The Cost of Privacy Failures: Evidence from Bank Depositors' Reactions to Breaches*

Christian Engels† Bill B. Francis‡ Dennis Philip§

This draft: June 2022

We study consumers' reactions in response to material privacy violations. Exploring privacy breach incidences of U.S. banks, where personally identifiable information (social security numbers, addresses, names) is exposed to unauthorized parties, we find that depositors reallocate significant wealth holdings away from breached banks – we note a 12% differential in total deposits growth between breached and control banks in the year following a breach of privacy. In response to the depletion in deposits, breached banks increase deposit rates just after breach incidences and draw on funding liquidity from the interbank market. The results highlight the importance of privacy for consumers.

Keywords: privacy, cyber-security, data breaches, banks, deposit institutions, depositors
JEL Codes: G21, D14, D18

*We are grateful to Irem Erten, Olivier De Jonghe, Gabriele Lattanzio, Nadya Malenko, Teodora Paligorova, Andre Silva, Dee Warmath and seminar and conference participants at the Bristol Workshop on Banking and Financial Intermediation, Edinburgh Economics of Financial Technology Conference, FINEST Autumn Workshop, Westpac Massey Fin-Ed Centre Conference for helpful comments and suggestions. Any errors remain our own.

†Centre for Responsible Banking & Finance, University of St Andrews, Gateway Building, North Haugh, St Andrews KY16 9AL, UK. E-mail: ce50@st-andrews.ac.uk.

‡Lally School of Management, Rensselaer Polytechnic Institute, Troy, NY, U.S.A. Email: francb@rpi.edu.

§Durham University Business School, Department of Economics and Finance, Mill Hill Lane, Durham DH1 3LB, UK. E-mail: dennis.philip@durham.ac.uk

I. Introduction

Corporations losing sensitive private data to data breaches is increasingly becoming a common occurrence, with the banking sector in the U.S. being the highest hit in recent years. In 2019, the banking and financial sector, from among all other sectors, exposed the largest share (61 percent) of sensitive personally identifying information to unauthorized parties, amounting to over 100 million compromised records (Identity Theft Resource Center, 2020). Such material privacy breaches from banks impose substantial personal and financial costs, both direct and indirect, on consumers.[1] However, whether consumers react to such material privacy violations is unclear.[2] The privacy literature recognizes the tension between consumers seeking privacy and the difficulties in achieving it. Acquisti, Brandimarte, and Loewenstein (2020) discuss several psychological factors that affect consumers' privacy-seeking and privacy-neglecting behaviors. Furthermore, privacy perceptions differ across cultures, creating different privacy preferences (Bélanger & Crossler, 2011).[3]

In this paper, we examine the economic costs arising from privacy breaches for the U.S. banking system and quantify the reactions of banking consumers when their personally identifiable sensitive data is unlawfully disclosed to third parties. To the extent that customers seek privacy, its violations will deteriorate firm reputation and customer-firm relationships, thereby leading to adverse economic consequences for breached firms (Martin, 2020). Examining data breaches of U.S. banks enables us to measure bank depositors' reactions through deposit level changes after privacy breach incidences. Additionally, we are able to observe any counter responses from breached banks through their deposit rate offerings to mitigate the adverse effects from privacy breaches.

[1] For example, there is a significant increase in identity theft over the past decade, most of which can be attributed to breach incidences where personally identifiable data has been lost. According to the 2019 Federal Trade Commission (FTC) statistics, identity theft was the most common type of consumer complaint received, with a notable 46.4% increase from 2018 levels (Consumer Sentinel Network, 2020).

[2] A privacy violation in this paper refers to a data breach incident where sensitive personally identifiable data is disclosed unlawfully. This can be due to a crime (e.g., hacking, skimming, theft) or an accidental disclosure (e.g., lost physical devices with personal data, data handling errors).

[3] See recent examples of the value of privacy and the interplay with convenience among Asian consumers (Tang, 2019; Agarwal, Ghosh, Ruan, & Zhang, 2022).

As such, we exploit the features of the banking system for identification. Recent regulations such as know-your-customer (KYC) and anti-money-laundering (AML) require customers to disclose verifiable personal and financial data to banks. This contrasts with other industries in which customers are prompted to manage their privacy and decide how much information to reveal (see, e.g., Montes, Sand-Zantman, & Valletti, 2019; Adjerid, Acquisti, & Loewenstein, 2019). Thus, banking customers have stronger expectations of their banks to keep their personally identifiable information secure. If we were to observe depositors reacting to privacy breaches by moving their funds away from breached banks, we would be able to estimate the (potentially) significant knock-on effects of violating customer privacy expectations. These estimates would constitute a lower bound on the cost of data privacy failures because banks can draw on deposit insurance and fraud insurance policies to compensate depositors for monetary losses due to breaches.

Beyond the general importance of our findings for firms handling of customer data, analyzing the effect of privacy breaches in the banking sector per se is fundamentally important for several reasons. First, with the increase in technology, banks have increasingly become prime targets for cyberattacks and data breaches, and given the importance of safe banking systems, ascertaining the response of customers is vital for bank shareholders, managers, regulatory bodies and policymakers. Second, as is well known (e.g., Kashyap, Rajan, & Stein, 2002; Cornett, McNutt, Strahan, & Tehranian, 2011), banks in their liquidity-transformation role are important for the provision of capital for economic growth. Therefore, any action that threatens the viability of the banking system also puts at risk the prospects of firms' growth and, consequently, the growth of the economy. And third, because banks keep up-to-date records of some of the most sensitive data, such as comprehensive personally identifiable information and payment card data, their data breaches are particularly detrimental for their customers.

We obtain information on banks' privacy breaches from Privacy Rights Clearinghouse for the period 2005-2018. We focus on privacy breach events where sensitive personally iden-

tifiable information, including social security numbers, addresses, names or other personal information, have been unlawfully compromised. Such privacy breach events increase identifiability of individuals (Sweeney, 2000; Golle, 2006) and pose serious risks to consumers, as the personally identifiable information revealed can be misused for the perpetration of fraud and identity theft. Appendix B provides descriptions of a sample of privacy breaches. Typically, breached banks vary in size and, by definition, the breach involves the disclosure of sensitive private data. For example, in July 2014, Total Bank (Miami, FL) notified its customers that their computer network was subject to phishing attacks that allowed hackers to obtain customers' personally identifiable information, including names, addresses, account numbers, account balances, social security numbers and driver's license numbers.

We analyze the effect of privacy breaches vis-à-vis a propensity-score-matched control group to attenuate the influence of differences in bank characteristics of breached vs. non-breached banks. We employ a standard difference-in-differences regression framework to quantify the effect of a breach of privacy in the presence of cross-sectional, temporal and spatial confounders. We find that for the breached banks, total deposits decrease by, on average, 12% more in the year after a breach as compared to the control banks with no privacy breaches. This differential effect is apparent across different types of depositor clienteles, as evidenced by the significant decline in deposits such as insured, interest-bearing, money market, savings and, notably, time deposits.

For comparison, we also examine the effects of data breaches that did not result in a breach of personally identifiable information (such as those with only financial or card information loss) and find no significant depositor reactions. The results confirm that banking customers react to privacy concerns after their sensitive personally identifiable data fall into unauthorized hands. We build on these results using a dynamic difference-in-differences approach and find that there is a significant decline in deposit levels post-breach which remain below the pre-breach level over the next three quarters following privacy breach incidences. An upward trend begins about a year after a privacy breach, but it remains well below the

pre-breach deposit growth path.

We conduct several additional tests to validate the notion that depositors react in response to privacy concerns. First, we examine the effect of the type of information breached. We find that there are sizeable depositor reactions when individuals' social security numbers are breached – the economic magnitude of the drop in deposits in the four quarters after a breach is around 17%. Second, we analyze the effects of breaches of purely financial information and find that these type of breaches insignificantly impact depositors reactions. This is important because it provides strong support for the notion that it is the violations of depositor privacy that drive depositor reactions. And third, we find that depositors tend to react to all privacy breaches, irrespective of whether the breach was intended for harm (e.g., hacking, skimming, theft) or was accidental (e.g., lost physical devices with personal data). This indicates that the reactions of withdrawing funds away from breached banks cannot be attributed as a response to cyber risk or security failures, per se. Rather, the results substantiate the importance depositors place on the maintenance of privacy.

To delve further deeper into understanding depositors' reactions, we study the flow of deposit funds from breached to non-breached banks after privacy breach incidences. We find significant deposit reallocations at the headquarters counties of breached banks towards non-breached savings banks and minority depository institutions (MDIs), and to a lesser extent to commercial banks. This suggests that after data breaches depositors choose safer institutions with stronger local ties.

We also examine if a bank's reputation for caring for its stakeholders is an important factor for depositors in their reallocation decision after privacy breaches. We proxy bank reputation by its ESG ratings. Consistent with our expectations, we find that non-breached banks with high ESG ratings are the banks that benefit from deposit inflows after privacy breach events. This suggests that depositors, when reallocating their funds after a negative privacy shock, are willing to trust their deposits with banks with a higher reputation to cater to their customer needs.

Losses in deposits can have substantial repercussions for banks by limiting lending capacities. We therefore examine how breached banks react to losses in deposits relative to non-breached banks. In fact, any responses in deposit rates offered by breached banks after a privacy breach will be suggestive of the real cost of privacy breaches to the affected banks. Utilizing the RateWatch database of branch-level deposit rates, we examine whether breached banks raise deposit rates more, relative to their non-breached counterparts, subsequent to a privacy breach. Interestingly, the rates on several new deposits accounts such as certificate of deposits, fixed-rate IRAs, variable-rate IRAs and interest-bearing checking accounts show an upward trend after a privacy breach, relative to their non-breached counterparts, after a privacy breach. The highest increases are observed in time deposit rates, with the certificate of deposit accounts being the most affected. These increases in rates coincide with the significant decreases in deposit levels following a privacy breach, suggesting that breached banks increase deposit rates to mitigate the outflow in time deposits after a breach. The findings corroborate the recent literature on reputation effects (Akey, Lewellen, & Liskovich, 2021; Kamiya, Kang, Kim, Milidonis, & Stulz, 2021), which posits that reputational losses necessitate higher compensation for consumers to continue transacting with the affected firms.

Because privacy breaches induce unexpected liquidity shocks, affected banks may seek short-term funding from peers to mitigate the liquidity shortfalls and avoid additional costs. In the interbank market, we find evidence of increased liquidity demand by breached banks after privacy breaches. More specifically, relative to control banks, privacy-breached banks substantially increase (approximately 59% more) the amount of money raised from other banks in the four quarters following a breach incident, as well as drawing on bank capital.

Our study contributes to the important literature on whether consumers care about their data privacy and whether they act to protect it. This is of significant public interest, as consumers are increasingly required to digitalize and share verifiable personal information online, as for example, in the fight against the COVID-19 pandemic. Academic research

highlights the complexities in consumers' privacy decision-making and the reasons for the dissonance between stated privacy beliefs by consumers and their behaviors (see, for example, Athey, Catalini, & Tucker, 2017; Acquisti et al., 2020; Agarwal et al., 2022; Solove, 2021). Our research shows that banking consumers care about and react strongly to breaches of their sensitive personal data, while such reactions are not seen when only financial information is lost. This indicates that privacy concerns and financial risks are distinct for consumers, where the latter are usually covered by insurance. The findings highlight the value for privacy in the banking sector, a sector where consumer privacy really matters.

The paper also contributes to the rapidly growing literature evaluating the economic effects of breaches on corporations (e.g., Campbell, Gordon, Loeb, & Zhou, 2003; Cavusoglu, Mishra, & Raghunathan, 2004; Acquisti, Friedman, & Telang, 2006; Akey et al., 2021; Amir, Levi, & Livne, 2018; Kamiya et al., 2021). These studies show declines in market values, sales growth and profitability for breached firms following cyberattacks. Florakis, Louca, Michaely, and Weber (2020) estimate a firm-level cybersecurity risk exposure measure using textual analysis and show that cyber risk is priced in the cross section of stock returns. As such, firms showing that they respond to protect consumer privacy by investing in newer security technologies lower their cost of capital (Havakhor, Rahman, & Zhang, 2021). Our work is distinct in that we focus on the economic costs of privacy violations due to any type of data breach, not just cyberattacks as in previous studies. Further, we are able to estimate directly how consumers react to privacy violations by studying depositor behavior in the U.S. banking sector, a sector hardest hit by breaches in recent years. Banking customers, who are legally obliged to disclose personally identifiable information to their banks, demand information privacy and this enables us to comprehensively quantify the economic value customers place of privacy and assess the economic costs to corporations. In this regard, our study contributes to the broader evaluation of customer privacy expectations.

Our study also contributes to the banking literature in that it provides evidence on depositors' reactions to non-financial disclosures and news releases. As banks' balance-sheet

and off-balance-sheet risks are inherently more opaque, relative to firms in other industries (Beatty & Liao, 2014), disclosures play an important role in revealing important information to the public. For example, Chen, Hung, and Wang (2019) find that information releases of banks' negative social performance reduce depositors' willingness to finance the bank by decreasing their trust in banks. In a similar vein, Homanen (2022) shows that depositors negatively react when banks do not address tractable ESG (environmental, social and governance) risks. Our study complements and adds to this literature by introducing privacy breaches and documenting their costs to banks. Finally, our paper also enables us to provide evidence of the steps taken by breached banks to mitigate the associated depositors fall-out. This evidence has been lacking in prior work.

The remainder of the paper proceeds as follows. Section II describes the data and sample characteristics. Sections III and IV present the results. Section V concludes.

II. Data sample and research design

A. Data

We retrieve records on financial institutions data breaches occurring between 2005 and 2018 from the Privacy Rights Clearinghouse's public database. The breach data include the names of breached institutions, their geographic locations, public breach announcement dates, the nature of the breaches (hacking or skimming, lost or stolen physical devices with private data, insider or employee involvement, accidental disclosure) as well as, brief contextual information, among other details. We hand-match the full universe of breach data to the population of FDIC-insured depository institutions, resulting in a sample of 99 breached banks and depository institutions. We note that several of these breaches resulted in the release of sensitive private details that enable identifiability of individuals, including social security numbers, addresses, names and other personal information. We denote such incidences as privacy breaches. To study comparative effects, we also consider breach events that

did not result in the loss of personally identifiable information, labelled as other breaches. Appendix A contains the word cloud associated with keywords describing all breach events, while Appendix B provides examples of privacy breaches contained in our sample.

To study depositor behavior over time, we obtain bank- and branch-level data from three data sources. First, we make use of the Statistics on Depository Institutions (SDI) from the Federal Deposit Insurance Corporation (FDIC). The SDI provide comprehensive information on bank deposit quantities as well as bank financial data at a quarterly frequency for depository institutions filing the Report of Condition and Income (Call Report) and for thrift banks (Savings and Savings & Loans banks) filing the OTS Thrift Financial Report (TFR). The expanded SDI coverage of U.S. depository institutions over Call Reports, including thrifts that comprise circa 10% of all depository institutions, allows for a wider study of bank depositor responses at a granular level, both cross-sectionally and across time. Given the design of our study, similar to (Berger & Bouwman, 2013), we exclude bank and depository institutions with missing data on domestic deposits, commercial and industrial loans, and real estate loans. Further, we exclude banks subject to mergers & acquisitions (M&A), purchases & assumptions (P&A) activities or any lawsuits in the event window around the respective breaches, resulting in 70 breached banks and depository institutions in our sample.

Second, we collect annual data of branch office deposits from the FDIC's Summary of Deposits (SOD). These annual data capture trends in domestic deposits and market concentration each year as of June 30 for all FDIC-insured institutions with branch offices. Beyond information on total branch deposits, the SOD data also include detailed geographical information on the surveyed branches, facilitating spatial analysis of depositor reactions. Combining the SOD and the SDI data enables us to relate privacy breaches to depositor reactions at the branch-level.

Third, we source branch-level deposit rate information from RateWatch. The data contain weekly rates, fees and account details for various new deposit product offerings by bank

branches. We focus on a series of deposit products, including money market, savings, retirement, checking and certificate of deposit, of varying account sizes and maturities. These are representative of the broad range of deposit products offerings by banks. In addition to the geographic characteristics of the branches, the RateWatch data also includes information on whether a branch office actively sets deposit. We follow the standard approach of limiting our analysis to rate-setting branches to avoid introduction of redundancies in the empirical analysis (Drechsler, Savov, & Schnabl, 2017). Table I reports sample summary statistics of all the variables constructed from the three datasets and utilized in our empirical analysis. Appendix C provides detailed definitions of all the variables.

B. Empirical framework and Identification

Our main empirical methodology is a difference-in-differences approach. To identify depositors' reactions to privacy violations, we exploit the following characteristics of privacy breach events. First, when a privacy breach becomes publicly known, arrival of the information is exogenous to the bank depositors' information set. Second, the public announcement of a breach equals an unexpected expansion of what the depositors know about their bank. And third, when depositors chose their bank, it is unlikely that they had priced the possibility of their data being illicitly accessed by third parties in the future. Thus, the main source of identification comes from affected depositors unexpectedly learning about the disclosure of their sensitive personally identifiable data by their banks, whereas depositors of non-breached banks receive no new information contrary to their expectation that their private information remains private.

Exploiting this exogenous variation, our empirical framework takes the following form:

$$D_{it} = \alpha_i + \lambda_{s(i)t} + \beta(Post_{it}) + \gamma_1(Post_{it} \times PrivacyBreach_i) + X'_{it-1}\theta + \epsilon_{it},$$
$$\forall i, \forall t \in \{B_i^{-H} \leq t \leq B_i^{+H}\}, \tag{1}$$

where D_{it} is the natural logarithm of one plus the level of deposits for bank i at time t, measured at a quarterly frequency. We consider the level of total, insured, time, interest, retirement, money market, demand or savings deposits in separate regressions. Bank fixed effects (α_i) are included to measure the effect on changes in deposits. Further, we include state-time fixed effects ($\lambda_{s(i)t}$) to capture for state-level differences over time. As such, we account for differences in laws surrounding privacy and data breaches between states, and any changes in the privacy laws for consumers over time (Romanosky, Hoffman, & Acquisti, 2014).[4] $PrivacyBreach_i$ is an indicator variable that is equal to one for banks subject to privacy breaches, and zero otherwise, where privacy breaches are breach incidences in which customer addresses, social security numbers or any financial information are exposed to third parties. $Post_{it}$ is an indicator variable, which becomes equal to one for bank i after its actual (for breached banks) or matched (for control banks) data breach has become public. Included observations are those that fall in the four ($H = 4$) quarters before and after a data breach (B_i^{-H} and B_i^H, respectively). The matrix of controls, X_{it}, includes the banks' return on assets (ROA), the liquidity and noninterest income ratios, as well as the overall deposit rate. All standard errors are clustered by bank and time to allow for serial correlation across time as well as cross-sectional shocks. The key parameter of interest is γ_1, measuring the average treatment effect of a privacy breach on bank deposits. It captures the difference in deposit levels between banks subject to a privacy breach and those in the control group of banks, after the breach becomes publicly known.

To comparatively assess the differential impact of privacy breaches from other types of breaches, we also extend Equation (1) with treatment indicator variables for banks subject to non-privacy breaches. This provides insights on the value of privacy placed by depositors. Later, we also disaggregate breach incidences based on the actual information content being breached such as breach of social security number, name, address, personal information, card details, account numbers or financial information. This disaggregation enable us to better

[4] Alternative model specifications with fixed effects at the bank level, at the bank and time levels, and at the bank, time and state levels produce similar results.

understand the factors distinguishing bank depositors' reactions to data breaches.

In all empirical specifications we use a propensity score matching approach, as it reduces potential biases in estimating causal effects stemming from correlations between the public announcements of data breaches and important bank characteristics, such as bank size. The matching procedure is implemented as follows. First, a logistic model is estimated to predict a bank's breach status based on important bank-level characteristics four quarters prior to a breach event. Second, we use the predicted latent propensities from this model, i.e., the propensity scores, to construct a nearest-neighbor matched sample of banks, where the control with the closest propensity score to a breached bank is chosen from the pool of non-breached banks in the same U.S. state. In essence, this matching procedure assigns with replacement each breached institution to a control bank of the same U.S. state and effectively discards non-breached banks that are too dissimilar to the breached banks based on the observed characteristics employed in the matching model.

To assess whether our matching procedure results in balancing the covariates between breached and control banks – and whether it succeeds in constructing a sample in which data breaches occur as-if randomly between breached and control banks given these covariates – we conduct the following test: we estimate logit models on (1) the full sample of banks and on (2) the matched sample of banks, again predicting the breach status of banks in the respective samples. If the matching is accurate, we expect none of the covariates to predict a bank's breach announcement.

Table II reports the results. In contrast to the pre-matched full sample logit model results reported in Columns (1) and (2), we observe that in Columns (3) and (4) the magnitudes of the logit regressions coefficients decline substantially for the matched sample of banks. Importantly, none of the sources of heterogeneity continue to play any role in explaining the breach status across banks, whereas the bank characteristics including size, deposit rate, non-interest income ratio and ROA are statistically significant in the pre-matched full sample. The fit of the models in Columns (3) and (4) also indicates that the matching

procedure achieved its objective. Specifically, the p-value of the χ^2 test for overall model fit is 0.731 and 0.911 for the models contained in Columns (3) and (4), respectively, indicating that the null hypothesis that all coefficient estimates are zero cannot be rejected. Further, in the sample resulting from our matching procedure, the baseline predicted probabilities are one-half (0.50). We can therefore conclude from these test results that the matching procedure removes meaningful differences along the observable dimensions between the two groups of banks and that banks in the matched sample experience data breaches with an average probability of 50%.

III. Depositors' reactions to privacy breaches

A. Descriptive evidence of deposit withdrawals from breached banks

We begin our analysis by visually inspecting the total deposit trends around the privacy breach events in our sample of banks. Because our definition of privacy breaches entails the loss of key personally identifiable information, to the extent that depositors care about privacy violations, we expect to see a sizeable effect of privacy breaches on deposit levels subsequent to the public announcement of a breach. Accordingly, Figure I plots the changes in total deposits (in logs) from four quarters before a breach became public to four quarters after. We consider separately the deposit level changes for banks with privacy breaches, banks subject to other (non-privacy) breaches and for control banks. The grey area represents the time when the breach event is made public, which happens between banks' quarterly reporting dates.

We highlight several characteristics of the data. First, in the four quarters prior to the announcement of a breach deposit levels of privacy-breached and control banks follow a common trend. This provides evidence that the conditions required for identification that were discussed in Section II.B are being met; namely, that the announcement of a privacy breach is unexpected information to depositors and that, relatedly, no withdrawals of deposits

occur before the breach information becomes public. Second, a large drop in breached banks' deposits with a reversal in the growth trend is visible in the first quarter after a privacy breach announcement. Further depletion in deposits occurs in the subsequent quarters, with no recouping to the pre-breach growth trend in deposits. Third, in contrast to privacy breaches, banks subject to other (non-privacy) breaches do not incur major withdrawals in deposits.

B. Depositor reactions: baseline results

Although Figure I shows evidence of strong depositor reactions to privacy breach events, the statistical significance of depositor reactions cannot be taken for granted due to potential confounding factors. For instance, the environments in which banks operate can differ significantly across regions and time. Regulatory changes affecting the operating environment of banks can introduce further challenges for causal identification. To address these concerns, our baseline fixed effects difference-in-differences model in Equation (1) includes various bank, time and spatial fixed effects. The regression specification captures the causal effects of privacy breaches by estimating changes in the average deposit levels following banks' privacy breach incidences. We consider the potential heterogeneous effects of privacy breaches on the different deposit account types by exploiting the detailed deposits information provided by the FDIC.

Table III reports the baseline regression results. In the various specifications, we include the effects from other (non-privacy) breaches by including $OtherBreach_i$, which is an indicator variable that is equal to one for banks subject to other types of non-privacy breaches, and zero otherwise. Columns (1) to (8) show the results for the different types of deposit accounts, with the first column focusing on total deposit levels. We find that the effect of privacy breaches on deposit levels is statistically significant and economically important with respect to most deposit accounts, with only retirement and demand deposits not significant. This is in stark contrast to other types of breaches where there is no evidence of

depositor reactions. To be specific, with the natural logarithm of deposits as the dependent variable, the estimates in Column (1) suggest that privacy breaches cause an average drop of total deposits of about 12% for breached banks relative to the non-breached banks in the control group. Among the (strongly) significant deposit types, the relative effects range from approximately –9% for interest-bearing deposits to –22% for time deposits. The effect of privacy breaches is seen to be not statistically different from zero only for the case of retirement and demand deposits. Importantly, Column (2) shows that the effect is not only limited to uninsured deposits but that deposits covered by the FDIC deposit insurance are also strongly affected. Column (3) further reveals that the relative effect is most pronounced for time deposits.

These findings show that when it comes to data breaches, personally identifiable information loss is of critical importance. As such, drawing a distinction among data breaches along the vector of information that is breached enables us to shed light on the economic impact of data breaches on corporations. To quantify the economic impact of privacy breaches, we obtain the predicted deposit levels for affected banks implied for each event time quarter. We then compare the predicted total deposit levels before privacy breaches with those after. Such comparisons indicate that, vis-à-vis banks in the control group, total deposits of breached banks are depleted by around $41 million in the first quarter after the breach. It should be noted that $19 million of this depletion can be attributed to missed growth opportunities (comparing against the counter-factual growth path) and $22 million to deposits leaving breached banks.

In order to mitigate concerns of omitted variables threatening the stability of our results, in Table A1 of the Online Appendix, we present evidence from the Oster (2019) procedure assessing the magnitude of bias stemming from unobserved factors. The procedure assumes that selection on observables is informative about selection on unobservables and tests for the degree of omitted variable bias (δ) necessary to fully eradicate the established privacy breach effects (i.e., Post $\times$ Privacy breach $= 0$) in the model estimates of Table III. For our

model of total deposits, we find that the influence of potential omitted variables would have to be 3.752 times as important as the effect of our currently included covariates to make the privacy breach effects immaterial. Additionally, the estimated effect bounds indicate qualitatively similar results, suggesting that the baseline estimates are robust to omitted variable bias.

C. Timing of depositor responses

In this subsection we build on the findings of a causal relationship in the previous section by investigating whether there are also dynamic causal effects. The baseline results provide evidence of the average causal effects and as such, they can understate the magnitude of depositors' response. For instance, this could occur when the effects of the breach manifest with strong reactions immediately after the public announcement of a privacy breach and appears to quickly taper off. Moreover, it could be the case that there is a continuation of deposit withdrawals, rather than a one-off adjustment when the breaches become public which would be indicative of significant depositor memory, where depositors that once shun a breached bank continue to do so in the future (analogous to Iyer & Puri, 2012, for the case of solvent bank runs). To provide evidence on the dynamic effects of privacy breaches, we focus on banks with privacy breaches and their control banks, and modify the regression model in Equation (1). Specifically, to capture deposit levels at different event time quarters before and after a breach is made public, we replace the variable $Post_{it}$ with a set of indicator variables that capture the event time quarter relative to the bank's (matched) privacy breach, and zero at any other event time quarter. The base case in this regression is the quarter before the breach occurs. The remainder of the specification stays unaltered. The parameters of are denoted by the set $\{\delta_h\}$, capturing the relative deposit levels at the respective event time quarter h. These measure the magnitudes of the differences in total deposit levels between the breached and non-breached banks at the respective event time quarter relative to the first quarter before privacy breaches have become public. In particular, the parameters for

which h is positive would indicate the causal effect of privacy breaches on deposit levels at the respective quarter in the post-breach period.

Figure II visualizes the estimated set $\{\delta_h\}$ with the corresponding 95% confidence interval for each parameter First, we observe that all parameter estimates for event time quarters before the privacy breach announcements are statistically insignificant. In addition to the descriptive evidence provided in Figure I, this speaks to the satisfaction of a common trend before a privacy breach, indicating that public release of information on privacy breaches is an unexpected event to depositors and no detectable deposit movements take place based on private information ahead of the public breach announcement. We confirm this more formally in an F-test reported in Table A2 of the Online Appendix, where we find no pre-breach differential in trends between the control and treated banks.

Second, the δ_h parameters for positive h indicate negative causal effects that are statistically significant and incrementally increase in magnitude as time progresses, providing strong evidence for deposit depletions increasing over time in response to privacy breaches. Importantly, we observe a drop of around 11% in total deposit levels in the next two quarters for breached banks, relative to the control group. The downward trend is observed to reverse in a year after the breach. These findings show the existence of important economic consequences from privacy breaches, which grow more severe as time passes, instead of an immediate bank run.[5]

D. Information content in privacy breaches

The threat to privacy arises when there is disclosure of person-related uniquely identifiable information (Sweeney, 2000; Golle, 2006), which makes affected individuals more susceptible to identity theft and other forms of victimization. In this section, we examine in more detail the reactions of bank depositors to breaches of different types of information. Specifically, we

[5]As an extension of the baseline model, we consider longer term effects of two and three years after a privacy breach event. The results are reported in Table A3 of the Online Appendix. We find that the privacy breach effects remain significant, while it tapers off three years post breach.

distinguish between loss of personal details from loss of financial information such as payment card details. Parsing the breach descriptions given in the Privacy Rights Clearinghouse (PRC) database, we classify breach events into those that have led to the disclosure of (i) social security number (SSN), (ii) address, (iii) personal information, (iv) name, (v) financial information, (vi) card details, and (vii) account numbers. Items (i)-(iv) are classed as breaches of personal details, while (v)-(vii) are classed as breaches of financial details.

For each breach event type, we estimate its effect on total deposit levels in our difference-in-differences approach. The estimates enable us to gauge the differential impact for banks subject to breaches of specific types of depositor information. Table V reports the results. Columns (1) to (7) present the separate regression results for the different types of breaches. We find strongly significant depositor reactions to breach events resulting in loss of personal details. To be specific, the economic magnitude of the drop in deposits is between 11% when individuals' names are breached and 18% when their social security numbers are breached. In contrast, depositors' reactions are observed to be less significant for breaches involving financial details and the magnitude drop in deposits is around 8%. Column (8) shows the results by amalgamating the different breach types into four non-overlapping categories according to whether personal and/or financial details have been breached. They show that there is significant deposit withdrawals when personal details are breached. In addition, we see that the economic magnitudes are the same whether or not the breached data contained financial details. Importantly, and somewhat unexpectedly, when financial details are breached with no personal details being lost, there are no significant effects on deposits levels. This provides strong support for the argument that depositors react more strongly to their privacy violations. These results corroborate our earlier findings that data breaches elicit strong depositor reactions only when depositor privacy is at stake, rather than the case of financial details being lost.

E. Does intention to harm rather than privacy per se drive depositor reactions?

In light of the cognate literature investigating the effect of cyber risk on firms (Akey et al., 2021; Kamiya et al., 2021), in which attackers successfully infiltrate firms to gain access to data through hacking and malware, it is natural to ask whether our findings are driven by depositors responding to cyber risk and security failures, rather than loss of privacy. We therefore examine any differential effects in depositor reactions from privacy breaches between those that intend harm and those that are accidental. Both classes of breaches lead to a loss of sensitive information; however, the negative ramifications from breaches that are motivated by criminal intentions are much larger than accidental disclosures of private data by banks. A priori, if bank depositors are responding to the breach of their privacy, we expect to see little difference between whether the breach of privacy occurred with or without criminal intentions. It should also be noted that this analysis also enables us to disentangle whether depositors are reacting to cyber risk per se, rather than loss of sensitive private information.

For this investigation, we classify all breaches into (i) harm intended breaches, which include loss of data due to hacking, skimming and theft, and (ii) accidental breaches, where loss of data occurred due to accidental disclosures or loss of personal data from physical devices without harmful intentions. We estimate a difference-in-differences model analogous to the Equation (1) in our empirical framework, while including interaction terms that capture differential effects for banks subject to breaches with harmful intentions and accidental breaches. In the specification, we consider both privacy breaches as well as other (non-privacy) breaches, such as only financial details being lost. All other specification details in the model remain the same as previously.

The findings are reported in Table V. We observe significant depositor reactions after privacy breaches, irrespective of whether the loss of privacy occurred due to a breach with intention to harm or by accident. The magnitude of the effects is very similar for these

two cases. When we consider other breaches, which have not led to the loss of personal information, we do not observe strongly significant depositor reactions for harm intended breaches or accidental breaches. Overall, these results provide strong support for the notion that the loss of personal data is the important driver for depositors' reaction of withdrawing their funds from breached banks. As such, it validates our argument that depositors respond to violations of their privacy rather than renewed salience of cyber risk.

F. Reallocation of branch deposits: where does the money go?

Our results so far suggest that depositors reallocate sizeable amounts of their funds away from banks that disclose privacy breaches. We now examine the characteristics of non-breached banks that exhibit measurable increases in their deposit levels following privacy breaches in their local banking system. We focus our assessment on bank branches of non-breached banks and exclude banks in our control group to alleviate any econometric concerns.

F.1. Reallocation effects after breaches in the local banking system

We first explore the reallocation effects observed in the primary local area of breached banks (i.e., in the headquarters counties of the breached banks). We expect that customers can pay closer attention to data breaches when they occur close-by due to reasons such as higher local news coverage in the case of a local bank breach, which consequently elicits stronger local depositor reactions. Moreover, closer proximity to bank headquarters elevates the role of information diffusion, which will likely be greater due to rumors and local peer effects. In contrast, we expect such localized effects not to be as pronounced among depositors located away from the bank's headquarters, and as such, are likely to react less strongly to privacy breaches as compared to locally domiciled depositors.

Accordingly, we test whether the deposit levels of non-breached banks located in the breached banks' headquarters increase after announcement of breaches. We use branch-level deposits information, obtained on an annual frequency from the FDIC Summary of Deposits.

For each branch of a bank, we introduce the *HQcounty* indicator variable, which takes the value of one if a bank's branch is located in the same county as its headquarters, and zero otherwise, and interact it with *Post* as defined in Equation (1) in our empirical framework. The double interaction term captures the reallocation effects in the headquarters counties of breached banks after a privacy breach. Standard errors are clustered by branch and time to account for serial correlation of branches' deposit levels across time as well as for cross-sectional shocks.

Table VI reports the regression result in Column (1). The coefficient on the double interaction term is seen to be strongly significant and positive, indicating that there are important localized effects where non-breached banks experience a significant increase in their deposits after local privacy breaches. Such effects are insignificant outside the headquarters counties of breached banks.[6] That is, local depositors react more negatively and reallocate their deposits after a privacy breach. These results are similar to those reported in prior studies such as Homanen (2022) which find that local depositors take significantly stronger measures in disciplining banks involved in scandals.

F.2. Heterogeneous reallocation effects: types of depository institutions

We further explore aggregate reallocation effects for different types of non-breached depository institutions. Following a privacy breach, if depositors orientate their banking business towards banks with a stronger focus on more traditional banking services provided in their local communities, then deposit increases should be pronounced for savings banks and minority depository institutions (MDIs). We therefore investigate the heterogeneous reallocation effects across depository institution types. The evidence so far shows that deposit withdrawals are significant in headquarters counties of breached banks. We therefore estimate the following fixed effects regression to examine the reallocation effects in the breached

[6]The results suggest localized spillover effects within breached banks' headquarters. As such, to alleviate any identification concerns to the baseline results due to spillovers to control banks, we perform additional robustness checks where control banks do not have any branches situated within breached banks' headquarters. All the findings remain unchanged.

banks' headquarters counties:

$$\begin{aligned}\log(D_{jt}) = \alpha_j + \lambda_{s(j)t} + \phi(Post_{jt}) + \beta_0(Post_{jt} \times HQCounty_{b(j)}) \\ + \beta_1(Post_{jt} \times HQCounty_{b(j)} \times BankType_{i(j)}) + \epsilon_{jt}, \\ \forall j, \forall t \in \{PrivacyBreach_{b(j)} = 0\},\end{aligned} \tag{2}$$

where $BankType$ captures the types of depository institutions. Specifically, we estimate the changes in deposits of non-breached savings versus commercial banks, and in a separate regression, consider the deposit changes in MDIs versus non-MDIs (because both savings and commercial banks can qualify as MDIs). In the regressions, $Post_{jt}$ takes the value of one for a non-breached bank branch j in the first year following a privacy breach announcement, and zero otherwise. $HQcounty_{b(j)}$ takes the value of one if branch j is located in the same headquarters as any breached bank b, and zero otherwise. Consistent with previous approaches, fixed effects at the branch level (a_j) and state-time ($\lambda_{s(j)t}$) level are included and standard errors are clustered by branch and time.

In Table VI, Column (2) shows the estimates for savings versus commercial banks, and Column (3) those for MDIs versus non-MDIs. We find that both commercial and savings bank branches that are in the headquarters counties of breached banks experience significant deposit increases. The deposit level increase for commercial banks is around 3.9%, while that for savings banks is around 6.4%. These findings support the notion that savings banks experience disproportionally greater deposit increases as depositors seek out institutions with stronger local ties. Further, we find a significant increase of around 7.9% in deposit levels for non-breached MDIs located in the headquarters county of breached banks. The results suggest that breached banks' depositors reallocate funds more towards safer and minority-owned banks, which primarily operate in local market areas.

F.3. Heterogeneous reallocation effects: ESG ratings

Prior work shows that firms disclosing data breaches suffer reputational losses (e.g., Akey et al., 2021; Kamiya et al., 2021). Thus, depositors affected by privacy breaches may seek to place their funds with banks that provide signals of acting responsibly and upholding values important to their stakeholders, such as environmental, social and governance (ESG) issues. Therefore, to the extent that reputation is indeed an important consideration for depositors in their reallocation decisions, non-breached banks with a high ESG rating should experience greater deposit inflows following the disclosure of privacy breaches. We examine the role of ESG ratings empirically by estimating the following fixed effects regression model:

$$\begin{aligned}\log(D_{jt}) = \alpha_j + \lambda_{s(j)t} + \phi(Post_{jt}) + \beta_0(Post_{jt} \times HQCounty_{b(j)}) \\ + \beta_1(Post_{jt} \times HQCounty_{b(j)} \times HighESG_{i(j)}) + \epsilon_{jt}, \\ \forall j, \forall t \in \{PrivacyBreach_{b(j)} = 0\},\end{aligned} \tag{3}$$

For each bank, the ESG rating is defined as the bank's number of ESG strengths subtracted by the number of ESG concerns. $HighESG$ takes the value of one for branches with ESG rating above the median. $Post_{jt}$ takes the value of one for a non-breached bank branch j for the first year following a privacy breach event, and zero otherwise. $HQcounty_{b(j)}$ takes the value of one if branch j is in the same county as any breached bank b's headquarters, and zero otherwise. Again, we include fixed effects at the branch level (a_j) and state-time level ($\lambda_{s(j)t}$), and cluster standard errors by branch and time.

Table VI reports the coefficient estimates of the regression in Column (4). We find that branches of banks with high ESG ratings situated in the headquartered county of a breached bank are the ones experiencing significant deposit inflows – the order of magnitude increase in deposits is estimated approximately 7.1% and significant at the 1% level. Non-breached banks with low ESG ratings have negative gains in deposit levels. These results indicate

that depositors exhibit a preference for banks with higher ESG ratings after breaches.

Our findings suggest that privacy breaches significantly reduce consumers' trust in a firm due to reputation loss and can be interpreted in the light of economic interactions built on trust. When privacy norms are violated, individuals' trust and willingness to engage with the firm diminishes. This can be even more pertinent for bank depositors, who are legally required to entrust their personally identifiable and financial information to banks on a regular basis (e.g., for bank know-your-customer and anti-money-laundering compliance). This trust-based reading of our findings originates in the privacy paradox literature, which documents that consumers retain strong privacy expectations after disclosing personal information. For example, Martin (2020) utilizes a series of factorial vignette surveys to show a higher negative impact on consumers' trust and willingness to engage with a market actor after security violations emerge in which an outsider gains access to their private information. In this context, our results contribute causal evidence with external validity on the adverse effect of privacy breaches resulting from a loss of trust in breached firms.

IV. Banks' responses to privacy breaches

A. Deposit rate responses by breached banks

Deposits are an essential and stable source of funding for banks and difficult to substitute with other funding sources. Therefore, banks subject to privacy breaches have strong incentives to take steps to mitigate the resulting depletion of deposit levels and to prevent possible negative effects on bank operations arising from a paucity of funds. One possible response open to breached banks is to increase deposit rates and provide preferential interest terms on new deposit products, relative to the market. This can foster the inflow of deposits from new customers, provide incentives for existing customers to take out these new products rather than moving funds elsewhere – such as to competitor banks or away from the banking system – and as a result, stabilize bank deposit levels. Further, if breached banks indeed react to

privacy breaches with significant deposit rate increases (compared to control banks), it is symptomatic of the severity of breached banks' deposits outflows after the breach incident.

Accordingly, we study movements in branch-level deposit rates, obtained from the RateWatch database, to test whether breached banks increase their rates after a privacy breach. Our difference-in-differences regression specification for average deposits rates, R_{jt}, offered by branch j in month t is:

$$\begin{aligned} R_{jt} = \alpha_j + \lambda_{s(j)t} + \phi(Post_{i(j)t}) + \beta(Post_{i(j)t} \times PrivacyBreach_j) \\ + \delta(Post_{i(j)t} \times HQCounty_j) \\ + \gamma(Post_{i(j)t} \times PrivacyBreach_{i(j)} \times HQCounty_j) + \epsilon_{it}, \\ \forall j, \forall t \in \{B_{i(j)}^{-H} \leq t \leq B_{i(j)}^{+H}\} \end{aligned} \tag{4}$$

where the dependent variable, R_{jt} is the deposit rate and we include branch (α_j) and state-time ($\lambda_{s(j)t}$) fixed effects. $Post_{i(j)t}$ is an indicator variable, which is equal to one for branch j of bank i for the time period after the breach is announced. $PrivacyBreach_{i(j)}$ is an indicator variable equal to one for branch j if its bank i is subject to privacy breach, and zero otherwise. We include observations covering twelve months before and after the breach of the bank. In this specification the data are at the monthly interval, so the set $\{B_i^h\}$ collects event time months and $H = 12$ in Equation (4). We estimate the triple interaction of the variables $Post_{i(j)t}$, $PrivacyBreach_{i(j)}$ and $HQcounty_j$ to capture rates charged by the bank branches situated in the headquarters counties, that are most affected by deposit outflows after a privacy breach. We consider rates of different types of deposit products, including money market, savings, retirement, checking and certificate of deposit (CD), of varying account sizes and maturities. Appendix C lists and provides the definitions for all the deposit products used. We cluster standard errors at the branch level to account for serial correlation over time as well as for cross-sectional shocks.

The causal effect of privacy breaches on branch-level deposit rates in counties with the

bank's headquarters is captured by γ. For cases in which banks respond to privacy breaches by disproportionally increasing deposit rates relative to the non-breached banks in the control group, we would expect a positive causal effect ($\gamma > 0$). Table VII reports the estimation results for privacy breaches, where we report the results for various types of rate products, including certificates of deposits of size \$10K with maturities 6, 12, 24 and 36 months, savings account of size \$2.5K, fixed IRA, variable IRA, checking account, and money market accounts of sizes \$2.5K and \$25K.

The estimates of γ in Columns (1) to (4) suggest that, subsequent to privacy breaches, breached banks in counties with banks headquarters disproportionally raise deposit rates on certificates of deposits relative to control banks. The increases are stable across CD maturities and range from 23 to 39 basis points. We also find greater increases in rates for new variable IRA accounts as compared to control banks, with an estimated magnitude of 95 basis points. Current accounts, \$2.5K savings accounts and fixed IRAs show no significant increase, while \$25K money market rates exhibit a marginally significant decrease. Taken together, we find that breached banks, as compared to control banks, offer higher rates on new certificate of deposit products, which are high yield fixed maturity products for depositors. These substantial increases in certificate of deposit rates coincides with the significant drops in time deposit levels found in Table III. These findings suggest that breached banks react to deposit depletions and compensate depositors with higher rates to attenuate the negative effects from privacy breaches.

B. Timing of deposit rate responses

In this subsection we build on the findings of the previous subsection on rate responses from branches of breached banks, and investigate the timing with which these deposit rate responses occur by focusing on the dynamic causal effects for a representative set of deposit products. Specifically, we analyse the deposit rates for new 6-month \$10K CD, 36-month \$10K CD, fixed IRA, variable IRA, checking and \$25K money market accounts.

Accordingly, we modify the Equation (4) to capture deposit rates at each event time month before and after a privacy breach is made public:

$$
\begin{aligned}
R_{jt} = \alpha_j + \lambda_{s(j)t} + \phi(HQCounty_j) \\
+ \sum_{h=-H}^{H} \psi_h(I_{jt}^h) + \sum_{h=-H}^{H} \beta_h(HQCounty_j \times I_{jt}^h) + \sum_{h=-H}^{H} \gamma_h(PrivacyBreach_{i(j)} \times I_{jt}^h) \\
+ \sum_{h=-H}^{H} \delta_h(PrivacyBreach_{i(j)} \times HQCounty_j \times I_{jt}^h) + \epsilon_{jt} \\
\forall j, \forall t \in \{B_{i(j)}^{-H} \leq t \leq B_{i(j)}^{+H}\}
\end{aligned} \tag{5}
$$

where R_{jt} denotes the different deposit rates offered by branch j at time t. We interact $PrivacyBreach_{j(i)}$ and $HQCounty_j$ with a set of indicator variables, $\{I_{jt}^h\}$, that takes the value of one if at time t the branch of the breached (control) bank is h months away from its (matched) privacy breach, and zero at any other event time month. The base case in this regression is the month before the breach becomes public. The remainder of the specification remains unaltered.

We focus on privacy breaches, for which the parameters of interest in this regression are collected in the set $\{\delta_h\}$, capturing the relative deposit rates at the respective event time month. These constitute estimates of the difference in deposit rates between the branches of breached and non-breached banks at event time month h, relative to the first month that privacy breaches have become public. The parameters for which h is positive indicate the causal effect of privacy breaches on deposit rates at the respective month in the post-breach period. Figure III visualizes the set $\{\delta_h\}$ with the corresponding 95% confidence interval for each parameter estimate. It displays the estimates for the case of deposit rates on certificate of deposits, fixed-rate and variable-rate IRAs, checking accounts and money market accounts.

We summarize the results as follows. First, certificate of deposit rates offered by branches belonging to breached banks are statistically indistinguishable from their controls in the period leading up to the public announcement of a privacy breach but begin to rise steadily in

the post breach period. The increase stabilizes four months later and the resulting positive rate difference relative to branches in the control group is sustained throughout the remainder of the estimation period. Second, variable IRA rates are also observed to steadily rise just after the breach, while fixed IRA rates show an immediate spike followed by insignificant differences between breached and control group. Third, no clear trend emerges for checking account rates and money market rates, where the rates are seen to be flat after the breach. Overall, these results depict some significant upward movements in deposit rate responses by branches of breached banks compared to their control group, especially with the certificate of deposit rates of different maturities. It is also consistent with the argument that banks may be strategically adjust their rates to remain attractive to depositors following the announcement of a privacy breach.

C. Liquidity demand in the interbank market

The sizeable deposit depletions that arise from privacy breaches constitute significant liquidity shocks for the breached banks. These shocks can generate real economic costs as they could lead to reserve shortfalls or insufficient balances to settle daylight overdrafts. To avert these adverse scenarios and avoid additional costs, breached banks can increase their activity on the interbank market for deposits by either decreasing interbank assets or increasing interbank liabilities (Bhattacharya, Gale, Barnett, & Singleton, 1985; Allen & Gale, 2000; Angelini, Nobili, & Picillo, 2011; Castiglionesi, Feriozzi, Lóránth, & Pelizzon, 2014; Dietrich & Hauck, 2020). Note that banks that are unable to draw on this coinsurance mechanism can instead opt to raise additional bank capital.

We examine banks' interbank and capital-raising activities following privacy breaches in a difference-in-differences fixed effects regression framework adapted from the preceding sections. The regression equation is estimated separately for the dependent variables total bank capital, interbank assets, and interbank liabilities (all on the log scale). Total bank capital is defined as the book value of bank capital according to banks' submissions to

the FDIC. Interbank assets comprise deposits at other banks, which can be drawn down to satisfy deposit withdrawals, while interbank borrowing positions (liabilities) capture the money raised from other banks. The Appendix provides the exact variable definitions. We include the same controls as in the preceding sections, while also including the log of last periods' total assets to control for potential confounding effects due to bank size. The remainder of the regression framework is unchanged.

Table VIII reports the results. Columns (1) to (3) contain the estimates for the dependent variables total bank capital, interbank assets and interbank liabilities, respectively. We make three observations based on the results. First, after the public announcement of a banks' privacy breach, total bank capital is significantly reduced as shown by the negative and significant coefficient of the interaction variable, $Post \times Privacy\ breach$. Second, the corresponding coefficient estimate for interbank assets is marginally positive, albeit also insignificant. In short, breached banks do not draw on money they have deposited with other banks. However, it should be noted that the interpretation of the estimated coefficient is complicated by the fact that in contrast to interbank liabilities, not all banks and depository institutions are required to disclose the levels of interbank assets they hold. Third and of key importance, the relevant coefficient in the interbank liabilities' regression is large in magnitude and significant at the 5% level. The coefficient estimate of 0.462 implies an increase of approximately 59% relative to the interbank fundraising activities of banks in the control group. This finding suggests that banks substantially increase the amount of money they raise from other banks following a privacy breach announcement.

In sum, this analysis shows that following a privacy breach, affected banks significantly draw on funds from other banks to mitigate the ensuing deposit outflows. This points to the effectiveness of banks coinsurance abilities through the interbank market for deposits in the case of the idiosyncratic risk arising from banks that are subjected to privacy violations.

V. Conclusion

With the quantity and value of sensitive personal data shared in the information economy ever increasing, particularly within the banking and financial network, consumers privacy concerns arising from data breaches have surged in recent years. This paper evaluates the cost of privacy failures for corporations by studying major privacy breach incidences of U.S. banks and other depository institutions, involving the loss of personally identifiable information such as social security numbers, addresses, names, and personal information to unauthorized third parties. Such breach events involve a comprehensive loss of private information and place affected individuals at risks of victimization, as their personal details can be misused for the perpetration of fraud and identity theft.

We study depositors' reactions following privacy breach incidences. The banking sector operates in an ideal environment to study the importance of privacy for customers, as regulatory requirements (know-your-customer and anti-money-laundering) necessitate customers to proactively disclose verifiable personal and financial information to their banks; this leads banking customers to retain strong privacy expectations towards their banks. To assess the effect(s) of privacy violations on bank depositors' behaviors, we conduct difference-in-differences analyses on a propensity-score matched sample of banks and study any deposit reallocation trends in the depositor clientele of breached and non-breached banks.

The analysis shows that following privacy violations, breached banks' deposit levels progressively decline over time, with average total deposit levels dropping by $41 million in the first quarter after the breach, where $19 million of this depletion can be attributed to missed growth opportunities (comparing against the counter-factual growth path) and $22 million to deposits leaving breached banks. We find significant declines in the level of deposits post-breach across the depositor clientele, as reflected notably in depletions of time deposits, insured deposits and savings deposits. The results suggest that depositors' trust in banks diminishes due to reputation loss resulting from privacy breaches. Breached banks are seen to respond to privacy breaches by offering higher deposit rates to their depositors subsequent

to a privacy breach, as compared to control banks. Thus, depositors require higher compensation, reflected in increased deposit rates, to continue transacting and funding breached banks.

The effects are observed predominantly when personal information, including social security numbers, addresses and names, are disclosed illicitly to third parties. The effects are non-existent after breaches of only financial information. This highlights that consumers care about their sensitive private data being breached and it is not a matter of reducing financial risks. Overall, our findings establish the value consumers place on their data privacy and, in turn, we document the economic costs of privacy breaches for the U.S. banking system.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2020). Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology*, *30*(4), 736–758.

Acquisti, A., Friedman, A., & Telang, R. (2006). Is there a cost to privacy breaches? An event study. *ICIS 2006 Proceedings*, 94.

Adjerid, I., Acquisti, A., & Loewenstein, G. (2019). Choice architecture, framing, and cascaded privacy choices. *Management Science*, *65*(5), 2267–2290.

Agarwal, S., Ghosh, P., Ruan, T., & Zhang, Y. (2022). Privacy versus convenience: Customer response to data breaches of their information. *SSRN Working Paper*.

Akey, P., Lewellen, S., & Liskovich, I. (2021). Hacking corporate reputations. *Rotman School of Management Working Paper 3143740*.

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

Amir, E., Levi, S., & Livne, T. (2018). Do firms underreport information on cyber-attacks? Evidence from capital markets. *Review of Accounting Studies*, *23*(3), 1177–1206.

Angelini, P., Nobili, A., & Picillo, C. (2011). The interbank market after august 2007: What has changed, and why? *Journal of Money, Credit and Banking*, *43*(5), 923–958.

Athey, S., Catalini, C., & Tucker, C. (2017). The digital privacy paradox: Small money, small costs, small talk. *National Bureau of Economic Research*.

Beatty, A., & Liao, S. (2014). Financial accounting in the banking industry: A review of the empirical literature. *Journal of Accounting and Economics*, *58*(2-3), 339–383.

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS quarterly*, 1017–1041.

Berger, A. N., & Bouwman, C. H. (2013). How does capital affect bank performance during financial crises? *Journal of Financial Economics*, *109*(1), 146–176.

Bhattacharya, S., Gale, D., Barnett, W., & Singleton, K. (1985). Preference shocks, liquidity, and central bank policy. *Liquidity and crises*, *35*.

Campbell, K., Gordon, L. A., Loeb, M. P., & Zhou, L. (2003). The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. *Journal of Computer Security*, *11*(3), 431–448.

Castiglionesi, F., Feriozzi, F., Lóránth, G., & Pelizzon, L. (2014). Liquidity coinsurance and bank capital. *Journal of Money, Credit and Banking*, *46*(2-3), 409–443.

Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce*, *9*(1), 70–104.

Chen, Y.-C., Hung, M., & Wang, L. L. (2019). Depositors' responses to public nonfinancial disclosure. *Working Paper, Hong Kong University of Science and Technology*.

Consumer Sentinel Network. (2020). Consumer sentinel network data book 2019. https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2019.

Cornett, M. M., McNutt, J. J., Strahan, P. E., & Tehranian, H. (2011). Liquidity risk management and credit supply in the financial crisis. *Journal of Financial Economics*, *101*(2), 297–312.

Dietrich, D., & Hauck, A. (2020). Interbank borrowing and lending between financially constrained banks. *Economic Theory*, *70*(2), 347–385.

Drechsler, I., Savov, A., & Schnabl, P. (2017). The deposits channel of monetary policy. *The Quarterly Journal of Economics*, *132*(4), 1819–1876.

Florakis, C., Louca, C., Michaely, R., & Weber, M. (2020). Cybersecurity risk. *National Bureau of Economic Research*.

Golle, P. (2006). Revisiting the uniqueness of simple demographics in the us population. In *Proceedings of the 5th acm workshop on privacy in electronic society* (pp. 77–80).

Havakhor, T., Rahman, M. S., & Zhang, T. (2021). Disclosure of cybersecurity investments and the cost of capital. *SSRN Working Paper*.

Homanen, M. (2022). Active depositors. *Journal of Banking & Finance*, *136*.

Identity Theft Resource Center. (2020). 2019 end-of-year data breach report. https://www.idtheftcenter.org/2019-data-breaches.

Iyer, R., & Puri, M. (2012). Understanding bank runs: The importance of depositor-bank relationships and networks. *American Economic Review*, *102*(4), 1414–45.

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*(3), 719–749.

Kashyap, A. K., Rajan, R., & Stein, J. C. (2002). Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking. *The Journal of Finance*, *57*(1), 33–73.

Martin, K. (2020). Breaking the privacy paradox: The value of privacy and associated duty of firms. *Business Ethics Quarterly*, *30*(1), 65–96.

Montes, R., Sand-Zantman, W., & Valletti, T. (2019). The value of personal information in online markets with endogenous privacy. *Management Science*, *65*(3), 1342–1362.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204.

Romanosky, S., Hoffman, D., & Acquisti, A. (2014). Empirical analysis of data breach litigation. *Journal of Empirical Legal Studies*, *11*(1), 74–104.

Solove, D. J. (2021). The myth of the privacy paradox. *The George Washington Law Review*, *89*, 1.

Sweeney, L. (2000). Simple demographics often identify people uniquely. *Health (San Francisco)*, *671*(2000), 1–34.

Tang, H. (2019). *The value of privacy: Evidence from online borrowers*. Working Paper.

Figure I
Total deposits around privacy breaches and other breaches

The figure plots average total deposits (in logs) from event time quarter -4 to h for banks announcing privacy breaches or other breaches, together with their propensity score matched controls. The time of public breach announcements is indicated by the vertical grey area. To construct the control group, breached banks are matched to non-breached banks in the same state based on the characteristics size, deposit rate and noninterest income ratio at event time quarter -4. The data are at a quarterly frequency and are sourced from the FDIC SDI.

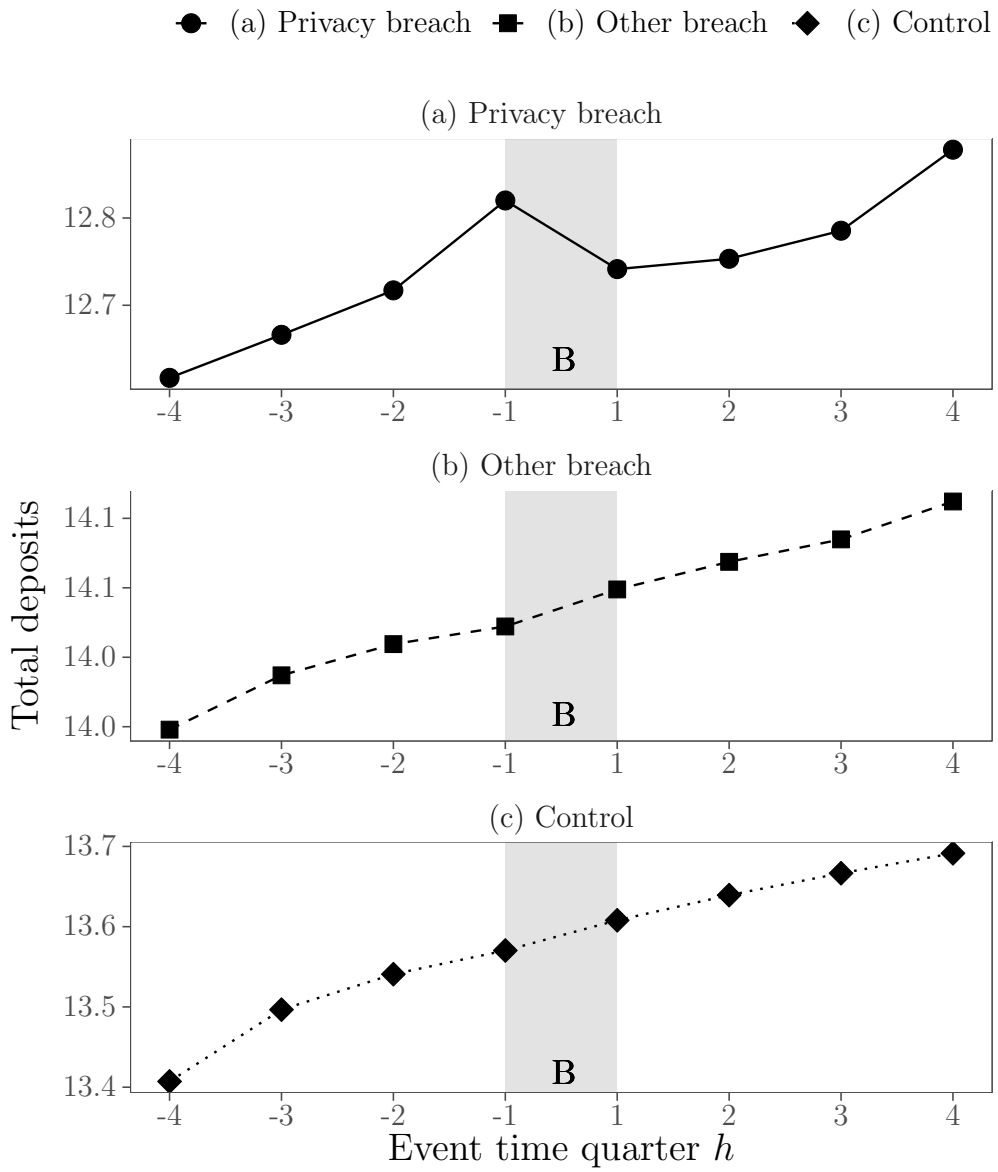


Figure II
Timing of deposit depletions after privacy breaches

The figure plots the estimated difference-in-differences coefficients (δ_h), which captures the relative deposit levels at the respective event time quarter h, together with their corresponding 95% confidence intervals. The dependent variable is total deposits in logs. The times of data breach announcements are indicated by the vertical grey area. The estimates capture effects relative to event time quarter -1. The data are at a quarterly frequency and are sourced from the FDIC SDI. Standard errors are clustered by both bank and time.

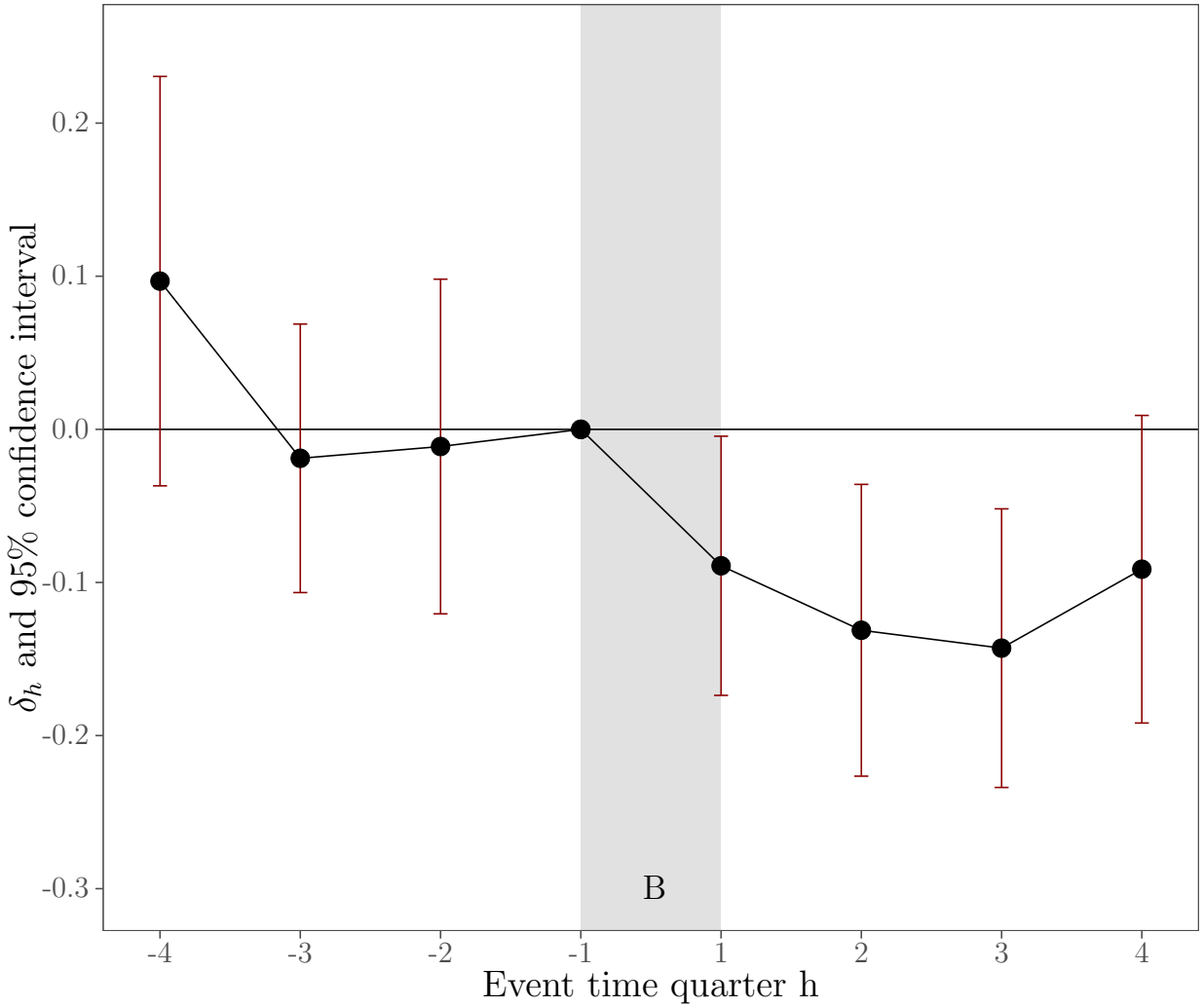


Figure III
Timing of deposit rate responses

The figure plots the estimated difference-in-differences coefficients (δ_h) of Equation (5) with their corresponding 95% confidence intervals for rate-setting branches belonging to privacy breached and control banks, respectively, that are domiciled in the same county as their institutional headquarters. The dependent variable in Panel (a) is deposit rates on new 6-month certificates of deposits with an account size of $10K; in Panel (b) it is deposit rates on new 36-month certificates of deposits with an account size of $10K; in Panel (c) it is deposit rates on new fixed-rate IRA accounts; in Panel (d) it is deposit rates on new variable-rate IRA accounts; in Panel (e) it is deposit rates on new current accounts; and in Panel (f) it is deposit rates on new money market accounts with an account size of $25K. The vertical line marks the months of public breach announcements. The estimates capture effects relative to event time month −1. The data are at a monthly frequency and are sourced from RateWatch. Standard errors are clustered by both branch and time.

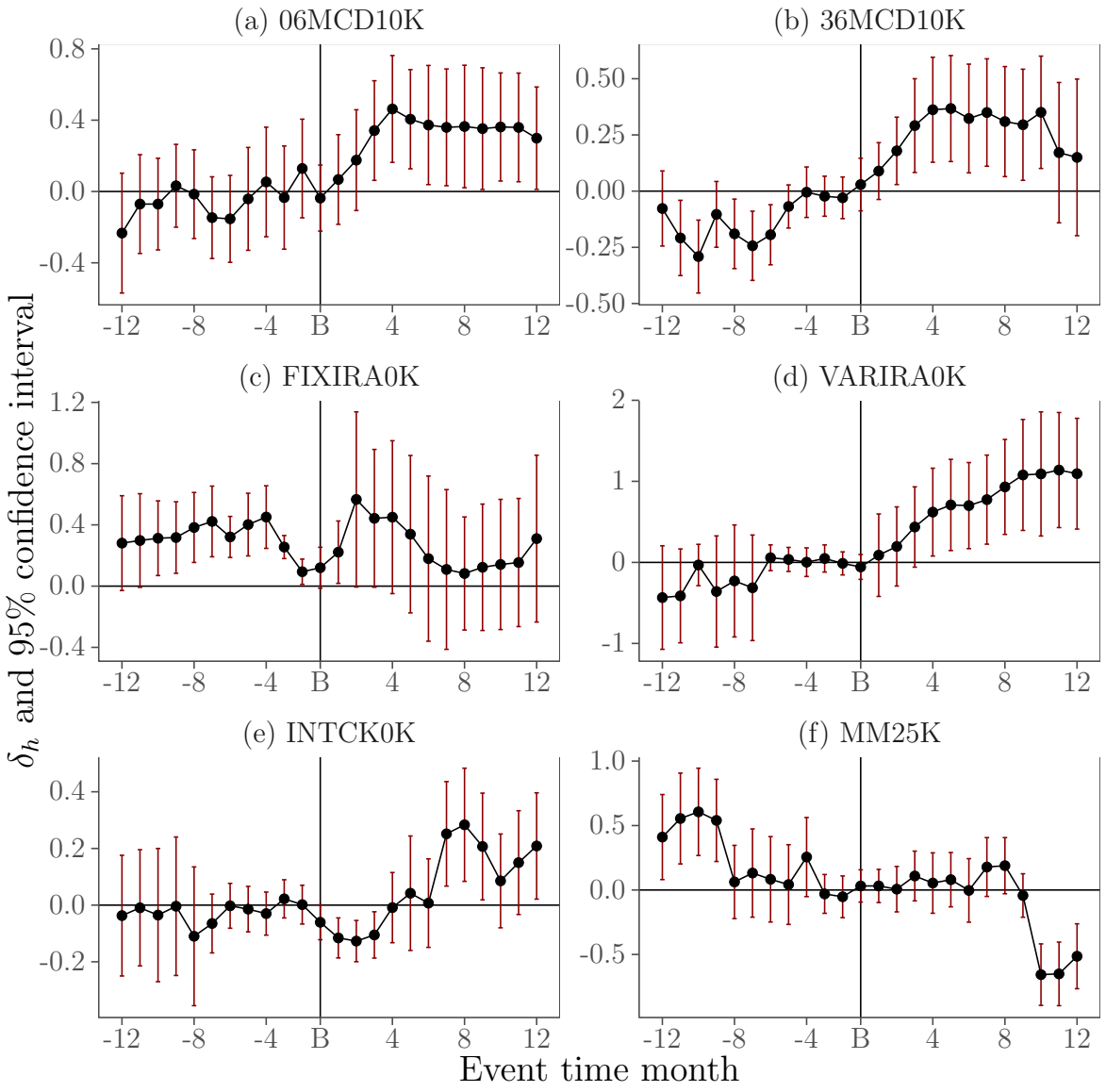


Table I
Summary statistics

The table reports the summary statistics for our data sample from 2014-2019. Panel A variables are at a quarterly frequency from the FDIC Statistics on Depository Institutions (SDI) for all U.S. depository institutions insured by the FDIC. Panel B variables are at an annual frequency from the FDIC Summary of Deposits (SOD). Finally, Panel C variables are deposit rates obtained from RateWatch and are at a monthly frequency. Appendix C provides exact definitions of all variables.

	Mean	p25	p50	p75	SD	Obs.
Panel A: Statistics on Depository Institutions (bank level, quarterly)						
Total deposits	11.79	10.97	11.70	12.51	1.37	436054
Insured deposits	11.55	10.77	11.49	12.26	1.36	436012
Interest deposits	11.60	10.78	11.51	12.31	1.36	434305
Retirement deposits	8.64	7.80	8.62	9.47	1.45	409553
Money market deposits	9.74	8.60	9.69	10.78	1.86	410976
Demand deposits	9.68	8.92	9.70	10.44	1.37	427064
Savings deposits	11.14	10.18	10.97	11.96	1.48	422147
Size	12.02	11.16	11.89	12.70	1.33	436159
ROA	0.47	0.20	0.45	0.83	0.83	435542
Liquidity ratio	25.43	13.85	21.81	33.24	16.06	411017
Noninterest income ratio	12.93	6.28	10.35	15.68	12.29	435472
Deposit rate	3.36	1.10	2.31	4.70	3.09	435438
Panel B: Summary of Deposits (branch level, annually)						
Branch deposits	10.30	9.71	10.42	11.05	1.28	1091608
HQ county	0.32	0.00	0.00	1.00	0.47	1134658
Panel C: RateWatch deposit rates (branch level, monthly)						
06MCD10K	1.27	0.25	0.74	2.00	1.28	1246180
12MCD10K	1.53	0.40	1.00	2.33	1.37	1253285
24MCD10K	1.79	0.65	1.35	2.71	1.34	1195978
36MCD10K	2.01	0.90	1.64	3.00	1.33	1127496
SAV2.5K	0.40	0.10	0.25	0.50	0.42	1246225
FIXIRA0K	1.25	0.40	0.75	1.74	1.22	887775
VARIRA0K	1.37	0.30	0.95	2.05	1.32	671661
INTCK0K	0.25	0.05	0.15	0.30	0.32	1188055
MM2.5K	0.50	0.10	0.25	0.75	0.58	1147279
MM25K	0.71	0.17	0.40	1.00	0.78	1181018

Table II
Propensity score matching

The table reports the coefficient estimates of logit regressions. The dependent variables are indicators taking the value one for banks that breach four quarters later, and zero otherwise. Columns (1) and (2) report the results for the full sample of banks; while Columns (3) and (4) report results for the sample obtained after matching. The model in Column (1) generates the propensity scores. The variables used in propensity score matching (with replacement) are size, deposit rate and noninterest income ratio, measured four quarters before a breach. Matched control banks are drawn from the set of non-breached banks in the same state as the breached banks. Appendix C provides exact definitions of all variables. Standard errors are reported in parentheses. ***, ** and * denote levels of significance at 1, 5 and 10 percent, respectively.

	Breached			
	Full sample		Matched sample	
	(1)	(2)	(3)	(4)
Size	0.687***	0.695***	0.064	0.060
	(0.05)	(0.05)	(0.07)	(0.08)
Deposit rate	0.138***	0.121***	0.043	0.030
	(0.03)	(0.03)	(0.05)	(0.06)
Noninterest income ratio	0.031***	0.027***	-0.002	-0.003
	(0.01)	(0.01)	(0.01)	(0.01)
ROA		0.286*		0.114
		(0.16)		(0.25)
Liquidity ratio		0.001		0.000
		(0.01)		(0.01)
Constant	-18.340***	-18.526***	-1.026	-0.982
	(0.76)	(0.83)	(1.12)	(1.42)
Baseline predicted probability	0.000	0.000	0.500	0.500
χ^2	183.247	186.789	1.291	1.516
Prob $> \chi^2$	0.000	0.000	0.731	0.911
Adjusted R^2	0.136	0.139	0.007	0.008
Observations	344098	344098	140	140

Table III
Effects of data breaches on deposits

The table reports the coefficient estimates of fixed effects regressions. The dependent variables in the various columns are different types of deposits (in logs). Privacy breaches are defined as data breaches that lead to loss of social security numbers, addresses, names or any personal information; all non-privacy breaches in our sample are classed as other breaches. Observations in the year before and after data breaches are included for the respective breached banks and controls. Appendix C provides exact definitions of all variables. Standard errors are clustered by both bank and time and are reported in parentheses. ***, ** and * denote levels of significance at 1, 5 and 10 percent, respectively. The data are obtained from the FDIC SDI and are at a quarterly frequency.

	Total deposits	Insured deposits	Time deposits	Interest deposits	Retirement deposits	Money market deposits	Demand deposits	Savings deposits
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	0.110*	0.159**	0.346**	0.125	-0.048	0.060	-0.130*	0.079***
	(0.06)	(0.07)	(0.16)	(0.08)	(0.09)	(0.05)	(0.07)	(0.03)
Post × Privacy breach	-0.126***	-0.191***	-0.248**	-0.095*	0.083	-0.099*	0.073	-0.127***
	(0.05)	(0.05)	(0.12)	(0.06)	(0.13)	(0.06)	(0.06)	(0.05)
Post × Other breach	0.053*	0.097	0.015	-0.104	0.663	0.233	-0.041	0.130*
	(0.03)	(0.09)	(0.05)	(0.13)	(0.42)	(0.19)	(0.10)	(0.07)
ROA	0.016	-0.016	0.061	0.001	0.088	0.017	0.029	0.022
	(0.03)	(0.03)	(0.07)	(0.03)	(0.08)	(0.04)	(0.03)	(0.02)
Noninterest income ratio	0.003	0.002	-0.009	0.004	0.018	0.016	0.002	0.009*
	(0.00)	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)
Liquidity ratio	-0.006*	-0.004	-0.009*	-0.007	0.000	-0.001	-0.003	-0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Deposit rate	0.044*	0.047**	0.087**	0.071**	0.050**	0.065	0.038*	0.044*
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.04)	(0.02)	(0.02)
Bank f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-time f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.115	0.0857	0.0740	0.0485	0.0396	0.0503	0.0126	0.0889
Observations	1120	1120	1120	1120	1120	1120	1120	1120

Table IV
Breaches of personal and financial information

The table reports the coefficient estimates of fixed effects regressions. The dependent variables are total branch deposits in logs. Breach events are separately classified into those that have led to the disclosure of (1) social security number (SSN), (2) address, (3) personal information, (4) name, (5) financial information, (6) card details and (7) account numbers. (1)–(4) are classed as breaches of personal details, while (5)-(7) are classed as breaches of financial details. Column (8) amalgamates the different breach types into four non-overlapping categories according to whether personal and/or financial details have breached. Observations in the year before and after data breaches are included for the respective breached banks and controls. Appendix C provides exact definitions of all variables. Standard errors are clustered by bank and time. ***, ** and * denote levels of significance at 1, 5 and 10 percent, respectively. The data are sourced from FDIC SDI and are at a quarterly frequency.

	Total deposits							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	0.116*	0.087	0.113*	0.079	0.082	0.057	0.081	0.110*
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Post × Breach (SSN)	-0.184***							
	(0.06)							
Post × Breach (address)		-0.173***						
		(0.06)						
Post × Breach (personal information)			-0.128***					
			(0.05)					
Post × Breach (name)				-0.119**				
				(0.06)				
Post × Breach (financial information)					-0.087*			
					(0.05)			
Post × Breach (card details)						-0.009		
						(0.04)		
Post × Breach (account numbers)							-0.109*	
							(0.06)	
Post × Breach								
× *Any personal but no financial details breached*								-0.127*
								(0.07)
× *Any personal and financial details breached*								-0.125**
								(0.06)
× *No personal and no financial details breached*								0.076
								(0.05)
× *No personal but any financial details breached*								0.037
								(0.03)

(Continued)

	Total deposits							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ROA	0.018	0.015	0.016	0.017	0.018	0.017	0.018	0.016
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Noninterest income ratio	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Liquidity ratio	-0.006*	-0.006*	-0.006*	-0.006*	-0.006*	-0.007*	-0.006*	-0.007*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Deposit rate	0.044*	0.043*	0.043*	0.043*	0.045*	0.044*	0.045*	0.044*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Bank f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-time f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.126	0.117	0.114	0.103	0.101	0.0890	0.103	0.116
Observations	1120	1120	1120	1120	1120	1120	1120	1120

Table V
Intention to harm and accidental breaches

The table reports the coefficient estimates of fixed effects regressions. The dependent variables in the various columns are different types of deposits (in logs). Privacy breaches are defined as data breaches that lead to loss of social security numbers, addresses, names, or any personal information; all non-privacy breaches in our sample are classed as other breaches. Harm and Accidental are two indicator variables taking the value of one for banks subject to a breaches with harmful intentions and accidental breaches, respectively, and zero otherwise. Observations in the year before and after data breaches are included for the respective breached banks and controls. Appendix C provides exact definitions of all variables. Standard errors are clustered by both bank and time. ***, ** and * denote levels of significance at 1, 5 and 10 percent, respectively. The data are sourced from FDIC SDI and are at a quarterly frequency.

	Total deposits	Insured deposits	Time deposits	Interest deposits	Retirement deposits	Money market deposits	Demand deposits	Savings deposits
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	0.112*	0.166**	0.358**	0.116	-0.031	0.066	-0.126*	0.079***
	(0.06)	(0.07)	(0.16)	(0.08)	(0.09)	(0.05)	(0.07)	(0.03)
Post × Privacy breach × Harm	-0.149**	-0.256**	-0.501*	0.138	-0.139	-0.140*	-0.058	-0.025
	(0.07)	(0.11)	(0.26)	(0.20)	(0.09)	(0.08)	(0.08)	(0.07)
Post × Privacy breach × Accidental	-0.119**	-0.172***	-0.173	-0.164***	0.150	-0.087	0.112	-0.157***
	(0.05)	(0.05)	(0.12)	(0.05)	(0.16)	(0.07)	(0.07)	(0.05)
Post × Other breach × Harm	0.002	-0.122	-0.154	-0.059	0.176	-0.024	-0.043	-0.033
	(0.04)	(0.10)	(0.21)	(0.07)	(0.13)	(0.10)	(0.13)	(0.07)
Post × Other breach × Accidental	0.063**	0.141	0.056	-0.120	0.764	0.284	-0.037	0.158**
	(0.03)	(0.09)	(0.04)	(0.16)	(0.49)	(0.21)	(0.11)	(0.08)
ROA	0.016	-0.017	0.056	0.006	0.085	0.017	0.027	0.025
	(0.03)	(0.03)	(0.06)	(0.03)	(0.08)	(0.04)	(0.03)	(0.02)
Noninterest income ratio	0.003	0.002	-0.008	0.003	0.019	0.016	0.003	0.009*
	(0.00)	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)
Liquidity ratio	-0.007*	-0.004	-0.009*	-0.007	-0.000	-0.002	-0.003	-0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Deposit rate	0.044*	0.048**	0.092**	0.065**	0.055**	0.066	0.041*	0.042*
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.04)	(0.02)	(0.02)
Bank f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-time f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.116	0.091	0.080	0.056	0.044	0.052	0.014	0.095
Observations	1120	1120	1120	1120	1120	1120	1120	1120

Table VI
Reallocation of branch deposits

The table reports the coefficient estimates of fixed effects regressions. The dependent variable in all columns is total branch deposits (in logs). The sample consists of non-breached bank branches in counties in which a bank experienced a privacy breach. Privacy breaches are defined as data breaches that lead to loss of social security numbers, addresses, names or any personal information. The variable Post takes the value of one for the year after a privacy breach event, and zero otherwise. Observations one year before and after the breach for branches of the respective privacy breached and control banks are included. Appendix C provides exact definitions of all variables. Standard errors are clustered by both branch and time and are reported in parentheses. ***, ** and * denote levels of significance at 1, 5 and 10 percent, respectively. The data are obtained from the FDIC SOD and are at a yearly frequency.

	Non-breached branch deposits			
	(1)	(2)	(3)	(4)
Post	-0.004	-0.003	-0.004	0.001
	(0.01)	(0.01)	(0.01)	(0.02)
Post × HQ county	0.046***	0.038***	0.041***	-0.044*
	(0.01)	(0.01)	(0.01)	(0.03)
Post × HQ county × Savings bank		0.024**		
		(0.01)		
Post × HQ county × Minority depository institution			0.076***	
			(0.02)	
Post × HQ county × High ESG rating				0.110***
				(0.03)
Branch f.e.	Yes	Yes	Yes	Yes
State-time f.e.	Yes	Yes	Yes	Yes
Observations	338524	338524	338524	328190

Table VII
Banks' rate responses

The table reports the coefficient estimates of fixed effects regressions. The dependent variables are deposit rates (in percentage points) on new accounts. Privacy breaches are defined as data breaches that lead to loss of social security numbers, addresses, names or any personal information. HQ county captures branches that are located in the same county as their institutional headquarters. Observations in the year before and after the breach are included for branches of the respective privacy breached and control banks. Appendix C provides exact definitions of all variables. Standard errors are clustered by both branch and time and are reported in parentheses. ***, ** and * denote levels of significance at 1, 5 and 10 percent, respectively. The data are sourced from RateWatch and are at a monthly frequency.

	06MCD10K	12MCD10K	24MCD10K	36MCD10K	SAV2.5K
	(1)	(2)	(3)	(4)	(5)
Post × Privacy breach	-0.321***	-0.160*	-0.358***	-0.354***	-0.006
	(0.109)	(0.094)	(0.109)	(0.095)	(0.027)
Post × HQ county	-0.273**	-0.193**	-0.164	-0.183**	-0.062
	(0.127)	(0.093)	(0.105)	(0.090)	(0.045)
Post × Privacy breach × HQ county	0.387***	0.230*	0.350***	0.377***	-0.004
	(0.140)	(0.117)	(0.130)	(0.107)	(0.044)
Branch f.e.	Yes	Yes	Yes	Yes	Yes
State-time f.e.	Yes	Yes	Yes	Yes	Yes
R^2	0.070	0.034	0.135	0.165	0.026
Observations	2985	3015	2991	2905	2947
	FIXIRA0K	VARIRA0K	INTCK0K	MM2.5K	MM25K
	(6)	(7)	(8)	(9)	(10)
Post × Privacy breach	0.078	-0.234	0.004	-0.054	0.156
	(0.197)	(0.217)	(0.060)	(0.068)	(0.180)
Post × HQ county	-0.016	0.516*	-0.049	-0.113**	-0.059
	(0.015)	(0.266)	(0.051)	(0.052)	(0.107)
Post × Privacy breach × HQ county	-0.014	0.950***	0.098	-0.017	-0.308*
	(0.163)	(0.346)	(0.077)	(0.077)	(0.180)
Branch f.e.	Yes	Yes	Yes	Yes	Yes
State-time f.e.	Yes	Yes	Yes	Yes	Yes
R^2	0.007	0.203	0.032	0.036	0.042
Observations	1830	1892	2870	2502	2637

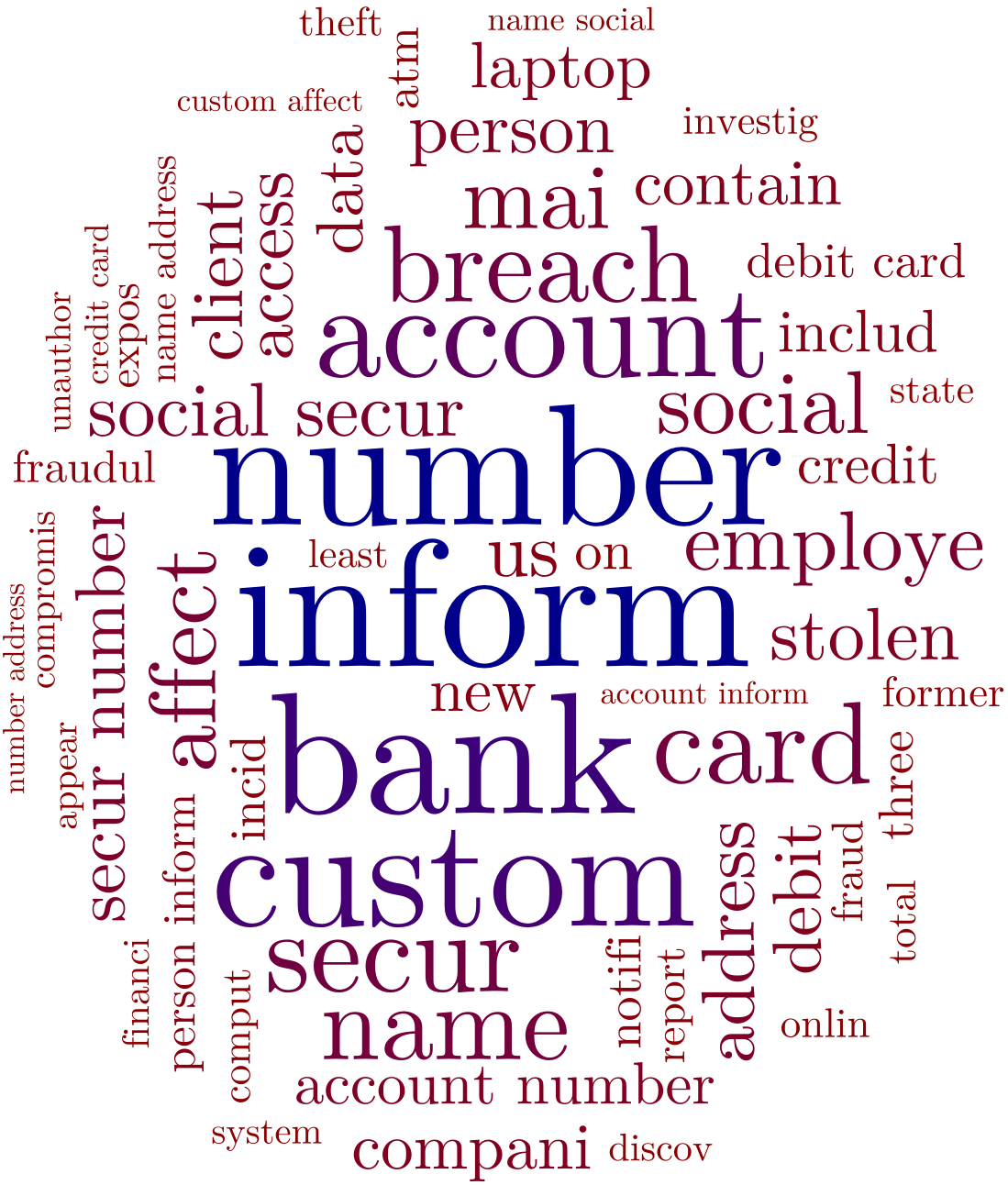
Table VIII
Bank capital and interbank activity

The table reports the coefficient estimates of fixed effects regressions. The dependent variables in Columns (1) to (3) are total bank equity, interbank assets and interbank liabilities, respectively, and are measured on the log scale. Privacy breaches are defined as data breaches that lead to loss of social security numbers, addresses, names, or any personal information; all non-privacy breaches in our sample are classed as other breaches. Observations in the year before and after data breaches are included for the respective breached banks and controls. Standard errors are clustered by both bank and time. ***, ** and * denote levels of significance at 1, 5 and 10 percent, respectively. The data are sourced from FDIC SDI and are at a quarterly frequency.

	Total bank capital	Interbank assets	Interbank liabilities
	(1)	(2)	(3)
Post	0.041**	0.222	-0.262
	(0.02)	(0.16)	(0.20)
Post × Privacy breach	-0.045**	0.004	0.462**
	(0.02)	(0.25)	(0.19)
Post × Other breach	0.043	-0.334	0.633
	(0.03)	(0.33)	(0.44)
Controls	Yes	Yes	Yes
Bank f.e.	Yes	Yes	Yes
State-time f.e.	Yes	Yes	Yes
R^2	0.100	0.018	0.037
Observations	1120	703	1120

A. Word cloud: data breach descriptions

This appendix provides a word cloud of Privacy Rights Clearinghouse (PRC) breach descriptions for breaches occurring in our sample of U.S. depository institutions. Stop words are removed and the stems of the remaining words are extracted using the Porter Stemming Algorithm. Frequencies are computed for single or at most two contiguous words. Words occurring with a frequency of less than 5 are omitted for clarity of display.



B. Privacy breaches in depository institutions

This appendix shows a sample of privacy breaches of U.S. depository institutions in our data. Privacy breaches are defined as data breaches with criminal intentions that lead to loss of sensitive customer data (customer addresses, social security numbers and financial information). The information is sourced from the database of breaches maintained by the Privacy Rights Clearinghouse (PRC).

PRC-recorded public announcement date	Institution or holding name	Breach type	Short breach description
29 Mar 2017	CFG Community Bank	Phishing	Phishing attack in which names, addresses, social security numbers and W2 tax information were compromised.
09 Sep 2014	Ameriprise Financial Services, Inc.	Device	Illicit access to desktop computer through which customer names, birth dates, social security numbers, account number and address were obtained.
17 Jul 2014	Total Bank	Hacking	Hack of bank's computer network which exposed customer names, addresses, account numbers, account balances, social security numbers and driver's license numbers.
04 Mar 2014	Capital One	Insider involvement	Ex-employee illicitly accessed customer accounts and obtained names, account numbers, social security numbers, payment information and other account information.
26 March 2008	Bank of New York Mellon	Accidental loss	The bank lost a box of computer data tapes storing personal information such as names, Social Security numbers and possibly bank account numbers.
22 Jan 2008	Target National Bank	Insider involvement	Three employees illicitly accessed customer accounts and stole names, social security numbers, addresses, account numbers and telephone numbers.
06 Nov 2007	Butte Community Bank	Device	Laptop was stolen from bank that stored names, addresses, social security numbers and bank account numbers.
05 May 2006	Wells Fargo	Device	Desktop computer stolen in transit that contained names, addresses, social security numbers and mortgage loan deposit numbers of existing and prospective customers.

C. Variable definitions

This appendix provides detailed definitions of the variables used throughout the tables.

Variable	Definition	Source
Panel A: Bank variables (quarterly frequency)		
Total deposits	Natural logarithm of one plus the variable *dep*.	FDIC SDI
Insured deposits	Natural logarithm of one plus the variable *depins*.	FDIC SDI
Time deposits	Natural logarithm of one plus the variable *ntrtime*.	FDIC SDI
Interest deposits	Natural logarithm of one plus the variable *depidom*.	FDIC SDI
Retirement deposits	Natural logarithm of one plus the variable *irakeogh*.	FDIC SDI
Money market deposits	Natural logarithm of one plus the variable *ntrmmda*.	FDIC SDI
Demand deposits	Natural logarithm of one plus the variable *ddt*.	FDIC SDI
Savings deposits	Natural logarithm of one plus the sum of $ts - ntrtime$ and $ntrsoth$, where ts captures time and savings deposits, $ntrtime$ captures time deposits and $ntrsoth$ captures other savings deposits.	FDIC SDI
Size	Natural logarithm of the total assets variable *asset*.	FDIC SDI
ROA	Calculated as $roaptx/100$. Winsorized at the 0.01 and 0.99 percentiles. Expressed in percentage points.	FDIC SDI
Liquidity ratio	Calculated as $(chbal+sc-scabs-scmtgbk)/asset5$ and winsorized to lie in the interval $[0, 1]$, where *chbal* equals cash & balances due from depository institutions, *sc* equals total securities, *scabs* equals asset backed securities, *scmtgbk* equals mortgage-backed securities and $asset5$ equals average total assets. Expressed in percentage points.	FDIC SDI
Non-interest income ratio	Calculated as the ratio of non-interest expenditures over the sum of non-interest and interest expenditures, $nonii/(intinc+nonii)$, and winsorized to lie in the interval $[0, 1]$. Expressed in percentage points.	FDIC SDI
Deposit rate	Calculated as the ratio of total interest expenditures over total deposits, $eintexp/dep$. Winsorized at the annual 0.01 and 0.99 percentiles. Expressed in percentage points.	FDIC SDI
Panel B: Breach indicator variables		
Privacy breach	Computed as taking the value one if a bank is breached and the data breach involves loss of depositors' social security numbers, addresses, names or any personal information; and zero for all other banks.	PRC
Other breach	Computed as taking the value one if a bank is breached and the data breach involves loss of sensitive data, however, does not involve the loss of depositors' social security numbers, addresses, names and any personal information; and zero for all other banks.	PRC
Post	Computed as taking the value one for observations of breached (control) banks after the actual (matched) data breach has been made public, and zero otherwise.	PRC

(Continued)

Variable	Definition	Source
Panel C: Branch variables (yearly frequency)		
Branch deposits	Natural logarithm of the total branch deposits variable *DEPSUMBR*.	FDIC SOD
HQ county	Computed as taking the value one if a bank's branch is located in the same county as its institutional headquarters, and zero otherwise.	FDIC SOD
Panel D: Branch deposit rate variables (weekly frequency)		
06MCD10K	Deposit rate (in percentage points) offered by rate-setting branches on new 6-month certificates of deposit with a minimum account size of $10,000.	RateWatch
12MCD10K	Deposit rate (in percentage points) offered by rate-setting branches on new 12-month certificates of deposit with a minimum account size of $10,000.	RateWatch
24MCD10K	Deposit rate (in percentage points) offered by rate-setting branches on new 24-month certificates of deposit with a minimum account size of $10,000.	RateWatch
36MCD10K	Deposit rate (in percentage points) offered by rate-setting branches on new 36-month certificates of deposit with a minimum account size of $10,000.	RateWatch
SAV2.5K	Deposit rate (in percentage points) offered by rate-setting branches on new savings accounts with a minimum account size of $2,500.	RateWatch
FIXIRA0K	Deposit rate (in percentage points) offered by rate-setting branches on new fixed-rate IRA accounts with no limits on the minimum account size.	RateWatch
VARIRA0K	Deposit rate (in percentage points) offered by rate-setting branches on new variable-rate IRA accounts with no limits on the minimum account size.	RateWatch
INTCK0K	Deposit rate (in percentage points) offered by rate-setting branches on new interest-bearing checking accounts with no limits on the minimum account size.	RateWatch
MM2.5K	Deposit rate (in percentage points) offered by rate-setting branches on new money market deposit accounts with a minimum account size of $2,500.	RateWatch
MM25K	Deposit rate (in percentage points) offered by rate-setting branches on new money market deposit accounts with a minimum account size of $25,000.	RateWatch

The Cost of Privacy Failures: Evidence from Bank Depositors' Reactions to Breaches

Online Appendix

Christian Engels* Bill B. Francis† Dennis Philip‡

This draft: June 2022

*We are grateful to Irem Erten, Olivier De Jonghe, Gabriele Lattanzio, Nadya Malenko, Teodora Paligorova, Andre Silva, Dee Warmath and seminar and conference participants at the Bristol Workshop on Banking and Financial Intermediation, Edinburgh Economics of Financial Technology Conference, FINEST Autumn Workshop, Westpac Massey Fin-Ed Centre Conference for helpful comments and suggestions. Any errors remain our own.

†Centre for Responsible Banking & Finance, University of St Andrews, Gateway Building, North Haugh, St Andrews KY16 9AL, UK. E-mail: ce50@st-andrews.ac.uk.

‡Lally School of Management, Rensselaer Polytechnic Institute, Troy, NY, U.S.A. Email: francb@rpi.edu.

§Durham University Business School, Department of Economics and Finance, Mill Hill Lane, Durham DH1 3LB, UK. E-mail: dennis.philip@durham.ac.uk

Table A1
Oster (2019) test of selection on unobservables

The table reports the test results from the Oster (2019) procedure. δ measures how influential omitted factors would need to be (proportional to the included controls) to fully subsume the causal effect of privacy breaches (i.e., Post × Privacy breach = 0). The effect bounds show the range of plausible effects of privacy breaches on the dependent variable. The upper bound of the causal effect is derived from the case of $\delta = 1$, in which excluded factors are as influential as included ones. The lower bound corresponds to the causal effect from the baseline models reported in Table III.

	Dependent variable	Key independent variable	δ	Effect bounds
Model (1)	Total deposits	Post × Privacy breach	3.752	[-0.095,-0.126]
Model (2)	Insured deposits	Post × Privacy breach	3.140	[-0.138,-0.191]
Model (3)	Time deposits	Post × Privacy breach	1.529	[-0.090,-0.248]
Model (8)	Savings deposits	Post × Privacy breach	6.384	[-0.111,-0.127]

Table A2

Parallel trends between privacy breached and control banks

The table reports the dynamic difference-in-differences coefficient estimates obtained in Section III.C together with the F-test of all pre-trend coefficients jointly equating to zero. Privacy breaches are defined as data breaches that lead to loss of social security numbers, addresses, names or any personal information. The dependent variable equals total deposits (in logs) and the key independent variables are the interactions between the privacy breach and the respective event time quarter indicator. Observations in the year before and after data breaches are included for the respective breached banks and controls. Exact definitions of the variables are provided in Appendix C of the paper. Standard errors are included in parentheses and clustered at the bank and time level. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent levels, respectively. The data are obtained from the FDIC SDI and are at a quarterly frequency.

	Total deposits
	(1)
Event time quarter -4 × Privacy breach (δ_{-4})	0.097
	(0.07)
Event time quarter -3 × Privacy breach (δ_{-3})	-0.019
	(0.04)
Event time quarter -2 × Privacy breach (δ_{-2})	-0.011
	(0.06)
Event time quarter 1 × Privacy breach	-0.089**
	(0.04)
Event time quarter 2 × Privacy breach	-0.131***
	(0.05)
Event time quarter 3 × Privacy breach	-0.143***
	(0.05)
Event time quarter 4 × Privacy breach	-0.091*
	(0.05)
Event time indicators	Yes
Controls	Yes
Bank f.e.	Yes
State-time f.e.	Yes
F ($H_0 : \delta_{-4} = \delta_{-3} = \delta_{-2} = 0$)	0.968
Prob > F ($H_0 : \delta_{-4} = \delta_{-3} = \delta_{-2} = 0$)	0.414

Table A3
Longer-term effects after a privacy breach

The table reports the coefficient estimates of fixed effects regressions for different time horizons after a breach incident. The dependent variable is total deposits (in logs). Column (1) includes observations in the year before and 2 years after a breach for the respective breached banks and controls; while Column (2) extends the observations to include 3 years after a breach. Privacy breaches are defined as data breaches that lead to loss of social security numbers, addresses, names or any personal information; all non-privacy breaches in our sample are classed as other breaches. Exact definitions of the variables are provided in Appendix C of the paper. Standard errors are clustered by both bank and time and are reported in parentheses. ***, ** and * denote levels of significance at 1, 5 and 10 percent, respectively. The data are obtained from the FDIC SDI and are at a quarterly frequency.

	Total deposits	
	2 years post breach	3 years post breach
	(1)	(2)
Post	0.062	0.105
	(0.08)	(0.15)
Post × Privacy breach	-0.151**	-0.152*
	(0.06)	(0.08)
Post × Other breach	0.025	0.016
	(0.03)	(0.04)
Controls	Yes	Yes
Bank f.e.	Yes	Yes
State-time f.e.	Yes	Yes
R^2	0.028	0.037
Observations	1652	2148