MIS Quarterly

RESEARCH ARTICLE

BREACHED AND DENIED: THE COST OF DATA BREACHES ON INDIVIDUALS AS MORTGAGE APPLICATION DENIALS[1]

Min-Seok Pang
Wisconsin School of Business, University of Wisconsin-Madison
Madison, WI U.S.A. {mpang9@wisc.edu}

Anthony Vance
Pamplin College of Business, Virginia Tech
Blacksburg, VA U.S.A. {anthony@vance.name}

While a large body of information systems (IS) literature has investigated the antecedents and consequences of data breaches in organizations, we do not have a good understanding of whether a data breach has a material impact on individuals whose private information is compromised and how much damage it causes. We overcome empirical challenges in investigating the impact of data breaches on individual victims by utilizing a unique natural experimental setting that allows us to credibly identify treated and controlled populations—the breach of South Carolina (SC) taxpayer records in 2012. With residents in SC as the treatment group and those in Georgia and North Carolina as the control group, our difference-in-differences estimations find that after the breach at the SC Department of Revenue, there was a significant increase in denials to SC residents' residential mortgage applications for refinance and home improvement. We also find that the adverse impact of the breach was more profound for Black and Hispanic residents. Our study provides significant theoretical and policy implications with respect to the harm and costs of a large-scale data breach.

Keywords: Data breach, South Carolina breach, individual impacts, mortgage refinance, natural experiment

Introduction

How damaging is a data breach to affected individuals? Or does it really matter to them? As data breaches occur with increasing frequency and magnitude (McKinsey, 2022; Security Magazine, 2022), a large body of information systems (IS) literature has investigated the factors that affect the likelihood of data breaches at the organization level and the consequences of data breaches at organizations (Hui et al., 2016). Regarding the former, prior work has identified several antecedents for cybersecurity incidents including cybersecurity investments, information technology (IT) management practices, IT modernization, or meaningful use of electronic health records (e.g., Kwon & Johnson, 2014; Angst et al., 2017; Wang et al., 2019; D'Arcy et al., 2020; Pang & Tanriverdi, 2022). Regarding the latter, the literature has primarily examined the consequences of data breaches on firm performance and stock market valuation (e.g., Kannan et al., 2007; Goel & Shawky, 2009; Malhotra & Malhotra, 2011; Ali et al., 2021; Foerderer & Schuetz, 2022).

However, the cost of data breaches to individuals remains unclear. Although the emotional cost to individuals in the wake of a data breach has been explored (e.g., Bachura et al., 2022; Novak & Viceanu, 2019), the tangible, monetary costs to those whose personal information is compromised remain difficult to identify. What is the magnitude of harm, if any, to those who are affected? Is the impact significant enough to cause material changes in individuals' welfare? For example,

[1] John D'Arcy was the accepting senior editor for this paper. Liangfei Qiu served as the associate editor. The transparency materials for this article are available at https://osf.io/c4a2w/

consider a pending class action lawsuit against Marriott International, Inc. for its 2018 data breach, which affected more than 130 million individuals (Schaffer, 2022). Marriott claims that the lawsuit should not proceed because economic harm to the affected individuals cannot be demonstrated (Schwartz & Cohn, 2022). In contrast, privacy rights organizations argue:

> *The information stored in these databases influences critical life events: whether someone is approved for a mortgage, hired or fired, provided friendly loan terms, able to rent a home, accepted to an educational institution, verified as themselves to confirm important transactions, or subjected to increased police surveillance and investigation. The injury due to the loss of privacy and control over data that has been negligently disclosed is real and concrete. It has clear impacts on us, even when we cannot easily trace them.* (Electronic Frontier Foundation, 2022, pp. 3-4)

To the best of our knowledge, the IS literature on cybersecurity does not provide conclusive answers to this question, which is critical from both theoretical and policy perspectives. Previous studies on the consequences of data breaches have generally found that the financial impact of breaches on firms is minimal, inconsistent, or short-lived in cases of significantly negative impacts (Kannan et al., 2007; Goel & Shawky, 2009; Martin et al., 2017; Ali et al., 2021; Foerderer & Schuetz, 2022). Firms lose market value in the short run after breaches become publicly known, but in most cases, their stock prices recover to pre-breach levels after a relatively short period of time. For example, after its massive data breach in 2013, Target lost considerable market value in early 2014, but its stock price regained its pre-breach level in late 2014. This might be because breached firms do not pay for the full extent of costs (Romanosky & Acquisti, 2009; Moore, 2010; Simon & Omar, 2020), and affected individuals may unfairly bear some of the costs. Yet the extent of these costs to individuals remains unclear. Furthermore, the literature is silent on individual-level heterogeneity in the impact of data breaches, such as by race or ethnicity, and we do not have a good understanding of which populations are most affected by data breaches. Hence, our research questions are as follows:

RQ1: *What is the financial impact on individuals affected by a large-scale data breach?*

RQ2: *Does the impact of a data breach vary by race or ethnicity?*

There are several challenges that hamper rigorous analyses on the matter. First, it is difficult to identify who is affected by a breach and who is not. For privacy and legal reasons, when a data breach occurs, it is challenging, if not impossible, for researchers to know whose information has been compromised. More importantly, it is even more difficult to identify those who were unaffected by the breach. In other words, it is hard to find a control group of individuals who were not affected but are comparable to a treatment group. Even if it were possible to conduct randomized field experiments, it would be unethical to do so. Second, it is not straightforward to measure the consequences of data breaches. It would be hard, if not infeasible, to collect personal financial information about those affected by a breach. Furthermore, the damages, if observable, do not occur immediately but materialize over a period of time. As we explain below, the impacts of data breaches on individuals can be both direct and indirect (Agrafiotis et al., 2018), with the latter effects taking longer to occur.

We overcome these challenges by taking advantage of a unique setting and official datasets. In October 2012, the State of South Carolina (SC) in the U.S. announced that its Department of Revenue (DOR) suffered a massive data breach, in which 6.4 million records of taxpayers and their dependents were stolen by Russian attackers (Knell, 2012; Krebs, 2024). They penetrated 44 DOR systems and downloaded 74.8 gigabytes of taxpayer data (Mandiant, 2012). The entire information of all SC residents who had ever paid state taxes since 1998 was breached. This offers a unique opportunity to investigate the impact of data breaches on individuals. It is possible for us to identify those affected by the breach (SC residents) as the treated group and those outside SC (e.g., the neighboring states of North Carolina and Georgia) as the control group.

This study examines the impact of this data breach on home mortgage applications using a large-scale dataset from the U.S. Consumer Financial Protection Bureau (CFPB). In this paper, we specifically focus on applications for refinancing and home improvement mortgage loans. We obtained individual mortgage application records from 2007-2017 from the CFBP. In the U.S. and many other countries, residential properties are the largest personal asset for most ordinary individuals, most of whom finance their real estate purchases through mortgages (Faig & Shurn, 2002). Therefore, if a data breach compromises their personal information and damages their credit score, they may have difficulty securing loans on their primary residence. For most ordinary citizens, their primary residence is by far their largest personal asset, and housing ownership is an essential means of wealth accumulation and financial stability (Kim, 2000; Blanchett, 2017). Therefore, being denied for a mortgage application has serious consequences not only for individuals but also for society at large. Using this dataset, we examined how the breach at the SC DOR affected the home mortgage applications of SC residents.

We adopted a difference-in-differences (DID) approach to estimate the impact of the SC DOR breach. For credible causal identification, we took the following considerations into account. First, we used residents from two neighboring states—Georgia (GA) and North Carolina (NC)—as a control group. As southern states, their historical, political, and economic environments are similar to the treated state (SC). We further limited the sample to residents living in the counties in GA, NC, and SC located along the state borders, following a spatial regression discontinuity design (Egger & Lassmann, 2015; Keele & Titiunik, 2015; Gonzalez, 2021). Second, we only considered mortgage applications for refinancing or home improvement for owner-occupied properties, excluding applications for new home purchases. Third, we utilized a variety of econometric methods such as linear time trends, matching, and counterfactual estimation techniques to rule out as many confounders, endogeneity concerns, and alternative explanations as possible.

Our analyses revealed that following the SC DOR breach in 2012, there was a significant increase in denials to mortgage applications for refinancing/home improvement in SC in 2013-2017, compared to before the breach and in the neighboring states. Economically, we found that there was an increase of $179 per capita per year in the monetary amount of denied mortgage applications, or a 22.1% increase compared to before the breach. Strikingly, the results indicate that the breach had a greater adverse impact on Black and Hispanic applicants. Due to historical injustices perpetuated by discriminatory lending practices over the centuries, their loans are more likely to be denied than others (LaCour-Little, 1999; Ross & Turner, 2005; Pager & Shepherd, 2008), and we found that the SC DOR breach in 2012 appears to have exacerbated this troubling trend. We show that our core findings remain stable to an extensive series of robustness checks. We also provide supplementary evidence for potential underlying mechanisms by utilizing data on the credit scores of approved loans, wire fraud incidents, and consumer complaints against financial institutions.

Our study significantly contributes to the IS literature by finding that a large-scale data breach can have a tangible, material impact on the well-being of victims. While the IS literature has developed research streams on the drivers of cybersecurity incidents and their firm-level consequences, we do not have a sufficient understanding of individual-level outcomes. This is a significant gap in the literature that may explain why organizations with cybersecurity failures are not held fully accountable for the costs of breaches. Our work also significantly contributes to the literature by theorizing and empirically demonstrating the discriminative impacts of data breaches on individuals, which are new insights in IS research. We hope that this study sparks interest in further research on uncovering other types of harm to individual victims.

This research also provides significant managerial and policy implications. By examining the degree of harm to data breach victims, we offer crucial insights for a variety of audiences such as policymakers, cybersecurity managers, the insurance industry, and the legal profession. As noted above, short-lived losses to market value after data breaches indicate that firms do not bear the full price of breaches, suggesting the existence of market failures that warrant policy interventions. In devising policy measures to hold breached firms accountable for their failures, policymakers can take our findings into consideration to ensure that the cost of cybersecurity failures is properly reflected in the market value (Ali et al., 2021; Foerderer & Schuetz, 2022). Our findings may also be informative to cybersecurity insurers. While there is increasing interest in cybersecurity insurance, a lack of concrete data on cybersecurity risks and damages is hampering the effective development of cybersecurity insurance policies. Therefore, our research can serve as a yardstick for the industry. Lastly, our findings could be used in data breach litigations to adjudicate compensation for data breach victims.

Theoretical Development

Prior Work

In the IS and cybersecurity literature, a large body of work has investigated the impact of cybersecurity incidents on breached firms (e.g., Kwon & Johnson, 2014; Angst et al., 2017; Wang et al., 2019; Kim & Kwon, 2019; D'Arcy et al., 2020; Haislip et al., 2021; Pang & Tanriverdi, 2022). However, the literature has paid limited attention to the consequences for individuals who are affected by data breaches. Among the few studies that do, Ke et al. (2022) found that following a data breach at the University of Washington Medical Center, there was a significant decline in hospital visits by the affected patients, compared to the unaffected control patients, and this decline was more pronounced for outpatient visits and for those who had alternative access to healthcare. Janakiraman et al. (2018) investigated customers' reactions to a data breach at a multichannel retail firm and found that those affected by the breach reduced their spending with the firm and migrated to an alternative, unaffected channel.

Similar to our study, Mikhed and Vogan (2018) examined the impact of the SC DOR breach in 2012. They found that immediately following the 2012 breach, many SC residents signed up for identity protection and credit monitoring services offered by the state government, but this change lasted for less than one year. In addition, their behaviors in credit use did not change significantly. This finding differs from ours, as we found that an increase in mortgage application denials for SC residents persisted until 2017. We revisited the findings of Mikhed and Vogan (2018) in comparison to ours below.

Hoehle et al. (2022) studied the effectiveness of firms' compensation strategies for customers affected by the Target data breach in 2013. Their surveys of Target customers found that a compensation strategy that meets consumers' expectations is associated with greater perceptions of justice (distributive, procedural, and interactive), which results in greater shopping intentions and more positive word-of-mouth. Bachura et al. (2022) examined the emotional reactions of data breach victims by analyzing thetweets labeled with "#OPMHack" posted following a breach at the U.S. Office of Personnel Management (OPM) in 2015. They found that affected individuals demonstrated different emotional responses (e.g., anxiety, anger, and sadness) at the different stages of breach revelations. Once again, our work differs from prior work because it investigates direct monetary harm to individual victims after a large-scale breach.

The 2012 Breach at the South Carolina Department of Revenue (SC DOR)

According to an investigative report provided by Mandiant, a cybersecurity consulting firm contracted by the state following the 2012 breach, multiple employees of the SC DOR received a malicious phishing email on August 13, 2012. One of the recipients clicked the link to a malicious file and installed malware, which stole the recipient's username and password. On August 27, attackers used the stolen credential to log on to the employee's workstation via a remote access service. In the following weeks, the attackers obtained further credentials to several servers, installed backdoors to them, and exfiltrated a large number of database backup files. The SC DOR was not aware of these activities until the U.S. Secret Service notified it of the breach on October 10. On October 26, the governor publicly announced the breach to the public.

In total, the attackers accessed 44 systems using 33 malicious components and acquired confidential files totaling 74.7 gigabytes. The entire tax records, including Social Security numbers, of as many as 3.8 million individuals with state tax records going back to 1998 were compromised, as were the tax records of 1.9 million dependents and 700,000 businesses in SC. A total of 387,000 credit and debit card numbers were also exposed. At the time, it was the largest breach at a state government agency (Holdman, 2022). The state spent $20 million to provide fraud protection services to those who were affected (Self, 2018). Although the attackers remain uncharged, evidence implicates Russian cybercriminals (Krebs, 2024).

The Effect of Data Breaches on Individuals' Finances

A data breach can have an adverse effect on the credit and finances of victims in a variety of ways. First, the private information of individuals, such as names, addresses, and personal identification numbers (e.g., social security numbers) can be used by hackers for fraud or other malicious acts, damaging victims' credit and financial records. Criminals can use this information to open new bank accounts, credit cards, online payment services, or loan accounts without victims' knowledge (Anderson, 2006; Anderson et al., 2008; Harrell, 2019), and they may make large purchases or obtain loans without repaying them. Such fraudulent actions can significantly deteriorate victims' credit scores. They might not even be aware that their information has been used for identity theft (Hoar, 2001), and while they can request that financial institutions correct or remove such records, it is often cumbersome and time-consuming to do so (Newman & McNally, 2005). Damage to victims' credit scores can make it difficult for them to secure loans or mortgages in the future (Hoar, 2001).

Second, victims of data breaches may also experience direct monetary losses (Anderson, 2006; Anderson et al., 2008). Hackers may use victims' personal information to access their bank accounts and withdraw funds. As mentioned above, hackers might also use victims' credit card numbers for fraudulent purchases or use victims' information to secure loans illegally. While victims can dispute bogus transactions with financial institutions, they may end up paying for fraudulent purchases and loans themselves, resulting in direct and unwarranted financial losses (Hoar, 2001; Newman & McNally, 2005; Harrell, 2019). Victims' personal information may be used in other illegal activities such as larceny or drug trafficking, resulting in further financial costs (Newman & McNally, 2005; Anderson et al., 2008).

Third, victims may also experience indirect costs (Harrell, 2019). According to surveys on identity theft victims, some have had to deal with a range of issues beyond financial troubles—involving criminal investigations or civil lawsuits, for example—requiring them to hire lawyers or private investigators (Newman & McNally, 2005; Anderson, 2006; Harrell, 2019). Victims may also experience housing difficulties, such as being turned down for a lease, resulting in more financial losses in the form of moving or other expenses (Newman & McNally, 2005; Anderson et al., 2008; Electronic Frontier Foundation, 2022). Such indirect costs can be formidable for individuals with low incomes or little savings.

Direct and indirect costs for victims of a data breach can also prevent individuals from being able to refinance their home mortgages, which is often a costly endeavor. According to

Freddie Mac, the average amount of closing costs for refinancing was as much as $5,000 in 2022.[2] Therefore, a refinancing application would more likely be denied if an applicant does not have sufficient cash available for the closing costs.

Discriminative Impacts of Data Breaches by Race or Ethnicity

We propose that a data breach causes greater harm to minority populations. To the best of our knowledge, there is little research on the discriminative impacts of data breaches on minorities in the IS or cybersecurity literature. We contribute to both by theorizing the heterogeneous impacts of data breaches on individuals.

A 2021 report titled "Demographics of Cybercrime Report" provides supportive data (Malwarebytes, 2021). Its survey of over 5,000 respondents across the U.S., the U.K., and Germany found that "BIPOC (Black, Indigenous, and people of color) people had the lowest rate of successfully avoiding any financial impact due to cybercrime. Only 47 percent of BIPOC respondents avoided any financial impact compared to 59 percent of all respondents." In addition, criminals appear to be more likely to target minority victims. According to the report, "5 percent of BIPOC consumers had their social media accounts hacked, as opposed to 40 percent of White consumers, and more BIPOC consumers said they had their identities stolen (21 percent) than White consumers (15 percent)."

Prior research documents that minorities are more likely to have limited credit histories and subpar credit scores for several reasons (Henderson et al., 2015; Ards & Myers, 2001; Mack, 2018). Due to historic discriminative treatments by governments, financial institutions, and others (e.g., real estate agents), minorities have been deprived of economic opportunities to accumulate wealth via entrepreneurship, education, or home ownership (Markkanen & Harrison, 2013; Olzak & Shanahan, 1996). For example, exclusionary housing policies have prevented minorities from obtaining mortgages and purchasing homes, which would have allowed them to establish good credit histories (Whittemore, 2021; Freeman, 2000). School segregation, a lack of quality public education, and other discriminative policies hampered the education of minority pupils, limiting their future job prospects (Sikkink & Emerson, 2008; Logan et al., 2008). This discrepancy in wealth, education, and housing has persisted over generations. Prior research on the digital divide also indicates that minorities have limited access to broadband or mobile internet access (Prieger, 2015; Campos-Castillo, 2015).

To understand how a data breach affects individual victims differently by race or ethnicity, we lay out what steps individuals can take to prevent or mitigate damages to their credit after a breach (Figure 1). Once an incident occurs (Step 1 in Figure 1), victims should seek to understand whether their information has become compromised (Step 2) (Chua et al., 2021; Bachura et al., 2022), which can done through news reports or notifications from the breached organization, the government, or law enforcement agencies. Victims can take preventative measures such as acquiring credit monitoring or identity theft protection services (Step 3), which are often offered by the breached organization as a corrective measure (Mikhed & Vogan, 2018; Archer, 2012). Victims should remain vigilant and constantly monitor their financial and credit scores so that if a criminal uses a victim's information for fraudulent purposes, the victim will become aware of this (Step 4). Then, impacted victims should report the incident to law enforcement agencies (Step 5) and take remedial actions with their financial institutions (Step 6). We propose that at each step of the way, minorities face a range of barriers, which can be categorized into (1) disparity in resource endowments, (2) distrust of authority, and (3) discriminative treatment.

First, with respect to resource disparity, because minority populations are more likely to lack broadband or mobile internet access (*a* in Figure 1) (Prieger, 2015; Campos-Castillo, 2015; Fairlie, 2004), it may take longer for them to learn that they have been victimized by a data breach and are vulnerable to resulting fraud (Step 1). For example, if a breached firm notifies affected individuals electronically, victims with little digital access would not become aware of the breach immediately. This would also be the case if their information had been used for fraudulent purposes. Therefore, victims lacking appropriate connectivity may miss opportunities to take quick corrective actions to mitigate damages (Step 3).

Second, in the case of the SC DOR breach in 2012, as mentioned above, the state offered credit monitoring and fraud protection services from a third-party vendor to all SC residents (Step 3; Mikhed & Vogan, 2018). However, because of historic unjust treatment and distrust of authority (*b* in Figure 1) (Arnett, 2020; Gofen et al., 2021; Pass et al., 2020; Howell & Fagan, 1988; Weitzer, 2014; Nuño, 2018), minority residents in SC may have viewed such credit "monitoring" services skeptically, as an attempt by the government to surveil them. Therefore, minority residents in SC may have been less inclined than other residents to utilize free credit protection services that could have mitigated damages to their credit.

[2] https://myhome.freddiemac.com/refinancing/costs-of-refinancing

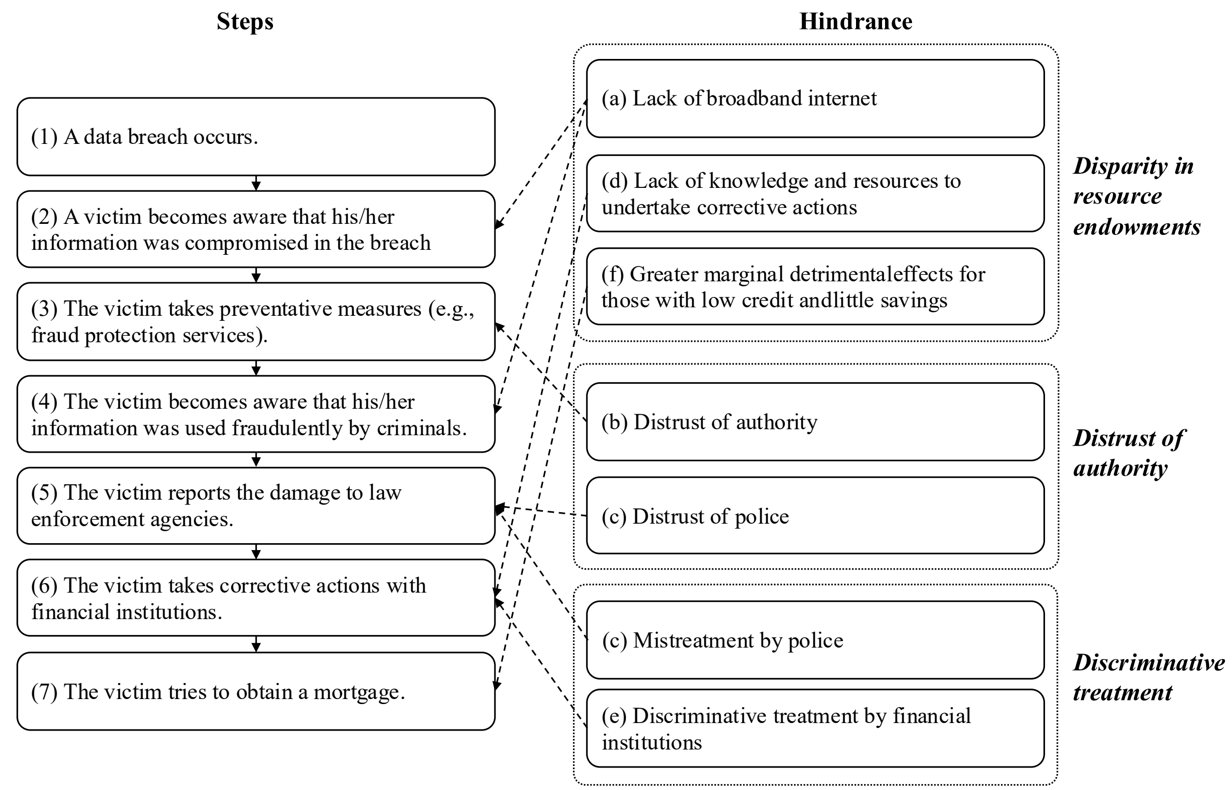


Figure 1. Impact of Data Breaches

Third, the criminology literature documents that minority crime victims are less likely to call the police (Avakame et al., 1999; Vargas & Scrivener, 2021; Wu & Miethe, 2022), due to distrust of authority and perceived threat from the police (McDonald & Stokes, 2006; Pass et al., 2020; Tyler, 2005). In addition, the literature finds that law enforcement agencies often exert different efforts to solve crimes against minorities compared to others (Bachman, 1994; Howerton, 2006). For example, police response time to crimes can differ by victims' race or ethnicity. This may also apply to cybercrimes. Therefore, minority victims may not believe cybercrimes are worth reporting to the police because they may believe that the police would not take their complaints seriously and try to apprehend the offenders (*c* in Figure 1). They may also fear "victim-blaming," in which the police attribute the problems to victims' carelessness and neglect (Dukes & Gaither, 2017; Clark, 2021; Lave, 2022). Thus, the perception that minority victims may be less likely to report crimes and that law enforcement agencies may treat minority victims discriminatively makes them an attractive target for criminals, who may believe that by targeting minority victims they can reduce their chance of apprehension (Becker, 1968).

Fourth, there is a disparity in resources among different populations regarding financial literacy. When individuals are victimized by identity thefts or other frauds, they are entitled by the Fair Credit Billing Act to dispute fraudulent transactions with financial institutions and request corrections to their records. However, this process requires time, effort, and resources, requiring data breach victims to (1) assemble evidence or documentation to demonstrate that they were targeted by fraud, (2) find out how to resolve the problem, whom to contact, and how to contact them, and (3) communicate with financial institutions. Angrisani et al. (2021) and Al-Bahrani et al. (2019) report a racial gap in financial literacy, indicating that minority victims may have limited financial literacy and may thus face greater challenges in reversing damages. As mentioned above, many minorities do not have adequate broadband or mobile internet access, making it more difficult to rectify damages. Furthermore, minorities are more likely to have jobs that are unstable or require long hours to make ends meet (Lu et al., 2023; McDaniel & Kuehn, 2013), resulting in less time to engage in the necessary activities to recover from fraud victimization. Because of these reasons, minorities may encounter greater difficulty in limiting the damages from fraud (*d* in Figure 1).

Fifth, the literature on financial inclusion documents suggests that, like law enforcement agencies, financial institutions such as banks, credit card companies, or credit agencies treat minority customers discriminatively (*e* in Figure 1) (Ambrose et al., 2021; Black et al., 1978; Lindley et al., 1984; LaCour-Little, 1999; Ross & Turner, 2005; Pager & Shepherd, 2008). Accordingly, we expect that when victims of data breaches attempt to resolve credit and financial problems, financial institutions treat minority victims less favorably than others. Haendler and Heimer (2021) found that consumer complaints against financial institutions from low-income or minority-concentrated zip codes are less likely to be resolved. Therefore, when breach victims register complaints that their personal information has been used for frauds or when they are charged for purchases or loans they did not initiate, financial institutions may handle these claims from minority customers differently. They may question the credibility of their complaints, require more onerous justifications, or take more time to resolve the complaints. Such unequal treatment of minority victims by financial institutions makes it more difficult for them to recover from damages from data breaches.

Lastly, minorities may be more negatively impacted by a data breach because they tend to have lower credit scores and smaller savings accounts (*f* in Figure 1) (Henderson et al., 2015; Ards & Myers, 2001). Given that a mortgage lender often makes approval decisions based on a certain credit score cutoff (Laufer & Paciorek, 2022; Bubb & Kaufman, 2014), someone with a high credit score may not be significantly affected by the breach. On the other hand, for an individual with a low credit score, a breach may cause their credit score to fall below the cutoff point, making it less likely for their loan application to be approved. Further, while following a breach, any victim may have to pay a higher interest rate or higher closing costs; however, individuals with a high credit score and more savings would be better able to manage these costs, whereas a loan may become unaffordable for individuals with little savings and a low credit score, even if they are approved for a loan.

For these reasons, which are illustrated in Figure 1, we expect that the SC DOR breach in 2012 had disproportionately harmful impacts on minority residents in SC with respect to mortgage refinance and home improvement loan applications.

Empirical Methods

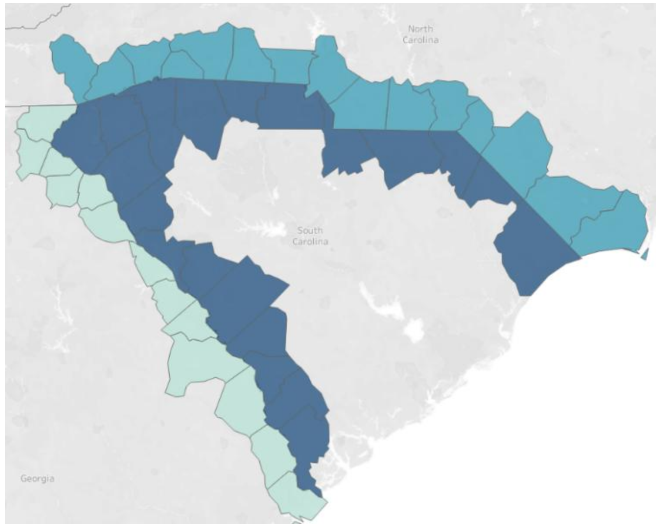
Data

To test the impact of the SC DOR breach on home mortgage application decisions, we utilized the Home Mortgage Disclosure Act (HMDA) dataset from the Consumer Financial Protection Bureau (CFPB). Enacted in 1973 to shed light on discriminatory lending practices, the HMDA requires financial institutions that originate residential mortgage loans to periodically submit all loan application records to the CFPB, which in turn makes anonymized records publicly available. We downloaded the records for 2007-2017 from the CFPB, yielding more than 187 million application-level records. The information available in this dataset included property location as a census tract, property type (e.g., single-family, multi-family), loan purpose (e.g., home purchase, refinancing), applicant information (e.g., race, ethnicity, gender), and decisions (approved or denied). It did not provide information on interest rates or credit scores.

With the SC DOR breach in 2012 as the treatment, we conducted difference-in-differences (DID) estimations at two levels: census tracts-year and loan applications. For the former, we aggregated application records to the census tract-year level. Note that the HMDA dataset only provides the loan application year, not the month or day. The census tract-year-level analyses allowed us to better control for various observed and unobserved heterogeneity, but we also utilized individual loan-level analyses to uncover the moderating effects of applicant race or ethnicity.

For the treatment group, we used loan applications in SC; for the control group, we used those of two neighboring states—North Carolina (NC) and Georgia (GA) (Figure 2). We chose these two states because they are all southern U.S. states and thus share many commonalities in terms of historical, demographic, and economic characteristics.

For more credible causal identification, we further narrowed down our sample in several ways, as specified in Table 1. First, we considered the counties at the borders of SC-NC and SC-GA (Figure 2) only because these border counties in SC are more likely to be similar to neighboring counties in NC and GA in terms of the demographic and socioeconomic conditions that affect mortgage decisions. Through this approach, we sought to ensure as much geographical homogeneity as possible. We followed a spatial regression discontinuity design (RDD) to limit the effect of unobserved heterogeneity, which has been widely used in recent economics studies. For example, Gonzalez (2021) used areas with cellular coverage to examine the impact of cellphone access on corruption in Afghanistan by comparing observations inside and outside of the mobile coverage area. Keele and Titiunik (2015) used the boundary between the New York and Philadelphia media markets to examine voter turnout in New Jersey. Hidano et al. (2015) adopted a similar approach to examine housing prices in high and low seismic-risk zones in Tokyo.



Note: SC = 280 census tracts in 20 counties; GA = 108 census tracts in 13 counties; NC = 174 census tracts in 15 counties

Figure 2. The Border Counties in South Carolina, North Carolina, and Georgia

Table 1. Sample Construction Procedures

Total loan application records in 2007-2017	187,462,446
Loan applications in South Carolina, North Carolina, and Georgia	14,803,622
Loan applications in the border counties (Figure 1)	3,247,810
Loan applications for owner-occupied properties	2,848,094
Loan applications for refinancing or home improvement	1,621,529

Second, we limited our sample to loan applications for the purpose of refinancing or home improvement and excluded applications for home purchases. We did so to better meet the stable unit treatment value assumption (SUTVA) (Rubin, 1980), which requires that the outcome of the treatment unit be unaffected by the control unit and vice versa. We excluded home purchases to rule out the possibility of residents in SC who were affected by the 2012 breach moving to GA or NC after 2012 and purchasing real estate properties there and vice versa. It is relatively unlikely that the residents of these states would have moved across the border, purchased a new home, and subsequently refinanced their mortgage all within the 2013-2017 period. We also limited the sample to loan applications for owner-occupant properties, excluding cases in which SC residents owned properties in NC or GA for investment purposes or vice versa. For the census tract-level panel, we only used tracts that had refinancing applications in all years between 2007 and 2017, so that we could utilize a balanced panel.

We acknowledge a few issues that could threaten causal identification. For example, SC does not have tax reciprocity agreements with NC and GA,[3] meaning that residents of NC or GA who have income from SC are required to pay state income taxes to SC. The breach might have compromised such GA and NC residents who filed income or business taxes with SC, and their credit scores might have been affected as well. In these cases, however, we expected the treatment effect to be underestimated. Additionally, as discussed above, after the DOR breach, the state government offered a credit monitoring service to all SC residents. While such a service cannot completely prevent damages to individuals, it could mitigate the likelihood or degree of damages to some extent. Had it not provided such services, we would have presumably observed a stronger treatment effect.

[3] https://www.ncdor.gov/taxes-forms/individual-income-tax/credit-income-tax-paid-another-state-or-country and https://dor.georgia.gov/filing-residents-nonresidents-and-part-year-residents-faq

Table 2. Census-Tract Level: Variable Definitions and Data Sources		
Variables	**Descriptions**	**Data sources**
Dependent variables		
Denial per capita	Number of loan application denials per thousand population	HMDA database
Denial amount per capita	Amount ($) of denied loan applications per capita (in 2012 dollars)	
Independent variable		
Treatment	1 for census tracts in South Carolina	
Post	1 for 2013-2017, 0 for 2007-2012	
Control variables		
Log(# of applications)	Log(the number of applications in census tract)	HMDA database
Log(total amount)	Log(total amount ($) of applications)	
Black applicants	Ratio of Black loan applicants	
Hispanic applicants	Ratio of Hispanic loan applicants	
Male applicants	Ratio of male applicants	
Applicant income	Average applicant's annual gross income in thousand $	
Log(population)	Log(total population in census tract)	ACS
Median age	Median age of census tract population	
Household per capita	Number of households per capita	
Median income	Household median income in thousand $	
Housing unit per capita	Number of housing units per capita	
Owner-occupied housing	Ratio of owner-occupied housing units	
Housing unit age	Housing unit median age	

Table 3. Loan Level: Variable Definitions and Data Sources		
Variables	**Descriptions**	**Data sources**
Dependent variables		
Denial	1 if a loan is denied	HMDA database
Independent variable		
Treatment	1 for loan applications in South Carolina	
Post	1 for 2013-2017, 0 for 2007-2012	
Control variables		
Black applicant	1 if applicant is Black	HMDA database
Hispanic applicant	1 if applicant is Hispanic	
Male applicant	1 if applicant is male	
Applicant income	Applicant's annual income in thousand $	

Note: Other control variables are the same as the census-track level analyses (Table 2).

Measures

Table 2 lists the variables for the census tract-level analyses. The primary dependent variables are the number of loan application denials for refinancing/home improvement per capita in a census tract (*denial per capita*) and the total monetary amount of denied loan applications per capita (in 2012 dollars) (*denial amount per capita*). We use these variables for ease of interpretation of coefficients. For the DID estimations, *treatment* is represented by a dummy variable of 1 for census tracts in SC, and *post* is represented by 1 for observations after 2012.

To account for observable heterogeneity in the estimations, we used various covariates using the HMDA dataset and the American Community Survey (ACS) from the U.S. Census Bureau, as listed in Table 2. Using the HMDA dataset, we counted the number of total mortgage applications in each census tract-year and the total amount (in 2012 dollars) of loan applications. The HMDA records also provided the race/ethnicity of loan applicants, with which we measured the ratio of minority applicants (Black and Hispanic) over total applicants in each census tract. Using the ACS data, we controlled for several census tract-level demographic and socioeconomic characteristics that explain housing market conditions, such as residence age, income, and housing unit characteristics. Tables A1 and A2 in Appendix A provide summary statistics and correlations, respectively.

Table 3 lists the individual loan application-level variables. The dependent variable (*denial*) is a dummy variable that is 1 if a loan application is declined. The *treatment* and *post* variables were measured in the same manner as in the census tract-level analyses. We also used similar control variables for the loan-level analyses to measure census tract-level characteristics (e.g., median age or income). Using the HMDA records, we used the dummy variables of applicants' race/ethnicity as well as their income. Tables B1 and B2 in Appendix B provide summary statistics and correlations of the loan-level variables, respectively.

Estimation Approaches

For the census tract-level analyses, we estimated the following model.

$$Y_{ij} = \alpha + \beta\, Treatment_i \times Post_t + \beta_X X_{it} + \gamma_i + \delta_t + \varepsilon_{ij}, \qquad (1)$$

where *i* and *t* indicate census tracts and years, respectively. Y_{it} is the dependent variable (*denial per capita* or *denial amount per capita*), and X_{it} are the observable covariates. The coefficient of *Treatment* × *Post* shows the average treatment effect (ATE), which is the difference in outcomes between the treated and the control units after the treatment. If the breach in 2012 adversely affected SC residents financially, as argued in Section 2, the ATE would be expected to be positive and significant. To control for unobserved time-invariant heterogeneity, Equation (1) includes census tract fixed effects (γ_i). In addition, year fixed effects (δ_t) capture unobserved nationwide temporal changes in economic and housing market conditions. To control for unobserved time-variant heterogeneity, the estimations included metropolitan statistical area (MSA)-specific linear and quadratic time trends.

In addition to Equation (1), we estimated the following leads-and-lags model to test the parallel trend assumption (Angrist & Pischke, 2009).

$$Y_{ij} = \alpha + \sum \beta_{-k} Post_{t-k} \times Treatment_t + \sum \beta_{+k} Post_{t+k} \times Treatment_t + \beta_X X_{it} + \gamma_i + \varepsilon_{ij} \qquad (2)$$

If the parallel trend assumption holds, we would expect the coefficients of leading variables (β_{-k}) to be insignificant, indicating that there is no significant difference in mortgage denial patterns between the treated (SC) and the control units (GA, NC) before the breach.

For the loan-level analyses, we used a similar estimation model (Equation 3).

$$Y_i = \alpha + \beta\, Treatment_i \times Post_t + \beta_X X_{it} + FE + \varepsilon_{ij} \qquad (3)$$

Y_i is a dummy variable that is equal to 1 if a loan application is denied. The ATE is measured again by *Treatment* × *Post*. In addition, to measure heterogenous treatment effects with respect to applicants' race/ethnicities, we included three-way interaction variables with *Treatment* × *Post* and Black/Hispanic indicators (Table 3). In addition to census tract and year fixed effects, the loan-level analyses included a range of other fixed effects including MSA, loan types (conventional, FHA insured, VA guaranteed), loan purposes, loan originators, loan purchasers, preapproval status, and lien status. We also included county-specific linear time trends.

Our estimations took careful empirical approaches, such as limiting the sample to the border counties and refinancing/home improvement loans (Table 1) and utilizing multiple fixed effects and linear/quadratic time trends to rule out as much unobserved heterogeneity as possible and to mimic a randomized experiment to the extent possible. Nonetheless, there are other endogeneity threats that can potentially hinder causal identification. There may still have been remaining unobserved heterogeneity that affected both the treatment and the outcomes. A significant treatment effect can be driven by chance, or there may be other alternative explanations behind the effect. To address these concerns, we presented a variety of robustness checks in the following sections, including counterfactual estimations, coarsened exact matching, random shuffle tests, falsification tests, and nonlinear estimations.

Results

Main Estimation Results

Table 4 presents the results of the census tract-level estimations with the border counties, as depicted in Figure 2. For both dependent variables, the coefficients of Treatment × Post are positive and significant, indicating that after the breach at the SC DOR in 2012, there was a significant increase in denials of refinancing/home improvement loan applications from SC residents. The coefficient of *Treatment* × *Post* in Column 2 indicates that after the breach, there was a $179 increase in the total denied monetary amount of loan applications per capita. The monetary amount of denied applications per capita in SC before 2012 totaled $810.1. Therefore, $179 represents a 22.1% increase in the denial amount.

Table 5 presents the leads-and-lags model estimations; as illustrated in Figures 3, we did not observe significant pre-treatment effects. Column 1 (*denial per capita*) of Table 5 shows that the impact of the DOR breach became significant only after 2015, while in Column 2 (*denial amount per capita*), the treatment effect became significant immediately after 2012. As noted previously, Mikhed and Vogan (2018) found a change in SC residents' credit usage behaviors after 2012, which did not last very long. Their finding would explain the insignificant impact in Column 1 in 2013-2014. Furthermore, criminals might not have utilized stolen personal information from the SC DOR in its entirety all at once; rather, they may have used these records incrementally over several years to avoid the attention of law enforcement. In an unreported loan-level analysis, we obtained similar findings showing that the likelihood of refinancing application denials in SC rose in 2013-2014 compared to the pre-breach period, with further increases in 2015-2017.

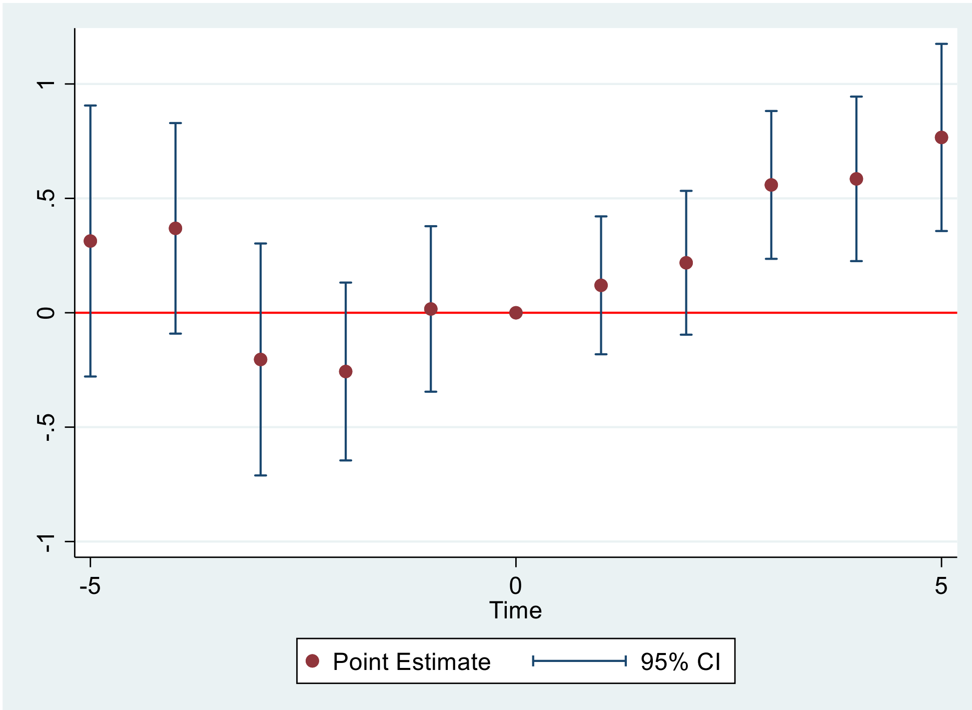
Table 4. Census-Tract Level: Main Difference-in-Differences Estimation Results

Dependent variables	**Denial per capita**	**Denial amount per capita**
	(1)	**(2)**
Treatment × Post	**0.361*** **(0.143)**	**179.0***** **(38.48)**
Log(# of applications)	3.554*** (0.185)	485.9*** (51.83)
Log(total amount)	-0.406*** (0.0937)	15.11 (19.70)
Black applicants	0.636+ (0.345)	182.9** (59.67)
Hispanic applicants	0.525 (0.712)	39.52 (162.9)
Male applicants	0.120 (0.260)	106.6+ (56.86)
Applicant income	-0.000474 (0.00112)	2.363*** (0.457)
Log(population)	-7.435*** (1.407)	-1,600*** (325.4)
Median age	-0.0555** (0.0197)	-11.42* (4.546)
Household per capita	-6.176* (2.510)	-870.3 (742.9)
Median income	0.0104 (0.00853)	3.812+ (2.069)
Housing unit per capita	-6.148* (2.601)	-1,635** (500.2)
Owner-occupied housing	0.936 (1.206)	-70.21 (260.4)
Housing unit age	-0.1591 (0.0117)	-3.434 (3.111)
Census tract FE	YES	YES
Year FE	YES	YES
MSA linear and quadratic time trends	YES	YES
No. of obs.	6,179	6,179
No. of census tract	562	562
R^2	0.6923	0.7281
Adjusted R^2	0.6587	0.6984
F	37.45***	15.41***
Root MSE	1.6669	415.3238

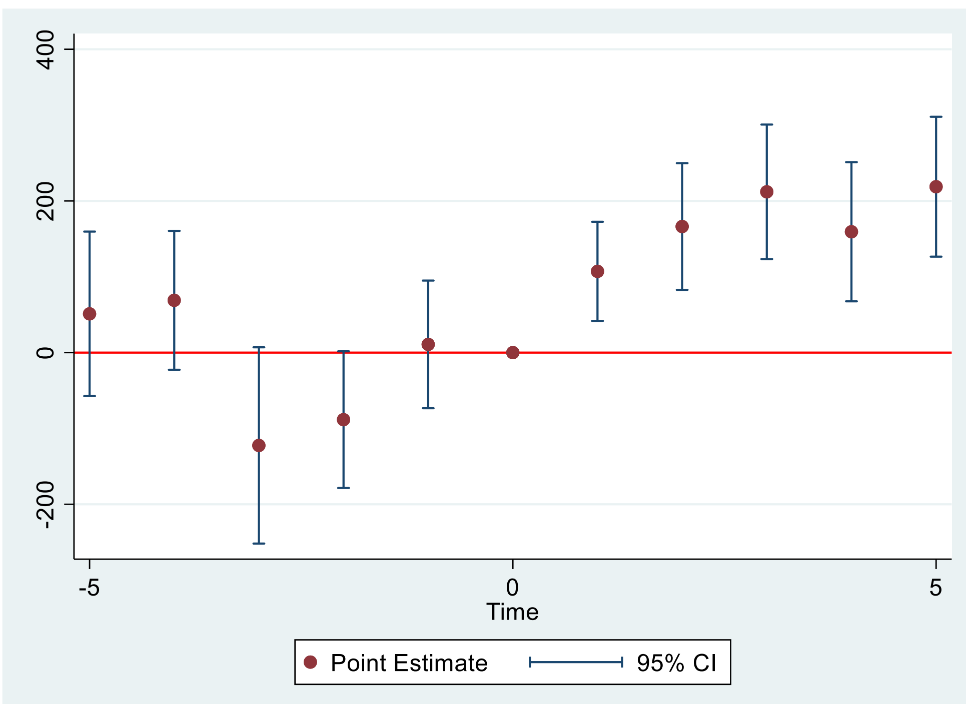
Note: Robust standard errors clustered by census tracts in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. *Treatment* = 1 for census tracts in South Carolina; *post* = 1 for observations after 2012. Unit of analysis = census tract-year (2007-2017); three singleton observations were dropped.

Table 5. Census-Tract Level: Leads-and-Lags Model Estimation Results		
Dependent variables	**Denial per capita**	**Denial amount per capita**
	(1)	**(2)**
t - 5	0.314	51.12
	(0.302)	(55.20)
t - 4	0.369	68.94
	(0.234)	(46.61)
t - 3	-0.204	-122.4+
	(0.258)	(65.88)
t - 2	-0.257	-88.43+
	(0.198)	(45.88)
t - 1	0.0166	10.80
	(0.184)	(42.85)
2012	**(base)**	**(base)**
t + 1	0.120	107.1**
	(0.153)	(33.28)
t + 2	0.218	166.3***
	(0.160)	(42.52)
t + 3	0.559***	212.0***
	(0.164)	(45.14)
t + 4	0.585**	159.3***
	(0.183)	(46.73)
t + 5	0.766***	218.7***
	(0.208)	(46.96)
Log(# of applications)	3.532***	479.5***
	(0.185)	(51.07)
Log(total amount)	-0.392***	19.60
	(0.0935)	(19.48)
Black applicants	0.621+	183.3**
	(0.344)	(59.76)
Hispanic applicants	0.522	24.54
	(0.714)	(162.4)
Other controls	YES	YES
Census tract FE	YES	YES
Year FE	YES	YES
MSA linear and quadratic time trends	YES	YES
No. of obs.	6,179	6,179
No. of census tract	562	562
R^2	0.6935	0.7294
Adjusted R^2	0.6954	0.6992
F	26.06***	11.54***
Root MSE	1.6651	414.6766

Note: Robust standard errors clustered by census tracts in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Unit of analysis = census tract-year (2007-2017)



Note: (a) Denial per capita (from Table 5, Column 1)



Note: (b) Denial amount per capita (from Table 5, Column 2)

Figure 3. Census-Level: Leads-and-Lags Model Estimation Results

Table 6. Loan Level: Main Estimation Results of Linear Probability Model

Dependent variables	**Denial**	**Denial**
	(1)	(2)
Treatment × Post	**0.0064**** **(0.0022)**	0.0054* (0.0022)
Treatment × Post × Black applicant		**0.0346***** **(0.0045)**
Treatment × Post × Hispanic applicant		**0.0228***** **(0.0046)**
Treatment × Black applicant		0.0336*** (0.0030)
Treatment × Hispanic applicant		0.0286*** (0.0037)
Post × Black applicant		0.0179*** (0.0027)
Post × Hispanic applicant		0.0219*** (0.0030)
Black applicant	0.0276*** (0.0014)	0.0289*** (0.0021)
Hispanic applicant	0.0296*** (0.0018)	0.0374*** (0.0028)
Male applicant	-0.0176*** (0.0007)	-0.0176*** (0.0007)
Applicant income	-0.0001*** (0.0000)	-0.0001*** (0.0000)
Other controls	YES	YES
Census tract FE	YES	YES
MSA FE	YES	YES
Respondent (Lender) FE	YES	YES
Loan type FE	YES	YES
Loan purpose FE	YES	YES
Preapproval req FE	YES	YES
Loan purchaser FE	YES	YES
Lien FE	YES	YES
Year FE	YES	YES
County linear time trends	YES	YES
No. of obs.	1,621,529	1,621,529
R^2	0.279957	0.279979
Adjusted R^2	0.278397	0.278417
F	132.25***	90.52***
Root MSE	0.307331	0.307327

Note: Robust standard errors clustered by census tracts in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. *Treatment* = 1 for loan applications in South Carolina; *post* = 1 for applications after 2012. In addition to census tract and year fixed effects, the analyses include a range of other fixed effects for unobserved time-invariant heterogeneity, including MSA, loan types (conventional, FHA insured, VA guaranteed), loan purposes (refinancing or home improvement), loan originator codes, loan purchaser codes (e.g., Fannie Mae, Freddie Mac, or commercial banks), preapproval status, and lien status. We also include county-level linear time trends to account for unobserved time-variant heterogeneity.

Table 6 presents the results of the individual loan-level estimations. As in Table 4, the coefficient of *Treatment* × *Post* in Column 1 is positive and significant, showing that a refinance/home improvement loan application in SC was more likely to be denied after the 2012 breach, compared to the pre-breach period and the neighboring states.

In Column 1 of Table 6, the coefficients of Black and Hispanic applicants are positive and significant, indicating that loan applications filed by Black and Hispanic individuals were more likely than others to be declined. While the Fair Housing Act and the Civil Rights Act of 1964 outlawed discrimination in housing and lending, Column 1 shows that minority homeowners still face greater difficulties in obtaining residential financing than others. Column 2 shows that the treatment effect of the SC DOR breach was more significant for Black and Hispanic loan applicants. The sum of the coefficients of *Treatment* × *Post* × *Black applicant* and *Treatment* × *Post* (0.04) is 7.4 times larger than that of *Treatment* × *Post* (0.0054).

This result appears to indicate that the SC DOR breach in 2012 did more damage to minority residents in SC than to others, supporting our theoretical discussion on the discriminative impacts of data breaches. In a census tract-level analysis presented in Column 1 of Table A3 (Appendix A), we interacted *Treatment* × *Post* with *minority majority*, which is 1 when the share of minority population at a census tract is over 50%. The coefficient of *Treatment* × *Post* × *Minority majority* is positive and significant, providing similar results to Column 2 of Table 6. In an unreported loan-level analysis, we found that Black and Hispanic individuals in SC suffered from the breach to a greater extent than others with similar income levels, suggesting that both socioeconomic conditions and individual ethnicity/race play a role in shaping harm to data breach victims.

Robustness Checks

Table 7 presents a list of robustness checks for both census tract-level and loan-level analyses. While we utilized an exogenous event to investigate the impact of a data breach and took a careful approach to credible causal identification, there are other endogeneity concerns and alternative explanations to address. Specifically, while our estimations included a range of observable control variables, multiple fixed effects, and region-specific time trends, there might have been other unobserved heterogeneity that could have biased the estimations and cloud causality.

There are two key assumptions in the spatial RDD approach. First, the eligibility of the treatment is continuous where the discontinuity occurs. In our case, it is the distance to the SC-NC-GA border. Second, the units in the treatment and control groups are similar to each other to the extent that it is as good as a random assignment. While we used the border counties in SC, NC, and GA to develop the treatment and control groups, which are well-balanced and comparable to each other, one might argue that counties are still large geographical areas and that some areas in these border counties are nonetheless far from the state border. To address this concern, we limited the sample to census tracts specifically within 5-12 miles from the border. In Table A4 in Appendix A, we see that the coefficients of *Treatment* × *Post* for *denial amount per capita* remained positive and significant when we used the census tracts within 7-12 miles from the borders (Columns 2-4).

To further reduce potential imbalance between the treated and control units, we took a recently developed counterfactual estimation approach—a matrix completion (MC) method (Athey et al., 2021). Based on the pre-treatment trends in the covariates and the outcome variables of both the treatment and control groups, this approach synthetically generates the counterfactuals of the treated units (Liu et al., 2022; Xu, 2017). These counterfactual observations indicate what would have happened to the treated units if they had not been treated but had followed a parallel trend with the control units. This approach generates ATE from the differences between the actual outcomes and the counterfactuals. Table A5 in Appendix A presents the MC method estimation results. While the estimations generated statistically significant pre-treatment trends in some periods, the magnitude of the post-treatment effects was much more significant, stable, and sizable. In addition, we ran coarsened exact matching (CEM) (Blackwell et al., 2009; Iacus et al., 2012) at both the census tract and loan levels. We used the observable covariates in Tables 2 and 3 to match the treated and control observations. Column 2 of Table A3 (Appendix A) and Column 1 of Table B3 (Appendix B) present the CEM estimations at the census tract and loan application levels, respectively, yielding similar results to the main estimates.

To rule out the possibility that the observed treatment effects were driven by chance or other unobserved factors, we conducted placebo and falsification tests (Cheng et al., 2020; Park et al., 2021). For the placebo tests, we conducted random shuffle tests with fake treatment variables that were randomly shuffled multiple times. We found that the effect of the placebo treatments was around zero. We conducted falsification tests with dependent variables that were unlikely to have been affected by the SC DOR breach; Columns 3-4 of Table A3 show that the treatment effect was not found to be statistically significant. These analyses demonstrate that our main results were unlikely to have been observed by chance or driven by unobserved factors.

Table 7. List of Estimations for Robustness Checks and Alternative Explanations

Census tract level (Appendix A)	Limiting the sample to census tracts closer to the state border	Table A4
	Matrix completion counterfactual estimation (Athey et al., 2021)	Table A5
	Coarsened exact matching (Iacus et al., 2012)	Table A3, Column 2
	Random shuffle test	
	Falsification tests	Table A3, Columns 3-4
	Excluding 2012 (the year of the breach)	Table A3, Column 5
	Alternative dependent variables	Table A3, Columns 6-8
	Estimations with all loan applications (including new home purchases)	Table A3, Column 9
	Excluding the financial crises (2007-2009)	
	Separating Charlotte metropolitan area	
	Controlling amendments to state mortgage laws	Table A6, Columns 1-2
	Controlling the number of other local-level breach incidents from Privacy Rights Clearinghouse	Table A6, Columns 3-4
	Controlling Google Trends of data breaches and identity theft	Table A6, Columns 5-6
Loan application level (Appendix B)	Coarsened exact matching	Table B3, Column 1
	Excluding 2012 (the year of the breach)	Table B3, Column 2
	Nonlinear (logit) estimation	Table B3, Column 3
	Random shuffle test	
	Estimations examining loan applications from credit unions only	Table B3, Column 4

Our time frame is 2007-2017. Given that the breach occurred and was announced in 2012, that year can be considered to be partially treated. We conducted additional analyses that excluded 2012 and observed similar findings (Column 5 of Table A3 and Column 2 of Table B3). In addition, a concern could be raised that the financial crisis in 2008-2009 significantly disrupted the housing market and mortgage supplies. When we excluded 2008-2009 in our sample as a robustness test, we observed similar results.

Further, the Charlotte metropolitan area is located at the SC-NC border and has a population of 2.7 million across five border counties in SC and NC. The effect of the breach might have differed in this large urban area from other counties. As robustness checks, we separate the census tracts in this area and the others outside the area. We observed similar finding in two separate samples.

In our main census tract-level analyses, we used the number of application denials per capita and the total monetary amount of denied mortgages per capita as dependent variables. In Column 6 of Table A3, we used the log-transformed number of denied applications without normalization as a dependent variable. While the coefficient of *Treatment* × *Post* was insignificant in an estimation with MSA-specific time trends, it became positive and significant without the time trends, as shown in Column 6. For Column 7, we used the log-transformed number of denials per capita and obtained similar results.

In our main analyses, we limited our sample to loan applications for refinancing and home improvement in the border counties to obtain more causal estimates. As a robustness check, we reran our estimation with a larger sample—loan applications of all types (including home purchases) in the border counties and refinancing/home improvement loans in all counties of SC, NC, and GA. In Column 9 of Table A3, we used a sample with all mortgage applications in the border counties of SC, NC, and GA, including those for new home purchases, and observed similar results. The treatment effect was still significant when we used refinancing/home improvement loans in all counties of SC, NC, and GA. Since the dependent variable for the loan-level analyses was dichotomous, we reran Equation (3) with a logit regression. Column 3 of Table B3 shows a significant treatment effect.

Alternative Explanations

There are a few possible alternative explanations for our main findings. First, there might have been changes in state-specific mortgage regulation policies. For example, SC might have enacted policy changes after 2012 that require a more stringent evaluation of loan applications, or GA or NC could have eased such restrictions. To address this concern, we measured the number of amendments to state mortgage lending statutes in 2007-2017: the Mortgage Lending Act in

SC, the Safe and Fair Enforcement Mortgage Licensing Act in NC, and the Fair Lending Act in GA. We looked up state legislature websites to find the amendments to these laws that were enacted each year. We measured the cumulative number of amendments to the mortgage laws since 2012 and included it as an additional control variable to the census tract-level estimations. Columns 1-2 in Table A6 of Appendix A show that the results remained similar with this additional control.

Second, there might have been other cybersecurity incidents that affected residents in the three states. While the impact of high-profile breaches at large national firms such as Target in 2013 or Home Depot in 2014 would be captured by year fixed effects, data breaches at smaller regional firms or local institutions such as hospitals or municipal governments could also have influenced residents in GA, SC, or NC. To account for this possibility, we obtained data breach records from Privacy Rights Clearinghouse, one of the popular data sources for cybersecurity research in IS (Pang & Tanriverdi, 2022; Zhang et al., 2024). In Columns 3-4 of Table A6, we controlled for the number of local breach incidents in each state of GA, SC, and NC in each year, and the results remained quantitatively similar.

Third, our results might have been driven by changes in lending policies by national and regional financial institutions. After 2012, large financial institutions serving the three states might have changed their lending policies in ways that would have tightened mortgage supplies in SC or loosened them in NC or GA. To address this issue, we reran the loan-level estimation with a sample of applications from local credit unions. A credit union is defined by the U.S. National Credit Union Administration as "a member-owned and controlled, not-for-profit, cooperative financial institution formed to provide its members with affordable and safe financial services." Given their structure, they typically serve a smaller, local geographic area, such as a city or county. Column 4 of Table B3 (Appendix B) presents an estimation examining loan applications from credit unions only in the border counties along SC, NC, and GA. The coefficients of *Treatment* × *Post* × *Black* and *Treatment* × *Post* × *Hispanic* remain positive and significant, as in Table 6.

Fourth, our observed results could have been driven by the heightened awareness of the SC DOR breach after 2012 among the SC residents. While Mikhed and Vogan (2018) found this awareness to be rather short-lived, it nonetheless may have affected SC borrowers' behaviors after 2012. To investigate this possibility, we conducted two analyses: (1) We measured the awareness of a data breach using Google Trend search indices with the two keywords "data breach" and "identity theft." For Columns 5-6 of Table A6 (Appendix A), we included this Google Trend measure as an additional control, and the estimates remained similar. (2) There is a possibility that SC residents may have avoided applying for mortgage refinancing after 2012, due to a concern that their personal information was compromised. In Column 8 of Table A3, we used the number of loan applications per capita as alternative dependent variables, and we did not find evidence of a decrease in refinancing loan applications from SC residents.

Underlying Mechanisms

So far, we have established that the breach at the SC DOR taxpayer database in 2012 caused an increase in denials of loan applications for refinancing and home improvement in SC. We argue that this is because the breach led to an increase in identity theft or other fraudulent activities that compromised SC residents' financial records and credit scores. One might ask, however, if the SC DOR breach actually did result in an increase in fraud incidents and damages to SC residents' credit scores.

To answer this question, we conducted four additional analyses. First, as the HMDA dataset did not provide credit score information for loan applicants, we instead utilized the records of approved applications purchased by Fannie Mae and Freddie Mac, two government-sponsored firms that provide liquidity and stability to the housing market. We downloaded approved loan records from the two firms. Unlike the HMDA, these records only included applications that were approved; they also provided credit scores, interest rates, and other information that was not available in the HMDA database. As in our main analyses, we used refinancing and home improvement loan applications in the counties along the SC-NC-GA border in 2007-2017.

Table C1 in Appendix C provides the estimation results. Column 1 shows that after the breach, the credit scores of approved refinancing applicants in SC decreased significantly. For Column 2, we used a dummy variable for whether an applicant's score was higher than 650, which is a common threshold for a good credit score. This analysis yielded similar findings. While the breach in 2012 did not significantly affect interest rates (Column 3) or debt-to-income ratios (Column 4), the loan-to-value ratio (Column 5) declined in SC after 2012, meaning that leaders originated smaller loans to SC borrowers given the property value. We conducted similar analyses with nationwide data from the National Mortgage Database Aggregate Statistics from the Federal Housing Finance Authority, which provides aggregate statistics of all approved loans beyond Fannie Mae and Freddie Mac. We saw a decrease in approved loan applicants' credit scores in SC after 2012, similar to the results reported in Table C1.

Second, we collected fraud incident records from the National Incident-Based Reporting System (NIBRS) databases provided by the Federal Bureau of Investigation. The NIBRS collects records of individual crime reports to law enforcement agencies in the U.S. Specifically, we focused on "wire fraud" incidents (offense code 26E). Wire fraud is defined in 18 U.S. Code 1343 as "having devised or intending to devise any scheme or artifice to defraud, or for obtaining money or property by means of false or fraudulent pretenses, representations, or promises, transmits or causes to be transmitted by means of wire, radio, or television communication in interstate or foreign commerce."[4] It includes a broad range of criminal actions that utilize communication means such as phone calls, emails, and messaging. Criminals can take advantage of SC taxpayer information, such as addresses, emails, and dependent information, to commit wire fraud in various manners—for example, scam phone calls, phishing emails, or fake check frauds.

From the NIBRS database, we built a panel dataset at the level of law enforcement agency-month (local police department, state police, or other specialized police) in 2007-2017. As in the main analyses, *treatment* is 1 for law enforcement agencies in SC, and *post* is 1 for observations after 2012. The dependent variable is the number of wire fraud incidents reported to an agency in each month. The estimations include several county-level control variables, agency/year/month fixed effects, and state-specific linear time trends.

Table C2 in Appendix C presents fixed effects negative binomial estimation results. Column 1 shows that after 2012, there was a significant increase in the number of wire fraud incidents reported to law enforcement agencies in SC, compared to other states. For Columns 2-4, we interacted *Treatment* × *Post* with the three variables for demographic and socioeconomic characteristics, and the results show that the SC DOR breach led to a greater increase in wire fraud in SC counties with greater minority populations, lower median income, and more residents living below the poverty level. This is in accordance with our findings in Column 2 of Table 6 that the SC DOR breach hurt loan applications by Black and Hispanic residents to a greater extent.

Third, we obtained records of consumer complaints against financial institutions from the CFPB. Created by the Dodd-Frank Wall Street Reform and Consumer Protection Act after the 2008 financial crisis, the CFPB is mandated to collect and report individual consumers' complaints about financial products and services. If the SC DOR breach indeed compromised SC residents' financial and credit records due to identity thefts and other frauds (Newman & McNally, 2005; Anderson et al., 2008; Harrell, 2019), they would have filed more complaints against financial institutions than before.

The CFPB provides consumer complaint records from December 2011, which did not enable us to observe sufficient pre-treatment trends prior to 2012. Hence, we were unable to provide causal evidence with respect to the impact of the SC DOR breach on consumer complaints. Nonetheless, we utilized the CFBP database to provide supplementary evidence for the temporal trends in consumer complaints since 2011. The database provides complaint records with zip codes and filing dates, and we aggregated these records to the zip code-month level from December 2011 to December 2017. We used the number of total complaints filed from each zip code-month as a dependent variable. Furthermore, the database also provides information on associated product categories and underlying issues; using this information, we counted the number of complaints related to credit reporting and credit cards.

Table C3 in Appendix C provides fixed-effects negative binomial regression results with year and zip code fixed effects. In these estimations, *SC* is equal to 1 for zip codes in SC, and the coefficients of *SC* × *Year* indicate temporal trends in the number of consumer complaints from SC. Column 1 shows that there was a significant increase in the number of total consumer complaints to the CFPB from SC residents in the 2011-2017 period. Column 2 indicates that this trend was stronger for complaints on credit cards and credit reporting. Like in Table C2 for wire fraud incidents, we interacted *SC* × *Year* with the share of the minority population, finding that the number of consumer complaints in SC grew faster after 2012 in counties with larger minority populations (Table C3, Columns 3-4).

Lastly, we utilized a public dataset from Lending Club, one of the largest peer-to-peer (P2P) lending platforms (Alyakoob et al., 2021; Guo et al., 2016). We expected that if the SC DOR breach undermined the credit scores of SC residents, they would have had greater difficulty in securing P2P loans through Lending Club. We obtained P2P application records from 2007-2017. We observed that after 2012, there was a significant increase in the number and monetary amount of rejected applications in SC compared to NC and GA (Table C4 in Appendix C).

The above findings regarding credit scores, wire fraud incidents, customer complaints against financial institutions, and P2P lending records are in line with our proposed mechanism that the breach of SC taxpayer data damaged SC residents' credit scores and financial records, resulting in greater denials of their mortgage applications after 2012.

[4] https://uscode.house.gov/view.xhtml?req=(title:18%20 section:1343%20edition:prelim)%20OR%20(granuleid:USC-prelim-title18-section1343)&f=treesort&edition=prelim&num=0&jumpTo=true

Discussion and Conclusion

How much does a data breach hurt affected individuals? While this question carries great significance for IS researchers, practitioners, and policymakers alike, the IS literature to date has largely been silent on this important matter. Utilizing a unique setting that allowed for a natural experiment and a large official dataset, we uncovered a significant and detrimental impact of data breaches on victims' welfare. Specifically, we find that a large-scale breach of taxpayer records in South Carolina (SC) in 2012 was associated with a significant increase in denials to refinancing and home improvement mortgage applications by SC residents after 2012, compared to those in the neighboring states of North Carolina and Georgia. This increase was about 22%, compared to the previous years. We observed this effect from both aggregated census tract-level and individual loan application-level analyses. Strikingly, we found that the SC DOR breach had more adverse impacts on mortgage applications by Black and Hispanic residents in SC. Our series of robustness checks indicated that this effect is likely to be causal. In addition, the supplementary analyses with approved loan datasets (Fannie Mae and Freddie Mac), national datasets of crime reports (NIBRS), consumer complaints against financial institutions (CFPB), and P2P lending data (Lending Club) revealed potential underlying mechanisms.

This study offers significant contributions to the IS literature by identifying the material monetary impact of cybersecurity incidents on individual victims. While prior work has studied affected individuals' emotional responses or behavioral changes (e.g., Bachura et al., 2022; Hoehle et al., 2022), ours is the first to investigate the direct financial damage inflicted on individuals due to data breaches. The extant literature indicates that the adverse impact of data breaches on firms' market value or performance is minimal or short-lived (Ali et al., 2021; Martin et al., 2017; Foerderer & Schuetz, 2022); these findings may be influenced by the absence of concrete evidence on the consequences suffered by individual victims. Our study fills this important void in the literature with credible empirical evidence. We also substantially contribute to the literature by theorizing and empirically demonstrating that a data breach causes greater harm to minority populations. We expand the body of IS knowledge by revealing the underlying mechanisms in the impact of data breaches on individuals, as evidenced by decreases in credit scores of approved loan applicants and increases in wire fraud incidents and consumer complaints against financial institutions.

Our work also provides meaningful implications from managerial and policy perspectives, with findings that are generalizable to other cybersecurity incidents. While our focal breach involved a government agency, its technical and substantive nature is not so different from other massive data breaches, such as those of Home Depot and Capital One, and our findings can be generalized to other incidents in for-profit settings. However, we acknowledge that since every breach incident differs in terms of causes, types of compromised information, and organizational settings, there is a limit to the extent to which our findings may be generalized. Nevertheless, given the theoretical arguments above, we expect that past large-scale data breaches have also hurt affected individuals materially and have impacted minority victims to a greater extent. We also note that while the SC data breach occurred in 2012, our findings still carry present-day relevance for two principal reasons. First, since 2012, there has been no comprehensive privacy or cybersecurity legislation enacted at the federal level in the U.S. that would aim to reduce harm to consumers who have been victimized by data breaches. Although 16 states have adopted privacy legislation, none of these laws include provisions that would protect individuals from data breaches such as the SC DOR breach (IAPP, 2024). Second, even though cybersecurity defense mechanisms have advanced and are now more available to individuals, the skills and methods of attackers have also advanced; in other words, there is an "arms race" between attackers and defenders (Madnick, 2024; Murphy & Bennett, 2023). Identity theft has increased substantially since 2012, suggesting that fraudsters' methods remain at least as effective as they were in 2012 (IdentityTheft.org, 2024).

As mentioned above, we do not have an adequate understanding of how harmful data breaches are to individual victims; for this reason, some argue that organizations responsible for cybersecurity failures should not bear the full costs of damages to victims (Romanosky & Acquisti, 2009; Moore, 2010; Simon & Omar, 2020). This lack of liability hardly incentivizes managers to sufficiently invest in strong cybersecurity defense and ensure that individuals' private information is well secured. Rather, they may simply regard data breaches as "the cost of doing business" (Manworren et al., 2016; Havakhor et al., 2022) and be less likely to consider cybersecurity protection a key need for consumers, shareholders, or other stakeholders. The finding that data breaches do cause material harm to individual victims sends a clear message to practitioners and policymakers that organizations not underestimate the damages to those affected by their cybersecurity failures. We also urge policymakers to develop more stringent laws and regulations to hold firms more accountable for failures and incentivize them to bolster their cybersecurity investments.

Our findings can also be utilized in data breach litigation to support compensation for data breach victims. Policymakers could provide official figures and statistics to help adjudicate legal disputes in data breaches and develop cybersecurity insurance policies. Given our findings that data breaches can be more damaging to minority victims, policymakers could also develop initiatives targeting minority populations to educate them on how to deal with data breaches and identity thefts. Such initiatives should also seek to alleviate potential mistrust of authority among minorities (Figure 1). In addition, policymakers could strengthen policies that prohibit financial institutions from discriminative treatment of minority customers.

Our work is by no means free from limitations. Our investigations include a battery of robustness checks and other analyses for causal identification. Nonetheless, since we rely on secondary datasets, there might be endogeneity issues or alternative explanations that we do not address. Given that randomized controlled experiments are not feasible for practical and ethical reasons, our research provides the most credible evidence possible on the adverse effect of data breaches on affected individuals. In addition, for reasons related to privacy or data limitations, we were unable to observe other personal information of loan applicants or why, exactly, each loan application was denied. The HMDA database only provides information on mortgage applications by year, without months or exact dates. For this reason, we could not control for seasonal or other temporal factors in our estimations. In the mechanism analyses, we did not observe the details or causes of wire fraud incidents or the reasons for consumer complaints against financial institutions. As mentioned above, since the SC government offered fraud protection services to all of its residents in 2012, we were unable to observe the impact of the breach in the absence of such services. However, since most large-scale data breaches in recent years have been accompanied by the offer of protection services by breached firms, our observed effect of the data breach in combination with protection services still provides meaningful implications for policy and practice.

As cybersecurity is an ever-growing risk for organizations and individuals, there is a plethora of future research opportunities for IS researchers. Future studies could examine other material, financial impacts on individual victims of cybersecurity incidents, particularly in for-profit contexts or concerning breaches stemming from different causes (e.g., ransomware). There may be other unintended consequences or collateral damage from data breaches experienced by individuals or entities that were not targets themselves. Future research could uncover further underlying mechanisms in the effects of data breaches for a more complete understanding of data breach consequences.

References

Agrafiotis, I., Nurse, J. R., Goldsmith, M., Creese, S., & Upton, D. (2018). A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate. *Journal of Cybersecurity, 4*(1). https://doi.org/10.1093/cybsec/tyy006

Al-Bahrani, A., Weathers, J., & Patel, D. (2019). Racial differences in the returns to financial literacy education. *The Journal of Consumer Affairs, 53*(2), 572-599. https://doi.org/10.1111/joca.12205

Ali, S. E. A., Lai, F.-W., Dominic, P. D. D., Brown, N. J., Lowry, P. B., & Ali, R. F. (2021). Stock market reactions to favorable and unfavorable information security events: A systematic literature review. *Computers & Security, 110*, Article 102451. https://doi.org/10.1016/j.cose.2021.102451

Alyakoob, M., Rahman, M. S., & Wei, Z. (2021). Where you live matters: Local bank competition, online marketplace lending, and disparity in borrower benefits. *Information Systems Research, 32*(4), 1390-1411. https://doi.org/10.1287/isre.2021.1043

Ambrose, B. W., Conklin, J. N., & Lopez, L. A. (2021). Does borrower and broker race affect the cost of mortgage credit? *The Review of Financial Studies, 34*(2), 790-826. https://doi.org/10.1093/rfs/hhaa087

Anderson, K. B. (2006). Who are the victims of identity theft? The effect of demographics. *Journal of Public Policy & Marketing, 25*(2), 160-171. https://doi.org/10.1509/jppm.25.2.160

Anderson, K. B., Durbin, E., & Salinger, M. A. (2008). Identity theft. *Journal of Economic Perspectives, 22*(2), 171-192. https://doi.org/10.1257/jep.22.2.171

Angrisani, M., Barrera, S., Blanco, L. R., & Contreras, S. (2021). The racial/ethnic gap in financial literacy in the population and by income. *Contemporary Economic Policy, 39*(3), 524-536. https://doi.org/10.1111/coep.12507

Angrist, J. D., & Pischke, J.-P. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. https://doi.org/10.2307/j.ctvcm4j72

Angst, C. M., Block, E. S., D'Arcy, J., & Kelley, K. (2017). When do IT security investments matter? Accounting for the influence of institutional factors in the context of healthcare data breaches. *MIS Quarterly, 41*(3), 893-916. https://doi.org/10.25300/MISQ/2017/41.3.10

Archer, N. (2012). Consumer identity theft prevention and identity fraud detection behaviors. *Journal of Financial Crime, 19*(1), 20-36. https://doi.org/10.1108/13590791211190704

Ards, S. D., & Myers, S. L. (2001). The color of money: Bad credit, wealth, and race. *American Behavioral Scientist, 45*(2), 223-239. https://doi.org/10.1177/00027640121957141

Arnett, C. (2020). Race, surveillance, resistance. *Ohio State Law Journal, 81*(6), 1103-1142.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., & Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association, 116*(536), 1716-1730. https://doi.org/10.1080/01621459.2021.1891924

Avakame, E. F., Fyfe, J. J., & Mccoy, C. (1999). "Did you call the police? What did they do?" An empirical assessment of Black's theory of mobilization of law. *Justice Quarterly, 16*(4), 765-792. https://doi.org/10.1080/07418829900094361

Bachman, R. (1994). Victim's perceptions of initial police responses to robbery and aggravated assault: Does race matter? *Journal of Quantitative Criminology, 12,* 363-390. https://doi.org/10.1007/BF02354424

Bachura, E., Valecha, R., Chen, R., & Rao, H. R. (2022). The OPM data breach: An investigation of shared emotional reactions on Twitter.

MIS Quarterly, 46(2), 881-910. https://doi.org/10.25300/MISQ/2022/15596

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76*(4), 169-217. https://doi.org/10.1086/259394

Black, H., Schweitzer, R. L., & Mandell, L. (1978). Discrimination in mortgage lending. *The American Economic Review, 68*(2), 186-191. https://doi.org/10.1257/jep.12.2.41

Blackwell, M., Iacus, S., & Porro, G. (2009). CEM: Coarsened exact matching in Stata. *The Stata Journal, 9*(4). https://doi.org/10.1177/1536867X0900900402

Blanchett, D. (2017). The home as a risky asset. *Journal of Personal Finance, 16*(1), 7-28.

Bubb, R., & Kaufman, A. (2014). Securitization and moral hazard: Evidence from credit score cutoff rules. *Journal of Monetary Economics, 63*(1), 1-18. https://doi.org/10.1016/j.jmoneco.2014.01.005

Campos-Castillo, C. (2015). Revisiting the first-level digital divide in the United States: Gender and race/ethnicity patterns. *Social Science Computer Review, 33*(4), 423-439. https://doi.org/10.1177/0894439314547617

Cheng, A. Z., Pang, M.-S., & Pavlou, P. A. (2020). Mitigating traffic congestion: The role of intelligent transportation systems. *Information Systems Research, 31*(3), 653-674. https://doi.org/10.1287/isre.2019.0894

Chua, H. N., The, J. S., & Herbland, A. (2021). Identifying the effect of data breach publicity on information security awareness using hierarchical regression. *IEEE Access, 9*, 121759-121770. https://doi.org/10.1109/ACCESS.2021.3107426

Clark, V. E. (2021). Victim-blaming discourse underpinning police responses to domestic violence: A critical social work perspective. *Social Work & Policy Studies, 4*(1), 1-15.

D'Arcy, J., Adjerid, I., Angst, C. M., & Glavas, A. (2020). Too good to be true: Firm social performance and the risk of data breach. *Information Systems Research, 31*(4), 1200-1223. https://doi.org/10.1287/isre.2020.0939

Dukes, K. N., & Gaither, S. E. (2017). Black racial stereotypes and victim blaming: Implications for media coverage and criminal proceedings in cases of police violence against racial and ethnic minorities. *Journal of Social Issues, 73*(4), 789-807. https://doi.org/10.1111/josi.12248

Egger, P. H., & Lassmann, A. (2015). The causal impact of common native language on international trade: Evidence from a spatial regression discontinuity design. *The Economic Journal, 125*(584), 699-745. https://doi.org/10.1111/ecoj.12253

Electronic Frontier Foundation. (2022). 2022-11-22—In re Marriott (4th Cir.)—Amicus brief of EFF and EPIC. https://www.eff.org/document/2022-11-22-re-marriott-4th-cir-amicus-brief-eff-and-epic

Faig, M., & Shum, P. (2002). Portfolio choice in the presence of personal illiquid projects. *Journal of Finance, 57*(1), 303-328. https://doi.org/10.1111/1540-6261.00423

Fairlie, R. W. (2004). Race and the digital divide. *Contributions in Economic Analysis & Policy, 3*(1), 1-38. https://doi.org/10.2202/1538-0645.1263

Foerderer, J., & Schuetz, S. W. (2022). Data breach announcements and stock market reactions: A matter of timing? *Management Science, 68*(10), 7298-7322. https://doi.org/10.1287/mnsc.2021.4264

Freeman, L. (2000). Minority housing segregation: A test of three perspectives. *Journal of Urban Affairs, 22*(1), 15-35. https://doi.org/10.1111/0735-2166.00037

Goel, S., & Shawky, H. A. (2009). Estimating the market impact of security breach announcements on firm values. *Information & Management, 46*(7), 404-410. https://doi.org/10.1016/j.im.2009.06.005

Gofen, A., Cohen-Blankshtain, G., & Ibraheem, M. (2021). It takes a village to build illegality: Minorities' noncompliance as manifestation of distrust. *Governance, 34*(4), 983-1000. https://doi.org/10.1111/gove.12528

Gonzalez, R. M. (2021). Cell phone access and election fraud: Evidence from a spatial regression discontinuity design in Afghanistan. *American Economic Journal: Applied Economics, 13*(2), 1-51. https://doi.org/10.1257/app.20190443

Guo, Y., Zhou, W., Luo, C., Liu, C., & Xiong, H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research, 249*(2), 417-426. https://doi.org/10.1016/j.ejor.2015.05.050

Haislip, J., Lim, J.-H., & Pinsker, R. (2021). The impact of executives' IT expertise on reported data security breaches. *Information Systems Research, 32*(2), 318-334. https://doi.org/10.1287/isre.2020.0986

Haendler, C., & Heimer, R. (2021). *The financial restitution gap in consumer finance: Insights from complaints filed with the CFPB.* SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3766485

Havakhor, T., Rahman, M. S., & Zhang, T. (2022). *Disclosure of cybersecurity investments and the cost of capital.* SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3553470

Harrell, E. (2019). *Victims of identity theft, 2016.* Bureau of Justice Statistics. https://bjs.ojp.gov/library/publications/victims-identity-theft-2016.

Henderson, L., Herring, C., & Thomas, M. (2015). Credit where credit is due? Race, gender, and discrimination in the credit scores of business startups. *The Review of Black Political Economy, 42*(4), 459-479. https://doi.org/10.1007/s12114-015-9215-4

Hidano, N., Hoshiro, T., & Sugiura, A. (2015). The effect of seismic hazard risk information on property prices: Evidence from a spatial regression discontinuity design. *Regional Science and Urban Economics, 53*, 113-122. https://doi.org/10.1016/j.regsciurbeco.2015.05.005

Hoar, S. B. (2001). Identity theft: The crime of the new millennium. *Oregon Law Review, 80*, 1423-1448.

Hoehle, H., Venkatesh, V., Brown, S. A., Tepper, B. J., & Kude, T. (2022). Impact of customer compensation strategies on outcomes and the mediating role of justice perceptions: A longitudinal study of Target's data breach. *MIS Quarterly, 46*(1), 299-340. https://doi.org/10.25300/MISQ/2022/14740

Holdman, J. (2022). SC's massive data breach 10 years later: Questions linger as investigation remains open. *The Post and Courier.* https://www.postandcourier.com/columbia/news/scs-massive-data-breach-10-years-later-questions-linger-as-investigation-remains-open/article_29dd8164-4025-11ed-9433-73cafd23fadb.html

Howell, S. E., & Fagan, D. (1988). Race and trust in government: Testing the political reality model. *Public Opinion Quarterly, 52*(3), 343-350. https://doi.org/10.1086/269111

Howerton, A. (2006). Police response to crime: Differences in the application of law by race. *Journal of Ethnicity in Criminal Justice, 4*(3), 51-66. https://doi.org/10.1300/J222v04n03_03

Hui, K.-L., Vance, A., & Zhdanov, D. (2016). *Securing digital assets* (*MIS Quarterly* research curation). https://www.misqresearchcurations.org/blog/2017/5/10/securing-digital-assets-1.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis, 20*(1), 1-24. https://doi.org/10.1093/pan/mpr013

IAPP (International Association of Privacy Professionals). (2024). *US state privacy legislation tracker* https://iapp.org/resources/article/us-state-privacy-legislation-tracker.

IdentityTheft.org. (2024). *2024 identity theft facts and statistics.* https://identitytheft.org/statistics

Janakiraman, R., Lim, J.-H., & Rishika, R. (2018). The effect of a data breach announcement on customer behavior: Evidence from a multichannel retailer. *Journal of Marketing, 82*(1), 85-105. https://doi.org/10.1509/jm.16.0124

Kannan, K., Rees, J., & Sridhar, S. (2007). Market reactions to information security breach announcements: An empirical analysis. *International Journal of Electronic Commerce*, *12*(1), 69-91. https://doi.org/10.2753/JEC1086-4415120103

Ke, J., Wang, W., & Foutz, N. Z. (2022). *"My data or my health? Heterogenous patient responses to healthcare data breach*. SSRN. https://ssrn.com/abstract=4029103

Keele, L. J., & Titiunik, R. (2015). Geographic boundaries as regression discontinuities. *Political Analysis, 23*(1), 127-155. https://doi.org/10.1093/pan/mpu014

Kim, S. (2000). Race and home price appreciation in urban neighborhoods: Evidence from Milwaukee, Wisconsin. *The Review of Black Political Economy, 28*(2), 9-28. https://doi.org/10.1007/s12114-000-1015-8

Kim, S. H., & Kwon, J. (2019). How do EHRs and a meaningful use initiative affect breaches of patient information? *Information Systems Research, 30*(4), 1184-1202. https://doi.org/10.1287/isre.2019.0858

Knell, N. (2012). *South Carolina breach compromises millions of records.* Government Technology. https://www.govtech.com/security/south-carolina-breach-compromises-records.html

Krebs, B. (2024). *Who stole 3.6M tax records from South Carolina?* KrebsonSecurity. https://krebsonsecurity.com/2024/04/who-stole-3-6m-tax-records-from-south-carolina

Kwon, J., & Johnson, M. E. (2014). Proactive versus reactive security investments in the healthcare sector. *MIS Quarterly, 38*(2), 451-471. https://doi.org/10.25300/MISQ/2014/38.2.06

LaCour-Little, M. (1999). Discrimination in mortgage lending: A critical review of the literature. *Journal of Real Estate Literature, 7*(1), 15-49. https://doi.org/10.1080/10835547.1999.12090078

Lave, T. R. (2022). Blame the victim: How mistreatment by the state is used to legitimize police violence. *Brooklyn Law Review, 87*(4), 1161-1196.

Laufer, S., & Paciorek, A. (2022). The effects of mortgage credit availability: Evidence from minimum credit score lending rules. *American Economic Journal: Economic Policy, 14*(1), 240-276. https://doi.org/10.1257/pol.20180229

Lindley, J. T., Selby, E. B., & Jackson, J. D. (1984). Racial discrimination in the provision of financial services. *The American Economic Review, 74*(4), 735-741.

Liu, L., Wang, Y., & Xu, Y. (2022). A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *American Journal of Political Science, 68*(1), 160-176. https://doi.org/10.1111/ajps.12723

Logan, J. R., Oakley, D., & Stowell, J. (2008). School segregation in metropolitan regions, 1970-2000: The impacts of policy choices on public education. *American Journal of Sociology, 113*(6), 1611-1644. https://doi.org/10.1086/587150

Lu, P., Kezios, K., Milazzo, F., Jawadekar, N., Shelley, M., & Al Hazzouri, A. Z. (2023). Racial differences in employment and poverty histories and health in older age. *American Journal of Preventive Medicine, 64*(4), 543-551. https://doi.org/10.1016/j.amepre.2022.10.018

Mack, J. B. (2018). Born in the red: How affirmative action could cure the race-credit divide. *Houston Law Review, 55*(5), 1157-1186.

Madnick, S. (2024). What's behind the increase in data breaches? *The Wall Street Journal.* https://www.wsj.com/tech/cybersecurity/why-are-cybersecurity-data-breaches-still-rising-2f08866c.

Malhotra, A., & Malhotra, C. K. (2011). Evaluating customer information breaches as service failures: An event study approach. *Journal of Service Science, 14*(1), 44-59. https://doi.org/10.1177/1094670510383409

Malwarebytes. (2021). *Demographics of cybercrime report.* https://web.archive.org/web/20210927214415/https://www.malwarebytes.com/resources/2021-demographics-of-cybercrime-report/index.html

Mandiant. (2012). *South Carolina Department of Revenue public incident response report*. https://oag.ca.gov/system/files/Mandiant%20Report_0.pdf

Manworren, N., Letwat, J., & Daily, O. (2016). Why you should care about the Target data breach. *Business Horizon, 59*(3), 257-266. https://doi.org/10.1016/j.bushor.2016.01.002

Markkanen, S., & Harrison, M. (2013). "Race," deprivation and the research agenda: Revisiting housing, ethnicity, and neighborhoods. *Housing Studies, 28*(3), 409-428. https://doi.org/10.1080/02673037.2013.759180

Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data privacy: Effects on customer and firm performance. *Journal of Marketing, 81*(1), 36-58. https://doi.org/10.1509/jm.15.0497

McDaniel, M., & Kuehn, D. (2013). What does a high school diploma get you? Employment, race, and the transition to adulthood. *The Review of Black Political Economy, 40*(4), 371-399. https://doi.org/10.1007/s12114-012-9147-1

McDonald, J., & Stokes, R. J. (2006). Race, social capital, and trust in the police. *Urban Affairs Review, 41*(3), 358-375. https://doi.org/10.1177/1078087405281707

McKinsey. (2022). Cybersecurity trends: Looking over the horizon. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/cybersecurity/cybersecurity-trends-looking-over-the-horizon

Mikhed, V., & Vogan, M. (2018). How data breaches affect consumer credit. *Journal of Banking & Finance*, 88, 192-207. https://doi.org/10.1016/j.jbankfin.2017.12.002

Moore, T. (2010). The economics of cybersecurity: Principles and policy options. *International Journal of Critical Infrastructure Protection, 3*(3-4), 103-117. https://doi.org/10.1016/j.ijcip.2010.10.002

Murphy, M., & Bennett, D. (2023). *Teen gamers swiped $24 million in crypto, then turned on each other*. Bloomberg https://www.bloomberg.com/news/features/2023-08-04/teen-gamers-targeted-michael-terpin-in-major-sim-swapping-theft

Newman, G. R., & McNally, M. M. (2005). *Identity theft literature review.* The National Criminal Justice Reference Service Virtual Library. https://www.ojp.gov/ncjrs/virtual-library/abstracts/identity-theft-literature-review

Novak, A. N., & Viceanu, M. O. (2019). The internet is not pleased: Twitter and the 2017 Equifax data breach. *The Communication Review, 22*(3), 196-221. https://doi.org/10.1080/10714421.2019.1651595

Nuño, L. E. (2018). Hispanics' perceived procedural justice, legitimacy, and willingness to cooperate with the police. *Police Practice and Research, 19*(2), 153-167. https://doi.org/10.1080/15614263.2018.1418160

Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology, 34*(1), 181-209. https://doi.org/10.1146/annurev.soc.33.040406.131740

Pang, M.-S., & Tanriverdi, H. (2022). Strategic roles of IT modernization and cloud migration in reducing cybersecurity risks of organizations: The case of U.S. federal government. *Journal of Strategic*

Information Systems, 31(1), 101707. https://doi.org/10.1016/j.jsis.2022.101707

Park, J., Pang, M.-S., Kim, J., & Lee, B. (2021). The deterrent effect of ride-sharing on sexual assault and investigation of situational contingencies. *Information Systems Research, 32*(2), 497-516. https://doi.org/10.1287/isre.2020.0978

Pass, M. D., Madon, N. S., Murphy, K., & Sargeant, E. (2020). To trust or distrust? Unpacking ethnic minority immigrants' trust in police. *The British Journal of Criminology, 60*(5), 1320-1341. https://doi.org/10.1093/bjc/azaa023

Prieger, J. E. (2015). The broadband digital divide and the benefits of mobile broadband for minorities. *The Journal of Economic Inequality, 13*(3), 373-400. https://doi.org/10.1007/s10888-015-9296-0

Romanosky, S., & Acquisti, A. (2009). Privacy costs and personal data protection: Economic and legal perspectives. *Berkeley Technology Law Journal, 24*(3), 1061-1102.

Ross, S. L., & Turner, M. A. (2005). Housing discrimination in metropolitan America: Explaining changes between 1989 and 2000. *Social Problems, 52*(2), 152-180. https://doi.org/10.1525/sp.2005.52.2.152

Rubin, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American Statistical Association, 75*(371), 591-593. https://doi.org/10.2307/2287653

Schaffer, A. (2022). The Marriott lawsuit explores the conundrum of how to value stolen data? *The Washington Post.* https://www.washingtonpost.com/politics/2022/05/27/marriott-lawsuit-explores-conundrum-how-value-stolen-data/

Schwartz, A., & Cohn, C. (2022). Let data breach victims sue Marriott. EFF. https://www.eff.org/deeplinks/2022/11/let-data-breach-victims-sue-marriott

Security Magazine. (2022). Leading cyber risks & trends in 2022. https://www.securitymagazine.com/articles/98695-leading-cyber-risks-and-trends-in-2022.

Self, J. (2018). Your state tax info was hacked. Now SC has stopped paying for your protection. *The State*. https://www.thestate.com/news/politics-government/article220951355.html

Sikkink, D., & Emerson, M. O. (2008). School choice and racial segregation in US schools: The role of parents' education. *Ethnic and Racial Studies, 31*(2), 267-293. https://doi.org/10.1080/01419870701337650

Simon, J., & Omar, A. (2020). Cybersecurity investments in the supply chain: Coordination and a strategic attacker. *European Journal of Operational Research, 282*(1), 161-171. https://doi.org/10.1016/j.ejor.2019.09.017

Syed, R. (2019). Enterprise reputation threats on social media: A case of data breach framing. *The Journal of Strategic Information Systems, 28*(3), 257-274. https://doi.org/10.1016/j.jsis.2018.12.001

Tyler, T. R. (2005). Policing in black and white: Ethnic group differences in trust and confidence in the police. *Police Quarterly, 8*(3), 322-342. https://doi.org/10.1177/1098611104271105

Vargas, R., & Scrivener, L. (2021). Why Latino youth (don't) call police. *Race and Justice, 11*(1), 47-64. https://doi.org/10.1177/2153368718776056

Wang, J., Shan, Z., Gupta, M., & Rao, H. R. (2019). A longitudinal study of unauthorized access attempts on information systems: The role of opportunity contexts. *MIS Quarterly, 43*(2), 601-622. https://doi.org/10.25300/MISQ/2019/14751

Weitzer, R. (2014). The puzzling neglect of Hispanic Americans in research on police-citizen relations. *Ethnic and Racial Studies, 37*(11), 1995-2013. https://doi.org/10.1080/01419870.2013.790984

Whittemore, A. H. (2021). Exclusionary zoning: Origins, open suburbs, and contemporary debates. *Journal of the American Planning Association, 87*(2), 167-180. https://doi.org/10.1080/01944363.2020.1828146

Wu, Y., & Miethe, T. D. (2022). Race/ethnicity, area, and willingness to call police. *American Journal of Criminal Justice* 49, 121-144. https://doi.org/10.1007/s12103-022-09691-8

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis, 25*(1), 57-76. https://doi.org/10.1017/pan.2016.2

Zhang, L., Wattal, S., & Pang, M.-S. (2024). Does sharing make my data more insecure? An empirical study on health information exchange and data breaches. *MIS Quarterly*, 48(3), 873-898. https://doi.org/10.1111/10.25300/MISQ/2023/17479

About the Authors

Min-Seok Pang is Karen A. and William S. Monfre Professor in Business and a professor of information systems and analytics at Wisconsin School of Business, University of Wisconsin-Madison. He received a B.S. in industrial engineering and an M.S. in management from Korea Advanced Institute of Science and Technology (KAIST) and holds a Ph.D. in business administration from the University of Michigan. His research interests include, among others, cybersecurity management, strategic management of information technologies, and technology-enabled public policies. His research has been published in top-tier academic journals such as *Management Science, MIS Quarterly, Information Systems Research, Strategic Management Journal,* and *Organization Science.* His research has been featured at several news outlets such as *The Wall Street Journal*, Computerworld, Federal Computer Week, and TechCrunch. He currently serves as a senior editor for *Journal of the Association for Information Systems* and an associate editor for *Information Systems Research.*

Anthony Vance is a professor of business information technology at the Pamplin College of Business at Virginia Tech. He earned Ph.D. degrees in information systems from Georgia State University, the University of Paris-Dauphine, France, and the University of Oulu, Finland. His work has been published in outlets such as *MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Journal of the Association for Information Systems, European Journal of Information Systems*, and in the proceedings of the Symposium on Usable Privacy and Security and the ACM Conference on Human Factors in Computing Systems. His research focuses on behavioral, neuroscience, and organizational approaches to cybersecurity. He is currently a senior editor at *MIS Quarterly* and serves on the editorial board of *Journal of the Association for Information Systems.*

Appendix A

Census-Tract Level Additional Analyses

Table A1. Census-Tract Level— Summary Statistics

Variables		*N*	Mean	Std. dev.	Min	Max
Denial per capita	(1)	6,182	4.1532	2.8523	0	50.569
Denial amount per capita	(2)	6,182	571.7098	755.9729	0	16620.92
Treatment	(3)	6,182	0.4982	0.5	0	1
Post	(4)	6,182	0.4545	0.498	0	1
Log(# of applications)	(5)	6,182	4.2921	0.9302	0	6.869
Log(total amount)	(6)	6,182	9.0828	1.2885	1.0428	12.7419
Black applicants	(7)	6,182	0.1827	0.203	0	1
Hispanic applicants	(8)	6,182	0.0208	0.0326	0	0.6667
Male applicants	(9)	6,182	0.5606	0.1122	0	1
Applicant Income	(10)	6,182	74.7787	38.9757	9.4575	612.2478
Log(population)	(11)	6,182	8.2538	0.4895	5.7071	10.0213
Median age	(12)	6,182	38.1464	6.2217	15.3	63.1
Household per capita	(13)	6,182	0.2495	0.0425	0	0.3801
Median income	(14)	6,182	42.3568	19.4325	7.8348	228.6472
Housing unit per capita	(15)	6,182	0.4886	0.2092	0.122	4.3688
Owner-occupied housing	(16)	6,182	0.5332	0.1795	0.0556	0.9412
Housing unit age	(17)	6,182	36.3918	13.3457	5	78

Table A2. Census-Tract Level—Correlation Table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1)	1															
(2)	0.784	1														
(3)	-0.102	-0.146	1													
(4)	-0.315	-0.264	0	1												
(5)	0.556	0.495	0.001	-0.25	1											
(6)	0.523	0.578	-0.039	-0.214	0.941	1										
(7)	-0.158	-0.288	-0.134	0.051	-0.489	-0.584	1									
(8)	-0.015	-0.065	-0.003	0.089	-0.052	-0.088	0.044	1								
(9)	0.099	0.187	0.144	0.04	0.332	0.381	-0.484	-0.067	1							
(10)	0.227	0.573	-0.151	-0.06	0.358	0.555	-0.412	-0.103	0.27	1						
(11)	0.072	0.044	0.078	0.035	0.637	0.508	-0.218	0.047	0.235	0.007	1					
(12)	0.132	0.189	0.126	0.08	0.262	0.29	-0.316	-0.166	0.258	0.188	0.01	1				
(13)	0.223	0.158	0.162	-0.106	0.451	0.39	-0.267	-0.102	0.254	0.059	0.263	0.484	1			
(14)	0.388	0.581	-0.117	-0.048	0.608	0.691	-0.456	-0.042	0.3	0.665	0.236	0.261	0.293	1		
(15)	0.119	0.234	0.007	-0.021	-0.039	0.05	-0.127	-0.079	0.035	0.154	-0.251	0.406	0.108	-0.018	1	
(16)	0.281	0.186	0.161	-0.048	0.597	0.531	-0.405	-0.107	0.358	0.148	0.4	0.351	0.545	0.541	-0.363	1
(17)	-0.286	-0.26	-0.147	0.127	-0.488	-0.463	0.293	0.016	-0.326	-0.113	-0.464	-0.131	-0.323	-0.305	-0.014	-0.385

Table A3. Census-Tract Level—Additional Estimations for Robustness Checks and Alternative Explanations

	Interaction	CEM	Falsification test		Without 2012	Alternative dependent variables			Alternative sample - All loans
Dependent Variables	**Denial amount per capita**	**Denial amount per capita**	**Median value**	**Median rent**	**Denial amount per capita**	**Log(# of denials)**	**Log(denial per capita)**	**# of applications per capita**	**Denial amount per capita**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment × Post	297.8868*** (52.3466)	106.9314** (36.6885)	1,784.5970 (1,676.6128)	-7.2346 (8.2619)	189.1495*** (43.8042)	0.0751*** (0.0198)	0.0585* (0.0263)	0.0013 (0.0010)	152.4717*** (42.6971)
Treatment × Post × Minority majority	105.0082** (34.1985)								
Log(# of applications)	452.0700*** (50.0176)	325.7362*** (54.0175)	-495.7902 (1,453.4927)	-9.9045 (6.0765)	471.4026*** (55.0956)	0.9829*** (0.0315)	1.0880*** (0.0401)		166.0631** (59.6545)
Log(total amount)	32.3095 (19.7194)	77.2651* (38.4776)	432.9405 (745.8331)	10.1672* (4.0468)	17.2919 (21.4453)	-0.1547*** (0.0225)	-0.1725*** (0.0286)		26.3374 (38.9420)
Black applicants	146.2918** (55.0387)	331.9642** (114.3974)	-3,460.6785 (2,697.2374)	-3.3557 (12.9812)	159.6310* (65.6481)	0.1681* (0.0767)	0.1669+ (0.0907)	0.0066*** (0.0019)	190.1883* (73.7676)
Hispanic applicants	41.0315 (147.2457)	138.8562 (312.7445)	-1,544.0653 (6,636.8257)	-4.5136 (25.8670)	62.4018 (163.1511)	0.3590* (0.1640)	0.4486* (0.2074)	0.0002 (0.0045)	288.1178 (238.8630)
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Census tract FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
MSA linear and quadratic time trends	YES	YES	YES	YES	YES	NO	NO	YES	YES
No. of obs.	6,179	2,052	6,178	6,171	5,617	6,182	6,143	6,179	6,179
No. of census tract	562	432	562	562	562	562	562	562	563
R^2	0.736	0.774	0.977	0.861	0.719	0.852	0.731	0.793	0.6890
Adjusted R^2	0.707	0.705	0.974	0.845	0.685	0.837	0.702	0.770	0.5995
F	15.20***	10.03***	3.93***	3.54***	12.79***	131.45***	107.7***	8.97***	5.03***
Root MSE	409.0695	263.1424	14924	70.26	420.9	0.2969	0.353	0.010	576.3035

Note: Robust standard errors clustered by census tracts in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Treatment = 1 for census tracts in South Carolina; Post = 1 for observations after 2012. Unit of analysis = census tract-year (2007-2017).

Table A4. Census-Tract Level—Estimations with Census Tracts Closer to the State Border

Dependent variables	**Denial amount per capita**	**Denial amount per capita**	**Denial amount per capita**	**Denial amount per capita**
Distance	**5 miles**	**7 miles**	**10 miles**	**12 miles**
	(1)	(2)	(3)	(4)
Treatment × Post	192.9332+ (104.3345)	134.5850* (66.3727)	239.6439** (81.5832)	245.7019*** (61.6038)
Log(# of applications)	288.5052** (104.1938)	444.0378*** (92.7308)	642.3454*** (97.8257)	700.8106*** (114.2353)
Log(total amount)	-28.7316 (36.2919)	-4.3015 (25.5409)	-17.5263 (37.9152)	18.0952 (46.9759)
Black applicants	19.3265 (93.4479)	102.2251 (106.8923)	206.2933+ (118.2254)	225.9272 (141.3370)
Hispanic applicants	-1,066.4218 (764.0682)	-424.9305 (626.6265)	363.0902 (452.4556)	-119.0898 (456.4077)
Other controls	YES	YES	YES	YES
Census tract FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
MSA linear and quadratic time trends	YES	YES	YES	YES
No. of obs.	330	649	1,594	2,342
No. of census tract	30	59	145	213
R^2	0.865	0.845	0.807	0.714
Adjusted R^2	0.834	0.818	0.781	0.679
F	3.00**	4.20***	8.17***	9.92***
Root MSE	243.6	272.2	391.6	532.6

Note: Robust standard errors clustered by census tracts in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. *Treatment* = 1 for census tracts in South Carolina; *post* = 1 for observations after 2012. Unit of analysis = census tract-year (2007-2017)

Table A5. Census-Tract Level—Matrix Completion Counterfactual Estimation Results (Athey et al., 2022)

		Denial per capita			Denial amount per capita		
		ATT	**Std. dev.**	***p*-value**	**ATT**	**Std. dev.**	***p*-value**
Overall ATT	**Treatment × Post**	0.5259	(0.1228)	[0.0000]	149.266	(32.6111)	[0.0000]
Pre-treatment ATT	*t*-5	-0.0267	(0.0180)	[0.1374]	-0.3863	(5.3897)	[0.9429]
	t-4	0.022	(0.0201)	[0.2730]	14.9143	(4.8913)	[0.0023]
	t-3	-0.0372	(0.0180)	[0.0392]	-14.6118	(4.7484)	[0.0021]
	t-2	-0.0352	(0.0190)	[0.0637]	-13.2011	(5.1510)	[0.0104]
	t-1	0.0394	(0.0204)	[0.0535]	8.0999	(4.8837)	[0.0972]
	t	0.0362	(0.0185)	[0.0503]	4.7967	(5.5803)	[0.3900]
Post-treatment ATT	*t*+1	0.3798	(0.1346)	[0.0048]	110.4181	(28.7673)	[0.0001]
	t+2	0.3551	(0.1456)	[0.0147]	148.4571	(37.9132)	[0.0001]
	t+3	0.6306	(0.1445)	[0.0000]	183.5922	(39.4817)	[0.0000]
	t+4	0.5993	(0.1490)	[0.0001]	129.4583	(38.8059)	[0.0008]
	t+5	0.6649	(0.1539)	[0.0000]	174.4058	(36.3710)	[0.0000]

Note: Unit of analysis = Census Tract-Year; N = 6,182. Census tract and year fixed effects are included. The same control variables in the main estimations are included. Robust standard errors clustered by census tracts are in parentheses. *p*-values are in brackets.

Table A6. Census-Tract Level—Estimations to Address Alternative Explanations

Dependent variables	**Denial per capita**	**Denial amount per capita**	**Denial per capita**	**Denial amount per capita**	**Denial per capita**	**Denial amount per capita**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Treatment × Post	**0.3412* (0.1578)**	**186.7769*** (42.0125)**	**0.3390* (0.1448)**	**174.8441*** (38.5491)**	**0.3479* (0.1491)**	**176.4315*** (39.4353)**
Number of amendments	-0.0040 (0.0255)	1.5462 (5.2712)				
Number of incidents			0.0145** (0.0053)	2.6990* (1.2957)		
Google Trends					-0.0016 (0.0016)	-0.2976 (0.3737)
Controls	YES	YES	YES	YES	YES	YES
Census tract FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
MSA linear and quadratic time trends	YES	YES	YES	YES	YES	YES
No. of obs.	6,179	6, 179	6, 179	6, 179	6, 179	6, 179
No. of census tract	562	562	562	562	562	562
R^2	0.6923	0.7281	0.6925	0.7282	0.6924	0.7281
Adjusted R^2	0.6585	0.6983	0.6588	0.6984	0.6585	0.6983
F	34.19***	14.39***	35.20***	14.44***	35.45***	14.49***
Root MSE	1.6671	415.4	1.666	415.2	1.667	415.3

Note: Robust standard errors clustered by census tracts in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. *Treatment* = 1 for census tracts in South Carolina; *post* = 1 for observations after 2012. Unit of analysis = census tract-year (2007-2017).

Appendix B

Additional Loan-Level Analyses

Table B1. Loan Level—Summary Statistics

Variables		*N*	Mean	Std. dev.	Min	Max
Denial	(1)	1,621,529	0.1549	0.3618	0	1
Treatment	(2)	1,621,529	0.3942	0.4887	0	1
Post	(3)	1,621,529	0.346	0.4757	0	1
Black applicant	(4)	1,621,529	0.0483	0.2144	0	1
Hispanic applicant	(5)	1,621,529	0.0291	0.1682	0	1
Male applicant	(6)	1,621,529	0.6576	0.4745	0	1
Applicant income	(7)	1,621,529	84.1999	110.821	0.9281	1.06×10^4
Log(population)	(8)	1,621,529	9.059	1.2995	2.3026	13.716
Median age	(9)	1,621,529	38.6719	6.0746	15.3	67.9
Household per capita	(10)	1,621,529	0.2663	0.031	0	0.4783
Median income	(11)	1,621,529	57.297	23.3194	3.1728	228.6472
Housing unit per capita	(12)	1,621,529	0.4588	0.175	0	7.5682
Owner-occupied housing	(13)	1,621,529	0.6425	0.1557	0	1
Housing unit age	(14)	1,621,529	25.2718	11.7252	3	78

Table B2. Loan Level—Correlation Table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)	1												
(2)	-0.001	1											
(3)	0.011	-0.001	1										
(4)	0.026	-0.047	0.004	1									
(5)	0.03	-0.027	0.024	-0.013	1								
(6)	-0.017	0.023	-0.002	0.119	0.052	1							
(7)	-0.035	-0.048	-0.026	0.003	-0.017	0.109	1						
(8)	-0.009	-0.039	-0.301	0.012	0.008	0.018	0.026	1					
(9)	-0.023	0.125	0.115	-0.077	-0.05	0.045	0.036	-0.172	1				
(10)	-0.033	0.116	-0.095	-0.049	-0.029	0.054	0.038	-0.001	0.576	1			
(11)	-0.071	-0.193	0.003	-0.001	0.006	0.039	0.236	0.025	0.032	0.243	1		
(12)	0.006	0.061	-0.033	-0.04	-0.027	0	0.008	-0.1	0.445	0.154	-0.192	1	
(13)	-0.048	0.03	-0.002	-0.014	-0.001	0.053	0.078	0.021	0.098	0.428	0.583	-0.495	1
(14)	0.035	0.017	0.099	-0.024	-0.028	-0.037	-0.04	-0.232	0.075	-0.289	-0.391	0.094	-0.39

Table B3. Loan Level—Additional Estimations for Robustness Checks and Alternative Explanation				
	CEM	**Without 2012**	**Logit**	**With credit union only**
Dependent variables	**Denial**	**Denial**	**Denial**	**Denial**
	(1)	**(2)**	**(3)**	**(4)**
Treatment × Post	0.0056* (0.0026)	0.0048 (0.0035)	0.126*** (0.0173)	0.0664*** (0.0135)
Treatment × Post × Black applicant	0.0350*** (0.0044)	0.0352*** (0.0045)		0.0958*** (0.0202)
Treatment × Post × Hispanic applicant	0.0210*** (0.0048)	0.0230*** (0.0046)		0.0601*** (0.0178)
Black applicant	0.0332*** (0.0031)	0.0288*** (0.0023)	0.2984*** (0.0122)	0.0929*** (0.0147)
Hispanic applicant	0.0253*** (0.0041)	0.0388*** (0.0031)	0.4470*** (0.0141)	0.0887*** (0.0222)
Male applicant	0.0236*** (0.0039)	-0.0176*** (0.0007)	-0.0845*** (0.0060)	0.0069 (0.0140)
Applicant income	0.0223*** (0.0037)	-0.0001*** (0.0000)	-0.0008*** (0.0001)	0.0326* (0.0145)
Other controls	YES	YES	YES	YES
No. of obs.	1,130,282	1,446,259	1,622,027	57,550
R^2	0.2883	0.280	0.0141[1)]	0.1644
Adjusted R^2	0.2862	0.278		0.1366
F	107.62***	81.94***	6833.37***[2)]	20.58***
Root MSE	0.309	0.307	-689347.86[3)]	0.341

Note: Robust standard errors clustered by census tracts in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. *Treatment* = 1 for loan applications in South Carolina; *post* = 1 for applications after 2012. The same controls, fixed effects, and linear time trends from Table 6 are used in Columns 1, 2, and 4. County and year fixed effects are included in Column 3. (1) Pseudo R^2, (2) Wald χ^2, (3) Pseudo log L

Appendix C

Additional Tables for Underlying Mechanisms

Table C1. Fannie Mae and Freddie Mac Approved Loans—Impact of SC DOR Breach on Applicant Credit Scores and Other Measures

Dependent variables	FICO credit score	I(credit score > 650) [a]	Interest rate	Debt-to-income ratio	Load-to-value ratio
	(1)	(2)	(3)	(4)	(5)
Treatment × Post	-1.9877* (0.7425)	-0.0111** (0.0034)	0.0160 (0.0105)	0.2909 (0.1768)	-1.1627* (0.5365)
Loan amount			-0.0007*** (0.0000)	0.0112*** (0.0013)	0.0296*** (0.0052)
Loan term			0.0035*** (0.0000)	0.0083*** (0.0003)	0.0435*** (0.0012)
Credit score			-0.0021*** (0.0000)	-0.0471*** (0.0013)	-0.0535*** (0.0019)
Median income	0.0785 (0.2671)	0.0012 (0.0009)	-0.0033 (0.0037)	0.1000* (0.0360)	-0.4689 (0.3495)
Poverty level	-13.4899 (29.3615)	0.0409 (0.0972)	0.4553+ (0.2248)	6.1840 (5.9039)	40.4959 (28.1498)
Owner-occupied unit	6.5465 (36.4910)	-0.1384 (0.1217)	0.6339* (0.2995)	-5.6373 (5.8669)	77.0932* (35.2880)
Unit age	-0.5011 (0.2992)	-0.0026** (0.0009)	-0.0045 (0.0029)	0.1666* (0.0705)	-0.1720 (0.2403)
Single-family unit	6.1603 (18.5163)	-0.0749 (0.0837)	0.7443* (0.2721)	-9.6953** (3.3204)	11.7137 (25.6916)
Zip code (3-digit) FE	YES	YES	YES	YES	YES
Year-month FE	YES	YES	YES	YES	YES
Property type FE	YES	YES	YES	YES	YES
Acquisition channel FE	YES	YES	YES	YES	YES
State linear time trends	YES	YES	YES	YES	YES
No. of obs.	544,177	544,177	544,177	458,239	543,818
R^2	0.0972	0.0437	0.878	0.109	0.141
Adjusted R^2	0.0970	0.0434	0.877	0.108	0.141
F	2.31+	6.99***	3642.35***	1491.34***	321.59***
Root MSE	49.63	0.213	0.330	10.37	16.44

Note: Robust standard errors clustered by census tracts in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. *Treatment* = 1 for three-digit zip codes in South Carolina; *post* = 1 for observations after 2012. Unit of analysis = approved refinance loans; [a] 1 if an applicant's credit score is over 650.

Table C2. Law Enforcement Agency-Level—Negative Binomial Regressions for the Impact of SC DOR Breach on Wire Fraud Incidents and Moderation Effects by Socioeconomic Conditions

Dependent variables	# of wire fraud incidents	# of wire fraud incidents	# of wire fraud incidents	# of wire fraud incidents
	(1)	(2)	(3)	(4)
Treatment × Post	0.2199*** (0.0524)	-0.4330** (0.1475)	2.0316*** (0.2483)	-1.1886*** (0.2183)
Treatment × Post × Minority population		0.0175*** (0.0037)		
Treatment × Post × Median income			-0.0464*** (0.0063)	
Treatment × Post × Poverty level				7.0151*** (1.0485)
County population	-0.0471 (0.0317)	-0.0333 (0.0317)	-0.0252 (0.0315)	-0.0275 (0.0315)
Minority population	-0.0046* (0.0019)	-0.0050** (0.0019)	-0.0055** (0.0019)	-0.0054** (0.0019)

Median income	-0.0041 (0.0030)	-0.0049+ (0.0030)	-0.0045 (0.0030)	-0.0048 (0.0030)
Poverty level	1.0311* (0.4718)	0.9456* (0.4723)	0.8427+ (0.4717)	0.7735 (0.4725)
Agency FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Month FE	YES	YES	YES	YES
State linear time trends	YES	YES	YES	YES
No. of obs.	147,708	147,708	147,708	147,708
No. of agencies	1,119	1,119	1,119	1,119
Log L	-55676	-55665	-55647	-55654
χ^2	1079***	1104***	1138***	1126***

Note: Standard errors in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. *Treatment* = 1 for agencies in South Carolina; *post* = 1 for observations after 2012. Unit of analysis = Law enforcement agency-month (2007-2017) in all U.S. states

Table C3. Zip Code Level—Negative Binomial Regressions for the Impact of SC DOR Breach on Customer Complaints against Financial Institutions

Dependent variables	# of total complaints	# of credit-related complaints	# of total complaints	# of credit-related complaints
	(1)	(2)	(3)	(4)
SC × Year	0.0410*** (0.0071)	0.0460*** (0.0127)	0.0354*** (0.0070)	0.0368** (0.0127)
SC × Year × Minority population			0.0690*** (0.0051)	0.1555*** (0.0090)
Year × Minority population			0.0701*** (0.0051)	0.1565*** (0.0090)
Year	0.1472*** (0.0008)	0.2267*** (0.0015)	0.1292*** (0.0015)	0.1848*** (0.0028)
County population	0.1849*** (0.0038)	0.2494*** (0.0044)	0.1834*** (0.0038)	0.2489*** (0.0045)
Minority population	0.1429*** (0.0307)	1.0381*** (0.0393)	-141.0154*** (10.2905)	-314.2181*** (18.2043)
Median income	-0.0354***	-0.0307***	-0.0355***	-0.0309***
	(0.0002)	(0.0003)	(0.0002)	(0.0003)
Poverty level	-11.9062*** (0.0752)	-12.1829*** (0.1099)	-11.9151*** (0.0753)	-12.2153*** (0.1102)
Zip code FE	YES	YES	YES	YES
Month FE	YES	YES	YES	YES
No. of Obs.	2,039,579	1,561,768	2,039,579	1,561,768
No. of zip codes	26,307	19,735	26,307	19,735
Log L	-1.205×10^6	-571652	-1.205×10^6	-571481
χ^2	81457***	45285***	81601***	45458***

Note: Standard errors in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. *SC* = 1 for zip codes in South Carolina. Unit of analysis = zip codes in 2011-2017 (all U.S.).

Table C4. Zip Code (Three-Digit) Level—Estimations with Lending Club Dataset

Dependent variables	Log(rejects + 1)	Log(rejected amount + 1)	Log(accepts + 1)	Log(accepted amount + 1)
	(1)	**(2)**	**(3)**	**(4)**
Treatment × *Post*	0.1907** (0.0620)	0.2267* (0.0951)	0.1269 (0.0888)	0.4510 (0.3343)
Zip code (first three digits) FE	YES	YES	YES	YES
Year-month FE	YES	YES	YES	YES
No. of Obs.	5,596	5,596	5,596	5,596
R^2	0.942	0.857	0.902	0.726
Adjusted R^2	0.938	0.845	0.894	0.704
F	9.46**	5.68*	2.04	1.66
Root MSE	0.475	0.965	0.481	2.898

Note: Robust standard errors clustered by three-digit zip codes in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. *Treatment* = 1 for zip codes in South Carolina; *post* = 1 for observations after 2012. Unit of analysis = three-digit zip codes-month (2007-2017).