Cash Substitution and Deferred Consumption as Data Breach Harms

Lior Strahilevitz
Lisa Yao Liu
Cash Substitution and Deferred Consumption as Data Breach Harms

Lisa Yao Liu* & Lior Jacob Strahilevitz**

Abstract

In a series of federal court cases, judges have debated whether data breaches that expose consumer information satisfy Article III of the United States Constitution’s requirement that plaintiffs suffer an “injury in fact.” Judicial opinions find no constitutional standing in a narrow majority of such cases, and plaintiffs are likely to lose absent causal links to subsequent identity theft or the disclosure of embarrassing information. Consumers whose data are breached thus are left without a federal remedy, and firms’ incentives to invest in data security are diminished. Our paper identifies a novel injury that results from data breaches. Upon learning about local data breaches, consumers immediately reduce their purchases and shift from credit card purchases to cash transactions. These effects are more pronounced with respect to purchases that are characterized by greater elasticity of demand. The changes we identify are short-lived, though. After a data breach, many consumers forego both the benefits of a short-term loan from their credit card issuer, cash-back benefits, and other perks associated with card purchases. They also forego purchases they otherwise would have made. In light of our empirical results, the standing barrier that has thwarted so many data breach suits may be easily surmounted.

* Assistant Professor of Business, Columbia University (yl4689@gsb.columbia.edu).
** Sidley Austin Professor of Law, University of Chicago (lior@uchicago.edu). We appreciate helpful comments from Will Baude, Philip Berger, Shirsho Biswas, Thomas Bourveau, Curt Bradley, Hans Christensen, Adam Davidson, Dhammika Dharmapala, Bridget Fahey, Gilles Hilary, Dick Helmholz, William Hubbard, Alex Jensen, Bob Kaplan, Christian Leuz, Mark Maffett, Mike Minnis, Michael Morse, Paul Schwartz, Chad Syverson, Luigi Zingales, and workshop participants at American Law and Economics Association Annual Meeting, Annual Machine Lawyering Conference, Annual Symposium on Applications of Contextual Integrity at the University of Chicago, Chicago Booth, Colorado Boulder Summer Conference on Consumer Financial Decision Making, Columbia Business School, Harvard Business School, McGill University Law School, and the University of Chicago Law School. We are grateful for research support from Columbia Business School, the Kilts Center for Marketing at the University of Chicago Booth School of Business, and the Carl S. Lloyd Faculty Fund. We gratefully acknowledge financial support from Columbia Business School, the University of Chicago Booth School of Business, and the Deloitte Foundation. All errors are our own. We own the analyses, which were calculated (or derived) based in part on data from The Nielsen Company (US), LLC and on marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Electronic copy available at: https://ssrn.com/abstract=4187930
Introduction

Data breaches of consumer information have become commonplace. In some instances, as with Target, Equifax, Yahoo!, and Marriott, these breaches have been nationwide, involved tens of millions of consumers, and generated recriminations and congressional hearings (Spinello 2021). In most instances, however, these breaches do not result in litigation or government enforcement actions (Park 2019), perhaps because breaches tend to be relatively localized, with thousands of consumers, rather than millions, being affected. In California, for example, the median breach that triggers data-breach notification obligations involves between 2,000 and 3,000 consumer records (Harris 2016). Regardless of whether they make front-page news, such breaches of consumer information adversely affect both the reputations and the market capitalizations of the firms where they occur (Gatzlaff and McCullough 2010). Banks respond to breaches by offering less favorable terms to the companies that suffered the breach (Huang and Wang 2021). There is reason to believe that they affect consumers as well.

Though many of the economic costs of data breaches are borne by merchants and credit card companies, substantial associated negative externalities seem to exist. These externalities materialize as time consumers spend monitoring their credit history, cancelling accounts, waiting for new cards to be issued, or implementing credit freezes and fraud alerts. As we explain below, the difficulty of tracing identity theft to a particular breach exacerbates the externalities.

Class action litigation is one way to internalize some of the externalities associated with data breaches. Litigation resulting from data breaches is now common. Class action attorneys often quickly file suit after major breaches, seeking to represent a large class of consumers whose personal information has been compromised or disseminated. But the federal courts have been divided over whether these cases can proceed to trial. The basis for that division is the federal
Constitution’s Article III case or controversy requirement. As it’s understood today by the federal courts, to be able to proceed toward a resolution of the merits, the plaintiff must demonstrate that they have suffered an injury in fact that is concrete, particularized, and actual or imminent. That injury also must be fairly traceable to the defendant’s conduct and redressable by a favorable judicial ruling. California v. Texas, 141 S. Ct. 2104 (2021). A concrete injury means a harm that is real rather than abstract. Bare procedural violations of a statute or rule are not concrete injuries. Spokeo, Inc. v. Robins, 578 U.S. 330 (2016). A particularized injury is a harm that affects the plaintiff “in a personal and individualized way,” rather than as some fraction of a generalized social harm. Id. An actual or imminent injury is a harm that has occurred, is occurring, or is about to occur, rather than a speculative harm that may not ever materialize. Summers v. Earth Island Inst. 555 U.S. 488 (2009).

As we show below, this burden of demonstrating Article III standing is easily satisfied by plaintiffs who have suffered identity theft that was proximately caused by the data breach. In rare cases in which the breach results in embarrassment to the individual whose data were compromised – such as the Ashley Madison leak – standing should be easy to prove for the plaintiff. But more typically, the nature of the information breached is not necessarily embarrassing – it may contain account information, purchase information, email addresses, dates of birth, or passwords. In these instances, the most obvious downside is the potential for this information to be used to facilitate identity theft. Yet, in the months after a typical breach, many of the consumers whose information was breached will not become identity theft victims. Further, as breaches become more common, and as the same personal information for a consumer is mirrored across many sites, establishing a causal link between a particular instance of identity theft and a specific breach becomes more challenging.
To be sure, as we explain below, some federal courts have recognized that the increased risk of identity theft resulting from a data breach is itself a concrete injury. Though this view rests on sound economic logic – rational consumers would be willing to pay some amount to reduce a non-trivial risk of identity theft – some courts have been unsympathetic to this view or flatly reject the proposition. Most prominently, language in recent Supreme Court decisions suggests the heightened risk of identity theft resulting from a breach might not by itself be an injury that is adequate to demonstrate Article III standing. The Court has suggested that rather than letting consumers sue to recover damages based on a probability of injury, they should wait to see who gets injured (and who doesn’t) and then recover accordingly.

One obvious response to this judicial skepticism about whether a heightened risk of future identity theft resulting from a data breach is a concrete injury is to consider the out-of-pocket and other costs that a reasonable consumer exposed to such a heightened risk might take. Such a prudent consumer might begin paying for a credit-history monitoring service, or they might spend more time scrutinizing their credit report. Alternatively, they might place a fraud alert on their credit report, which will lower the likelihood that an unauthorized account will be opened in their name, but will also increase the transaction costs associated with their own legitimate efforts to engage in economic activity by requiring them to go to greater lengths than usual to verify their identities. These transaction costs, whether expressed in terms of money or time, all seem concrete to economists. Alas, not so to many judges. As we explain below, in both its Clapper case and its Ramirez case, the Supreme Court expressed worries that a plaintiff could incur such expenses solely for the purposes of manufacturing standing where it otherwise would not exist.

The difficulty that plaintiffs sometimes have demonstrating standing in federal data-breach cases leads to underdeterrence of breaches (Park 2019). Cases get transferred to state courts, some
of which have onerous standing requirements themselves, and most of which have resource-
strapped judiciaries. If plaintiffs will have difficulties tying a particular incident of identity theft
to a particular data breach, firms will not have to internalize the external costs they impose on their
customers. Moreover, a legal system that denies standing to consumers who have incurred
reasonable expenses to reduce their risk of identity theft after a breach also permits firms to
externalize those harms to their consumers. In this legal environment, firms may decide to take the
associated reputational hit from failing to compensate consumers for these kinds of harms. This
hit will be mitigated if most firms in the marketplace offer the same terms to consumers.

In this environment, if one can demonstrate that (1) large numbers of consumers do incur
economic costs as a result of data breaches, and that (2) these costs are not manufactured (i.e., a
consumer is not incurring them only to strengthen their position in litigation), class action litigation
risk will increase for firms. This class action exposure in turn will prompt firms that control
consumers’ data to take more appropriate precautions to safeguard those data. Below, we explain
why we have reason to believe such economic costs do arise in the wake of data breaches that are
publicized. First, we describe the legal lay of the land in more detail.

I. The Law of Standing in Data-Breach Cases

The typical data-breach suit is brought as a federal class action on behalf of a large number
of consumers whose data were compromised by a breach. The plaintiffs may assert a cause of
action for negligence, breach of contract, breach of a fiduciary duty, breach of covenants of good
faith and fair dealing, public disclosure of private facts, fraud, or, in the case of a federal defendant,
violations of the Privacy Act of 1974 (5 U.S.C. § 552a et seq.). It is uncontroversial that plaintiffs
whose data have been misused by identity thieves have standing to sue in federal court. See, e.g.,
the typical data breach, some consumers may have suffered identity theft at the time the lawsuit is filed, but many more have not (yet) been victimized, perhaps because no crime was attempted or because credit card issuers, government actors, or other entities thwarted such attempts. Many federal appellate courts have heard cases involving this general set of facts, but the courts are divided on whether those who haven’t yet suffered from identity theft in the wake of the breach can sue in federal court. The data-breach standing question is perhaps the most consequential unresolved issue in data security law, so it is not surprising that it has spawned dozens of litigated cases, with plaintiffs being granted standing roughly 47% of the time (Haley 2020). The topic has inspired more than a hundred scholarly papers in law reviews (e.g., Cofone 2022, Cox 2016, Mank 2017, Wells 2021, Wu 2017).

On one side of the issue are cases such as Reilly v. Ceridian Corporation, 664 F.3d 38 (3d Cir. 2011), which involved the hacking of a payroll processing firms’ database of employee information, including Social Security numbers, bank account information, and dates of birth. The Third Circuit affirmed the lower court’s dismissal of the suit on the grounds that speculating whether the hacker had copied and understood the plaintiffs’ personal information, intended to use it to commit future criminal acts, and would be able to do so if they had such intent would be inappropriate. The court also rejected arguments put forward by the plaintiff that the future risk of identity theft and plaintiffs’ out-of-pocket expenditures to detect identity theft constituted harms.

By contrast, Remijas v. Neiman Marcus Corporation, 794 F.3d 688 (7th Cir. 2015), found standing after hackers breached a database containing customer information, including credit card numbers. Neiman Marcus customers whose information had been hacked but evidently had not been used to commit identity theft could establish standing based on the objectively reasonable likelihood that identity theft would later occur. As the court put it, “Why else would hackers break
into a store’s database and steal consumers’ private information? Presumably, the purpose of the hack is, sooner or later, to make fraudulent charges or assume those consumers’ identities.” Parting ways with the Reilly court, it also regarded the plaintiffs’ expenditure of time and money to protect themselves against future identity theft as concrete injuries that conferred standing on the plaintiffs (Strahilevitz 2020).

This basic divide between Reilly and Remijas has prompted other appellate courts to take sides. The Fourth and Eleventh Circuits follow the Reilly approach. See Beck v. McDonald, 949 F.3d 262 (4th Cir. 2017) and Tsao v. Captiva MVP Restaurant Partners, LLC, 986 F.3d 1332 (11th Cir. 2021). See also In re Equifax, Inc. Customer Data Security Breach Litigation, 999 F.3d 1247 (11th Cir. 2021) (when some customers suffer identity theft and others have not yet, the risk the latter group faces is sufficiently high to give rise to an injury in fact). The Second, Ninth, and D.C. Circuits have embraced logic similar to that of Remijas. See McMorris v. Carlos Lopez & Assoc., LLC, 995 F.3d 295 (2d. Cir. 2021); In re. Zappos.com, Inc., 888 F.3d 1020 (9th Cir. 2018); and Attias v. Carefirst, Inc., 865 F.3d 620 (D.C. Cir. 2017). Thus, when a class of people includes some identity theft victims, the question of whether those not victimized have Article III standing to sue in federal court is energetically disputed. As one federal appellate judge recently noted, the issue cries out for Supreme Court resolution. See Tsao, 986 F.3d at 1345 (Jordan, J., concurring in judgment).

The approach that some courts have taken – looking at the presence of identity theft victims to shed light on whether the non-victims face a reasonably high risk of injury – is unsatisfying. Identity theft occurs frequently, such that in a large group of people, some will become identity theft victims over the course of a few months regardless of whether a particular breach happened. The survey evidence suggests that roughly one in every twenty consumers are victimized by
identity theft every year (Harrell 2019; Strahilevitz 2020). Causation generally cannot be resolved at the earliest stage of legal proceedings, when motions to dismiss for lack of standing are typically brought. Courts have not grappled with this base-rate problem in a nuanced way, and some evidently assume implicitly the appropriate base rate for identity theft is zero, at least for the purposes of determining standing. See McMorris, 995 F.3d at 301-02.

Despite numerous opportunities during the past couple of decades, the Supreme Court has not resolved whether prudent precautions taken by consumers after a data breach are themselves injuries that create standing. In Friends of the Earth, Inc. v. Laidlaw Environmental Services, 528 U.S. 167 (2000), the Court found that environmental organizations and neighbors had standing to sue a polluter that had emitted excess mercury into a watershed. Neighbors testified that, prior to the pollution, they made recreational use of the river for fishing or canoeing, but after the pollution, they avoided the area because of a concern about contamination. Laidlaw suggested that if a defendant’s actions chilled plaintiffs’ willingness to engage in activities that provided them with tangible benefits, that deterrence would constitute an injury in fact for the purposes of Article III.

Monsanto Co. v. Geertson Seed Farms, 561 U.S. 139 (2010), further underscored that reasonable precautions, this time by the growers of conventional alfalfa in the wake of a governmental decision to de-regulate genetically modified alfalfa, gave rise to injuries in fact that established standing. More precisely, given the existence of genetically engineered alfalfa and the preference of many customers for natural alfalfa seeds, growers would need to begin testing their crops to show that their natural seeds had not been contaminated, and some of them would have to begin importing seeds from overseas, where the genetically modified alfalfa was still banned. The Court did not examine the rationality of the underlying belief systems of consumers opposed to genetically modified foods. Rather, it took those consumer reactions as a given and held that the
costs of addressing those concerns would injure alfalfa growers “even if their crops were not actually infected with the” genetically-modified alfalfa. Id. at 155.

In Clapper v. Amnesty International, 568 U.S. 398 (2013), the Court evaluated the standing of lawyers, journalists, and human rights advocates communicating with foreign persons who were likely subject to warrantless wiretaps by the US government. The Court emphasized that to satisfy Article III’s injury-in-fact requirement, the harm to the plaintiffs had to be “concrete, particularized, and actual or imminent; fairly traceable to the challenged action; and redressable by a favorable ruling.” The Court concluded the plaintiffs’ injuries were too speculative to constitute an actual or imminent injury in fact because they had not shown their conversations were ever intercepted. Of course, as commentators pointed out, the plaintiffs’ failure was hardly their fault – the government program itself was largely secret and the targets of foreign intelligence surveillance remained in the dark about whether the government was eavesdropping. Therefore, Amnesty International and the other plaintiffs were caught in a Catch-22 (Kwoka 2017; Re 2014). If the government exercised its authority pursuant to the statute, it might be violating someone’s rights, but the secretive nature of the surveillance program made determining who that someone was impossible. The plaintiffs tried to sidestep this issue by noting their communications had been chilled. Since learning about the secret surveillance program through media leaks, they had avoided certain email and telephone conversations, flying overseas instead so they could speak confidentially to their clients. This workaround, the Court said, was inadequate. The plaintiffs “cannot manufacture standing merely by inflicting harm on themselves based on their fears of hypothetical future harm that is not certainly impending. . . If the law were otherwise, an enterprising plaintiff would be able to secure a lower standard for Article III standing simply by making an expenditure based on a nonparanoid fear.” The Court distinguished Laidlaw by noting
it involved indisputable violations of the law.

Spokeo, Inc. v. Robins, 578 U.S. 330 (2016), was the next major standing case from the Court. In Spokeo, the defendant violated the federal Fair Credit Reporting Act (FCRA) by disseminating inaccurate information about Robins. The case turned on whether the dissemination of inaccurate information constituted a concrete injury. The Court held that just because Spokeo had apparently violated the FCRA, that “bare procedural violation” was not sufficient to establish a concrete injury as a matter of law. Rather, the inaccurate information disseminated needed to either produce a harm or a “material risk of harm.” On remand, the Ninth Circuit found that Spokeo’s misstatements of Robins’s education level, employment status, and marital status was a sufficiently concrete injury to warrant standing. The court noted these discrepancies between the profile and Robins’ own representations about himself might cause an employer to question his truthfulness or regard him as overqualified for the positions he was seeking. See Robins v. Spokeo, 867 F.3d 1108, 1117 (9th Cir. 2017).

Transunion LLC v. Ramirez, 141 S. Ct. 2190 (2021), represents the Supreme Court’s most recent pronouncement on standing. Ramirez involved a class action suit over a credit reporting agency’s false positives. Many people were erroneously flagged as potential terrorists or drug traffickers by Transunion because they shared the same surname as others who were on a government watch-list. For people who were misidentified as potential terrorists or criminals and had this information disseminated to third parties, the Court had “no trouble concluding that [they] suffered a concrete harm that qualifies as an injury in fact.” Other members of the plaintiff class whose misidentification was never disseminated to a third party could not demonstrate a concrete injury in fact. Of course, the possibility that the erroneous information would be disseminated later could have posed a risk for these other plaintiffs, but the Court said that this risk needed to
materialize for them to establish standing. The Court added a footnote that a plaintiff who was exposed to a risk of future harm might suffer “current emotional or psychological harm” as a result, though the Court dodged the question of whether such injuries alone would give rise to standing.

Reviewing the federal appellate cases, then, some legal conclusions are clear and others are murky. Someone whose private information is breached and is subsequently victimized by identity theft may have standing to sue, provided a causal link between the breach and identity theft is plausible. Yet it is uncertain whether someone whose information is breached but has not yet suffered identity theft has standing to sue if other consumers whose information was in the same breached database were victimized by identity theft. The courts are split on this question and are divided over whether sensible expenditures to monitor one’s credit history after a data breach constitute an injury in fact.

Moving to the Supreme Court cases, if data are breached and cause any reputational injury to a consumer – either because the underlying information is embarrassing or because of discrepancies between the disclosed information and the way a consumer accurately presents himself to others – an injury in fact has occurred. Spokeo tells us as much. And Laidlaw and Geertson Seed Farms make it clear that if a consumer prudently avoids doing things they enjoy as a result of an unlawful act by the defendant, that prudent avoidance constitutes an injury in fact, either for the consumer herself or for a producer catering to those consumer interests. But Clapper adds the key caveat that if it is not clear the defendant has engaged in unlawful conduct with respect to the plaintiffs, then no actual or imminent injury has taken place, and hence plaintiffs have no standing. Ramirez further clarifies that if personal data are exposed but no evidence exists that they were accessed by any third party that might use the data to harm the consumers whose information leaked, demonstrating an injury in fact from the mere exposure itself will prove challenging.
This legal uncertainty raises an intriguing question. What if data-breach litigation’s focus on identity theft injuries is largely misplaced? That is, where a breach has occurred and the integrity of consumers’ personal information has been compromised, are the kinds of harm-avoidance steps that the neighbors took in *Laidlaw* or that alfalfa growers took in *Geertson Seed Farms* sufficient to count as concrete injuries? As we show below, evidence indicates consumers respond to known data breaches by curtailing their spending in the short term and switching from credit cards to cash, thereby foregoing the benefits of an interest-free, short-term loan from their credit card issuer and the possibility of earning reward points from credit card purchases. Are these consequences concrete “injuries in fact” for the purposes of Article III standing? The Supreme Court has not decided any cases that squarely address that question, but a few recent lower-court precedents are instructive.

The Eleventh Circuit’s decision in *MSPA Claims 1, LLC v. Tenet Florida, Inc.*, 918 F.3d 1312 (11th Cir. 2019), examines whether a delay in receiving a financial benefit constitutes a concrete injury for the purposes of standing. The court held that a health insurer that was eventually reimbursed for a payment it made but did not owe had suffered an injury during the period while it was waiting to receive its money: “The question is whether delay alone is a ‘concrete’ injury. It is. . . . The inability to have and use money to which a party is entitled is a concrete injury.” *MPSA* could be read as suggesting that if consumers stop using their credit cards for a month or two after a data breach, this behavioral shift could constitute an injury in fact.

*Thomas v. John A. Youderian Jr., LLC*, 232 F. Supp.3d 656 (D. N.J. 2017) is especially instructive. In that case, the plaintiff alleged that a communication from the defendant stated that customers repaying their debts using a credit card would be charged a $3.00 fee over and above what customers paying by check would owe. Collecting such a surcharge was unlawful. *Thomas*
did not pay by credit card, so he did not pay a surcharge. Nonetheless, the court held that in those circumstances, causing consumers to “forego the convenience of paying the bill on credit” did amount to an injury in fact sufficient to establish standing under *Spokeo*. When it comes to cash there does not seem to be a *de minimis* exception for injuries in fact. As Justice Alito recently noted, “Even $1 in harm is enough to support standing.” California v. Texas, 141 S. Ct. 2104, 2124 (2021) (Alito, J., dissenting).

Putting *MPSA* and *Thomas* together, consumers whose confidential data were compromised in a breach and who temporarily shied away from using their credit cards afterwards could have suffered a concrete injury in fact. Such consumers would have been deprived, by circumstances, of utilizing their preferred payment option. In making that choice, consumers who paid off their credit card bills in full would have foregone the additional benefit of a roughly month-long interest-free loan from their credit card issuer until the bill’s due date. These “convenience users” of credit cards pay off their outstanding balances in full each month and account for approximately 56% of all households in the US with credit cards (Basnet and Donou-Adonsou 2016). Consumers who had signed up for rewards cards that provided cash back on all purchases would have lost income as well. One lingering question, of course, is whether it matters that consumers who curtailed credit card usage would be doing so to defend themselves against a threat that has suddenly become more salient. *Thomas* suggests so, though it is not quite on all fours.

Fifty-eight percent of Americans have credit cards that offer rewards in the form of cash back or air travel (Reales and O’Connell 2017), and these rewards significantly shift consumers from using cash to credit transactions during normal economic times (Carbo-Valverde and Linares-Zegarra 2011). Consumers who would benefit from the fraud-protection policies of credit
card issuers, a benefit that does not exist for cash, also suffer tangible losses (Basnet and Donou-Adonsou 2016). Thus, although consumers who carry substantial credit card balances plausibly obtain a net welfare benefit from reducing their consumption, these consumers also suffer corresponding concrete harms that plausibly constitute an injury in fact, above and beyond the thwarting of their (possibly misguided) preferences to incur credit card debt.

Assuming MPSA and Thomas have provided an accurate characterization of the contemporary law of standing, such losses for consumers would constitute injuries in fact. Laidlaw suggests the self-imposed nature of those losses would not prevent their characterization as concrete. Under Clapper, if the defendant’s unlawful conduct is established, plaintiffs injured in this way should be able to get their day in federal court. Although the injuries to each consumer would be small, the aggregate harm from a large data breach would be large, rendering class action litigation viable.

Having established the possibility that consumer mitigation efforts after a data breach could help resolve a contentious Article III standing problem in contemporary class action litigation, we turn to the empirical evidence of how consumers respond to such breaches.

II. The Effects of Data Breaches on Consumer Behavior

Recent data breaches have exposed the private information of many consumers. For example, in 2017, 20 million Social Security numbers were exposed through cyber breaches. In 2017, an Equifax breach resulted in the compromise of 143 million US consumer accounts. The 2013 and 2014 breaches of Yahoo! affected 3 billion and roughly 500 million people, respectively. The 2013 Target data breach exposed the credit card information of 40 million customers. In November 2018, when Marriott International identified a breach of its Starwood reservation system, approximately 500 million people were affected (Al-Muslim et al. 2018).
Corporate managers are also aware of the issues stemming from data breaches. In the 2018 Hiscox CFO Research survey, 204 senior finance and risk executives were asked to identify the risks that most concerned them. At the top of the list was “data breach,” chosen by 42% of respondents. A 2017 survey by PricewaterhouseCoopers found 67% of CEOs believed AI and automation would have a negative impact on stakeholder trust over the next five years.

Institutionally, data breaches have significant negative consequences. For instance, Target’s 2013 breach sent quarterly profits into a downward spiral, which ultimately cost the CEO his job. Additionally, the stock price dropped more than 2% on announcement day and insider trading was detected. Under the EU’s General Data Protection Regulation, fines for noncompliance can be up to 4% of a company’s worldwide revenues, due to the increased penalties for corporate privacy violations.

A. Literature Review

Prior research shows data breaches harm individuals (e.g., Janakiraman et al. 2018); up to this point, however, evidence on whether this individual harm can aggregate up to the local level is scant. Specifically, whether individuals can easily substitute to non-breached firms to mitigate the negative impacts of data breaches, how large the effect is at the regional level (if any), and how long the effect persists is unclear. Our paper investigates these questions and examines whether data breaches lead to a chilling effect on consumer spending at an aggregate level.

Prior work documents some firm responses to data breaches. Kamiya et al. (2021) show firms respond to online attacks by reducing CEO risk-taking incentives and strengthening risk management. Akey et al. (2021) reveal firms rebuild intangible capital through activities such as corporate social responsibility (measured by charitable contributions) after data breaches. Duffie and Younger (2019) discuss the economic harm of cyber-attacks on bank deposits, potentially
leading to cyber runs and turning an operational event into a liquidity event. Sheneman (2018) finds data breaches increase firms’ cost of debt. Spanos and Angelis (2016) conduct a systematic review of the literature on the impact of data-breach events on the stock market and find the majority of these studies (more than 75%) document significant negative effects of security breaches on the stock price of breached firms.¹ Haislip et al. (2019) find non-breached peers experience negative capital market reactions and have higher audit fees, and that affected insurers have negative market reactions. Janakiraman et al. (2018) use a single retailer store as their setting and find a significant decrease in consumer spending after the breach. This literature sheds little light, however, on the question of whether data breaches harm consumers at an aggregate level, a topic we explore below.

Prior research has examined heterogeneity in the marginal propensity to consume out of income and credit-availability shocks (Parker 1999; Johnson, Parker, and Souleles 2006; Jappelli and Pistaferri 2014). Given the persistence and concentration in household consumption choices documented in prior literature (e.g., Bornstein 2018; Neiman and Vavra 2019), it is less clear whether and how data breaches, which are direct negative news to consumers, are associated with household spending at an aggregate level and with the persistence of effects. For example, if consumption does not materially change after a breach, it would suggest consumers have

---

¹ See Campbell, Gordon, Loeb, and Zhou (2003), Cavusoglu, Mishra, and Raghunathan (2004), Acquisti, Friedman, and Telang (2006), Goel and Shawky (2009), and Gatzlaff and McCullough (2010). Cavusoglu, Mishra, and Raghunathan (2004), for example, find breached firms lost an average of 2.1% of their market value within two days of the announcement and an average of $1.65 billion in market capitalization per breach. Campbell et al. (2003) find a highly significant negative market reaction to information security breaches involving unauthorized access to confidential data, but no reaction when no confidential data are compromised. Rosati, Cummins, Deeney, Gogolin, Van der Werff, and Lynn (2017) conduct an event study on 74 data breaches between 2005 and 2014 and find data-breach announcements have a short-term positive effect on both bid-ask spread and trading volume. Recent studies, including empirical studies (Akey, Lewellen, Liskovich, and Schiller 2021) and theoretical studies (Kamiya, Kang, Kim, Milidonis, and Stulz 2021), corroborate these findings. Particularly, Kamiya et al. (2021) develop a rational model and find a successful attack results in a significant loss of shareholder wealth when personal financial information is compromised.
internalized the possibility of personal information leakage or that these shocks are substitutable, diminishing the distributional consequences and welfare losses associated with data breaches. By contrast, if consumer spending changes meaningfully, it would suggest the impact of data breaches is passing through to consumption, and hence that important distributional consequences for welfare exist. Our results below suggest consumers lose trust in the product market (Giannetti and Wang 2016).

Additionally, a relevant literature investigates the economic consequences of disclosing consumer-related information. Jin and Leslie (2003) study how disclosing information to consumers affects restaurants’ hygiene quality and consumers’ health outcomes. Other papers (e.g., Mathios 2000; Cawley et al. 2020) examine the effects of the mandatory disclosure of food nutrition and calorie counts on consumer behavior. We contribute to this literature by examining whether consumers who experience the unanticipated disclosure of their personal information “price protect” themselves, both at the individual and aggregate levels.

B. Current Study – Empirical Approach

Examining consumer spending is important for two reasons. First, data breaches reveal consumers’ personal (and sometimes sensitive) information, which means consumers are directly impacted economic agents. Consumers are important stakeholders, and their reactions are directly linked to firms’ revenue in financial statements, making consumer behavior an economically significant outcome (e.g., Berger et al. 2017). Second, unlike the capital market, where stock prices reflect long-term valuation impacts, we can use consumer responses to document a short-term reaction. Specifically, we use detailed consumer visit data to examine how consumers react to data breaches in the short term.

Data breaches can have a chilling effect on consumer spending through two underlying
mechanisms: trust reductions (the direct channel) and substitutions (the indirect channel). Under
the direct channel, consumers lose trust in companies’ data protection after a breach; this distrust
can spill over into the product market via correlated learning (e.g., Sridar et al. 2012; Che et al.
2015; Ching and Lim 2020). By scaling back their participation in the product market, consumers
can avoid providing personal information to companies.\(^2\) Trust reduction could result from
consumers’ intrinsic preference for privacy (e.g., Becker 1980; Farrell 2012; Goldfarb and Tucker
2012) or as a result of feelings of betrayal following the unexpected revelation of their personal
information (e.g., Guiso, Sapienza, and Zingales 2008). Under the indirect channel of substitutions,
consumers switch to the rivals of breached firms. However, based on consumers’ revealed
preferences, these rivals are not a consumer’s first-best choice. Therefore, when consumers face a
constrained optimization purchase decision, some of them may reduce their spending.
Furthermore, consumers can substitute between payment methods. For example, when purchase
amounts are relatively small (e.g., for groceries), consumers are more likely to use cash than a
credit card if they fear exposing identity information.

The magnitude and the persistence of this chilling effect are a priori unclear. If consumers
internalize the harm of data breaches in the information technology age over time, or if consumers
anticipate companies might become safer after a data breach due to the probable investment in
data protection (e.g., Kamiya et al. 2021), consumers may not react negatively to the breach news.
Additionally, because data breaches are not a fundamental shock predicted by the permanent-
income model or even a temporary change in wealth/income (e.g., changing house prices (Mian

\(^2\) For example, in their 2013 10-K, Target disclosed that “we believe that the greatest risk to our business arising out of the Data Breach is the negative impact on our reputation and loss of confidence of our guests.”
Some (small-n, convenience sample) survey research finds most consumers did not change their behavior even after a widely publicized, nationwide Equifax breach (Zhou et al. 2018). We use two datasets with granular consumer purchase information to study the effect of data breaches on consumer spending. The first dataset is R.L. Polk, which has monthly ZIP-code data on new automobile purchases. The second dataset is the Nielsen Consumer Panel Dataset, which includes households’ purchasing information (i.e., when and where households make purchases and how much they spend). The Nielsen dataset is large, with 125,000 households in more than 20,000 ZIP codes. The combination of geographic detail and the high frequency of purchase information makes these datasets unique; these high frequency data give us a valuable perspective on how communities and households quickly react to unexpected, announced breach information. Furthermore, the detailed geographic and time-series information is instrumental in controlling for confounding economic shocks. Another feature of both datasets is that they cover consumer-facing industries, where failing to protect consumer information may lead to a significant loss in consumer trust. These advantages make these industries a convenient laboratory to examine consumer reactions to data breaches.

Using these two datasets, we also examine how the chilling effect varies by consumers’ demand elasticity. For example, if the demand elasticity of a certain product is zero, consumers will not change their purchasing behavior. Because the demand elasticity of automobiles is larger than the demand elasticity of groceries, we explore and compare the durable-goods (e.g., automobiles) consumption response to financial-institution breaches with the non-durable (e.g.,

---

3 The standard economic consumption model cannot explain responses in consumer spending, because data breaches are not income (e.g., permanent income hypothesis and life-cycle theory) or macroeconomy/policy news (e.g., precautionary saving for uncertainty), nor are they related to household liquidity/borrowing constraints.
groceries) consumption response to retailer breaches. We then examine the extensive margin (e.g., how many shopping trips consumers make), the aggregate household spending, and the substitution between payment methods.

To empirically assess the aggregate impact of data breaches, we exploit variation in a city’s exposure to data breaches and correlate it with city-level changes in consumption. Consumer monthly spending in a city is our unit of analysis. Using city-level outcomes allows us to estimate the total effect of data breaches, including any spillover effects (i.e., changes in consumer spending stemming from firms in the same city as breached companies). Furthermore, because city information is explicitly disclosed (in newspapers and/or on government websites), local consumers presumably would have the strongest reaction, increasing the power of the test. Thus, we exploit within-county variation across cities that experience data breaches in the same month. Throughout the analysis, we include time-varying city or county characteristics to capture any unobserved time-varying variation at the city or county level and to allow for different trends in each county or city. The empirical specification relies on the assumption that changes in the economy will have similar effects on households in the same county.

We find that data breaches have a significant impact on consumer spending in the short-term. Although our two datasets are from different sources and vary in granularity, our results consistently show data breaches reduce consumer spending at an aggregate (i.e., city) level in the short-run. We find consumers reduce the number of shopping trips and are more likely to use cash after the breach, both of which contribute to a decrease in spending. Specifically, consumer spending in Nielsen data declines by approximately 2% ($25 per household), number of trips declines by approximately 2-3%, and use of credit cards declines by roughly 1%. Further, to explicitly evaluate whether local economic conditions or other confounding factors lead to a
reduction in household spending, we test for and do not find a decline in household income in data-breach cities. The decrease in shopping trips is consistent with the trust-reduction channel, and the increased likelihood of cash payments is consistent with the payment-substitution channel. Furthermore, the reduction in automobile purchases is larger than the reduction in retail spending, a finding consistent with durable goods (e.g., automobiles) having a larger demand elasticity than nondurable goods (e.g., groceries).

With respect to the persistence of the chilling effect, we find the reduction in consumer spending is short-term (persisting for no longer than one month), after which the drop is reversed, suggesting consumers are likely to postpone instead of eliminating consumption. This temporary effect is consistent with the prediction that data breaches are not a fundamental source of income shocks, and consumer behavior is likely to be rational. In particular, after a data breach, which is a salient signal about revealing consumers’ personal information, they reduce their purchasing and shopping trips. However, nothing severe regarding their personal identities occurs for a period of time or consumers are satisfied with firms’ post-breach responses, causing consumers to return to their pre-breach behavior.

The temporally limited behavioral response we observe is also consistent with research from the social sciences finding that salient and sudden events may cause dramatic behavioral changes that prove to be short lived because of discounting and fading memories (Durbach and Montibeller 2018). Along these lines, McCoy and Walsh (2018) find that a major wildfire depresses real estate prices in fire-prone areas for a year, but after two or three years, the effect dissipates and prices return to their baseline. Bento et al. (2020) find that local internet searches for COVID-related information immediately spiked after the first case was reported in their jurisdiction, and then declined within a week or two. Abouk and Adams (2013) find traffic
fatalities decline immediately in response to prohibitions on texting while driving and then return to previous levels within a few months. Further, Kugler and Strahilevitz (2017) find that in the wake of a landmark Supreme Court decision expanding privacy rights in their cell phones, citizens’ expectations of privacy against cell phone searches immediately spike among those who heard about the decision, but privacy expectations had returned to baseline one year after the decision and remain there after two years as well.

One potential concern is that the widespread dissemination of massive breaches (e.g., the Target data breach) could underestimate the treatment effect, because branches of nationally recognized firms are negatively affected but may not be included in the treatment group. To examine this possibility, we exclude nationwide data breaches and repeat the analyses. We find a larger magnitude in the grocery-shopping analysis and a slight increase in the magnitude of the results for automobile spending. Another consideration is the location disclosure in our breach dataset. Some breaches are highly localized and occur at one location. For instance, Nyack Hospital in New York’s 2011 breach occurred at an organization that has just one place of business. In other cases involving larger companies with multiple offices or branches, the breach was recorded as occurring at either the corporate headquarters or the branch. To this end, in Table 8 we replicate our primary analysis using only the cities containing localized breaches and breaches that were limited to employee data (with the same control group as our main analysis). We find similar results.

To provide empirical evidence on the substitution channel, we examine the variability of retailers’ customer lists and market shares. The underlying assumption is that if the retailer-consumer relationship is stable, the variability of the retailer’s customer list and market share should not change (after controlling for consumers’ outside options and adjusting for seasonality).
However, if consumers switch retailers abnormally, we would find greater variability in retailers’ customer lists and market shares. We find this variability increases in breached cities with local newspaper coverage. This suggestive evidence supports the indirect-substitution channel whereby consumers substitute their spending from breached to non-breached companies.

To further explore whether information dissemination is an important channel for consumer awareness, we exploit pre-existing variation in the extent to which different cities have local newspapers. Media plays an important role in informing citizens about government policy and firm misconduct (Strömberg and Snyder 2010, Dyck et al. 2008), and local newspapers can facilitate information dissemination and consumer awareness of data breaches. We examine how the local dissemination of data breaches affects consumers’ subsequent monthly consumption decisions, relative to those in a control group. Consistent with information dissemination being an important channel for consumer awareness, we find that the chilling effect is larger in cities with a pre-existing local newspaper.

It stands to reason that consumers whose data were actually breached – as opposed to those who live in the same community where a breach occurred – would have their consumption chilled to a greater degree. For such consumers, the data breach notifications that are typically required by state law (Verstraete and Zarsky 2021), would provide an additional means of information dissemination. Local media could amplify or reinforce knowledge of a data breach from these individual notifications, making breaches more salient or reaching consumers who discarded data breach notification communications they were sent, while also raising awareness among consumers who are familiar with a local business but did not themselves experience a breach.

Finally, using SafeGraph individual consumer visit data, we conduct a short-window event study to document individual-level consumer behavior change. We find a significant decrease in
consumer visits to a breached store three to seven days after the announced breached event (“post-breach period”). We also perform quantile regressions to understand consumer responses across different distributions. The goal is to determine whether our findings are explained by a large number of consumers changing their behavior slightly or by a small number of consumers changing their behavior significantly. We see a decline in consumer visits to the breached firms in the post-breach period, across all quantile distributions. This finding strengthens the case for an Article III injury from deferred consumption by showing it not to be idiosyncratic.

Because of consumer privacy concerns researchers cannot readily obtain data that allows us to compare the behavior of consumers whose data was breached to those whose data was not breached. Our study takes the next most informative path – comparing communities where many consumers experienced a data breach, or where a firm that suffered a data breach was headquartered, to comparable locales where few consumers experienced a breach and no breached firms were located. While there are likely to be within-community information spillovers, such that knowing your neighbor was breached affects your purchasing behavior, it is likely that consumers whose data have been breached or whose employer’s databases were penetrated would experience a more pronounced behavioral shift. So city-level data would pick up behavioral changes by people whose data was breached and received individual notice, spillover effects among people whose data were not breached but heard about a local breach and were spooked, as well as less attentive consumers whose behavior did not change. As a matter of legal doctrine, it’s the behavior of the first of these groups that is relevant to the doctrinal question of Article III standing in federal court.

As we stated at the outset, our primary goal in this paper is to shed light on an important and unresolved legal question that affects the potential for data breach class actions to cause firms
to internalize the externalities associated with breaches: whether consumers suffer concrete and particularized harms that result from data breaches, apart from resulting identity thefts or an increased risk of identity theft. By showing evidence that consumers in communities where breaches have occurred curtail their spending in the short term and switch from credit card to cash transactions, we identify novel theories of data-breach harm that plausibly clear the doctrinal hurdles established by Article III.

III. Empirical Analyses

Data breaches are economically interesting in their own right. With the advancement of information and communication technology (especially AI and big data), the costs of collecting, parsing, and storing detailed personal information have decreased, motivating sellers to collect more data (Goldfarb and Tucker 2012). Consumer privacy can be viewed as an information asymmetry, not just within focal transactions, but also over time. In the economic theory, privacy is modeled as an information transfer from consumers to firms; many theories show consumers find this transfer undesirable (Posner 1981; Acquisti and Varian 2005; Hui and Png 2006). Information economics theories demonstrate both buyers and sellers have incentives to disclose or to withhold private information that can affect market efficiency. A seller may commit to a consumer-friendly data-usage policy ex ante, but could share the data with other entities in the future, creating ongoing information asymmetry (Jin 2018). Data breaches could also influence the extent and direction of firms’ data-based innovation activities (Goldfarb and Tucker 2012).

Data breaches decrease consumers’ trust in the product market, which negatively affects spending behavior. Breaches hurt the relationship between firms and consumers by exposing consumers to the possibility of identity theft. This potential exposure leads to a household’s fear that sensitive information could be stolen when the household purchases products. Additionally,
although sellers commit to careful data usage ex ante, third parties on the seller’s platforms could create a negative externality. Anticipating this externality, consumers are prompted to exit the market (at least temporarily) to mitigate the possibility that their personal information could be stolen. This ongoing information asymmetry discourages consumers from providing information ex ante, leading to hesitation at the time of purchase.

If consumers can perfectly substitute breached with non-breached firms, aggregate spending should not decrease. Therefore, we predict the demand curve is not perfectly inelastic. To capture a complete picture, in the consumer spending analyses, we explore durable (e.g., automobiles) and non-durable consumption (e.g., grocery shopping) at the extensive margin (e.g., how many shopping trips consumers make) as well as aggregate household spending, examining substitution between payment methods.

A. Data

Our information on data breaches comes from Privacy Rights Clearinghouse (PRC) and is available from 2005. Under state law, firms are required to disclose any breaches to impacted households promptly. Breached organizations include the following: business organizations (e.g., financial and insurance or retail/merchant); educational institutions; government and military organizations; and healthcare, medical providers, nonprofits, and medical insurance services. We mainly use the breaches related to business organizations for our analyses. The PRC dataset is granular at the city (rich cross-sectional variation) and monthly (rich time-series variation) levels. Thus, this dataset is our primary data source for conducting empirical analyses. We also augment

---

4 See http://www.privacyrights.org/data-breach. For inclusive data information, please see https://www.privacyrights.org/data-breach-FAQ. Kamiya et al. (2021) also provide details on data verification. 

5 Around 29% of the observations do not include city information. Because our paper relies heavily on refined geographic information, we eliminate these observations. This missing information likely leads to an underestimation of our results, because these observations could have been in our treated group.
Audit Analytics breach data for our analysis at the individual level described below. Figure 1 shows the number of data breaches by year; total breaches and breaches with real records have similar trends over the years. Figure 2 shows the geographic distribution of data breaches by state (in Panel A) and by city (in Panel B); these breaches are mainly in cities and states with large populations, consistent with the geographic distribution of consumers and corporate headquarters.

Our consumer spending information comes from two sources: Nielsen Consumer Panel Datasets (non-durable consumption) and R.L. Polk (durable consumption). Nielsen Consumer Panel Datasets provide detailed information on household-level shopping behavior. Participating households continually provide information to Nielsen about their demographic information and about when and where they make purchases. Panelists are geographically dispersed and demographically balanced. Around half of the expenditures are in grocery stores, and a third are in discount or warehouse club stores. The remaining small part of expenditures are split among pet, liquor, and electronic stores (Neiman and Vavra 2019). Thus, most household purchasing can be viewed as non-durable consumption. To mitigate the potential reporting bias (e.g., selective reporting or attrition), Nielsen has implemented the following measures: (1) using sweepstakes and other discount sales to incentivize reporting and (2) ensuring households have at least a 12-month purchasing history threshold. This dataset has detailed information on consumers’ spending amounts and the number of trips. Since 2013, the Nielsen Scanner Data Trips file has collected information on the method of payment used during each trip. Thus, our analyses on the substitution between payment methods starts in 2013, whereas our analyses on spending amounts

6 Panelists use in-home scanners to record all purchases intended for personal, in-home use.
7 Each panelist is assigned a projection factor to ensure the sample is representative of the US. Neiman and Vavra (2019) show the results are similar with and without the projection factor.
8 The annual attrition rate of panelists is roughly 20%, and new households are regularly added to the sample to replace existing households (Neiman and Vavra 2019).
and the number of trips span from 2004 to 2016. The Nielsen dataset has been widely used in prior research to measure household spending; see, for example, Neiman and Vavra (2019), Leung and Seo (2018), and Bornstein (2018).

The second data source, R.L. Polk, has monthly ZIP-code information on the total quantity of new purchased automobiles. R.L. Polk data are collected from automobile registrations, so the ZIP code comes from the address of the household that purchased the new automobile between 2004 and 2016. Because this dataset has accurate information on the number of new cars, we use these data to examine the extensive margin. Mian, Rao, and Sufi (2013) and Mian and Sufi (2014) use the R.L. Polk dataset to study the consumption response to changes in home values between 2006 and 2009. They use the share of quantity purchased to allocate total census retail-sales expenditures on autos every year. They acknowledge that this approach introduces measurement error, because they do not have information on the change in prices across ZIP codes unless prices changed equivalently across all ZIP codes from the sample period. Because section 2.2. predicts the extensive-margin results, we use the accurate information (i.e., the total quantity of new autos) in the dataset as the dependent variable.

For geographic-level economic information, we use data from the County Business Patterns of the Census (CBP) and the IRS Statistics of Income (IRS SOI). Both datasets capture information that is easy to measure and that provides variation at a narrow geographic dimension and at a high frequency. We obtain data on media discussions of data breaches from RavenPack and local newspaper information from Gentzkow and Shapiro (2010), who collected 434 English-language daily newspapers from ProQuest and NewsLibrary databases (representing 74% of the total circulation of daily newspapers in the US in 2001).

To examine consumer responses at the individual level for our research, we leverage
SafeGraph data on device mobility, which has been extensively used in prior literature (e.g., Goolsbee and Syverson 2020, Liu and Lu 2021). SafeGraph allows us to observe daily data on consumer foot traffic to commercial sites of businesses. It collects data from almost 45 million cellular phones (about 10% of all devices in the US) and calculates the number of visits to millions of “points of interest” throughout the country. SafeGraph captures data by triangulating the precise GPS coordinates of a smartphone at various points during the day using pings against cell phone towers. The longitudinal and latitudinal data are sufficiently detailed that SafeGraph can determine the location of a device within a few meters and determine whether an individual visits a firm’s business location based on the detailed geolocation data. To conduct our study, we combine SafeGraph’s foot traffic data with our breached datasets by matching the breached organization names and the breach announcement dates. The SafeGraph data are at the establishment level, so our analysis includes consumer visits to the individual stores. The dataset spans the months of January 2018 and December 2019, since the earliest SafeGraph data are available in 2018. To increase the sample size for our individual analyses, we supplement Audit Analytics breach data with PRC data to obtain 2019 breach events of business organizations.

B. The Chilling Effect of Data Breaches on Consumer Spending

B.1. Household Participation in the Product Market

We employ tight empirical strategies and granular data to examine the negative economic consequences of data breaches on consumer spending using the R.L. Polk and Nielsen Consumer Panel Dataset. Our baseline regression, suppressing time and city subscripts, is

\[
\text{Consumer Spending} = \alpha_0 + \alpha_1 \text{Breach} + \sum \alpha_i \text{Fixed Effects} + \gamma_i \text{City Controls}_i + \epsilon.
\]

(1)

9 SafeGraph collects anonymised data from a variety of smartphone applications that make use of location services, such as weather, dating, or local news. These applications run on Apple and Android smartphones.
The dependent variable includes a number of consumer spending proxies, including $Std(Auto\ Purchases)$, $#\ of\ Trips$, $Cash\ Payment\ Ratio$, and $Std(Spending)$. Breach, the variable of interest, is an indicator coded as one when companies in a city experience data breaches in a certain year-month, and zero otherwise. All of the breach types identified in our analyses are business-related. Moreover, it is probable that we document the impact for cities that experience many breaches in the same year-month (or different months).

Because we are primarily interested in the aggregate effect of data breaches on consumer spending, the unit of analysis is at the city level. The use of macro-level outcomes (i.e., consumer spending at the city level) allows us to estimate the total effect of data breaches, including any spillover effects (i.e., changes in consumer spending at peer firms in the same city as breached companies). Additionally, because city information is explicitly disclosed (e.g., in the newspaper and/or in government websites), consumers in that city would presumably have the strongest reaction, thus increasing test power.

One major concern is whether our results are driven by the state-/county-level economic conditions that are associated with both the revelation of data breaches and with household spending. For example, if the revelation of data breaches occurs during an economic downturn, the downturn itself may independently drive households’ decisions to reduce their spending. To address this concern, our empirical specification exploits within-county heterogeneity across different cities in the same year and month. Another reason for analyzing the data at the monthly level is that R.L. Polk data are available monthly, and analyzing on a monthly basis is consistent across empirical specifications. We examine the association between data breaches and average household spending and measure the overall treatment effect using a simple assumption:
individuals living in the same county and city should be similarly affected by any economic factors that are associated with local economic conditions.

The detailed data are instrumental in controlling for many confounding economic shocks and for isolating the variation in data breaches. For example, we absorb any city-level shocks by including interactions of city×year fixed effects to control for overall trends in the long-period analysis and city×month fixed effects to control for seasonality. These fixed effects are helpful to control for macroeconomic conditions and concurrent local shocks. Effectively, these fixed effects compare two cities within the same year-month and county and then capture any unobservable differences based on location or economic cycle. This empirical strategy mitigates the concern that our findings are driven by a deterioration in county economic conditions, because changes in county conditions should affect all cities in the county similarly, independent of their data-breach experiences.

We report descriptive statistics in Table 1. We aggregate the data at the city-year-month level, and there is skewness in the data, possibly due to city size heterogeneity, which drives the differences between the mean and the median. For the data sampled in Nielsen, there are around the average of 87 households in a city that participate in the Nielsen panel, and these households spend about $290 each trip (on average) in a month. Approximately 20% of these trips are paid with cash during 2013–2016 (when we have available datasets). For the RL Polk data, consumers in a city purchase 44 cars on average.10 As our two dependent variables are from different datasets, for comparative purposes, we standardize the dependent variables, auto purchases and consumer grocery spending, to have a mean of zero and a standard deviation of one. This standardization allows the coefficients on the independent variables to be compared directly across different

---

10 The Nielsen and RL Polk datasets have different sampling densities in each city, so the comparison may not be direct.
analyses regarding the effect of data breaches on consumer behavior change.

Table 2 reports the household extensive-margin spending behavior (i.e., the number of trips). We find unambiguous evidence that consumers’ product market participation decreases following data breaches in the city where the household is located. Panel A of Table 2 represents non-durable consumption (i.e., grocery shopping). In columns (1) to (4) in Panel A of Table 2, we find households in the panel reduce their number of monthly shopping trips by two to three (roughly from 78 trips to 75 or 76 trips) when the city suffers data breaches. Panel B of Table 2 represents the durable consumption of automobiles. Columns (1) to (4) in Panel B of Table 2 show that in the same county, year, and month, a city with data breaches experiences a 5% to 6% reduction in the standardized value of auto purchases immediately afterwards. Because the outcome variables are standardized, each coefficient corresponds to a decrease of one standard deviation. To interpret the economic magnitude, this number can be translated into around 20 (=333×6%) fewer monthly auto purchases per city.

Throughout the analysis, we include time-varying city characteristics (city×year and city×month fixed effects, depending on the specification), which capture any unobserved time-varying variation at the city/county level and allow for different trends in each county/city. While there may be a correlation between data-breach temporal trends and household consumption, it would be difficult to reconcile this interpretation with the results obtained using county×year×month fixed effects. The coefficient estimates of data breaches are almost identical with and without the inclusion of city×month fixed effects. Under the relatively weak assumption that changes in the economy will have similar effects on households living in the same county, it is difficult to argue that the opposite reaction of no-data breaches in the same county are due to changes in actual economic conditions or other policy changes. Furthermore, our results hold after
we include economic control variables throughout the specifications.

To explicitly evaluate whether some confounding shocks at the city level (e.g., deteriorating local economic conditions) may lead to a reduction in household spending, we examine household income in data-breach cities. Specifically, we use Nielsen data (household-level self-reported income) and IRS SOI (government income data) to repeat the same empirical analyses, and change the dependent variable from household spending to household income. In column (7) in Panel A and Panel B of Table 2, we find household income does not change in cities that experience data breaches. Therefore, based on the theoretical model that increased income should increase household spending, our results are not driven by any deterioration in economic conditions.\(^{11}\)

We also examine whether the negative effect on consumer spending is permanent, and we find the effect is short term. In columns (5) and (6) of Table 2, we find the average effect lasts for only one month. More specifically, the effect reverses quickly for grocery shopping (with expenditures significantly higher than baseline a month after a breach) but not for automobile purchases, consistent with consumers’ demand elasticity for products.

We next examine whether households increase their cash payments when grocery shopping. Repeating the analyses in Table 2 Panel A with the dependent variable of households’ cash-payment ratio in grocery shopping, in Table 3, we document that households are more likely to pay using cash in cities that have data breaches. Columns (1) and (4) in Table 3 show households are 0.8% to 1.2% more likely to pay for groceries using cash.\(^{12}\) Columns (5) to (7) show the

\(^{11}\) A caveat is that realized income is not equal to income expectations, which could also affect household purchase decisions.

\(^{12}\) In this analysis, we find significant results for breaches with real records leaked, though city×year fixed effects make the results weaker. Because our sample period is short for this analysis, adding stringent fixed effects may lead to tests with low power. Another explanation is that because cash-payment data start in 2013, consumers may be more aware of the data breaches, making the salience of data breaches weaker in the later period.
increase in cash payment is also temporary. The evidence shows that within the same county and month, households increase their cash payments in cities that suffer from data breaches more than in cities without data breaches. This finding is consistent with the view that households fear the theft of their identity information and rely more on physical cash payments when data breaches are publicized. An alternative interpretation could be that households have to use cash because their information has been stolen and their credit cards are frozen. Because we find this effect lasts for three months, that interpretation is less plausible.

Because we find consumers reduce the number of shopping trips in Table 2 and consumers are more likely to use cash after the breach in Table 3, both of which should contribute to a decrease in spending, we then examine how aggregate household spending changes. In Table 4, we find that within the same county, year, and month, a city with data breaches experiences a 1% to 2% reduction in the standardized value of total consumer spending. The economic magnitude of this coefficient can be interpreted as a $2000 (=99583×2%) reduction in monthly spending per city or roughly $25 dollars per Nielsen panel household.

Another related concern regarding the city-level observation is that, within the same county, if consumer spending is affected due to data breaches, it is likely that an adjacent city might be affected in either a positive or negative way. For example, consumers may travel to an adjacent city to purchase products, and the adjacent city is in our control group, so our documented magnitude might be overestimated. But if consumers’ chilling effect spills over to the adjacent city, our reported magnitude may be under-estimated. To empirically assess the possibility, in an untabulated analysis, we repeat our analysis by removing the adjacent city and find that our results are similar. Specifically, in the Nielsen spending analysis, our results are stronger after removing the adjacent city. This result indicates that the consumers seem to travel to the adjacent city to
shop, which is consistent with small demand elasticity. In the auto-purchase analysis, our results are weaker and qualitatively similar after removing the adjacent city. This result suggests that the consumers may cut their spending directly by not going to the surrounding area to shop, and is consistent with a larger demand elasticity for durable goods.

We further map out how the consumer spending changes for the grocery shopping and auto purchases in order to assess the parallel trends assumption and depict the dynamic effects in Figure 3. The parallel trends assumption holds. Consistent with low demand elasticity, the effect on grocery shopping is short-lived and soon reverses, with subsequent customer expenditure exceeding pre-breach levels to compensate for missed consumption. For automotive purchases, the effect is equally temporary and quickly returns to its pre-breach level, but there is no catch-up spending, which is consistent with a greater demand elasticity for these products.

This aggregate reduction within a city also suggests the presence of a spillover effect from breached to other firms; otherwise, the reduction from breached firms could be substituted for other, non-breached firms. While the city-level analysis provides aggregate impact at the regional level, to further understand the economic effect of data breaches on consumer spending, we also explore the variation in media coverage through which the breach information dissemination is broader and conduct individual analysis.

B.2. Cross-sectional Analyses on Media Coverage

The media is instrumental in informing citizens about government policy and firm misconduct (Strömberg and Snyder 2010, Dyck et al. 2008). A local newspaper can disseminate information and enhance consumers’ awareness of data breaches. The spread of such information further damages consumers’ trust in firms’ ability to protect their personal information. Thus, we exploit pre-existing variation in local newspapers and examine the impact of the local
dissemination of information about data breaches on subsequent monthly consumption decisions, relative to those in a control group. In column (1) of Panel A, Panel B, and Panel C of Table 5, we find the chilling effect is larger in cities with pre-existing local newspapers, consistent with information dissemination being an important channel for consumer awareness.

One concern is the extent to which nationally known data breaches could bias our estimates. Because the branches of those nationally known firms are negatively affected but are not included in the treatment group, the local variations we use make the estimates more conservative. In this regard, our estimates could be considered a lower bound of the true effect. To empirically examine this possibility, we use the media discussion data in RavenPack and conduct two analyses: (1) For the subsample of Not Visible Firms, we exclude nationally known data breaches, which are defined as breaches above the 75th percentile in terms of breach media discussions; and (2) for the subsample of Not Intensive National Coverage, we exclude months in which media mentions of data breaches are above the 75th percentile of all media discussions about data breaches across time. In columns (2) to (5) of Panel A, Panel B, and Panel C of Table 5, we repeat the analyses and find a larger magnitude of results in the grocery shopping analyses, and we find a slight increase in the magnitude of the results for automobile spending. For instance, in Panel A of Table 5, Panel A, we find that the trip decline is larger in those communities that have local newspapers.

If certain consumer spending is not discretionary, consumers may switch to close competitors of breached firms, despite the fact that these competitors may not have been the consumers’ first-choice alternatives. In Table 6, we observe that the variability of retailers’ customer lists and market shares increases in breached cities with local media coverage, indicating that consumers switch their shopping to non-breached businesses. The underlying assumption is
that if the retailer-consumer relationship is stable, the variability of a particular retailer’s customer list and market share (after controlling for customers’ outside options and adjusting for seasonality) should not change. However, consumers’ abnormally high rate of retailer churn may result in greater variation in a retailer’s client list and market share. Our results provides suggestive evidence on consumers’ spending substitution by switching to competitors.

Of course, media reports are hardly the only way that consumers learn about data breaches. As a result of data breach notification laws in all fifty U.S. states, consumers regularly receive letters and emails alerting them to breaches of their own data. We are unable to obtain data on data breach notifications at the household level so we cannot examine how receiving such a notification affects consumer behavior. Nor do we know which individual employees have been notified by their employers that their data were breached. Yet our findings with respect to localized breaches and the effects of local media in disseminating information about those breaches provide inferential support for the proposition that people whose data are breached and who are subsequently notified would experience short-lived but significant changes in their consumption patterns. This would be a kind of particularized harm, even in the absence of identity theft or the revelation of embarrassing consumer information. One can interpret the local media findings to increase awareness and salience of a breach among all kinds of consumers. Those whose data were breached may have discarded a data breach notification unread but could see media coverage. Those whose data were not breached could experience heightened anxiety after learning of the local breach through the media or word-of-mouth responses to local media reports. Employees at local firms where a breach happened would learn about the breaches through their employers or media reports and could experience similar anxiety.
B.3 Individual Analysis (Average and Quantile Results)

We use a short-window event study to examine whether consumer store visits change after a breach. We matched the breach event with the SafeGraph dataset by the breached business organization name and the event announcement date so that the individual analysis is at the business store-day level. We compare the consumer visit changes in the pre-period (three days before the breach announcement date), during the breach-announcement period (three days right after the announcement date), and post-period (day four until day six after the announcement). We include day-of-week fixed effects to account for the variation in activity that is associated with a given day of the week and that captures the variation between weekend and weekday visits. We include ZIP-code fixed effects to control for time-invariant characteristics that may affect foot traffic, for example, population size. We control for day fixed effects to account for any time-varying macroeconomic factors across different geographic areas. In all specifications, we cluster standard errors by ZIP code to allow for correlation across different firms and time periods within a given ZIP code.

For our average results at the individual level, we find evidence consistent with a decrease in the consumer visits to breach incidents in 2018 and 2019, which is when SafeGraph data are available. In Table 7, we show that in all specifications, foot traffic significantly decreases following firms’ data breaches. We present the average and quantile regression results for the (-7, +7) and (-10, +10) windows in Table 7 Panel A and Panel B, respectively. Column (1) reports the average results, and columns (2) through (9) report the quantile regression results for the 20th to the 90th percentiles. We do not report quan10 (the 10th percentile) results, due to the limited sample size and power. For the (-7, +7) window, the benchmark is (-7, -4) days, and for the (-10, +10) window, the benchmarks are (-10, -4) days and (8, 10) days.
On average, we see a significant decrease in consumer visits after the breach. More specifically, compared with the 7 or 10 days before a data breach, consumers decrease their visits to stores operated by the breached firm in the 8 or 11 days following the breach. The coefficient in Post (4,7) is -0.27 or -0.18 for the average results, depending on the event window, which suggests daily store visits decrease by approximately 27% or 18%.

Next, we examine the prevalence of the consumer behavioral change. That is, we study whether our results are driven by many consumers changing their behavior a little or by a few consumers changing their behavior a lot. For this reason, we conduct a quantile regression to understand consumer responses across different distributions; the results are presented in Table 7, Panels A and B, columns (2) through (9). The results from the quantile regressions are very similar to the average results. We find a significant decrease in consumer visits in the post-breach period across different quantile distributions at the breached firm’s places of business. Specifically, for the median consumer response in the quantile regression, the coefficient in Post (4,7) is -0.32 or -0.20, depending on the event window, which suggests median daily store visits decrease by approximately 32% or 20%.

B.4 Application to Article III Standing

The findings in our analysis, such as a 1% reduction in credit card usage, a 2% reduction in consumer spending at stores, and a 5% reduction in new automobile purchases, at first blush, may not seem especially meaningful, even though they are statistically significant. To contextualize the magnitudes of the effects, though, it is important to understand that the majority of the consumers in the municipalities where these breaches occurred were untreated by the breach. That is, they lived in communities where a breach occurred, but their own data were not compromised, they did not receive an individual data breach notification letter or a notice from
their employer, and they probably missed any local news coverage of the breach. The fact that we nonetheless observe significant changes in aggregate local consumer behavior causes us to believe that the behavioral change among consumers and employees who were notified that their data were breached would be more dramatic. This interpretation of our results is buttressed by the quite substantial decline in foot traffic at the places of business that experienced a breach following the breach notifications’ dissemination.

The results of our analysis suggest consumers whose data are exposed in a data breach may suffer the sort of injury in fact that gives rise to Article III standing, even after Supreme Court decisions such as Clapper and Ramirez that interpreted the Constitution’s standing requirement to present a substantial hurdle. Whereas existing litigation focuses on identity theft and embarrassment, our new theory of harm relies on chilled consumption and the related loss of benefits. Consumers appear to substitute from credit card purchases to cash purchases and to decrease their consumption during the period following a data breach in their community. This effect appears to be driven by a large number of customers or employees of a breached firm changing their purchasing behavior slightly rather than by a small number of consumers or employees sharply curtailing credit card purchases. The Safegraph data, showing broad-based avoidance of breached firms’ places of business immediately after a breach, increases our confidence that the aggregate effects we identify are not being driven by a small number of consumers changing their behavior a great deal. These effects are larger when local media enables the efficient dissemination of information about a localized data breach.

To date, no federal appellate court has considered whether prompting consumers to reduce their spending or shift it from credit to currency constitutes a concrete and particularized injury in fact. But the best reading of the cases suggests the behavior we identify is widespread and
consequential enough to satisfy the existing test. In economic terms, the shift to cash purchases reflects a sacrifice of a zero-interest loan from a credit card issuer between the date of the purchase and the due date of the subsequent credit card bill. For the large number of consumers with “rewards cards,” shifting to cash also constitutes a loss of monetary benefits. Since it is well established that even relatively small economic injuries of this sort constitute injuries in fact, these harms do seem to satisfy the standing threshold in current doctrine.

To be clear, combining our data with our doctrinal analysis, we reject the idea that consumers and employees whose data were not breached have Article III standing. While our data indicate that there are plausibly local spillover effects from data breaches that cause people whose data weren’t breached to change their consumption practices, these behavioral effects are better characterized under current standing doctrine as generalized community harms, not particularized harms to individuals. Similarly, other doctrinal hurdles would prevent car dealerships or grocery stores from suing a company whose data were breached, resulting in a short-term loss of sales. Recall that an injury in fact must be fairly traceable to the defendant’s conduct for a plaintiff to have Article III standing. As a statistical matter we can be confident that a city will lose a few car sales on the margins after a data breach, but we can’t determine which car dealers in the city were affected and which car-purchasers were on the margins. This analysis suggests, however, that state or local governments that saw a dip in sales tax collections at car dealerships after a breach might be in a better position to sue the local firms whose data were breached.

The final fly in the doctrinal ointment arises as a result of Clapper. In that case the Supreme Court suggested that prudent measures to avoid government surveillance, such as substituting away from phone calls and towards air travel to communicate confidentially, did not constitute an injury in fact. But the Clapper Court’s analysis turned on the speculative nature of the
government’s surveillance and the plaintiff’s claims that it might be unlawful. Where a data breach results from external hackers or an inside job, there is little question that consumers whose data become available become attractive targets for mischief. The relatively widespread nature of the response we document here suggests that consumers and employees are harmed even if they aren’t victimized by identity theft.

Deferred gratification is also a plausible injury in fact. Market economies regularly put a price tag on permitting a consumer to enjoy a desired good or service as rapidly as possible. Delay has an opportunity cost. Consider expedited shipping fees, line-skipping fees at theme parks (e.g., Disney theme parks’ Lightning Lanes), or premiums for air versus bus travel. The interesting legal question, though, is whether consumers’ decisions to postpone desired purchases out of an abundance of caution after a breach is of the same character. To draw an analogy, if a highly publicized plane crash causes consumers to take Greyhound rather than United Airlines to reach their next vacation destination, has the consumer been harmed? Or is it more appropriate to say there was a shift in preferences that are now being satisfied via a substitute provider of travel? We think that query presents a close question, though it is plausible under MPSA that the extra delay and inconvenience associated with the service that is no longer deemed as safe as it used to be constitutes the kind of injury that the Supreme Court expects to see for someone seeking their day in federal court.

More broadly, our findings shed light on questions relevant to the administrative enforcement of data security laws. In the United States consumer class actions are the primary means of using the legal system to deter breaches, with administrative enforcement by the Federal Trade Commission and state attorneys general playing a secondary role. In other developed countries, administrative enforcement plays a leading role. To the extent that a clear-eyed
assessment of the costs associated with data breaches informs that administrative enforcement, our results suggest that the commitment of additional resources towards enforcement may be warranted.

IV. Conclusion

This paper documents the short-term chilling effect of data breaches on household product market participation in the regional economy. Our paper considers whether and how data breaches (direct negative news to consumers) are associated with household spending at an aggregate level and with the persistence of the effects.

This paper also has important legal and policy implications. Two of the widespread consumer responses to local data breaches – shifting toward cash transactions and deferring purchases – are the sort of injuries in fact that may give rise to standing to sue in federal court. Data breaches have a marked effect on consumption patterns, with the changes most plausibly concentrated among the people who have been best informed of a recent breach (i.e., those whose data have been compromised and have received notification). As a result, existing legal controversies over whether data breaches that do not result in identity theft or embarrassment can be resolved in a relatively straightforward manner. Our empirical results then suggest that the legal system may have erroneously mistaken an easy doctrinal question for a difficult one.
References


Cox, C. M. (2016). Risky Standing: Deciding on Injury. 8, 75-134.


Electronic copy available at: https://ssrn.com/abstract=4187930


Electronic copy available at: https://ssrn.com/abstract=4187930
Figure 1: Total Data Breach by Year

Notes: This figure presents the number of total data breaches and breaches with actual records leaked by year. Data breaches with actual records refer to the breaches in which personal data are leaked. Total data breaches also include cases in which no personal data are leaked, but the breach incident happened (e.g., the system was hacked, but there were no actual data leaked).

Electronic copy available at: https://ssrn.com/abstract=4187930
Figure 2 Total Breaches by Geography

Panel A: Distribution by State

Panel B: Distribution by City

Notes: This figure presents the geographic distributions of total data breaches by state (Panel A) and by city (Panel B).
Figure 3: A Visual Representation of Post-Breach Consumption

Panel A: Total Spending

Panel B: Auto Purchases

Notes: This figure plots the mapping out of the breach event. The event-year 0 is defined as being the month when data breach occurs. Panel A reports the total spending and Panel B reports the auto purchases.
Table 1: Descriptive Statistics for Consumer Spending

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Median</th>
<th>P10</th>
<th>P90</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nielsen Spending</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spending Total ($)</td>
<td>1,509,224</td>
<td>23,272.143</td>
<td>99,583.232</td>
<td>2,357.590</td>
<td>319.610</td>
<td>49,682.801</td>
</tr>
<tr>
<td>Household Size</td>
<td>1,509,224</td>
<td>86.815</td>
<td>367.289</td>
<td>10.000</td>
<td>2.000</td>
<td>184.000</td>
</tr>
<tr>
<td>Income Total ($)</td>
<td>1,509,224</td>
<td>744.370</td>
<td>3,324.936</td>
<td>78.000</td>
<td>15.000</td>
<td>1,565.000</td>
</tr>
<tr>
<td># of Trips Total</td>
<td>1,509,224</td>
<td>78.053</td>
<td>224.947</td>
<td>30.000</td>
<td>7.000</td>
<td>159.000</td>
</tr>
<tr>
<td>Ratio of Cash Payment (2013 - 2016)</td>
<td>486,076</td>
<td>0.202</td>
<td>0.251</td>
<td>0.121</td>
<td>0.000</td>
<td>0.562</td>
</tr>
<tr>
<td><strong>Auto Purchase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of Auto Sales</td>
<td>4,216,626</td>
<td>44.475</td>
<td>333.723</td>
<td>6.000</td>
<td>0.000</td>
<td>84.000</td>
</tr>
<tr>
<td>Taxable Income Growth</td>
<td>4,216,626</td>
<td>0.002</td>
<td>0.712</td>
<td>-0.081</td>
<td>-0.343</td>
<td>0.393</td>
</tr>
</tbody>
</table>

Notes: This table presents summary statistics for the city-year-month-level datasets used in the analysis. For the Nielsen spending data, the income data are based on the resident and the spending considers the entire family. The sample period is from 2004 to 2016 (except the ratio of cash payment sample is from 2013 to 2016).
Table 2: Effect of Data Breaches on Shopping Trips (Extensive Margin)

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Nielsen Spending Trips</th>
<th>Panel B: Auto Purchase Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Extensive Margin: # of Trips</td>
<td>Std(Auto Purchases)</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Breach</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.256***</td>
<td>-.63***</td>
</tr>
<tr>
<td></td>
<td>(-2.92)</td>
<td>(3.05)</td>
</tr>
<tr>
<td>Fixed Effects:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>State×Year×Month</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>County×Year×Month</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>City×Year</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>City×Month</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>1,509,224</td>
<td>1,509,224</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>Cluster</td>
<td>State</td>
<td>State</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breach</td>
<td>-.06***</td>
<td>-.06***</td>
</tr>
<tr>
<td></td>
<td>(-2.95)</td>
<td>(-2.92)</td>
</tr>
<tr>
<td>Fixed Effects:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>State×Year×Month</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>County×Year×Month</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>City×Year</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>City×Month</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>4,216,626</td>
<td>4,216,626</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.961</td>
<td>0.961</td>
</tr>
<tr>
<td>Cluster</td>
<td>State</td>
<td>State</td>
</tr>
</tbody>
</table>

Notes: This table reports the results on the effect of data breaches on shopping trips (extensive margin). Panel A presents results using Nielsen data, and Panel B presents results using data from R.L. Polk. The variable of interest, Breach, is an indicator variable coded as one for cities that have data breaches in that year-month, and zero otherwise. Dependent variables in Panel A include Extensive Margin: # of Trips, F(t+1), F(t+2), and Placebo: Income. Extensive Margin: # of Trips is defined as the number of shopping trips at a city-year-month level. F(t+1)(F(t+2)) are defined as the number of shopping trips after one month (two months). Placebo: Income is household income at the city-year-month level. Dependent variables in Panel B include Std(Auto Purchases) and Placebo: Income. Std(Auto Purchases) is defined as the standardized value of household auto purchases at a city-year-month level. F(t+1)(F(t+2)) are defined as the standardized value of household auto purchases after one month (two months). Placebo: Income household taxable income growth at the city-year level. We cluster standard errors by state and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.

Electronic copy available at: https://ssrn.com/abstract=4187930
Table 3: Effect of Data Breaches on Cash Payment

<table>
<thead>
<tr>
<th>Dependent Variable: Cash Payment Ratio</th>
<th>Current (t)</th>
<th>F(t+1)</th>
<th>F(t+2)</th>
<th>F(t+3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breach</td>
<td>0.008**</td>
<td>0.010***</td>
<td>0.012***</td>
<td>0.011*</td>
</tr>
<tr>
<td></td>
<td>(2.34)</td>
<td>(3.18)</td>
<td>(3.04)</td>
<td>(2.75)</td>
</tr>
</tbody>
</table>

**Fixed Effects:**

- State: No Yes No No No No No
- State×Year: Yes No No No No No No
- County: Yes No No No No No No
- County×Year: No Yes No No No No No
- County×Year×Month: No No Yes Yes Yes Yes Yes
- City×Month: Yes Yes No Yes Yes Yes Yes
- City: No No Yes No No No No

Observations: 486,076 486,076 486,076 486,076 470,167 457,272 444,450
Adjusted R-squared: 0.642 0.700 0.705 0.621 0.622 0.623 0.624

Notes: This table reports the results on the effect of data breaches on consumer cash payment using the data from Nielsen. The variable of interest, Breach, is an indicator variable coded as one for cities that have data breaches in that year-month, and zero otherwise. The dependent variable, Cash Payment Ratio, is the ratio of the number of cash payments to the total number of payments. Current (t) are defined as the cash-payment ratio in the month t. F(t+1), F(t+2), F(t+3) are defined as the cash payment ratio one month (two months, three months) after month t. We cluster standard errors by state and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.
Table 4: Effect of Data Breaches on Consumer Spending

<table>
<thead>
<tr>
<th>Breach</th>
<th>$(t+1)$</th>
<th>$(t+2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>-0.02***</td>
<td>-0.02***</td>
<td>-0.01*</td>
</tr>
<tr>
<td>(-3.87)</td>
<td>(-3.87)</td>
<td>(-1.98)</td>
</tr>
</tbody>
</table>

**Fixed Effects:**
- State: Yes
- State×Year: No
- State×Year×Month: No
- County×Year: Yes
- County×Year×Month: Yes
- City×Year: Yes
- City×Month: Yes
- Observations: 1,509,224
- Adjusted R-squared: 0.994

Notes: This table reports the results on the effect of data breaches on consumer spending using Nielsen data. The variable of interest, Breach, is an indicator variable coded as 1 for cities that have data breaches in that year-month, and 0 otherwise. The dependent variable $Std(Spending)$ is defined as the standardized value of household spending at a city-year-month level. $F(t+1)(F(t+2))$ is defined as the standardized value of household spending after one month (two months). We cluster standard errors by state and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.
Table 5: Cross-sectional Variation in Media Coverage

<table>
<thead>
<tr>
<th>Panel A: Nielsen Spending Trips</th>
<th>Total Sample</th>
<th>Not Visible Firms</th>
<th>Not Intensive National Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable: % of Trips</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breach with Local Media Coverage</td>
<td>-3.82**</td>
<td>-5.71***</td>
<td>-5.84***</td>
</tr>
<tr>
<td></td>
<td>(-2.36)</td>
<td>(-3.21)</td>
<td>(-3.32)</td>
</tr>
<tr>
<td>Breach</td>
<td>-0.25</td>
<td>-3.23***</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(-0.76)</td>
<td>(-3.10)</td>
<td>(0.16)</td>
</tr>
</tbody>
</table>

**Fixed Effects:**
- County x Year x Month: Yes
- City x Year: Yes
- City x Month: Yes

| Observations | 1,509,224 | 1,084,076 | 1,084,076 | 1,105,062 | 1,105,062 |
| Adjusted R-squared | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 |
| Cluster | State | State | State | State | State |

<table>
<thead>
<tr>
<th>Panel B: Auto Purchase Trips</th>
<th>Total Sample</th>
<th>Not Visible Firms</th>
<th>Not Intensive National Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable: Std(Auto Purchases)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breach with Local Media Coverage</td>
<td>-0.09**</td>
<td>-0.09**</td>
<td>-0.10**</td>
</tr>
<tr>
<td></td>
<td>(-2.37)</td>
<td>(-2.34)</td>
<td>(-2.48)</td>
</tr>
<tr>
<td>Breach</td>
<td>-0.00</td>
<td>-0.05***</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(-0.24)</td>
<td>(-2.01)</td>
<td>(0.37)</td>
</tr>
</tbody>
</table>

**Fixed Effects:**
- County x Year x Month: Yes
- City x Year: Yes
- City x Month: Yes

| Observations | 4,216,626 | 3,129,816 | 3,129,816 | 3,130,003 | 3,130,003 |
| Adjusted R-squared | 0.954 | 0.954 | 0.955 | 0.955 | 0.954 |
| Cluster | State | State | State | State | State |

<table>
<thead>
<tr>
<th>Panel C: Nielsen Spending</th>
<th>Total Sample</th>
<th>Not Visible Firms</th>
<th>Not Intensive National Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable: Std(Total Spending)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breach with Local Media Coverage</td>
<td>-0.02**</td>
<td>-0.03***</td>
<td>-0.04**</td>
</tr>
<tr>
<td></td>
<td>(-2.27)</td>
<td>(-2.72)</td>
<td>(-2.54)</td>
</tr>
<tr>
<td>Breach</td>
<td>-0.00</td>
<td>-0.02**</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(-0.95)</td>
<td>(-2.25)</td>
<td>(0.90)</td>
</tr>
</tbody>
</table>

**Fixed Effects:**
- County x Year x Month: Yes
- City x Year: Yes
- City x Month: Yes

| Observations | 1,509,224 | 1,084,076 | 1,084,076 | 1,105,062 | 1,105,062 |
| Adjusted R-squared | 0.993 | 0.993 | 0.993 | 0.993 | 0.993 |
| Cluster | State | State | State | State | State |
Table 5 (Continued)

Notes: This table reports the results on the effect of data breaches on consumer spending. Panel A and Panel C present results using Nielsen data, and Panel B presents results using data from R.L. Polk. The variable of interest, Breach, is an indicator variable coded as one for cities that have data breaches in that year-month, and zero otherwise. Breach with Local Media Coverage is an indicator variable coded as one for cities with pre-existing local newspapers that have data breaches in that year-month, and zero otherwise. # of Trips is defined as the number of shopping trips at a city-year-month level. Std(Auto Purchases) is defined as the standardized value of household auto purchases at a city-year-month level. Std(Spending) is defined as the standardized value of household spending at a city-year-month level. Column (1) is for the total sample. Columns (2) and (3) are for the subsample of Not Visible Firms: we exclude nationally known data breaches, which are defined as breaches above the 75th percentile in terms of breach media discussions. Columns (4) and (5) are for the subsample of Not Intensive National Coverage: We exclude months in which media mentions of data breaches are above the 75th percentile of all media discussions about data breaches across time. We cluster standard errors by state and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.
Table 6: Suggestive Evidence on Consumers’ Switching-to-Competitors Behavior

<table>
<thead>
<tr>
<th></th>
<th>Stddev(Retailers’ # of Consumers)</th>
<th>Stddev(Retailers’ Market Share)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Breach (Retail) with Local Coverage</td>
<td>0.179</td>
<td>0.500*</td>
</tr>
<tr>
<td></td>
<td>(0.99)</td>
<td>(1.93)</td>
</tr>
<tr>
<td>Breach (Retail)</td>
<td>0.020</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.93)</td>
<td>(1.42)</td>
</tr>
</tbody>
</table>

**Fixed Effects:**
- County × Year × Month: Yes Yes Yes Yes
- City × Year: Yes No Yes No
- City: No Yes No Yes
- Consumers’ Outside Option within City-Year-Month: Yes Yes Yes Yes

Observations: 1,028,319 1,028,319 1,028,319 1,028,319
Adjusted R-squared: 0.984 0.959 0.707 0.429
Cluster: State State State State

Notes: This table reports the suggestive evidence on consumers’ switching-to-competitors behavior (proxied by the variability of retailers’ consumer base). Stddev(Retailers’ # of Consumers) is defined as the standard deviation of retailers’ number of customers at the city-year-month level. Stddev(Retailers’ Market Share) is defined as the standard deviation of retailers’ market share at the city-year-month level. Retailers’ market share is defined as a specific retailer’s number of consumers divided by the total number of consumers in a given city, year, and month. Consumers’ Outside Option within City-Year-Month is defined as the total number of retailers in a given city, year, and month. Breach (Retail) is an indicator coded as one if a retailer firm has a data breach, and zero otherwise. We cluster standard errors by state and report t-statistics in parentheses. Breach (Retail) with Local Media Coverage is an indicator variable coded as one for cities with pre-existing local newspapers that have retail firm data breaches in that year-month, and zero otherwise. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.
Table 7: Individual Analysis (Average and Quantile Results)

Panel A: SafeGraph (-7, +7) window

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log (1+Visit)</td>
<td>Average</td>
<td>Quan20</td>
<td>Quan30</td>
<td>Quan40</td>
<td>Quan50</td>
<td>Quan60</td>
<td>Quan70</td>
<td>Quan80</td>
<td>Quan90</td>
</tr>
<tr>
<td>Pre(-4,0)</td>
<td>-0.01</td>
<td>0.04**</td>
<td>0.13***</td>
<td>0.13***</td>
<td>0.04**</td>
<td>-0.03**</td>
<td>-0.07***</td>
<td>-0.12***</td>
<td>-0.20***</td>
</tr>
<tr>
<td></td>
<td>(-0.68)</td>
<td>(2.53)</td>
<td>(5.22)</td>
<td>(4.47)</td>
<td>(2.05)</td>
<td>(2.14)</td>
<td>(5.17)</td>
<td>(9.13)</td>
<td>(12.69)</td>
</tr>
<tr>
<td>(1,3)</td>
<td>0.12***</td>
<td>-0.06**</td>
<td>-0.09**</td>
<td>-0.20***</td>
<td>-0.17***</td>
<td>-0.15**</td>
<td>-0.13***</td>
<td>-0.13**</td>
<td>-0.11***</td>
</tr>
<tr>
<td></td>
<td>(-4.94)</td>
<td>(-2.77)</td>
<td>(-2.55)</td>
<td>(-4.26)</td>
<td>(-4.33)</td>
<td>(-5.25)</td>
<td>(-5.28)</td>
<td>(-5.45)</td>
<td>(-4.24)</td>
</tr>
<tr>
<td>Post(4,7)</td>
<td>-0.27***</td>
<td>-0.14***</td>
<td>-0.23***</td>
<td>-0.37***</td>
<td>-0.32***</td>
<td>-0.28***</td>
<td>-0.26***</td>
<td>-0.26***</td>
<td>-0.31***</td>
</tr>
<tr>
<td></td>
<td>(-7.56)</td>
<td>(-4.97)</td>
<td>(-4.54)</td>
<td>(-5.69)</td>
<td>(-6.22)</td>
<td>(-6.71)</td>
<td>(-6.94)</td>
<td>(-7.36)</td>
<td>(-7.75)</td>
</tr>
</tbody>
</table>

Observations 243,581
Adjusted R-squared 0.668

Panel B: SafeGraph (-10, +10) window

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log (1+Visit)</td>
<td>Average</td>
<td>Quan20</td>
<td>Quan30</td>
<td>Quan40</td>
<td>Quan50</td>
<td>Quan60</td>
<td>Quan70</td>
<td>Quan80</td>
<td>Quan90</td>
</tr>
<tr>
<td>Pre(-4,0)</td>
<td>-0.01</td>
<td>0.04***</td>
<td>0.10***</td>
<td>0.10***</td>
<td>0.00</td>
<td>-0.05***</td>
<td>-0.07***</td>
<td>-0.11***</td>
<td>-0.15***</td>
</tr>
<tr>
<td></td>
<td>(-1.62)</td>
<td>(4.32)</td>
<td>(6.95)</td>
<td>(5.59)</td>
<td>(0.08)</td>
<td>(4.84)</td>
<td>(8.18)</td>
<td>(13.29)</td>
<td>(14.67)</td>
</tr>
<tr>
<td>(1,3)</td>
<td>-0.02**</td>
<td>-0.02</td>
<td>-0.04**</td>
<td>-0.03**</td>
<td>-0.03**</td>
<td>-0.03**</td>
<td>-0.03**</td>
<td>-0.03**</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(-3.19)</td>
<td>(-1.81)</td>
<td>(-1.57)</td>
<td>(-2.10)</td>
<td>(-2.11)</td>
<td>(-2.97)</td>
<td>(-3.41)</td>
<td>(-4.06)</td>
<td>(-1.01)</td>
</tr>
<tr>
<td>Post(4,7)</td>
<td>-0.18***</td>
<td>-0.08***</td>
<td>-0.14***</td>
<td>-0.22***</td>
<td>-0.20***</td>
<td>-0.19***</td>
<td>-0.18***</td>
<td>-0.21***</td>
<td>-0.22***</td>
</tr>
<tr>
<td></td>
<td>(-17.86)</td>
<td>(-7.95)</td>
<td>(-7.95)</td>
<td>(-9.94)</td>
<td>(-12.25)</td>
<td>(-15.25)</td>
<td>(-15.89)</td>
<td>(-17.99)</td>
<td>(-15.90)</td>
</tr>
</tbody>
</table>

Observations 358,918
Adjusted R-squared 0.670

Notes: This table reports the average and quantile regression results of the individual analysis using SafeGraph data. Panel A reports the (-7, +7) window and Panel B presents the (-10, +10) window. Log (1+Visit) is defined as the natural log of 1 plus individual visits. Pre(-4,0) is -4 to -1 days before the breach announcement. (1,3) is the breach announcement date through three days after the breach announcement. Post(4,7) is 4 to 7 days after the breach announcement. Column (1) reports the average results, and columns (2) through (9) report the quantile regression results for the 20th to 90th percentiles. We do not report quantile results due to the limited sample size and variation. We cluster standard errors by ZIP code and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.
### Table 8: Results for Local Breaches and Employee-only Breaches

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Extensive Margin: # of Trips</th>
<th>Std(Auto Purchases)</th>
<th>Cash Payment Ratio</th>
<th>Std(Total Spending)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Breach</td>
<td>-2.1***</td>
<td>-2.25*</td>
<td>-0.10***</td>
<td>0.010***</td>
</tr>
<tr>
<td></td>
<td>(-1.67)</td>
<td>(-1.74)</td>
<td>(-3.32)</td>
<td>(3.68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-2.75)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-2.71)</td>
</tr>
</tbody>
</table>

**Fixed Effects:**

<table>
<thead>
<tr>
<th>State</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>No</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>State×Year</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>State×Year×Month</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>County</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>County×Year</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>City×Year</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>City×Month</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Notes:** This table reports the results for local breaches and employee-only breaches. The variable of interest, *Breach*, is an indicator variable coded as 1 for cities that have data breaches in that year-month, and 0 otherwise. *Margin: # of Trips* is defined as the number of shopping trips at a city-year-month level. *Std(Auto Purchases)* is defined as the standardized value of household auto purchases at a city-year-month level. *Cash Payment Ratio*, is the ratio of the number of cash payments to the total number of payments. *Std(Spending)* is defined as the standardized value of household spending at a city-year-month level. *cluster* standard errors by state and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.

Observations: 1,628,817, 1,628,661, 4,337,344, 4,337,326, 522,299, 522,169, 1,628,817, 1,628,661

Adjusted R-squared: 0.996, 0.996, 0.961, 0.961, 0.642, 0.700, 0.994, 0.994

Clustered: State, State, State, State, State, State, State, State

Electronic copy available at: https://ssrn.com/abstract=4187930