-1.76) | | | (-2.06) | | |
| Q1 | -1.76%** | -0.10% | 24 | -4.55%* | -0.39% | 24 | -1.66%* | 0.17% | 24 | -1.72% | -0.40% | 24 |
| | (-2.12) | | | (-1.93) | | | (-1.64) | | | (-1.54) | | |
| Q2 | -0.61%* | -0.37% | 24 | -1.14%* | -0.83% | 24 | -0.70% | -0.98% | 24 | -0.85%* | -1.04% | 24 |
| | (-1.67) | | | (-1.69) | | | (-1.55) | | | (-1.76) | | |
| Q3 | -0.29% | 0.03% | 25 | 0.75% | 0.60% | 25 | 0.06% | 0.05% | 25 | 0.13% | -0.23% | 25 |
| | (-0.61) | | | (0.75) | | | (0.10) | | | (0.21) | | |
| Q4 | 0.27% | 0.24% | 24 | 0.00% | 0.08% | 24 | 0.66% | 0.20% | 24 | 0.16% | 0.34% | 24 |
| | (0.72) | | | (-0.00) | | | (1.33) | | | (0.29) | | |
| Unknown | -0.35% | -0.39% | 74 | -0.32% | 0.13% | 74 | -0.40% | -0.34% | 75 | -0.46% | -0.48% | 75 |
| | (-1.62) | | | (-0.89) | | | (-1.38) | | | (-1.41) | | |

Panel C: Large Bank Sample

| | (0,+1d) | | | (0,+5d) | | | (-1,+1d) | | | (-2,+1d) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Median | N | Mean | Median | N | Mean | Median | N | Mean | Median | N |
| All Banks | -0.33%** | -0.18% | 91 | -1.26%** | 0.12% | 91 | -0.37%** | -0.13% | 91 | -0.36%* | -0.25% | 91 |
| | (-2.07) | | | (-2.13) | | | (-2.19) | | | (-1.85) | | |
| Large Banks | -0.39%* | -0.18% | 25 | -0.30% | 0.17% | 25 | -0.34%* | -0.47% | 25 | -0.38% | -0.28% | 25 |
| | (-1.86) | | | (-0.72) | | | (-1.84) | | | (-1.56) | | |

Table 17: Multivariate Analysis of Cumulative Abnormal Returns (CARs)

The dependent variables are different time windows for cumulative abnormal returns (CARs) on the data breach announcement date. 0 is the event date. For example, 0 to 1d is the window between the event date and one day after. -30 to -2d is the window between 30 days before the event date and 2 days after the event date. Please see the full definition for each control variables in Appendix A. ***,**,* are significant at 1, 5, 10 percent, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| VARIABLES | 0 to 1d | 0 to 5d | -1 to 1d | -2 to 1d | -30 to -2d | 0 to 90d |
| Firm Size | -0.002 | -0.001 | -0.001 | -0.000 | 0.014*** | 0.009 |
| | (-1.17) | (-0.70) | (-0.83) | (-0.17) | (2.76) | (0.83) |
| ROA | 0.148 | 0.091 | 0.298* | -0.051 | -0.506 | -0.629 |
| | (1.04) | (0.42) | (1.77) | (-0.26) | (-0.98) | (-0.55) |
| Leverage | -0.000 | -0.003 | -0.001 | 0.002 | -0.002 | -0.009 |
| | (-0.30) | (-1.22) | (-0.28) | (0.71) | (-0.31) | (-0.64) |
| Z-score | -0.043* | -0.061 | -0.017 | 0.027 | 0.168* | 0.156 |
| | (-1.72) | (-1.62) | (-0.58) | (0.80) | (1.87) | (0.79) |
| Unused Commitment Ratio | 0.011 | 0.008 | 0.044 | 0.065 | -0.010 | 0.058 |
| | (0.37) | (0.17) | (1.25) | (1.60) | (-0.10) | (0.25) |
| Net Wholesale Funding | 0.018 | 0.027 | -0.013 | -0.041* | -0.153*** | -0.190 |
| | (1.22) | (1.16) | (-0.73) | (-1.97) | (-2.80) | (-1.57) |
| NPL to Loans | 0.115 | 0.652 | 0.640 | 1.463*** | 2.935** | 5.501* |
| | (0.33) | (1.22) | (1.54) | (3.04) | (2.32) | (1.97) |
| Capital Ratio | -0.177 | -0.196 | -0.171 | -0.246 | -0.184 | -0.983 |
| | (-1.05) | (-0.76) | (-0.86) | (-1.06) | (-0.30) | (-0.73) |
| Large Bank Indicator | -0.009 | 0.001 | -0.018** | -0.019* | 0.020 | -0.060 |
| | (-1.28) | (0.07) | (-2.13) | (-1.90) | (0.75) | (-1.03) |
| Real Estate Loan Share | -0.015 | -0.002 | -0.003 | 0.018 | 0.106** | 0.183* |
| | (-1.22) | (-0.13) | (-0.21) | (1.07) | (2.41) | (1.89) |
| Constant | 0.063 | 0.058 | 0.041 | 0.003 | -0.318** | -0.217 |
| | (1.56) | (0.95) | (0.85) | (0.06) | (-2.17) | (-0.67) |
| Observations | 86 | 86 | 86 | 86 | 86 | 86 |
| R-squared | 0.4625 | 0.4862 | 0.3887 | 0.4367 | 0.5838 | 0.4640 |
| District FE | YES | YES | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES | YES | YES |

Table 18: Difference-in-Difference Analysis with Breach Disclosure Law

*Treat* is a dummy variable equal to one for firms with a data breach announcement and zero otherwise. *AfterLaw* is a dummy variable equal to one when data breach announcements are after the adoption date. We identify banks' location based on their zip code (RSSD9220) and state (RSSD9200) information from Call Reports. Dependent variables from columns (1) to (5) are different types of quarterly deposit growth. *Deposits* quarterly growth of all types of deposits. *Core* is quarterly growth of transaction, saving, and time deposits less than \$100,000. *Insured* is quarterly growth of insured deposits. *Brokered* is the quarterly growth of deposits from brokers. *Time* is the time deposit account. Please see the full definition for each type of deposit and other control variables in Appendix A. ***,**,* are significant at 1, 5, 10 percent, respectively.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | Deposits | Core | Insured | Brokered | Time |
| Treat*AfterLaw | -0.000 | -0.004** | -0.002 | -0.000 | 0.001 |
| | (-0.13) | (-2.30) | (-1.40) | (-0.36) | (0.70) |
| Unused Commitment Ratio | -0.008 | -0.010** | -0.008** | -0.005*** | 0.012*** |
| | (-1.56) | (-2.22) | (-1.99) | (-2.91) | (4.11) |
| Net Wholesale Funding | -0.024*** | -0.033*** | 0.000 | 0.001 | 0.013*** |
| | (-6.09) | (-9.64) | (0.08) | (0.86) | (5.74) |
| NPL to Loans | -0.214*** | -0.101*** | -0.169*** | -0.048*** | -0.091*** |
| | (-7.63) | (-4.07) | (-7.59) | (-4.63) | (-5.60) |
| Capital Ratio | -0.147*** | -0.083*** | -0.012 | 0.000 | 0.031*** |
| | (-8.06) | (-5.14) | (-0.82) | (0.03) | (2.91) |
| Large Bank Indicator | 0.011** | 0.009** | 0.007 | 0.001 | 0.009*** |
| | (2.22) | (2.00) | (1.59) | (0.70) | (3.08) |
| Real Estate Loan Share | -0.019*** | -0.003 | -0.008* | -0.006*** | 0.001 |
| | (-3.64) | (-0.61) | (-1.85) | (-2.90) | (0.26) |
| Constant | 0.059*** | 0.033*** | 0.017*** | 0.000 | -0.009* |
| | (7.12) | (4.51) | (2.63) | (0.03) | (-1.79) |
| Observations | 8,718 | 8,718 | 8,718 | 8,718 | 8,718 |
| R-squared | 0.1487 | 0.1438 | 0.2515 | 0.0835 | 0.3402 |
| Bank FE | YES | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES | YES |

Table 19: Data Breach Announcements and Deposits with Cumulative Breach

The results show the robustness test of data breach announcement effect on deposit growths by adding cumulative number of breaches, *Cumulative Breach*, as a control variable. Dependent variables from columns (1) to (5) are different types of quarterly deposit growth. *Deposits* quarterly growth of all types of deposits. *Core* is quarterly growth of transaction, saving, and time deposits less than $100,000. *Insured* is quarterly growth of insured deposits. *Brokered* is the quarterly growth of deposits from brokers. *Time* is the time deposit account. Please see the full definition for each type of deposit and other control variables in Appendix A. Q1before, Qevent, and Q1-4after, are a dummy variable equal to one if the interaction terms between a data breach bank dummy and a time dummy of a data breach announcement prior quarter, quarter, and 1-4 quarter after the announcement, respectively, are equal to one. ***,**,* are significant at 1, 5, 10 percent, respectively.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | Deposits | Core | Insured | Brokered | Time |
| Q1before | 0.001 | 0.001 | -0.001 | 0.002 | 0.001 |
| | (0.27) | (0.40) | (-0.54) | (1.25) | (0.44) |
| Qevent | -0.002 | -0.001 | -0.005** | 0.001 | -0.000 |
| | (-0.79) | (-0.24) | (-2.19) | (0.64) | (-0.15) |
| Q1after | -0.000 | 0.002 | -0.000 | -0.000 | -0.002 |
| | (-0.12) | (0.72) | (-0.16) | (-0.08) | (-0.98) |
| Q2after | 0.001 | 0.001 | 0.003 | 0.001 | -0.001 |
| | (0.33) | (0.52) | (0.79) | (0.75) | (-0.43) |
| Q3after | -0.004 | -0.002 | -0.003 | -0.001 | 0.000 |
| | (-1.36) | (-0.81) | (-1.04) | (-0.90) | (0.14) |
| Q4after | 0.002 | 0.001 | -0.002 | -0.003** | -0.001 |
| | (0.60) | (0.50) | (-0.82) | (-2.40) | (-0.38) |
| Cumulative Breach | -0.000 | -0.000 | 0.000 | 0.000 | 0.000 |
| | (-0.29) | (-0.47) | (0.66) | (0.29) | (0.82) |
| Unused Commitment Ratio | -0.007 | -0.009 | -0.009 | -0.005 | 0.011 |
| | (-0.52) | (-0.83) | (-0.89) | (-0.74) | (1.35) |
| Net Wholesale Funding | -0.024** | -0.033*** | 0.001 | 0.001 | 0.013** |
| | (-2.39) | (-3.87) | (0.13) | (0.29) | (2.00) |
| NPL to Loans | -0.213*** | -0.102* | -0.172*** | -0.049*** | -0.092*** |
| | (-4.26) | (-1.94) | (-4.57) | (-2.66) | (-2.90) |
| Capital Ratio | -0.147*** | -0.083*** | -0.012 | 0.000 | 0.030 |
| | (-3.75) | (-2.66) | (-0.35) | (0.02) | (0.97) |
| Large Bank Indicator | 0.012*** | 0.009*** | 0.006*** | 0.001 | 0.009*** |
| | (4.35) | (2.97) | (3.24) | (0.91) | (4.08) |
| Real Estate Loan Share | -0.020* | -0.003 | -0.006 | -0.005 | 0.002 |
| | (-1.83) | (-0.47) | (-1.18) | (-1.49) | (0.36) |
| Constant | 0.060*** | 0.033*** | 0.016** | -0.000 | -0.009* |
| | (5.06) | (4.00) | (2.10) | (-0.01) | (-1.77) |
| Observations | 8,718 | 8,718 | 8,718 | 8,718 | 8,718 |
| R-squared | 0.1488 | 0.1434 | 0.2519 | 0.0843 | 0.3404 |
| Bank FE | YES | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES | YES |
| Qtr FE | YES | YES | YES | YES | YES |

# References

Acharya, V. V., Mora, N., 2015. A crisis of banks as liquidity providers. The Journal of Finance 70, 1–43.

Allen, F., Gale, D., 2000. Financial contagion. Journal of Political Economy 108, 1–33.

Allen, L., Saunders, A., 1992. Bank window dressing: Theory and evidence. Journal of Banking & Finance 16, 585–623.

Arifovic, J., Jiang, J. H., 2019. Strategic uncertainty and the power of extrinsic signals–evidence from an experimental study of bank runs. Journal of Economic Behavior & Organization 167, 1 – 17.

Avery, R. B., Berger, A. N., 1991. Loan commitments and bank risk exposure. Journal of Banking & Finance 15, 173–192.

Baur, D. G., 2012. Financial contagion and the real economy. Journal of Banking & Finance 36, 2680–2692.

Berger, A. N., Curti, F., Mihov, A., Sedunov, J., 2022. Operational risk is more systemic than you think: Evidence from u.s. bank holding companies. Journal of Banking & Finance 143, 106619.

Berger, A. N., DeYoung, R., Flannery, M. J., Lee, D., Öztekin, Ö., 2008. How do large banking organizations manage their capital ratios? Journal of Financial Services Research 34, 123–149.

Brusco, S., Castiglionesi, F., 2007. Liquidity coinsurance, moral hazard, and financial contagion. The Journal of Finance 62, 2275–2302.

Bryant, J., 1980. A model of reserves, bank runs, and deposit insurance. Journal of Banking & Finance 4, 335 – 344.

Campbell, K., Gordon, L., Loeb, M., Zhou, L., 2003. The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. Journal of Computer Security 11, 431–448.

Campbell, T. C., Gallmeyer, M., Johnson, S. A., Rutherford, J., Stanley, B. W., 2011. Ceo optimism and forced turnover. Journal of Financial Economics 101, 695–712.

Chen, W.-D., Chen, Y., Huang, S.-C., 2021. Liquidity risk and bank performance during financial crises. Journal of Financial Stability 56, 100906.

Chernobai, A., Ozdagli, A., Wang, J., 2021. Business complexity and risk management: Evidence from operational risk events in u.s. bank holding companies. Journal of Monetary Economics 117, 418–440.

Cont, R., Schaanning, E., 2019. Monitoring indirect contagion. Journal of Banking & Finance 104, 85–102.

Core, J., Guay, W., 2002. Estimating the value of employee stock option portfolios and their sensitivities to price and volatility. Journal of Accounting Research 40, 613–630.

Curti, F., Gerlach, J., Kazinnik, S., Lee, M., Mihov, A., 2006. Cyber risk definition and classification for financial risk management. Journal of Operational Risk 1755, 2710.

DeYoung, R., Peng, E. Y., Yan, M., 2013. Executive compensation and business policy choices at us commercial banks. Journal of Financial and Quantitative Analysis 48, 165–196.

Diamond, D. W., Dybvig, P. H., 1983. Bank runs, deposit insurance, and liquidity. Journal of Political Economy 91, 401–419.

Diamond, D. W., Rajan, R. G., 2000. A theory of bank capital. The Journal of Finance 55, 2431–2465.

Diamond, D. W., Rajan, R. G., 2001. Liquidity risk, liquidity creation, and financial fragility: A theory of banking. Journal of Political Economy 109, 287–327.

Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. The Journal of Finance 64, 2023–2052.

Foerderer, J., Schuetz, S. W., 2022. Data breach announcements and stock market reactions: a matter of timing? Management Science 68, 7298–7322.

Garcia, L., Lewrick, U., Sečnik, T., 2021. Is window dressing by banks systemically important? BIS Working Papers .

Gatev, E., Schuermann, T., Strahan, P. E., 2009. Managing bank liquidity risk: How deposit-loan synergies vary with market conditions. The Review of Financial Studies 22, 995–1020.

Gopalan, R., Kadan, O., Pevzner, M., 2012. Asset liquidity and stock liquidity. Journal of Financial and Quantitative Analysis 47, 333–364.

Ho, P.-H., Huang, C.-W., Lin, C.-Y., Yen, J.-F., 2016. Ceo overconfidence and financial crisis: Evidence from bank lending and leverage. Journal of Financial Economics 120, 194–209.

Huang, H. H., Wang, C., 2021. Do banks price firms' data breaches? The Accounting Review 96, 261–286.

Imbierowicz, B., Rauch, C., 2014. The relationship between liquidity risk and credit risk in banks. Journal of Banking & Finance 40, 242–256.

Iyer, S. R., Simkins, B. J., Wang, H., 2020. Cyberattacks and impact on bond valuation. Finance Research Letters 33, 101215.

Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., Stulz, R. M., 2021. Risk management, firm reputation, and the impact of successful cyberattacks on target firms. Journal of Financial Economics 139, 719–749.

Karas, A., Pyle, W., Schoors, K., 2013. Deposit insurance, banking crises, and market discipline: Evidence from a natural experiment on deposit flows and rates. Journal of Money, Credit and banking 45, 179–200.

Kim, M., Kliger, D., Vale, B., 2003. Estimating switching costs: The case of banking. Journal of Financial Intermediation 12, 25–56.

Kishan, R. P., Opiela, T. P., 2000. Bank size, bank capital, and the bank lending channel. Journal of Money, credit and banking pp. 121–141.

Lending, C., Minnick, K., Schorno, P. J., 2018. Corporate governance, social responsibility, and data breaches. Financial Review 53, 413–455.

Mikhed, V., Vogan, M., 2018. How data breaches affect consumer credit. Journal of Banking & Finance 88, 192 – 207.

Park, S., Peristiani, S., 1998. Market discipline by thrift depositors. Journal of Money, Credit and Banking pp. 347–364.

Peng, J., Zhang, H., Mao, J., Xu, S., 2023. Repeated data breaches and firm value. Economics Letters 224, 111001.

Peria, M., Soledad, M., Schmukler, S. L., 2001. Do depositors punish banks for bad behavior? market discipline, deposit insurance, and banking crises. The Journal of Finance 56, 1029–1051.

Pérignon, C., Thesmar, D., Vuillemey, G., 2018. Wholesale funding dry-ups. The Journal of Finance 73, 575–617.

Piccotti, L. R., Wang, H., 2022. Informed trading in the options market surrounding data breaches. Global Finance Journal p. 100774.

Romanosky, S., Telang, R., Acquisti, A., 2011. Do data breach disclosure laws reduce identity theft? Journal of Policy Analysis and Management 30, 256–286.

Sharpe, S. A., 1990. Asymmetric information, bank lending and implicit contracts: A stylized model of customer relationships. The Journal of Finance 45, 1069–1087.

Sharpe, S. A., 1997. The effect of consumer switching costs on prices: A theory and its application to the bank deposit market. Review of Industrial Organization 12, 79–94.

Shin, H. S., 2009. Reflections on northern rock: The bank run that heralded the global financial crisis. Journal of economic perspectives 23, 101–119.

Spanos, G., Angelis, L., 2016. The impact of information security events to the stock market: A systematic literature review. Computers & Security 58, 216–229.

Stuart, E. A., Rubin, D. B., 2008. Best practices in quasi-experimental designs. Best practices in quantitative methods pp. 155–176.

Vesala, T., 2007. Switching costs and relationship profits in bank lending. Journal of Banking & Finance 31, 477–493.

Wang, H. E., Wang, Q. E., Wu, W., 2022. Short selling surrounding data breach announcements. Finance Research Letters p. 102690.

# Appendix A. Dependent Variables and Controls

*Unless otherwise noted, all variables are from Federal Reserve Call reports.

| Variables | Description |
|---|---|
| Quarterly growth of deposits | Quarterly change in deposits divided by assets, $\frac{\Delta Deposits_t}{Assets_{t-1}}$ |
| Quarterly growth of core deposits | Quarterly change in core deposits divided by assets. Core deposits are the sum of transaction deposits, saving deposits, and time deposits less than \$100,000, $\frac{\Delta CoreDeposits_t}{Assets_{t-1}}$ |
| Quarterly growth of insured deposits | Quarterly change in insured deposits divided by assets. Insured deposits are calculated from the accounts of \$100,000 or less before 2009Q3. After 2009Q3, the insured amount increased to \$250,000, $\frac{\Delta InsuredDeposits_t}{Assets_{t-1}}$ |
| Quarterly growth of brokered deposits | Quarterly change in brokered deposits divided by assets, $\frac{\Delta BrokeredDeposits_t}{Assets_{t-1}}$ |
| Quarterly growth of time deposits | Quarterly change in time deposits divided by assets, $\frac{\Delta TimeDeposits_t}{Assets_{t-1}}$ |
| Quarterly growth of loans | Quarterly change in loans divided by assets, $\frac{\Delta Loans_t}{Assets_{t-1}}$ |
| Quarterly growth of credit (loans+commitments) | Quarterly change in credit divided by asset size and unused commitment. Credit is the sum of loans and unused commitments, $\frac{\Delta Credit_t}{(Assets+Committment)_{t-1}}$ |
| Quarterly growth of CI loans | Quarterly change in commercial and industrial loans divided by asset size, $\frac{\Delta CILoans_t}{Assets_{t-1}}$ |
| Unused commitment ratio | Unused commitment divided by the sum of unused commitments and loans, $\frac{Committment_t}{(Loans+Committment)_t}$ |
| Capital Ratio | Book capital to asset size, $\frac{Bookcapital_t}{(Assets)_t}$ |
| Real estate loan share | Real estate loans divided by total loans, $\frac{Realestateloans_t}{(Loans)_t}$ |

| Variables | Description |
|---|---|
| Net wholesale funding | Wholesale funds less liquid assets to total assets. Wholesale funds are the sum of large time deposits, deposits booked in foreign offices, subordinated debt and debentures, gross federal funds purchased, repos, and other borrowed money. Liquid assets are cash, federal funds sold and reverse repos, and securities excluding MBS/ABS securities, $\frac{(Wholesalefunds-Liquidassets)_t}{(Assets)_t}$ |
| Large bank indicator | Dummy variable equal to one for the top five largest banks by asset size in each quarter |
| Firm Size | Natural log of total assets |
| Leverage | Debt divided by total assets |
| Tangibility | Gross PPE divided by total assets |
| Z-Score | (1.2× working capital+1.4× retained earnings+3.3× income before extraordinary items+0.999× sales)/sales |
| ROA | Net income divided by sales |
| Loss customer (LOSSCUS) | Dummy variable equal to one if a firm loses at least one of its major customers (Source: COMPUSTAT) |
| Market Share Growth (MKTSHARE) | Change in annual market share based on the size of the assets to total assets of all banks in each quarter. |
| CFO | Operating cash flow divided by total assets (Source: COMPUSTAT) |
| Asset growth (ASSETGR) | Quarterly growth of assets |
| Log of illiquidity (LOGILL) | Natural logarithm of the stock illiquidity measure from Gopalan, Kadan, and Pevzner (2012) (Source: CRSP) |
| Standard deviation of returns (STDRET) | Standard deviation of stock returns next 12 months (Source: CRSP) |

# Appendix B. Data Breach Announcement Examples

| Date Made Public | Company | City | State | Type of breach | Total Records | Description of incident |
|---|---|---|---|---|---|---|
| 17-Jan-18 | Ameriprise Financial, Inc. | Minneapolis | Minnesota | DISC | 56 | Ameriprise Financial suffered an inadvertent disclosure of 56 records, including SS numbers and names |
| 5-Feb-18 | 1st Mariner Bank | Baltimore | Maryland | HACK | 1500 | 1st Mariner Bank experienced a phishing attack that resulted in the exposure of the records of 1500 persons. Information exposed included Social Security Numbers, as well as names in combination with credit card or financial account information. |
| 26-Feb-18 | Southern National Bancorp of Virginia, Inc | Glen Allen | Virginia | HACK | 24999 | Southern National Bancorp of Virginia suffered a breach affecting 24,999 records, including social security numbers, driver's license number or non-driver identification card numbers, as well as financial account numbers or credit card numbers, in combination with the security code, access code, password or PIN for the account |
| 20-Apr-18 | SunTrust Banks, Inc. | Atlanta | Georgia | HACK | 1500000 | SunTrust Banks Inc. said an employee may have stolen the information of about 1.5 million customers and provided it to a criminal third party, the latest example of a potential breach that underscores the vulnerability of consumers, private data. The Atlanta-based bank on Friday said the employee, who no longer works at SunTrust, attempted to access customer information, although it has not identified significant fraudulent activity around the accounts involved |

# Appendix C. Variable Description (Call Reports)

Bank level data are from the quarterly Call Reports. We aggregate banks to top holder level (RSSD9348) when RSSD9348 is available. If RSSD9348 is not available for any bank, all variables are calculated at the bank level. The sample excludes non-US banks. To control for the merger effect following Acharya and Mora (2015), we also exclude banks with more than 10 percent growth in assets from the previous quarter. All the growth rates are the quarterly change divided by the beginning period assets. For the assets, we use RCFD2170 for bank holding level and RCON2170+RCFN2170 for bank level. All growth rates are winsorized at 1% tails.

When we download Call Report variables in the balance sheet focusing on bank holding level (RCFD), many variables are missing after 2011. To recover the missing information, we first need to understand how a variable is calculated. First, we need to understand that there are three types of Call Reports:[31]

1. FFIEC 031: Banks with domestic and foreign offices
2. FFIEC 041: Banks with domestic offices only
3. FFIEC 051: Banks with domestic offices only and total assets less than $5 billion

For example, if we need *total assets* variable from Call Reports, we can call RCFD2170. If RCFD2170 is missing, we can also call RCON2170. RCFD2170 refers to information from FFIEC 031. To access the information from FFIEC 041 and 051, we will call RCON 2170 for total assets. The prefix RCFD is specifically for FFIEC 031. It refers to banks with domestic and foreign offices. If the variable is still missing, we trace back how the variable is calculated. For example, total assets (RCFD2170) is calculated from the sum of domestic assets (RCON2170) and foreign assets (RCFN2170).

### Dependent and control variables

[31] https://www.ffiec.gov/ffiec_report_forms.htm

- Quarterly growth of deposits: RCFD2200. If RCFD2200 is missing, deposits are the sum of interest-bearing deposits (RCON6636) and non-interest bearing deposits (RCON6631).
- Quarterly growth of core deposits

  Core deposits are the sum of transaction deposits, saving deposits, and time deposits less than $100,000. RCON2215+RCON6810+RCON0352+RCON6648
- Quarterly growth of insured deposits

  The threshold for insured deposit was $100,000 or less for any account until 2009Q3 when the amount was increased to $250,000. Note retirement account increased the amount to $250,000 earlier in 2006Q2. Insured deposits before 2006Q2: RCONF049. Since 2006Q2, insured deposits: RCONF049+RCONF045
- Quarterly growth of brokered deposits

  Deposits received from brokers and dealers: RCON2365
- Quarterly growth of time deposits

  Deposits from time deposit accounts less than $100,000 and more than $100,000: RCON6648+RCON2604
- Quarterly growth of loans

  Loans are RCFD1400. If RCFD1400 is missing, we use RCON1400. If RCON1400 is missing, we use RCFD2122 - RCFD3123 which is total loans and leases held for investment and held for sale less allowance for loan and lease losses. If RCFD3123 is missing, we use RCFD2122. Next, if Loans variable is still missing, we use RCON2122. If RCON2122 is missing, we then use RCFDB528+RCFD5369 which are loans and leases held for investment and held for sale.

  Commercial and industrial (CI) loans are RCFD1766. If RCFD1766 is missing, we use RCON1766. If RCON1766 is missing, we use the sum of RCFD1763 and RCFD1764 which are CI loans to US addresses (domicile) and non-US addresses (domicile). Lastly, if CI loans variable is still missing, we use the sum of RCON1763 and RCON1764.

- Quarterly growth of credit

  Credits are the sum of loans (RCFD1400) and unused commitments (RCFD3814+RCFD3816+RCFD3817+RCFD3818+RCFD6550+RCFD3411). In this case, the denominator of the growth rate is the sum of beginning period assets and commitment.

- Unused commitment ratio

  Unused commitment consists of RCFD3814 +RCFD3816 +RCFD3817 +RCFD3818 +RCFD6550 +RCFD3411. If the variable is missing, we use RCON3814 +RCON3816 +RCON3817 +RCON3818 +RCON6550 +RCON3411

  Unused commitment ratio is unused commitments divided by the sum of unused commitments and loans.

- Net wholesale funding to asset ratio

  The ratio of wholesale funds (excludes liquid assets) to total assets

  Wholesale funds are the sum of deposits booked in foreign offices, large time deposits, subordinated debt and debentures, gross federal funds purchased, repos, and other borrowed money: RCFN2200 + RCON2604 + RCFD3200 + RCFD2800 (From 2002Q1: RCONB993+RCFDB995) + RCFD3190.

  If RCFD3200 is missing, we set it as RCON3200. If RCFD2800 is missing, we use RCON2800. If RCON2800 is missing, we use RCONB993+RCONB995. If RCFD3190 is missing, we use RCON3190.

  Liquid assets are the sum of cash (RCFD0010), federal funds sold and reverse repos (Before 2002Q1: RCFD1350, From 2002Q1: RCONB987 +RCFDB989), and securities excluding MBS/ABS securities (Before 2009Q2: RCFD1754 +RCFD1773 – (RCFD8500 +RCFD8504 +RCFDC026 +RCFD8503 +RCFD8507 +RCFDC027), From 2009Q2: RCFD1754 +RCFD1773 – (RCFDG300 +RCFDG304 +RCFDG308 +RCFDG312 + RCFDG316 + RCFDG320 +RCFDG324 +RCFDG328 +RCFDC026 +RCFDG336 +RCFDG340 +RCFDG344 +RCFDG303 +RCFDG307 +RCFDG311 +RCFDG315

+RCFDG319 +RCFDG323 +RCFDG327 +RCFDG331 +RCFDC027 +RCFDG339 +RCFDG343 +RCFDG347).

To maximize the availability of liquid assets, we use alternative IDs for each variable used to calculate liquid assets. For cash, if RCFD0010 is missing, we use RCON0010. For securities excluding MBS/ABS securities before 2009Q2, we use (RCON1754 + RCON1773) - (RCON8500 + RCON8504 + RCONC026 + RCON8503 + RCON8507 + RCONC027).

For 2009Q2 and after, we use (RCON1754 + RCON1773) - (RCONG300 + RCONG304 + RCONG308 + RCONG312 + RCONG316 + RCONG320 + RCONG324 + RCONG328 + RCONC026 + RCONG336 + RCONG340 + RCONG344 + RCONG303 + RCONG307 + RCONG311 + RCONG315 + RCONG319 + RCONG323 + RCONG327 + RCONG331 + RCONC027 + RCONG339 + RCONG343 + RCONG347)

- Nonperforming loans to loans

  Nonperforming loans (NPL) are loans past due 90 days or more and non-accruals: RCFD1407 + RCFD1403. If NPL is missing, we use RCON1407+ RCON1403. If NPL is still missing, we use RCONF174 + RCONF175 + RCON3494 + RCON5399 + RCONC237 + RCONC239 + RCON3500 + RCONF180 + RCONF181 + RCFNB573 + RCFD5378 + RCFD5381 + RCFD1597 + RCFD1252 + RCFD1255 + RCFDB576 + RCFDK214 + RCFDK217 + RCFD5390 + RCFD5460 + RCFDF167 + RCFDF170 + RCONF176 + RCONF177 + RCON3495 + RCON5400 + RCONC229 + RCONC230 + RCON3501 + RCONF182 + RCONF183 + RCFNB574 + RCFD5379 + RCFD5382 + RCFD1583 + RCFD1253 + RCFD1256 + RCFDB577 + RCFDK215 + RCFDK218 + RCFD5391+ RCFD5461 + RCFDF168 + RCFDF171

  If NPL is still missing, we use RCONF174 + RCONF175 + RCON3494 + RCON5399 + RCONC237 + RCONC239 + RCON3500 + RCONF180 + RCONF181 + RCONB835 + RCON1607 + RCONB576 + RCONK214 + RCONK217 + RCON5460 + RCON1227 + RCONF176 + RCONF177 + RCON3495 + RCON5400 + RCONC229 + RCONC230

+ RCON3501 + RCONF182 + RCONF183 + RCONB836 + RCON1608 + RCONB577 + RCONK215 + RCONK218 + RCON5461 + RCON1228

- Capital ratio

  Book capital (RCFD3210) to asset ratio. If RCFD3210 is missing, we use RCON3210.

- Indicator for large banks

  If a bank organization is in the top 5 largest bank organization by assets, the indicator is equal to 1; 0 otherwise.

- Real estate loan share

  Loans backed by real estate (RCFD1410) divided by total loans

  If RCFD1410 is missing, then we use RCON1410. If RCON1410 is missing, we use RCFDF158 + RCFDF159 + RCFD1420 + RCFD1797 + RCFD5367 + RCFD5368 + RCFD1460 + RCFDF160 + RCFDF161. if the real estate variable is still missing, we use RCONF158 + RCONF159 + RCON1420 + RCON1797 + RCON5367 + RCON5368 + RCON1460 + RCONF160 + +RCONF161

- District time trends

  Federal Reserve district (RSSD9170)

- Firm size

  Natural log of assets (RCFD2170)

- Leverage

  Debt/assets=(RCFD2948-RCFD2930)/RCFD2170

- Tangibility

  Gross PPE/assets = (RCFD2145+RCFD2150/RCFD2170)

- Z-score

  Z-score=(1.2× working capital+1.4× retained earnings+3.3× income before extraordinary items+0.999× sales)/sales. Working capital = RCD2170-(RCFD2160+RCFD2143+RCFD3656-(RCFD2948-RCFD2930). Retained earnings = RCFD3632. Income before extraordinary items = RIAD4301. Sales = RIAD4107+RIAD4079.

- ROA

  Net income/Assets=RIAD4340/RCFD2170

- Market Share Growth (MKTSHARE)

  Market share = assets/total assets of all banks in each quarter. Market share growth = quarterly change in market share each quarter. Assets = RCFD2170

- Asset Growth (ASSETGR)

  Quarterly change in assets (RCFD2170)

# Appendix D. Propensity Score Matching (PSM)

Table D.1: Mean Difference Test by PSM Methods

This table shows the results of mean differences for each variable by Propensity Score Matching (PSM) methods.

|  | Mean |  |  |  |
|---|---|---|---|---|
| Matching variables | Control | Sample | Differences | t-value |
| No matching |  |  |  |  |
| Unused commitment ratio | 0.067 | 0.263 | -0.196 | -180.27 |
| Net wholesale funding ratio | -0.121 | 0.069 | -0.190 | -63.29 |
| NPL to loans | 0.014 | 0.021 | -0.007 | -22.83 |
| Capital ratio | 0.111 | 0.115 | -0.004 | -7.51 |
| Real estate loan share | 0.701 | 0.469 | 0.231 | 80.02 |
| Number of observations | 312,244 | 4,249 |  |  |
| Probit 1:1 no replacement |  |  |  |  |
| Unused commitment ratio | 0.143 | 0.275 | -0.132 | -44.75 |
| Net wholesale funding ratio | -0.037 | 0.072 | -0.109 | -30.77 |
| NPL to loans | 0.013 | 0.021 | -0.008 | -22.31 |
| Capital ratio | 0.119 | 0.115 | 0.003 | 10.17 |
| Real estate loan share | 0.472 | 0.469 | 0.003 | 0.49 |
| Number of observations | 4,641 | 4,249 |  |  |
| Logit 1:1 no replacement |  |  |  |  |
| Unused commitment ratio | 0.141 | 0.275 | -0.134 | -42.54 |
| Net wholesale funding ratio | -0.040 | 0.072 | -0.112 | -28.85 |
| NPL to loans | 0.013 | 0.021 | -0.007 | -16.56 |
| Capital ratio | 0.115 | 0.115 | 0.000 | 0.80 |
| Real estate loan share | 0.511 | 0.469 | 0.042 | 8.02 |
| Number of observations | 4,693 | 4,249 |  |  |

Table D.1: Mean Difference Test by PSM Methods (Continued)

This table shows the results of mean differences for each variable by Propensity Score Matching (PSM) methods.

| | Mean | | | |
|---|---|---|---|---|
| Matching variables | Control | Sample | Differences | t-value |
| Local Linear Regression (LLR) | | | | |
| Unused commitment ratio | 0.134 | 0.275 | -0.141 | -44.75 |
| Net wholesale funding ratio | -0.049 | 0.072 | -0.122 | -30.71 |
| NPL to loans | 0.012 | 0.021 | -0.009 | -22.21 |
| Capital ratio | 0.124 | 0.115 | 0.009 | 10.14 |
| Real estate loan share | 0.473 | 0.469 | 0.004 | 0.50 |
| Number of observations | 4,641 | 4,249 | | |
| Radius with 0.25*std of pscore as caliper | | | | |
| Unused commitment ratio | 0.134 | 0.275 | -0.141 | -44.75 |
| Net wholesale funding ratio | -0.049 | 0.072 | -0.122 | -30.71 |
| NPL to loans | 0.012 | 0.021 | -0.009 | -22.21 |
| Capital ratio | 0.124 | 0.115 | 0.009 | 10.14 |
| Real estate loan share | 0.473 | 0.469 | 0.004 | 0.50 |
| Number of observations | 4,641 | 4,249 | | |
| Kernel Normal, biweight,uniform | | | | |
| Unused commitment ratio | 0.134 | 0.275 | -0.141 | -44.75 |
| Net wholesale funding ratio | -0.049 | 0.072 | -0.122 | -30.71 |
| NPL to loans | 0.012 | 0.021 | -0.009 | -22.21 |
| Capital ratio | 0.124 | 0.115 | 0.009 | 10.14 |
| Real estate loan share | 0.473 | 0.469 | 0.004 | 0.50 |
| Number of observations | 4,641 | 4,249 | | |