DO DATA BREACH NOTIFICATION LAWS WORK?

*Aniket Kesari**

Over 2.8 million Americans have reported being victims of identity theft in recent years, costing the U.S. economy at least thirteen billion dollars in 2020. In response to this growing problem, all fifty states have enacted some form of data breach notification law in the past twenty years. Despite their prevalence, evaluating the efficacy of these laws remains elusive. This Article fills this gap, while further creating a new taxonomy to understand when these laws work and when they do not.

Legal scholars have generally treated data breach notification laws as doing just one thing—disclosing information to consumers. But this approach ignores rich variation: differences in disclosure requirements to regulators and credit monitoring agencies; varied mechanisms for public and private enforcement; and a range of thresholds that define how firms should assess the likelihood that a data breach will ultimately harm consumers.

This Article leverages the Federal Trade Commission's Consumer Sentinel database to build a comprehensive dataset measuring identity theft report rates since 2000. Using staggered adoption synthetic control—a popular method for policy evaluation that has yet to be widely applied in empirical legal studies—this Article finds that whether identity theft laws work depends on which of these different strands of legal provisions are employed. In particular, while baseline disclosure requirements and private rights of action have small effects, requiring firms to notify state regulators reduces identity theft report rates by approximately 10%. And surprisingly, laws that fail to exclude low-risk breaches from reporting requirements are counterproductive, increasing identity theft report rates by 5%.

[*] Associate Professor, Fordham Law School. I thank Sabrina Arias, Elliott Ash, Stefan Bechtold, Elettra Bietti, Siv Cheruvu, Bill Eskridge, Kat Geddes, Brad Greenwood, Amit Haim, James Hicks, Chris Hoofnagle, Jiaying Jiang, Christine Jolls, Sonia Katyal, Jae Yeon Kim, Filippo Lancieri, Lawrence Liu, Tejas Narechania, Aileen Neilsen, Ira Rubinstein, Chris Sprigman, Gurpal Sran, David Stein, Katherine Strandburg, Paul Vaaler, Andrew Verstein, Mark Verstraete, Stav Zeitouni, and participants at NYU's Information Law Institute, UC Berkeley's Center for Long-term Cybersecurity Symposium on CCPA/GDPR, ETH Zurich's workshop series in Law, Economics, and Data Science, and the 2022 Conference on Empirical Legal Studies for helpful advice and feedback. I also thank Wendy Leben for excellent research assistance with data collection, editing, and legal analysis. Finally, I thank the N.Y.U. Journal of Legislation And Public Policy editors for their thorough and thoughtful work in preparing this piece for publication. Replication and code available at: *https://github.com/Akesari12/Do-Data-Breach-Notification-Laws-Work-NYU-JLPP-Replication-Code/tree/main*.

The Article ties together these results within an economic theory of data breach laws: namely, whether legal provisions (1) enable consumer mitigation of data breach harms, or (2) encourage organizations to invest in better data security. It explains how these results and theory provide lessons for current federal and state proposals to expand or amend the scope of breach notification laws. A new federal law that simply mimics existing baseline requirements is unlikely to have an additional effect and may preempt further innovations. At the state level, introducing private rights of action may help at the margins, but likely suffers from well-identified issues of adequately establishing standing and damages. States that close loopholes surrounding breach requirements for encrypted data see lower identity-theft report rates, which suggests that other states may be wise to tighten these requirements as well. Looking forward, states should experiment with solutions such as automatically enrolling consumers in identity theft protection services or providing direct incentives for strong data security.

Introduction

Do data breach notification laws work? Each year, more and increasingly severe data breaches are reported.[1] With these data breaches come in-

[1] *See, e.g.*, Nicole Perlroth, *All 3 Billion Yahoo Accounts Were Affected by 2013 Attack*, N.Y. Times (Oct. 3, 2017), https://www.nytimes.com/2017/10/03/technology/yahoo-hack-3-billion-users.html [https://perma.cc/F9EE-87JT] (describing data breach

creased rates of reported identity theft.[2] One estimate suggests that identity theft in 2014 cost consumers an average of $1,300 per incident and totaled over $15 billion.[3] In response, between 2003 and 2018, all fifty U.S. states adopted some form of a data breach notification law.[4] As a baseline, these laws require breached organizations to notify data subjects when a breach occurs. At the federal level, similar breach-disclosure requirements apply to organizations maintaining health data,[5] publicly traded companies,[6] and maintainers of critical infrastructure.[7] One central goal of these laws is to protect consumers by reducing identity theft.

Despite the popularity of these laws, scholars are pessimistic about their effectiveness. In a recent book, Daniel Solove and Woodrow

of Yahoo in 2013 that affected three billion accounts); *Equifax Data Breach Settlement*, Fed. Trade Comm'n (Dec. 2022), https://www.ftc.gov/enforcement/refunds/equifax-data-breach-settlement [https://perma.cc/N3A3-T9HV] (discussing 2017 Equifax data breach that affected 147 million records); Taylor Telford & Craig Timberg, *Marriott Discloses Massive Data Breach Affecting up to 500 Million Guests*, Wash. Post (Nov. 30, 2018, 1:03 PM), https://www.washingtonpost.com/business/2018/11/30/marriott-discloses-massive-data-breach-impacting-million-guests [https://perma.cc/97CX-YTCV] (discussing 2018 Marriott data breach that affected up to five hundred million guests); *Data on 540 Million Facebook Users Exposed*, BBC News (Apr. 4, 2019), https://www.bbc.com/news/technology-47812470 [https://perma.cc/6ZNZ-6XFK] (discussing 2019 Facebook data breach that affected 540 million users); Aaron Holmes, *How Hackers Breached IT Company SolarWinds and Staged an Unprecedented Attack That Left US Government Agencies Vulnerable for 9 Months*, Bus. Insider (Dec. 14, 2020, 12:01 PM), https://www.businessinsider.com/solarwinds-hack-us-government-agencies-cisa-fireeye-microsoft-2020-12 [https://perma.cc/9KVM-W8P6] (discussing 2020 SolarWinds data breach that affected an unknown number of records).

[2] *See* Press Release, Fed. Trade Comm'n, New Data Shows FTC Received 2.2 Million Fraud Reports From Consumers in 2020 (Feb. 4, 2021), https://www.ftc.gov/news-events/news/press-releases/2021/02/new-data-shows-ftc-received-22-million-fraud-reports-consumers-2020 [https://perma.cc/PB3L-RYKY].

[3] *See* Erika Harrell, *Victims of Identity Theft, 2014*, Dep't Just. 6–7, https://bjs.ojp.gov/content/pub/pdf/vit14.pdf (last revised Nov. 13, 2017).

[4] *See Security Breach Notification Laws*, Nat'l Conf. State Legislatures, https://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx [https://perma.cc/FA8Q-ZWSX] (last updated Jan. 17, 2022).

[5] Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936, 2009 (1996) (codified at 42 U.S.C. §§ 1320a–7e); Health Information Technology for Economic and Clinical Health Act, Pub. L. No. 111-5, 123 Stat. 115, 226 (2009) (codified at 42 U.S.C. §§ 300jj–11).

[6] *See* Press Release, Sec. & Exch. Comm'n, SEC Adopts Rules on Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure by Public Companies (July 26, 2023), https://www.sec.gov/news/press-release/2023-139 [https://perma.cc/RXV6-ZKBY].

[7] *See* Cyber Incident Reporting for Critical Infrastructure Act of 2022, Pub. L. No. 117-103, 136 Stat. 49, 1043 (2022) (codified at 6 U.S.C. § 681b); *see also* Colleen Theresa Brown et al., *Congress Passes Cyber Incident Reporting for Critical Infrastructure Act of 2022*, Sidley Austin LLP (Mar. 21, 2022), https://www.sidley.com/en/insights/newsupdates/2022/03/congress-passes-cyber-incident-reporting-for-critical-infrastructure-act-of-2022 [https://perma.cc/CU6X-Y4SZ] (providing an overview of the Act).

Hartzog argue that data security laws fail because of their focus on the targets of data breaches rather than the companies that write vulnerable software or make vulnerable devices. They then ask why breach notification laws do not seem to make any difference.[8] Empirical investigations of breach notification laws offer conflicting evidence about their effectiveness.[9] Policy proposals include creating federal privacy standards,[10] empowering administrative agencies to punish lax data security practices,[11] and reforming private causes of action to encourage more breach litigation.[12]

Are breach notification laws functionally useless, or can policymakers salvage something from them? This Article addresses a number of gaps in the current scholarly discourse. First, scholars typically treat breach notification laws as doing just one thing—notifying consumers about breaches.[13] This simplification masks important variation between

[8] Daniel J. Solove & Woodrow Hartzog, Breached! Why Data Security Law Fails and How to Improve It 9 (2022).

[9] *See* Sasha Romanosky, Richard Sharp & Alessandro Acquisti, Data Breaches and Identity Theft: When Is Mandatory Disclosure Optimal? 20–26 (Aug. 15, 2010) (unpublished manuscript), https://econinfosec.org/archive/weis2010/papers/session1/weis2010_romanosky.pdf; Sasha Romanosky, Rahul Telang & Alessandro Acquisti, *Do Data Breach Disclosure Laws Reduce Identity Theft?*, 30 J. Pol'y Analysis & Mgmt. 256, 260 (2011) (showing that breach notification laws reduce identity theft by about 6.1%); Sanjay Goel & Hany A. Shawky, *Estimating the Market Impact of Security Breach Announcements on Firm Values*, 46 Info. & Mgmt. 404, 406 (2009) (showing that breached firms' stock prices dip slightly after announcing a breach but quickly recover). *But see* Richard J. Sullivan & Jesse Leigh Maniff, *Data Breach Notification Laws*, 2016 Econ. Rev. 65, 75–76 (using a regression analysis to show that some data breach notification law provisions reduce identity theft and some increase it).

[10] *Proposed U.S. Privacy Legislation*, Elec. Priv. Info. Ctr., https://epic.org/issues/privacy-laws/proposed-legislation [https://perma.cc/R6GN-ZAYV] (last visited Sept. 26, 2023).

[11] For example, the FTC has received attention as the nation's primary privacy enforcer. *See, e.g.*, William R. Denny, *Cyber Center: Cybersecurity as an Unfair Practice: FTC Enforcement Under Section 5 of the FTC Act*, Am. Bar Ass'n (June 20, 2016), https://www.americanbar.org/groups/business_law/resources/business-law-today/2016-june/cyber-center-cyber-security-as-an-unfair-practice/ [https://perma.cc/H8A7-9HAB].

[12] For example, breach litigation has risen dramatically following the adoption of the California Consumer Privacy Act, Cal. Civ. Code § 1798.100 (West 2022). *See* Jena M. Valdetero & David A. Zetoony, *CCPA Litigation Up 44.1%*, Nat'l L. Rev. (Mar. 7, 2022), https://www.natlawreview.com/article/ccpa-litigation-441 [https://perma.cc/7RKF-YPQB].

[13] For a rich overview of the theoretical justifications underlying breach notification and how breach notification laws might be brought in closer alignment with them, see Mark Verstraete & Tal Zarsky, *Optimizing Breach Notification*, 2021 U. Ill. L. Rev. 803, 808–39 (2021). Much of the empirical literature to date focuses on the disclosure aspects of breach notification laws. *See* Brad N. Greenwood & Paul M. Vaaler, Do US State Data Breach Notification Laws Decrease Firm Data Breaches? (Mar. 6, 2023) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3885993 [https://perma.cc/Y2JB-3TRM] (analyzing the effect of states' first breach notification

state laws and the fact that these laws are routinely updated with new provisions. Second, scholars often question whether these laws are effective at all because we still see data breaches despite the popularity of these laws.[14] A better way to statistically formulate this question is whether we see *less identity theft than we would have in the absence of these laws*. Third, scholars talk about breach notification laws in theoretical and legal terms, but state variation provides avenues to assess these theories empirically.[15] Breach notification laws vary across states. States also vary in population characteristics, such as race, age, and Internet access.[16] These facts make it possible to study whether different provisions have heterogeneous effects across states and over time.

This Article addresses the question of whether data breach notification laws deter identity theft by implementing a staggered adoption synthetic control approach. Synthetic control is a popular method in the social sciences for conducting data-driven comparative case studies[17] and has been described as "arguably the most important innovation in the policy evaluation literature in the last 15 years."[18] The staggered version allows estimation of causal effects when different units (states, in this case) adopt similar laws at different time periods.[19] Using Federal

law on data breach incidents); Joshua Mitts & Eric Talley, *Informed Trading and Cybersecurity Breaches*, 9 HARV. BUS. L. REV. 1 (2019) (analyzing how capital markets react before and after a breach announcement).

[14] *See, e.g.*, Sullivan & Maniff, *supra* note 9, at 77; *see also* Jane K. Winn, *Are "Better" Security Breach Notification Laws Possible?*, 24 BERKELEY TECH. L.J. 1133, 1133–34 (2009); Mark Burdon, *Contextualizing the Tensions and Weaknesses of Information Privacy and Data Breach Notification Laws*, 27 SANTA CLARA HIGH TECH. L.J. 63, 63 (2011) (arguing that breach notification laws suffer from a lack of contextual approach); Jill Joerling, Note, *Data Breach Notification Laws: An Argument for a Comprehensive Federal Law to Protect Consumer Data*, 32 WASH. UNIV. J.L. & POL'Y 467, 468 (2010) (arguing that a federal standard could cover weaknesses in state laws).

[15] Verstraete & Zarsky, *supra* note 13, at 808–39 (discussing the theoretical dimensions of breach notification laws); *see* Carol M. Hayes, *Comparative Analysis of Data Breach Laws: Comprehension, Interpretation, and External Sources of Legislative Text*, 23 LEWIS & CLARK L. REV. 1221 (2020) (analyzing differences in statutory language among data breach notification laws).

[16] *See U.S. Census Bureau, Data Profiles*, U.S. DEP'T COM., https://data.census.gov/profile?q=United+States&g=010XX00US [https://perma.cc/SY27-DHH4] (last visited Sept. 23, 2023).

[17] The original paper that introduced the synthetic control approach has been cited over 6,200 times, demonstrating the popularity of the method. *See* Alberto Abadie et al., *Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program*, 105 J. AM. STAT. ASS'N 493 (2010) (introducing the synthetic-control method and applying it to California's tobacco tax).

[18] Susan Athey & Guido W. Imbens, *The State of Applied Econometrics: Causality and Policy Evaluation*, 31 J. ECON. PERSPS. 3, 9 (2017).

[19] *See* Eli Ben-Michael et al., *Synthetic Controls With Staggered Adoption*, 84 J. ROYAL STAT. SOC'Y SERIES B: STAT. METHODOLOGY 351 (2022).

Trade Commission (FTC) panel data,[20] this Article categorizes breach notification provisions in terms of their two primary economic functions: (1) enabling consumers to mitigate data breach harms and (2) encouraging organizations to adopt stronger data security practices. This Article builds on the author's previous study of California's 2016 breach notification disclosure requirements by extending the synthetic control method to all U.S. states across the entire twenty-year history of breach notification laws.[21]

The results of this novel empirical approach yield several important insights for data breach notification law. The most effective provisions are those that require disclosure to state regulators and those that apply breach notification requirements to encrypted data. Each of these provisions reduces identity theft reports by about 10%. Baseline breach notification provisions and providing a cause of action have smaller effects of reducing reports by about 3-5%. Identity theft reports increase by about 5% when states force firms to disclose even when they conclude a breach has a low risk of harm for consumers.

These quantitative findings inform several high-level policy implications and yield new insights about the political economy of data breach federalism. First, any new federal breach notification law should create a *floor* for further state action, as state innovation has been an important part of experimenting with more effective breach notification law provisions. Having said that, a federal law that simply adopts the baseline breach notification requirements that exist in every state will likely not be effective; therefore, lawmakers should pay attention to introducing new provisions that have not been adopted across the country. Second, this Article makes important progress on untangling the economic theory of data breach notification laws. Provisions consistent with consumer *mitigation* of data breach harms have the largest effects. In contrast, provisions that aim to punish organizations through reputational harms might be counterproductive and lead to estimated *increases* in identity theft reporting. Third, market mechanisms alone are unlikely to deter data breaches. Regulators play an important role in investigating data breaches and recovering damages for consumers. Fourth, this Article has implications for data-driven policy studies. Within privacy law, high-quality data about privacy violations is still elusive, and more should be made available to scholars and the public.[22]

[20] Panel data refers to a data structure where each unit is observed across time.

[21] Aniket Kesari, *Do Data Breach Notification Laws Reduce Medical Identity Theft? Evidence from Consumer Complaints Data*, 19 J. EMPIRICAL LEGAL STUD. 1222 (2022).

[22] *See* Chris Jay Hoofnagle, *Making the Known Unknowns Known*, 21 HARV. J.L. & TECH. 97, 101–03 (2007) (discussing the challenges posed by synthetic identity theft

Across empirical legal studies more generally, the methodology used here can be adapted to policy choices that vary across states or other sub-national units.

Overall, this Article provides a data-driven and nuanced look at how these laws have impacted identity theft over time and points to policy implications that non-adopting states and the federal government might consider as they continue to adopt new breach notification laws. This Article contributes to both the privacy law and empirical legal studies literature. It is the most comprehensive quantitative examination of data breach notification provisions across their entire history. As such, it bolsters some of the previous empirical findings in this space and yields new insights about previously unexplored aspects of breach notification laws. It also provides a thorough walkthrough of the synthetic control method and its staggered adoption extension, thus expanding its accessibility to applied legal and policy scholars. Privacy law scholarship can expand its policy relevance with empirical studies that investigate questions around what has and has not worked so far. This type of analysis will be critical as both states and the federal government are increasingly active in regulating this space.

I.
The Landscape of Data Breaches and Identity Theft

A. *Identity Theft*

Identity theft is a relatively recent crime. The FTC only started to track it in 1997.[23] Federal law defines identity theft as a crime "that involves the transfer or use of 1 or more means of identification if, as a result of the offense, any individual committing the offense obtains anything of value aggregating $1,000 or more during any 1-year period."[24] Most states have also passed an identity-theft criminal statute.[25] Identity

and lack of high-quality identity theft data). Recently there has been progress in analyzing privacy violations more broadly. *See, e.g.*, Karel Kubíček et al., *Checking Websites' GDPR Consent Compliance for Marketing Emails*, 2 Proc. on Priv. Enhancing Techs. 282, 288 (2022) (describing having recently conducted one of the first large-scale studies of privacy violations of the GDPR by generating a dataset of emails from registering for 1,000 websites).

[23] Fed. Trade Comm'n, Consumer Sentinel Network Data Book 2021, at 2 (2022), https://www.ftc.gov/system/files/ftc_gov/pdf/CSN%20Annual%20Data%20Book%202021%20Final%20PDF.pdf [hereinafter Data Book 2021].

[24] Identity Theft and Assumption Deterrence Act of 1998, Pub. L. 105-318, § 3, 112 Stat. 3007, 3007 (1998) (codified as amended at 18 U.S.C. § 1028).

[25] *See* Graeme R. Newman & Megan M. McNally, Identity Theft Literature Review 63–65 (2005); *State Identity Theft Statutes and Criminal Use of Personal ID*, Nat'l Conf. State Legislatures, https://www.ncsl.org/financial-services/identity-theft [https://perma.cc/LG2D-M2GH] (last updated May 31, 2017).

theft is treated as an economic crime where an entity uses someone else's persona for economic gain.[26]

The FTC collects data on both identity theft and identity fraud. This latter category makes up the lion's share of consumer complaints—in 2021, of the 5.7 million reports to the FTC, approximately 1.43 million were identity theft reports and the remainder were related to fraud or another category.[27] Fraud reports often deal with scams such as fake sweepstakes, imposter scams, and negative reviews.[28]

One main difference between identity fraud and identity theft is that victims are usually able to pinpoint the event that led to fraud, whereas identity theft can materialize months or even years after the precipitating event.[29] Nearly one-third of identity theft reports involve a victim discovering that someone attempted to apply for government benefits in their name, oftentimes months or years later.[30] Resolving identity theft problems takes an average of six months and one- to two-hundred hours of work.[31]

Identity theft itself has more subcategories. Much of identity theft is financial—thieves attempt to gain access to credit or other economic benefits by impersonating someone.[32] Medical identity theft is another serious problem where perpetrators use a stolen identity to make fraudulent claims against Medicare and other insurers.[33] In rare instances a victim will personally know the individual who stole their identity.[34] Among the most complicated cases to detect are cases of "synthetic identity theft" where perpetrators construct fake personas using a blend of stolen information about real people (e.g., Social Security Numbers) and fake biographical information.[35]

[26] *Identity Theft*, U.S. DEP'T JUST., https://www.justice.gov/criminal/criminal-fraud/identity-theft/identity-theft-and-identity-fraud [https://perma.cc/8MDC-AK6K] (last updated Aug. 11, 2023).

[27] DATA BOOK 2021, *supra* note 23, at 86.

[28] *Id.* at 7.

[29] Oftentimes victims only discover their identity has been stolen after they are contacted by a financial institution about suspicious activity. *See, e.g.*, Harrell, *supra* note 3, at 7.

[30] *See id.* at 3.

[31] *See How to Protect Yourself Against the Theft of Your Identity*, ECONOMIST (Sept. 14, 2017), https://www.economist.com/finance-and-economics/2017/09/14/how-to-protect-yourself-against-the-theft-of-your-identity [https://perma.cc/5EH8-NKFY].

[32] *See Identity Theft*, *supra* note 26.

[33] *See Medical Identity Theft*, HHS OFF. INSPECTOR GEN., https://oig.hhs.gov/fraud/consumer-alerts/medical-identity-theft [https://perma.cc/9A53-ZVA3] (last visited Sept. 24, 2023).

[34] Harrell, *supra* note 3, at 8 ("Overall, 6% of victims knew something about the identity of the offender in the most recent incident of identity theft.").

[35] *See* Hoofnagle, *supra* note 22, at 101 (discussing the challenges posed by synthetic identity theft).

B. Data-Breach Notification Laws

Identity theft is frequently a result of a data breach—an event where information is disclosed to an unauthorized party.[36] Some data breaches are attributable to accidental losses by employees of an organization, for example, if someone loses a company laptop or accidentally emails a file they should not to an unauthorized party. Others are the result of an external attack where a perpetrator intends to acquire data illegally.

Data breach laws attempt to shield consumers from identity theft following a breach. One potential problem is that breached organizations may not want to disclose the event. However, failure to disclose a data breach could leave consumers vulnerable, as it leaves them unaware that their data was compromised. To solve this problem, every state now has enacted some kind of data breach notification law.[37] These laws have two principal purposes: (1) encourage organizations to invest in cybersecurity so as to avoid damaging disclosures and (2) give consumers the opportunity to safeguard their identity.[38]

The evidence on whether these laws succeed is mixed. Brad Greenwood and Paul Vaaler conducted a two-way fixed effects analysis of data-breach notification laws and found that they do not reduce the number of data breaches.[39] Sanjay Goel and Hany A. Shawky found that publicly traded companies suffer temporary hits to their market caps after a breach announcement.[40] However, Joshua Mitts and Eric Talley found that this effect actually leads to insider trading.[41] Sasha Romanosky, Rahul Telang, and Alessandro Acquisti directly studied the question of what effect data breach laws have on identity-theft report rates and found about a 6% reduction in identity-theft report rates among states that adopted data breach notification laws in the mid-2000s.[42] In similar work, Romanosky, David A. Hoffman, and Acquisti explored federal data breach litigation and found that firms are more vulnerable

[36] However, the exact relationship between data breaches and identity theft is unknown. Not all data breaches result in identity theft, and the true number of data breaches and identity theft is difficult to quantify. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-07-737, PERSONAL INFORMATION: DATA BREACHES ARE FREQUENT, BUT EVIDENCE OF RESULTING IDENTITY THEFT IS LIMITED; HOWEVER, THE FULL EXTENT IS UNKNOWN (2007).

[37] *Security Breach Notification Laws*, *supra* note 4.

[38] *See* Romanosky, Telang & Acquisti, *supra* note 9, at 259 (discussing the goals of data breach notification laws).

[39] Greenwood & Vaaler, *supra* note 13, at 4.

[40] Goel & Shawky, *supra* note 9, at 406.

[41] Mitts & Talley, *supra* note 13, at 3–4.

[42] Romanosky, Telang & Acquisti, *supra* note 9, at 260.

to lawsuits when consumers suffer financial harm but less vulnerable when they offer credit monitoring services.[43]

Breach notification laws share a number of similar definitions. Nearly all breach notification laws define the following:

- **Personal Information**: Examples of personal information that are covered by the law. Sometimes the laws also define what combinations of information constitute a security breach (e.g., a social security number alongside a name).
- **Covered Entities**: Entities that are subject to breach disclosure requirements. Every law covers private businesses, but some regulate government agencies and insurers differently.
- **Notification Trigger**: The threshold that triggers a notification obligation. Examples include "substantial harm to individuals," "reasonable likelihood of harm," or "awareness of breach." Under the first two standards, organizations only need to provide notice if they believe there will be harm to their data subjects, whereas standards closer to "awareness of breach" remove this discretion.
- **Timing of Notification**: Some states only require that notifications are made within something like "the most expedient time possible" and "without unreasonable delay." For example, the relevant provision in the original California statute says:

> Any person or business that conducts business in California, and that owns or licenses computerized data that includes **personal information**, shall disclose any breach of the security of the system **following discovery** or notification of the breach in the security of the data to any resident of California whose unencrypted personal information was, or is reasonably believed to have been, acquired by an unauthorized person. The disclosure shall be made in the **most expedient time possible and without unreasonable delay**, consistent with the legitimate needs of law enforcement, as provided in subdivision (c), or any measures necessary to determine the scope of the breach and restore the reasonable integrity of the data system.[44]

Other states may use this exact language, alter it slightly, or use entirely different definitions of personal information, triggers, and

[43] Sasha Romanosky, David A. Hoffman & Alessandro Acquisti, *Empirical Analysis of Data Breach Litigation Empirical Analysis of Data Breach*, 11 J. EMPIRICAL LEGAL STUD. 74 (2014).

[44] California Consumer Privacy Act, CAL. CIV. CODE § 1798.82 (West 2003) (emphasis added).

timelines. State variations notwithstanding, these basic ingredients make up all breach notification laws.

II.
The Law & Economics of Breach Notification

Though every U.S. state has adopted a breach notification law, the interesting variation between these laws should not be understated.[45] States vary in what types of information and to whom information must be disclosed in a breach disclosure. Scholars have analyzed the role disclosures play in informing consumers and other stakeholders.[46] This exclusive focus on the information flow aspects of breach notification laws potentially misses other provisions that are commonly found in these laws. One way to characterize specific provisions of these laws is to look at how they relate to the core economic theory of data breach notification laws.[47] The two prongs of data breach notification law are that they can a) empower consumers to mitigate harms after a data breach occurs, and b) encourage organizations to invest in strong data security before a breach can occur.[48] Relating various provisions to these two prongs of the economic theory can illuminate which of these mechanisms are likely to be effective.

A. Consumer Mitigation of Harms

At their core, breach notification laws are about correcting information asymmetries between consumers and data-holding organizations. In the absence of a breach notification law, organizations have a strong incentive to keep the knowledge of a breach private. Doing this avoids possible fallout from a breach announcement including regulatory scrutiny and consumer backlash. Breach notification laws aim to make private information public, thereby addressing the information asymmetry problem. As discussed earlier, one of the economic justifications for these laws is to give consumers adequate opportunity to

[45] *Security Breach Notification Laws*, *supra* note 4.

[46] *See, e.g.*, Verstraete & Zarsky, *supra* note 13, at 808–39; Stephen Jackson et al., *An Investigation of the Impact of Data Breach Severity on the Readability of Mandatory Data Breach Notification Letters: Evidence From U.S. Firms*, 70 J. Ass'n for Info. Sci. & Tech. 1277 (2019) (analyzing the readability of data breach disclosures); Mitts & Talley, *supra* note 13 (discussing the role that breach notifications play in informing investors in capital markets as they investigate evidence of insider training).

[47] The information for the timeline of the adoption of each of these laws was constructed by using a combination of Bloomberg Law and LexisNexis sources. Bloomberg Law provides a spreadsheet of which state laws include each provision, and this information was checked against the text of the laws found on LexisNexis. Both the Bloomberg spreadsheet and text of the individual laws are on file with the author.

[48] Romanosky, Telang & Acquisti, *supra* note 9, at 260.

safeguard their data by investing in credit-monitoring services, identity theft protection, etc.[49]

One way that states might improve consumer opportunities to mitigate data breach harms is by imposing additional content requirements for disclosures. For example, since 2016, California requires that breach disclosures follow a particular format with clear headings describing what happened, how it happened, and mitigation steps.[50] States may also require that firms publicize the number of individuals affected and information about consumers' rights in the event of a breach. In some instances, states may actually restrict the types of information that firms can provide to avoid unnecessarily alarming consumers. In Massachusetts, for example, the statute forbids firms from disclosing how a breach occurred or the number of individuals implicated.[51]

Beyond regulating the content of breach notifications, states also differ in how they regulate who must be notified in the event of a breach. At baseline, every state law requires that businesses notify affected consumers.[52] Some states further require notification to the state's Attorney General, as seen in Figure 1.[53] Among these states, some further require that breach notices are posted online, generally on the Attorney General's website.[54] Attorneys general and other consumer protection regulators

[49] *Id.*

[50] California Consumer Privacy Act, CAL. CIV. CODE § 1798.100 (West 2022).

[51] *See Requirements for Data Breach Notifications*, MASS. OFF. CONSUMER AFFS. & BUS. REGUL., https://www.mass.gov/info-details/requirements-for-data-breach-notifications [https://perma.cc/2XD9-YHKP] (last visited Sept. 26, 2023).

[52] *Security Breach Notification Laws*, *supra* note 4.

[53] For example, Alabama's statute reads:

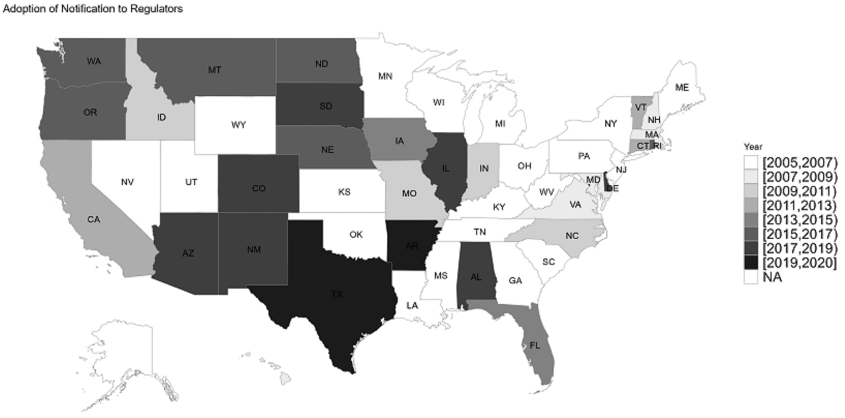
> (b) Written notice to the Attorney General shall include all of the following: (1) A synopsis of the events surrounding the breach at the time that notice is provided. (2) The approximate number of individuals in the state who were affected by the breach. (3) Any services related to the breach being offered or scheduled to be offered, without charge, by the covered entity to individuals, and instructions on how to use the services. (4) The name, address, telephone number, and email address of the employee or agent of the covered entity from whom additional information may be obtained about the breach. (c) A covered entity may provide the Attorney General with supplemental or updated information regarding a breach at any time. (d) Information marked as confidential that is obtained by the Attorney General under this section is not subject to any open records, freedom of information, or other public record disclosure law.

ALA. CODE § 8-38-6 (2022).

[54] *See, e.g.*, *Data Security Breach Reporting*, CAL. OFF. ATT'Y GEN., https://oag.ca.gov/privacy/databreach/reporting [https://perma.cc/644B-ZS6C] (last visited Sept. 27, 2023) (posting notices for California); *Data Breach Notifications*, WASH. STATE OFF. ATT'Y GEN., https://www.atg.wa.gov/data-breach-notifications [https://perma.cc/65QV-UYFC] (last visited Sept. 27, 2023) (posting notices for Washington).

may take a number of actions following a data breach such as pursuing injunctions requiring companies to take particular steps to safeguard consumer identities, requiring consumer restitution, or levying civil penalties for failure to disclose properly.[55] State attorneys general may also pursue multistate litigation, as was the case with the multistate investigation into the Uber data breach that led to a $148 million settlement.[56]

Figure 1
Timeline of State Adoption of Requiring Notification to State Regulators[57]



Others also require that firms report breaches to credit reporting agencies.[58] The three major credit reporting agencies are Equifax,

[55] *See Data Breaches*, Nat'l Ass'n Attys. Gen., https://www.naag.org/issues/consumer-protection/consumer-protection-101/privacy/data-breaches [https://perma.cc/3DL5-TF5F] (last visited Sept. 27, 2023).

[56] *See* News Release, N.J. Off. Atty. Gen., AG Grewal Announces Historic Settlement Resolving Uber Data Breach (Sept. 26, 2018), https://nj.gov/oag/newsreleases18/pr20180926a.html [https://perma.cc/GW8L-HK4F]; Press Release, Mass. Off. Atty. Gen., AG Healey Leads Multistate Coalition in Reaching $148 Million Settlement With Uber Over Nationwide Data Breach (Sept. 26, 2018), https://www.mass.gov/news/ag-healey-leads-multistate-coalition-in-reaching-148-million-settlement-with-uber-over-nationwide-data-breach [https://perma.cc/KZ4A-TE58]; Press Release, Off. Atty. Gen. for D.C., AG Racine Reaches $148 Million Nationwide Settlement Over Uber Data Breach (Sept. 26, 2018), https://oag.dc.gov/release/ag-racine-reaches-148-million-nationwide [https://perma.cc/BEA6-JBU4].

[57] Since 2005, several states required breach notifications to be filed with the Attorney General or other state regulator. In six of these states (California, Washington, Oregon, New Hampshire, Vermont, and Montana) these notices are also posted online on the AG's website.

[58] For example, Arizona law requires that:

> 2. If the breach requires notification of more than one thousand individuals, notify both: (a) The three largest nationwide consumer reporting agencies. (b) The attorney general, in writing, in a form prescribed by rule or order

Experian, and TransUnion. One immediate step that a credit reporting agency might take after being notified of a data breach is to offer consumers the ability to freeze their credit reports. This step essentially allows consumers to block anyone from trying to open a new credit account under their identity.[59] Certain states also require that firms sign up consumers for identity theft protection services that can include credit monitoring, identity theft insurance, and public-records searches.[60]

Finally, states vary in the prescribed timeline for notification. Most states include language in their statutes mandating that breaches are reported without an "unreasonable delay."[61] Other states require that notification be given within a specific timeframe such as forty-five, sixty, or ninety days.[62] These laws can also differ with regards to different timelines for notification to different parties. For instance, Alabama and Colorado require notification to consumers and the state Attorney General within thirty and sixty days respectively but both only require notification to credit reporting agencies "without unreasonable delay."[63] Vermont requires notification to consumers within forty-five days, but to the Attorney General within fourteen days.[64] Minnesota

> of the attorney general or by providing the attorney general with a copy of the notification provided pursuant to paragraph 1 of this subsection.

ARIZ. REV. STAT. ANN. § 18-552 (2022).

[59] *See What to Know About Credit Freezes and Fraud Alerts*, FED. TRADE COMM'N (May 2021), https://consumer.ftc.gov/articles/what-know-about-credit-freezes-fraud-alerts [https://perma.cc/DSV4-688Q].

[60] *See What Is Identity Monitoring or "Identity Theft Protection" Service?*, CONSUMER FIN. PROT. BUREAU, https://www.consumerfinance.gov/ask-cfpb/what-is-identity-monitoring-or-identity-theft-protection-service-en-1369 [https://perma.cc/DEZ4-5KLQ] (last reviewed Sept. 21, 2018).

[61] Text of state laws available at: *Data Breach Notification Laws*, GITHUB, *https://github.com/Akesari12/Data-Breach-Notification-Laws* [https://perma.cc/FQ22-7QE8] (last updated Nov. 1, 2023) [hereinafter GitHub].

[62] *Id.*

[63] For instance, Louisiana's statute reads:

> The notification required pursuant to Subsections C and D of this Section shall be made in the most expedient time possible and without unreasonable delay but not later than sixty days from the discovery of the breach, consistent with the legitimate needs of law enforcement, as provided in Subsection F of this Section, or any measures necessary to determine the scope of the breach, prevent further disclosures, and restore the reasonable integrity of the data system.

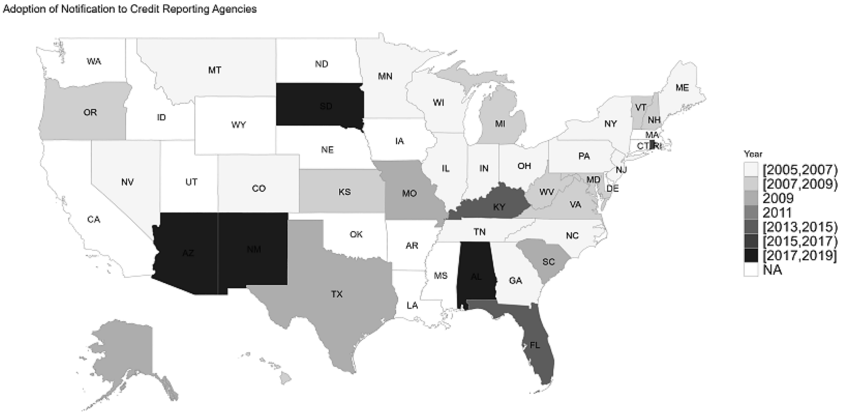
And further clarifies the law enforcement exception as: "If a law enforcement agency determines that the notification requirement under this section would impede a criminal investigation, such notification may be delayed until such law enforcement agency determines that the notification will no longer compromise such investigation."

LA. STAT. ANN. § 51:3074 (2022).

[64] GitHub, *supra* note 61.

requires notification to consumers “without unreasonable delay” but to credit reporting agencies within two days.[65] The main mechanism for these time limit requirements is that the sooner consumers know about a breach, the sooner they might be able to take actions to mitigate its consequences.

Figure 2
Timeline of State Adoption of Requiring Notification to Credit Reporting Agencies[66]



B. Encouraging Better Data Security

The second prong of the economic theory of data breach notification laws suggests that they should work in part by encouraging firms to avoid making disclosures in the first place. One way to encourage better data security across organizations might be to punish those that are caught being negligent. States may prescribe statutory damages or fines for failure to comply with breach notification laws. States also differ in who is allowed to bring legal action against negligent firms. Virtually every state allows the Attorney General to bring legal action, but only some provide a private cause of action for individuals or class actions.[67] Another way states might encourage better data security

[65] *Id.*

[66] Since 2005, several states have required breach notifications to be reported to consumer credit reporting agencies.

[67] For example, South Carolina’s law reads:

> (G) A resident of this State who is injured by a violation of this section, in addition to and cumulative of all other rights and remedies available at law, may: (1) institute a civil action to recover damages; (2) seek an injunction to enforce compliance; and (3) recover attorney’s fees and court costs, if successful.

S.C. Code Ann. § 1-11-490 (2022).

is by incentivizing the adoption of certain standards, such as strong encryption.

Among states that do allow a private cause of action, plaintiffs are generally entitled to actual damages if they can show negligence on the part of the breached firm.[68] One exception to this pattern is a provision in the California Consumer Privacy Act (CCPA) that provides for statutory damages of at least one hundred dollars per record.[69]

There is a rich literature surrounding whether public or private enforcement is desirable in various policy areas.[70] Consumers likely have a better sense of the harms they individually suffered from a breach, whereas regulators may have more information, or ability to gather information, about possible practices that constituted cybersecurity negligence. Within consumer protection law, and specifically in regulating data breaches, considering the role of information can be especially insightful. One principle for determining whether public or private enforcement might be preferable is to consider whether public regulators or private citizens are better positioned, in terms of available information, to punish wrongdoing. As discussed, breach notification laws differ in how they establish these information flows—almost all require notification to consumers,[71] and some also require notification to public regulators such as state attorneys general. Assuming both consumers and regulators receive a breach notification, there may still be differences in the information available to them. At least theoretically, there are plausible arguments for either private or public enforcement of data breach harms.

A private cause of action should theoretically empower consumers to recover damages even if their state Attorney General declines to pursue legal action. Indeed, there is evidence that firms might be especially sensitive to allegations of financial loss arising from data breaches.[72] However, these lawsuits might also falter when other types of harm are alleged, particularly psychological harms. Courts are loathe

[68] GitHub, *supra* note 61.

[69] California Consumer Privacy Act, CAL. CIV. CODE § 1798.150 (West 2022).

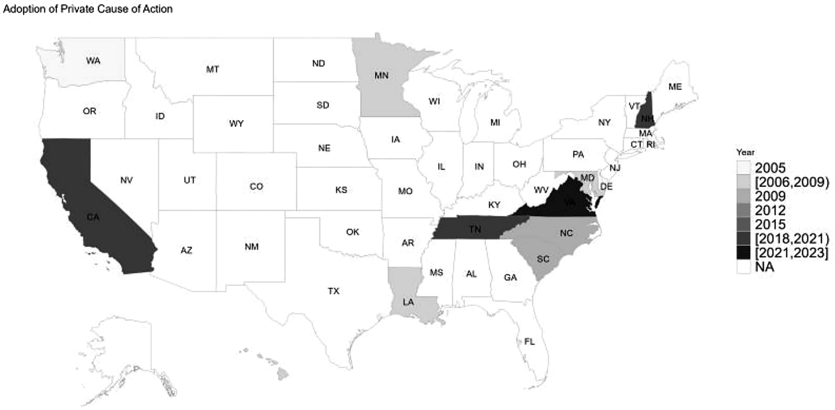
[70] *See, e.g.*, ROBERT A. KAGAN, ADVERSARIAL LEGALISM: THE AMERICAN WAY OF LAW (2002); John Fabian Witt, *Bureaucratic Legalism, American Style: Private Bureaucratic Legalism and the Governance of the Tort System*, 56 DEPAUL L. REV. 261 (2007); William M. Landes & Richard A. Posner, *The Private Enforcement of Law*, 4 J. LEGAL STUD. 1 (1975); Steven Shavell, *The Optimal Structure of Law Enforcement*, 36 J.L. & ECON. 255 (1993).

[71] Some exceptions, however, include good-faith exceptions, likelihood-of-harm exceptions, encryption exceptions, etc.

[72] *See* Romanosky, Hoffman & Acquisti, *supra* note 43, at 102 ("[O]ur results suggest that defendants settle 30 percent more often when plaintiffs allege financial loss from a data breach").

to rule in favor of plaintiffs when damages are difficult to calculate and generally dismiss claims alleging that data breaches cause consumers anxiety over potential future consequences and the prospect of invasive surveillance.[73]

Figure 3
Timeline of State Adoption of Private Cause of Action Provisions[74]



Another mechanism for encouraging better data security could be to provide a "risk-of-harm analysis" exception to disclosure requirements. Most states require that breached organizations give notice *unless* they have a reasonable belief that the breach is unlikely to cause consumer harms. The standards for these analyses can differ from state to state but generally require some kind of internal investigation of the breach to determine the risk of harm.[75] The exact mechanisms and standards that firms use to conduct these analyses is somewhat opaque and subjective, which has prompted criticism from consumer advocates.[76] There is likely variation across industries, however. For example, healthcare organizations are all subject to a similar standard by the

[73] *See* Daniel J. Solove & Danielle Keats Citron, *Risk and Anxiety: A Theory of Data-Breach Harms*, 96 Tex. L. Rev. 737, 753–54 (2018); Ido Kilovaty, *Psychological Data Breach Harms*, 23 N.C. J.L. & Tech. 1, 22–23 (2021) (citing to Solove & Citron).

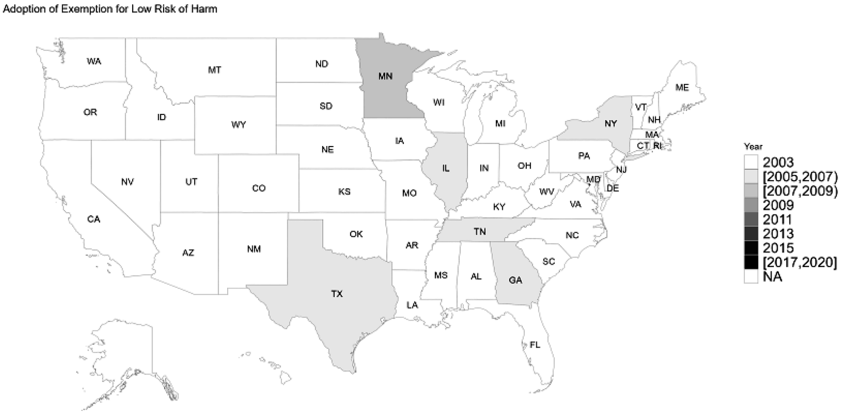
[74] Since 2005, several states have included explicit rights of action in their breach notification statutes. The latest statutes including them are California's CCPA and Virginia's Consumer Data Protection Act ("VCDPA"), which will go into effect in 2023.

[75] For example, New Jersey's statute requires that: "Disclosure of a breach of security to a customer shall not be required under this section if the business or public entity establishes that misuse of the information is not reasonably possible. Any determination shall be documented in writing and retained for five years." N.J. Stat. Ann. § 56:8-163 (West 2022).

[76] The FTC recommends that firms consider hiring cyber forensics experts to help conduct such investigations, but firms are not generally required to do so. *See Data Breach Response: A Guide for Business*, Fed. Trade Comm'n (Feb. 2021),

federal HITECH Act, which has a four-factor analysis for determining risk of harm in the event of a breach of personal health information.[77]

Figure 4
Timeline of State Adoption of Risk-of-Harm Exception Provisions[78]



Several states also provide a "good faith" exception.[79] These exceptions generally apply when a breached firm has identified a breach that occurs because of something akin to an honest mistake by an employee who inadvertently accessed data.[80] These good-faith exceptions require that the breach occurred during a legitimate activity related to the employee's duties and did not result in a disclosure of data to another party.[81] These good-faith exceptions are meant to give firms leeway to avoid reporting breaches that are extremely low risk and pose little danger to consumers.[82]

https://www.ftc.gov/system/files/documents/plain-language/560a_data_breach_response_guide_for_business.pdf.

[77] *See No Harm, No Foul? Companies Need a Better Way to Assess Risk of Harm*, IDX (July 20, 2015), https://www.idx.us/knowledge-center/no-harm-no-foul-companies-need-a-better-way-to-assess-risk-of-harm [https://perma.cc/Q5GV-MMC2].

[78] States vary in whether they allow firms to not give notice if they conclude that there is little risk of harm to consumers.

[79] GitHub, *supra* note 61.

[80] For instance, Washington state's statute says: "Good faith acquisition of personal information by an employee or agent of the agency for the purposes of the agency is not a breach of the security of the system when the personal information is not used or subject to further unauthorized disclosure." WASH. REV. CODE ANN. § 42.56.590 (West 2022).

[81] GitHub, *supra* note 61.

[82] *Id.*

Most states provide encryption "safe harbors" in their breach notification statutes.[83] These carveouts exempt firms from notification when encrypted data was breached and the firm does not believe there is a way for the external attacker to decrypt the data.[84] The idea here again is that the risk of harm is low if the data is encrypted and therefore not worth alarming consumers about. Some states exempt most encrypted data, except in some critical categories like personal health information.[85] A key variation in these laws is whether states explicitly say that breached encrypted data that is lost with the encryption key must be disclosed.[86]

Though the original motivation for provisions like risk-of-harm analyses, good faith exceptions, and encryption exceptions was to avoid alarming consumers with unnecessary breach notifications, they may also play an important role in encouraging better data security. To take advantage of these exceptions to breach disclosure requirements, organizations need to take adequate precautions before a breach can occur. For instance, if an organization knows that it will be exempted from disclosure requirements if it adequately encrypts its data, it might invest in this technology in advance.[87]

III.
A Staggered Adoption Synthetic Control Approach to Data Breach Notification Laws

A. Data

The final dataset for this Article is a panel dataset containing identity theft report rates for each state in each year, which flags whether the state was exposed to treatment in that year (meaning, whether a particular provision was in effect) and covariate information about the

[83] *Id.*

[84] *Id.*

[85] *Id.*

[86] For example, Colorado's statute reads: "The breach of encrypted or otherwise secured personal information must be disclosed in accordance with this section if the confidential process, encryption key, or other means to decipher the secured information was also acquired in the security breach or was reasonably believed to have been acquired." Colo. Rev. Stat. § 6-1-716 (2022).

[87] Indeed, this was the intention of California state legislators when they included an encryption provision. In the floor analysis of the bill, legislators noted that "Furthermore, under current law, if the personal information that was stolen was encrypted, businesses and agencies are not required to provide notice. This provision, serves to encourage businesses and agencies who store personal information to adopt encryption so that if information is stolen, that information would be deemed less vulnerable to abuse. However, encryption is not clearly defined in statue." Assembly Comm. on Priv. and Consumer Protection, Committee Report for 2015 California Assembly Bill No. 964, 2015-2016 Reg. Sess., at 2 (2015).

demographics of that state in that year. Assembling this dataset requires some data collection and preprocessing.

The main data source for this Article is the Federal Trade Commission's Consumer Sentinel. Consumer Sentinel is the most comprehensive dataset of identity theft reports from American consumers and draws from consumer complaints to local and state law-enforcement agencies, federal consumer agencies such as the FTC and Consumer Financial Protection Bureau (CFPB), and private entities such as the Microsoft Cybercrime Center.[88]

The Consumer Sentinel dataset is not available to researchers and is currently only available to law enforcement agencies.[89] However, the FTC does publish annual reports detailing identity-theft report rates.[90] The reports are in PDF format, and the relevant printed data tables were extracted using a Python script.[91]

Aside from this outcome data, state-level covariates are collected for inclusion in the synthetic control models. Internet and broadband access data use estimates from Arizona State University's Center for Technology, Data and Society.[92] This dataset uses a combination of the Current Population Survey (CPS) and American Community Survey (ACS) between 1997 and 2014.[93] These data are extended with ACS

[88] *See Consumer Sentinel Network Fact Sheet*, FED. TRADE COMM'N, https://www.ftc.gov/system/files/attachments/consumer-sentinel-network/sentinel_fact_sheet1_508.pdf (last visited Sept. 27, 2023).

[89] *See Consumer Sentinel Network*, FED. TRADE COMM'N, https://www.ftc.gov/enforcement/consumer-sentinel-network [https://perma.cc/YH9T-NW4N] (last visited Sept. 27, 2023).

[90] *See Consumer Sentinel Network Reports*, FED. TRADE COMM'N, https://www.ftc.gov/enforcement/consumer-sentinel-network/reports [https://perma.cc/R237-WFCL] (last visited Sept. 27, 2023).

[91] Data for 2000-2001 and 2004-2005 are further processed because the published formats do not match other years. For 2000-2001, aggregated data are available along with state proportions, so the raw numbers must be manually calculated. Data for 2004-2005 includes tables for identity theft reports among individuals aged 50 and older, but the reports also provide the proportion of total reports this group comprises, so these values are transformed. Data from 2006 onward contain data for all states and age groups.

[92] Caroline Tolbert & Karen Mossberger, *U.S. Current Population Survey & American Community Survey Geographic Estimates of Internet Use, 1997-2014*, HARV. DATAVERSE (Dec. 2015), https://doi.org/10.7910/DVN/UKXPZX [https://perma.cc/K3HE-HCGS].

[93] Both of these datasets are routinely updated by the U.S. Census Bureau. ACS data usually comes in either one-year or five-year estimate series, and this Article primarily relies on the five-year estimates. The CPS is a monthly survey that mainly focuses on characteristics of the labor market. For more on the methodology underlying ACS, *see Research and Methodology*, U.S. CENSUS BUREAU, https://www.census.gov/

five-year estimates on broadband and Internet use from 2014-2020. These data are further supplemented with ACS data of state-level demographics pertaining to age, college-education level, race, and credit card usage.[94]

Data about state provisions are collected from Bloomberg and LexisNexis. Bloomberg provides a dataset summarizing data breach notification provisions across all fifty states, and the legislative text for every data breach notification law in the past twenty years is collected from LexisNexis.[95] Each provision under study (e.g., cause of action, application to encryption, harm analysis, etc.) is cross-referenced against the legislative text of each state's breach notification laws to record when the provision came into effect. This information constitutes the "treatment" in the causal inference framework discussed below. These effective dates for various provisions are merged into the Consumer Sentinel and Census Bureau data to create the full panel dataset.

B. Causal Inference Framework Applied to Identity-Theft Report Rates

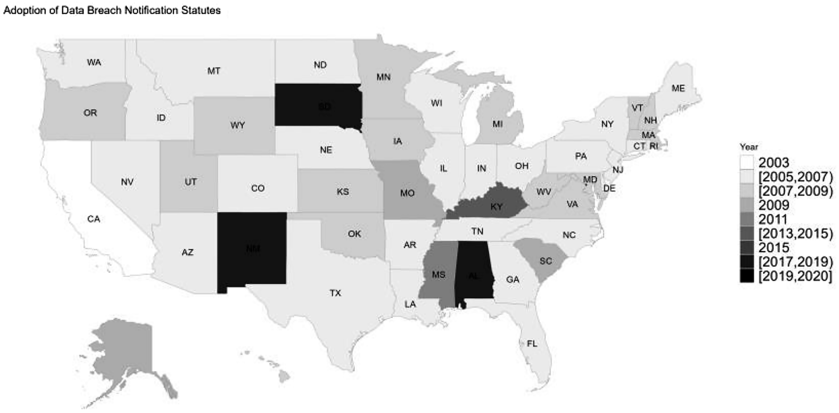
One challenge with any quantitative social science study is approaching the question of establishing causality. This question is especially relevant when dealing with *observational* data, meaning data that was not created by the researcher. One approach to quantitative analysis of this sort of data might be a regression-based approach that fits a statistical model of some outcome of interest with some covariates included as controls. However, these kinds of observational studies are

programs-surveys/acs/methodology.html [https://perma.cc/5VYN-7UJR] (last updated Aug. 11, 2023), and for CPS, see Labor Force Statistics From the Current Population Survey, U.S. Bureau of Lab. & Stat., https://www.bls.gov/cps/ [https://perma.cc/6DLS-EH39] (last visited Sept. 27, 2023).

[94] These data were collected and processed using the tidycensus package in R. Kyle Walker & Matt Herman, *Package 'Tidycensus'*, CRAN (June 3, 2022), https://cran.r-project.org/web/packages/tidycensus/tidycensus.pdf.

[95] See Figure 5 for a visual timeline of when each state adopted its first data breach notification law.

Figure 5
Timeline of State Adoption of Data Breach Notification Laws[96]



subject to numerous problems pertaining to replicability and validity of causal estimates.[97] In recent years, empirical legal studies scholars have called attention to the ways that design-based approaches to causal inference can provide more credible causal estimates and the dangers of how model-based observational studies can lead to erroneous conclusions.[98]

My analysis applies the Neyman-Rubin potential outcomes causal inference framework.[99] The potential-outcomes framework frames the causal effect of an intervention on a unit as the difference between the outcome of that unit under treatment and the outcome of that unit under control.[100] Because, in actuality, we can only observe that unit under either treatment or control,[101] but not both, we use methods for estimating the average treatment effect (ATE) across a population.[102]

[96] Between 2003 and 2020, every state and DC adopted a data breach notification law.

[97] *See* Daniel E. Ho & Donald B. Rubin, *Credible Causal Inference for Empirical Legal Studies*, 7 ANN. REV. L. & SOC. SCI. 17 (2011).

[98] *Id.* at 19.

[99] *See* Jasjeet S. Sekhon, *The Neyman-Rubin Model of Causal Inference and Estimation Via Matching Methods*, *in* THE OXFORD HANDBOOK OF POLITICAL METHODOLOGY 271 (Janet M. Box-Steffensmeier et al., eds., 2008).

[100] *Id.*

[101] Analogizing to experiments, "treatment" refers to units that get the intervention, and "control" refers to units that do not receive the intervention. In this case, "treatment" would be states that pass a data breach law, and "control" would be states that do not pass a data breach law.

[102] Formally, the ATE can be defined as:

$$\frac{1}{N}\sum_{i}(y_1(i)-y_0(i))$$

where y_1(i) is the outcome for the unit under treatment, and where y_0(i) is the outcome for the unit under control. The inability to observe any given unit under both treatment and control is known as the "fundamental problem of causal inference."

Another key assumption of the Neyman-Rubin framework is the Stable Unit Treatment Value Assumption (SUTVA), or non-interference between units.[103] Simply put, this assumption states that exposing one unit to treatment does not affect the outcomes in other units. Within the context of data breach notification, this assumption needs to be carefully evaluated. There are several potential ways state provisions can spill over into other states. Provisions that govern the *content* of notices might suffer from SUTVA problems, as firms may not adjust notices for different states.[104] A private cause of action may prompt multi-state litigation, notification to state regulators and credit reporting agencies may benefit consumers in other states, and mechanisms for encouraging better data security through encryption can make national firms comply everywhere. However, there are also several ways the relationship between data breaches and identity theft may be highly localized, as is the case with health data, breaches of small businesses, etc. More granular identity-theft report data could help untangle some of these local versus national effects.[105,106]

Applying this framework to the effect of data breach notification laws on identity theft, the question is *what would a state's identity-theft report rates be if it did not adopt a breach notification law?* For example, if we were interested in estimating the effect of California's original data breach notification law in 2003, we would frame the causal estimand as the difference between California's identity-theft report rates under treatment (having a data breach notification law) and California's identity-theft report rates under control (not having a data breach notification law). We can formulate this quantity as the *Average Treatment Effect on the Treated (ATT)*. Extending this idea to the staggered setting, this Article looks at the ATT of various breach notification provisions on state identity-theft report rates per 100,000 population.

[103] Ho & Rubin, *supra* note 97, at 21.

[104] Jens Frankenreiter characterizes these as "cost-based California Effects," where firms comply with stricter requirements in one jurisdiction everywhere they do business. He does not find evidence for these effects in privacy policies among U.S. firms that comply with the GDPR in Europe, but this effect should be evaluated for breach notifications. Jens Frankenreiter, *The Missing 'California Effect' in Data Privacy Law*, 39 YALE J. ON REG. (forthcoming 2022) (manuscript at 7), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3883728 [https://perma.cc/2NNH-K6GZ].

[105] Also see Appendix C for one approach for assessing whether a spillover occurred in the case of Texas's 2009 encryption provision.

[106] In a recent working paper, Lior Strahilevitz and Lisa Yao Liu examine cash substitution and deferred consumptions effects after localized breach notifications. They show that after a local data breach announcement, consumers shift from credit card transactions to cash purchases and defer certain purchases entirely in the short run. *See* Lior Strahilevitz & Lisa Yao Liu, *Cash Substitution and Deferred Consumption as Data Breach Harms* (U. Chi. Coase-Sandor Inst. for L. & Econ., Research Paper No. 963, 2022).

C. Staggered Synthetic Control

The main estimation method is the "staggered synthetic control." Synthetic control and its variants have become increasingly popular in the policy-evaluation literature.[107] Empirical legal studies scholars are likely familiar with regression-based methods for working with panel data.[108] Methods such as synthetic control provide useful data-driven alternatives to these regression-based methods.[109] At a high level, synthetic control estimates the effect of a policy by constructing a simulated version of a unit (state, city, country, etc.) and comparing this simulated unit to real-world outcomes.

Synthetic control is an extension of another popular method for evaluating policy effects: the difference-in-differences (DiD) method.[110] DiD analysis compares two units, one treated and one control, before and after a policy intervention.[111] For both the treated and control units, the analyst subtracts the pre-intervention outcome from the post-intervention outcome.[112] These differences are then subtracted from each other to estimate the Average Treatment Effect on the Treated.[113] Table 1 illustrates a simple example of how to calculate this effect assuming one pre-intervention period and one post-intervention period. One of the main assumptions for proceeding with a DiD analysis is the "parallel trends" assumption, which requires that the treated and control units should have similar pre-treatment trends.[114] Figure 6 illustrates the parallel trends concept, and Figure 7 shows an example where the parallel trends assumption may not be plausible.[115]

[107] *See, e.g.*, Ben-Michael et al., *supra* note 19, at 351; Athey & Imbens, *supra* note 18.

[108] For a comparison of regression and synthetic control methods in economic and policy research, see Orkideh Gharehgozli, *An Empirical Comparison Between a Regression Framework and the Synthetic Control Method*, 81 Q. Rev. Econ. & Fin. 70 (2021).

[109] See John J. Donohue et al., *Right-to-Carry Laws and Violent Crime: A Comprehensive Assessment Using Panel Data and a State-Level Synthetic Control Analysis*, 16 J. Empirical Legal Stud. 198, 199 (2019) for an example of synthetic control in empirical legal studies where the authors study the impact of right-to-carry laws on violent crime.

[110] For the original DiD study, see David Card & Alan B. Krueger, *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania*, 84 Am. Econ. Rev. 772 (1994).

[111] *Id.* at 778.

[112] *Id.*

[113] *Id.*

[114] *Id.*

[115] Note that this assumption does not require that the treated and control units have the same outcomes, just that the trends are moving in similar directions.

Table 1
Example of a 2 × 2 Difference-in-Differences Analysis[116]

	Colorado	New Mexico	Difference
2006	92.5	83	9.5
2007	89	87.5	1.5
Difference	3.5	-4.5	-8

Figure 6
Example Parallel Trends[117]

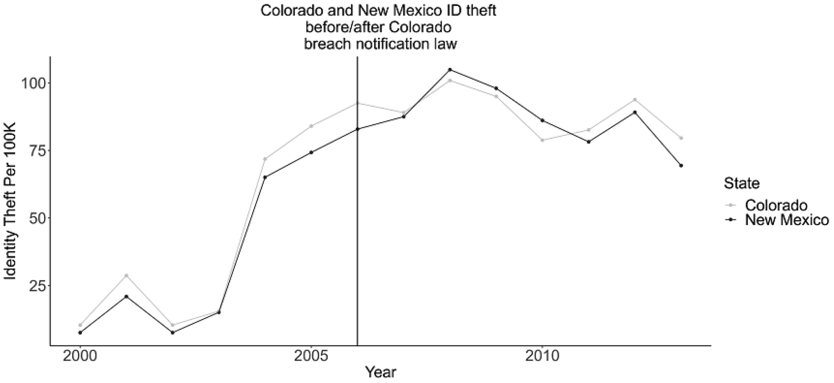
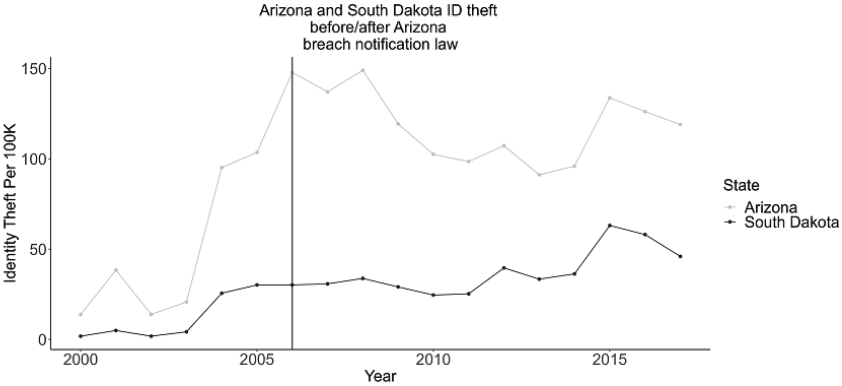


Figure 7
Example Non-Parallel Trends[118]



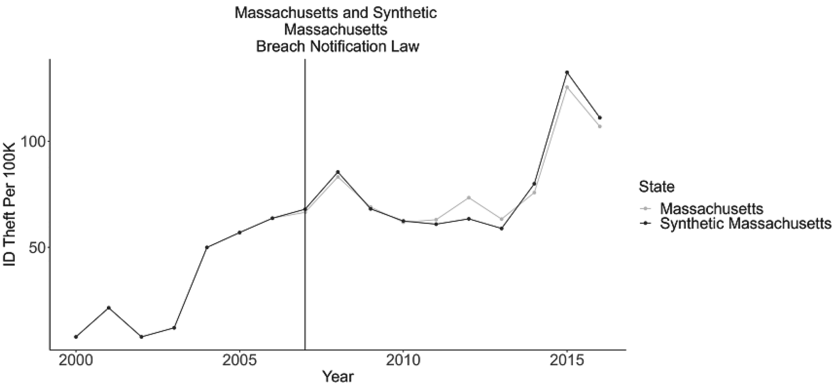
[116] In this example, Colorado adopted a breach notification law, and New Mexico is used as a control. The outcomes from each state are subtracted across time period 1 (2006) and time period 2 (2007), then these differences are subtracted to get the estimated treatment effect of eight fewer identity theft reports per 100,000 population.

[117] An example of parallel trends between two states, one treated (Colorado) and one control (New Mexico).

[118] An example where parallel trends is less plausible, using Arizona (treated) and South Dakota (control).

In the absence of parallel trends, synthetic control can be a useful alternative. In the simplest case, synthetic control estimates the ATT by constructing a "synthetic" version of a unit that experienced a policy change.[119] The synthetic control unit is created by constructing a weighted average of the non-treated units that closely matches the treated unit's pre-treatment outcomes.[120] If the pre-treatment synthetic control closely matches the pre-treatment observed unit, we can assume that it acts as a good approximation for what the unit's outcomes would have been but for exposure to treatment.[121] The ATT (here, the increase or reduction in identity theft report rates per 100,000 population) is then calculated by averaging the differences between the post-treatment observed outcomes and post-treatment synthetic outcomes.[122] See Figure 8 for an example synthetic control.[123]

Figure 8
Example Synthetic Control[124]



[119] Alberto Abadie & Javier Gardeazabal, *The Economic Costs of Conflict: A Case Study of the Basque Country*, 93 AM. ECON. REV. 113 (2003), originally proposed the method to estimate the effect of terrorism in the Basque Country on GDP. The method was extended and popularized by Alberto Abadie, Alexis Diamond & Jens Hainmueller, *supra* note 17, where the authors demonstrated the method on California's tobacco tax.

[120] *See* Abadie et al., *supra* note 17, at 494.

[121] Abadie & Gardeazabal, *supra* note 119, at 116.

[122] *Id.* at 118.

[123] Formally using the Abadie/Diamond/Hainmueller notation, imagine that there are J + 1 regions of interest (states in this case). Y^{N}_{it} is the outcome for region i at time t for units i = [1 : J + 1] and time period t = [1 : T]. Let T_0 be the number of preintervention periods, with $1 \leq T_0 < T$. Let Y^{I}_{it} be the outcome if unit i at time t was exposed to the intervention. The ATT for one treatment regime could be expressed as $ATT = Y^{I}_{1t} - Y^{N}_{1t}$, where i = 1 refers to the region of interest.

[124] Figure 8 presents an example of a synthetic control for Massachusetts's data breach notification law. The synthetic version of Massachusetts should match the observed Massachusetts closely in the pre-treatment period.

The basic setup for synthetic control and DiD estimates assume one treated unit at one time period;[125] however, we might be interested in situations where different units adopt the same policy at different times.

Several extensions have been proposed to estimate treatment effects in these "staggered adoption" scenarios.[126] One way to approach staggered adoption is to fit separate synthetic controls for each treated unit.[127] One drawback of this approach is that it requires that each unit achieves good pre-treatment fit, meaning the synthetic unit closely tracks the treated one in the pre-treatment period *for every unit*.[128] This may not always be possible, and it also involves a tradeoff with minimizing global pre-treatment imbalance across all treated units.[129] More simply, we are concerned with making sure each synthetic control for each state closely fits its corresponding state, but also that the synthetic controls have good fits on average. Achieving both is difficult. To address this limitation, this Article uses the new multisynth method proposed by Ben-Michael, Feller, and Rothstein.[130] Multisynth finds the optimal tradeoff between individual pre-treatment imbalance and global pre-treatment imbalance to fit a staggered synthetic control model.[131] Figure 9 shows an example staggered adoption synthetic control for states' first data breach notification laws.[132] Specifically, we can see the individual fits for each state, and the partially pooled synthetic average across non-treated states. This strategy allows the estimate of

[125] Abadie et al., *supra* note 17, at 494.

[126] While this Article focuses on staggered synthetic control, there are also other options. Two-way fixed effects (TWFE) estimators are a standard DiD extension to staggered adoption scenarios. However, these estimators suffer from several drawbacks that can bias results, as explained by Andrew C. Baker et al., *How Much Should We Trust Staggered Difference-in-Differences Estimates?*, 144 J. Fin. Econ. 370 (2022). The synthetic difference-in-differences method, *see* Dmitry Arkhangelsky et al., *Synthetic Difference-in-Differences*, 111 Am. Econ. Rev. 4088 (2021), and Matrix Completion methods, *see* Susan Athey et al., *Matrix Completion Methods for Causal Panel Data*, 116 J. Am. Stat. Ass'n 1716 (2021), have also been proposed as alternatives.

[127] *See* Donohue et al., *supra* note 109, at 200 (using this approach to estimate the effect of right-to-carry gun laws on shootings).

[128] Ben-Michael et al., *supra* note 19, at 352.

[129] *Id.* at 353.

[130] Ben-Michael et al., *supra* note 19.

[131] It does so by tuning a hyperparameter, υ, that controls the amount of pooling of units. At $\upsilon = 0$, each treated unit is fit separately with pre- and post-treatment imbalance being minimized per unit. At $\upsilon = 1$, all treated units are pooled together and pre- and post-treatment imbalance is minimized for the average of all treated units. Partially pooled synthetic control utilizes the fact that the tradeoff between these two imbalances is often convex and therefore searches for the balance between them that minimizes the balance possibility frontier.

[132] This analysis excludes late adopting states such as Alabama, South Dakota, and New Mexico because once all fifty states adopted a data breach notification law, there were no more states to use in the donor pool to construct synthetic controls.

both the overall Average Treatment Effect on the Treated (ATT) for all treated states, as well as individual state-level estimates for increases and decreases in identity theft rates, as seen in Figure 10. This type of approach is particularly advantageous when states are heterogenous in the estimated effects of adopting data breach notification laws – some saw increases in identity theft reporting and some saw decreases.

Figure 9
Example Staggered Synthetic Control[133]

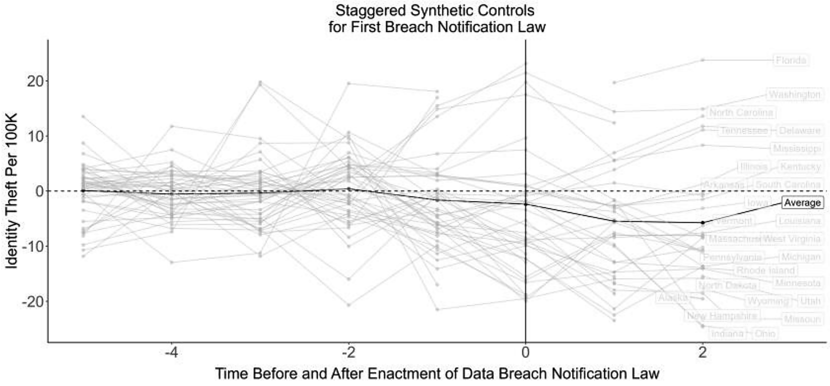
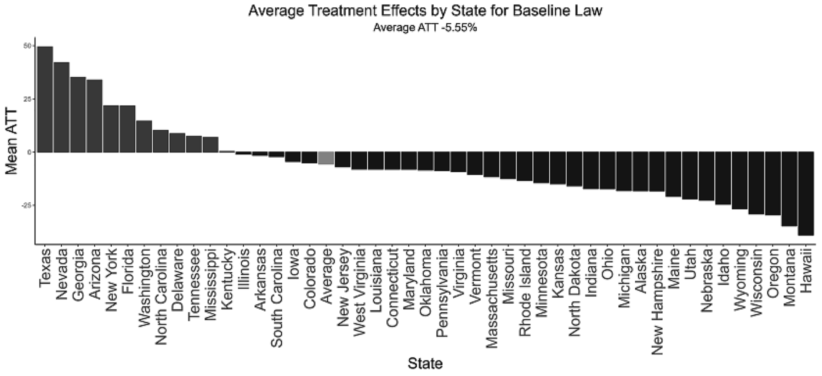


Figure 10
Treatment Effects by State for First Breach Notification Law[134]



[133] An example of a staggered synthetic control using the multisynth method for states' first data breach notification laws. In this plot, only the difference between the synthetic control and the observed unit is visualized (rather than both units actual outcomes as in the previous figure). Ideally, the value should be as close to 0 as possible in the pre-treatment period (before the vertical line), with the trend after intervention being the estimated treatment effect.

[134] The Average Treatment Effect on the Treated (ATT) by state. The positive bars indicate increases in estimated identity theft reports, negative bars indicate decreases in

The multisynth method is the basis for the results that follow. There are a few important caveats when interpreting the results. Credible estimates require sufficient pre-treatment data to construct the synthetic controls.[135] Synthetic controls can also be biased if there are relatively few post-treatment donor[136] units to draw from.[137] Because the multisynth method is trading off between two different types of imbalances (local and global), excellent pre-treatment fit may not be possible in all cases.

To fit the synthetic controls, it is common practice to include a vector of covariates that improve the prediction of the observed unit in the pre-treatment period.[138] This process is similar to including covariates in a regression model to control for them. Rather than interpret the correlation between these covariates and the outcome of interest, synthetic control uses them to improve the prediction of the pre-treatment outcome.[139] For each state in each year the model includes the following:

- Treatment indicator
- Percent of state with Internet access (household level)
- Percent of state with broadband access (household level)
- Percent of state age 60 and older
- Percent of state that uses a general-purpose credit card (household level)
- Percent of state that is White
- Percent of state that is Black or African American
- Percent of state that is Asian
- Percent of state that is Hispanic or Latino (of any race)
- Percent of state, 25 years or older, with a Bachelor's degree or higher

D. Results

This section discusses the principal results, organized by the typology introduced. The main outcomes of interest are (1) requirement to notify a state regulator, (2) requirement to notify a credit reporting agency, (3) imposing a specific timeline for when a breach must be reported, (4) explicit grant of a private right of action to consumers, (5) no exception to breach reporting duties for a firm that concludes a breach has a low risk of harm for consumers, and (6) applying breach notification requirements

estimated identity theft reports, and the bar labeled "Average" indicates the average. In this case the average is a decrease of 5.55% or 4.50 reports per 100,000 people.

[135] Ben-Michael et al., *supra* note 19, at 356.

[136] "Donor" refers to non-treated states that are weighted to create the synthetic state.

[137] Ben-Michael et al., *supra* note 19, at 356.

[138] *Id.* at 371.

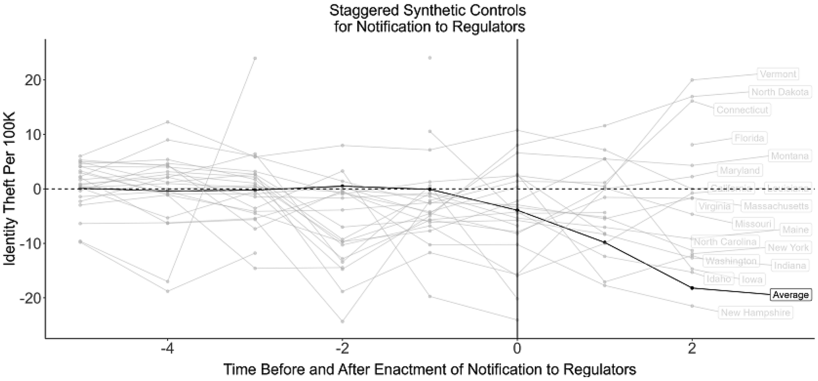
[139] *Id.*

to encrypted data. These outcomes are presented alongside the baseline breach notification requirement presented in the previous section.

1. Consumer Mitigation of Harms

All data breach notification laws require notification to data subjects, but in some states, disclosures must be made to additional parties as well. Some states require notification to a state Attorney General or similar regulator (usually the state consumer protection agency).[140] Others require notification to consumer credit reporting agencies.[141] Figures 11 and 12 show the results for requiring notification to state Attorneys General.[142] The average ATT across all treated units is a reduction of 10.67 per 100,000 (9.58%) identity theft reports.

Figure 11
ATT for Notification to State Regulator[143]



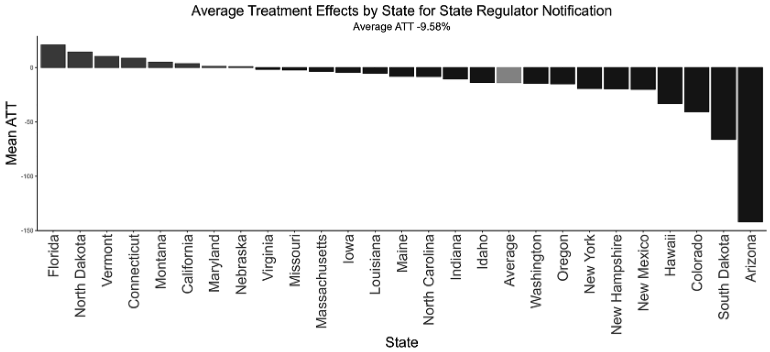
[140] GitHub, *supra* note 61.

[141] *Id.*

[142] These results exclude states that adopted AG notification after 2017 because there is insufficient post-intervention data. These states are Texas (2020), Alabama (2018), Delaware (2018), Arkansas (2019), Illinois (2018), and Rhode Island (2019).

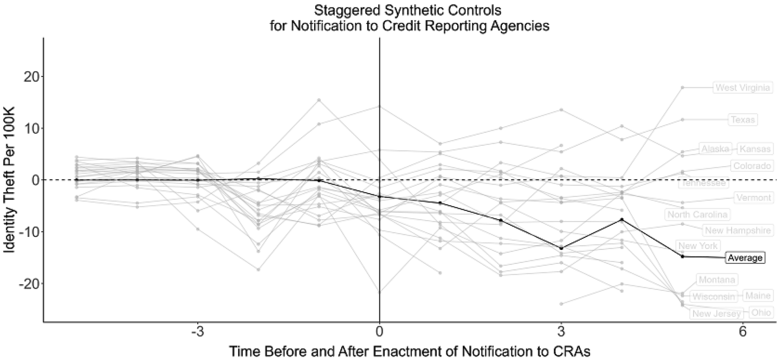
[143] Synthetic control fit for required notification to state regulator. The black line indicates the average synthetic control and grey lines indicate the fits for each individual state. The black line closely matches 0 before treatment (enactment of the law), indicating a strong fit. The slope downward after treatment indicates an estimated decrease in identity theft reports.

Figure 12
Treatment Effects by State for Notification to State Regulator[144]



Figures 13 and 14 tells a similar story about notification requirements to credit reporting agencies. Here, the average ATT is a reduction of 10.43 per 100,000 (7.71%) identity theft reports. In short, notification to parties other than consumers deters identity theft in states that require such disclosures. While the synthetic control analysis cannot disentangle the exact mechanisms for each of these requirements, we can conclude that disclosure to intermediaries with more sophistication and resources than the average consumer makes a difference.

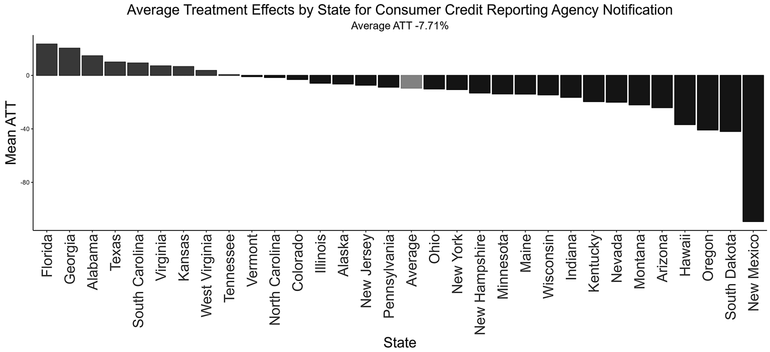
Figure 13
ATT for Notification to Credit Reporting Agencies[145]



[144] The Average Treatment Effect on the Treated (ATT) by state for required notification to a regulator. The positive bars indicate increases in estimated identity theft reports, negative bars indicate decreases in estimated identity theft reports, and the bar labeled "Average" indicates the average. In this case the average is a decrease of 9.58% or 10.67 reports per 100,000 population.

[145] Synthetic control fit for required notification to a credit reporting agency. The black line indicates the average synthetic control and grey lines indicate the fits for each individual state. The black line closely matches 0 before treatment (enactment of the law), though with a poor fit in the last time period. The slope downward after treatment indicates an estimated decrease in identity theft reports.

Figure 14
Treatment Effects by State for Notification to Credit Reporting Agency[146]



The final requirement enabling consumer mitigation of harms is the specification of a time limit for when to notify consumers, state regulators, and credit reporting agencies. Figures 15 and 16 illustrate the effects of a time limit for notifying consumers. Time limits for notification vary, with most being forty-five, sixty, or ninety days;[147] this analysis simply looks at whether there is any time limit requirement, instead of the "without unreasonable delay" standard that tends to be used as the default.[148] This provision also has a sizable effect, decreasing estimated identity theft reports by about 9.32%.

[146] The Average Treatment Effect on the Treated (ATT) by state for required notification to a credit reporting agency. The positive bars indicate increases in estimated identity theft reports, negative bars indicate decreases in estimated identity theft reports, and the bar labeled "Average" indicates the average. In this case the average is a decrease of 7.71% or 10.43 reports per 100,000 population.

[147] GitHub, *supra* note 61.

[148] *Id.*

Figure 15
Treatment Effects by State for Timeline Requirement[149]

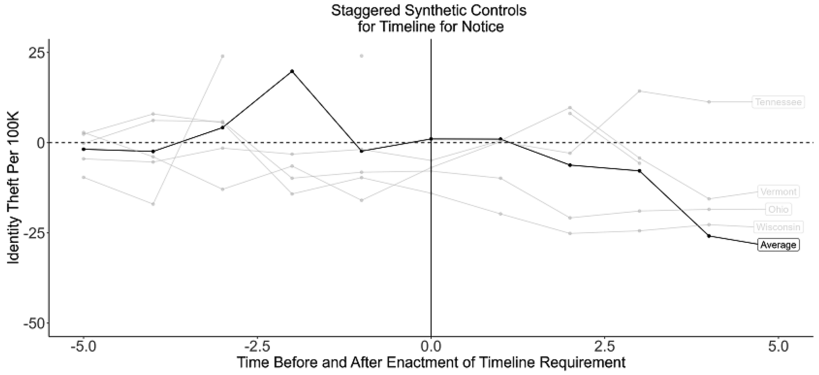
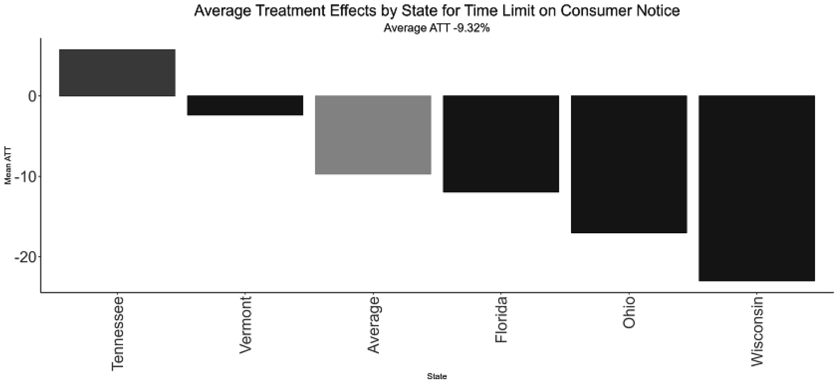


Figure 16
Treatment Effects by State for Timeline Requirement[150]



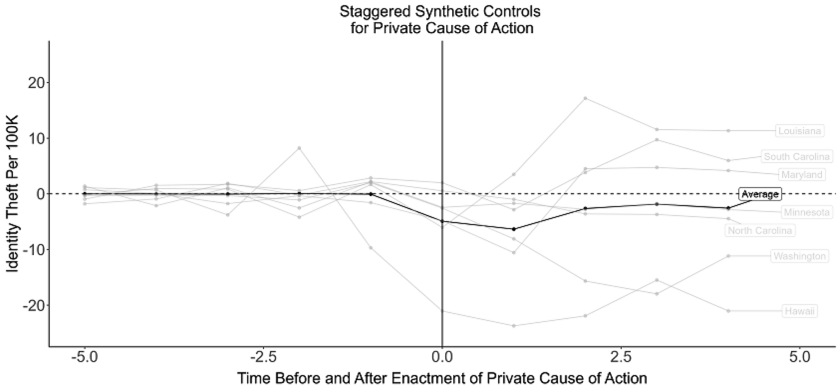
[149] Synthetic control fit for mandating a time limit for when to notify consumers. The black line indicates the average synthetic control and grey lines indicate the fits for each individual state. The black line closely matches 0 before treatment (enactment of the law), indicating a strong fit. The slope downward after treatment indicates an estimated decrease in identity theft reports.

[150] The Average Treatment Effect on the Treated (ATT) by state for mandating a time limit for when to notify consumers. The positive bars indicate increases in estimated identity theft reports, negative bars indicate decreases in estimated identity theft reports, and the bar labeled "Average" indicates the average. In this case the average is a decrease of 9.32% or 7.88 reports per 100,000 population.

2. *Encouraging Better Data Security*

Do provisions that permit more private or public enforcement have an effect on identity theft report rates? A handful of states include provisions allowing consumers a private cause of action in the event of a data breach.[151] Figures 17 and 18 show the results of this staggered synthetic control for states with such provisions.[152] The average ATT across all treated units is a reduction of 3.66 identity theft reports per 100,000 population (3.40%) following adoption. There is variation in how this provision affected each state's estimate, with Maryland, Louisiana, and South Carolina experiencing estimated increases.

Figure 17
ATT for Private Cause of Action[153]

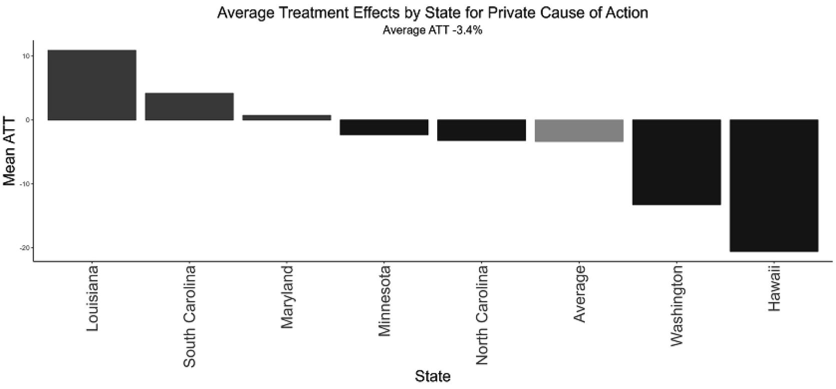


[151] GitHub, *supra* note 61.

[152] These estimates exclude states that adopted private cause of action provisions after 2017 because there is insufficient post-intervention data. These states are Tennessee (2017), New Hampshire (2020), California (2023), and Virginia (2023).

[153] Synthetic control fit for providing a private cause of action. The black line indicates the average synthetic control and grey lines indicate the fits for each individual state. The black line closely matches 0 before treatment (enactment of the law), indicating a strong fit. The slope downward after treatment indicates an estimated decrease in identity theft reports, though the effect lessens over time.

Figure 18[154]
Treatment Effects by State for Private Cause of Action



State data breach notification laws vary in how much discretion they permit firms in deciding how to respond to a data breach. Nearly every state provides a "good faith" exception that allows organizations to not make a disclosure when the breach was a result of a good faith mistake.[155] Other variations include whether organizations can be exempted from disclosure if they conclude there is no risk of harm to consumers or if data was encrypted.[156] They may also differ in requirements for the timing of the notification.

Figures 19 and 20 show the staggered synthetic control results for states that do *not* allow organizations to exempt themselves from disclosure after analyzing the risk of harm and concluding it is minimal for consumers.[157] Theoretically, we might expect that in states without a harm-analysis exception, organizations would be unable to hide damaging breaches by cooking the books with a harm analysis.

The results here are counter-intuitive and suggest that average identity theft report rates *increased* by about 5.15 (4.91%) identity-theft reports per 100,000 population. Some states did see a decline in identity theft report rates, but overall, there is little evidence that not providing exceptions decreased identity theft.[158]

[154] The Average Treatment Effect on the Treated (ATT) by state for providing a private cause of action. The positive bars indicate increases in estimated identity theft reports, negative bars indicate decreases in estimated identity theft reports, and the bar labeled "Average" indicates the average. In this case the average is a decrease of 3.4% or 3.66 reports per 100,000 population.

[155] GitHub, *supra* note 61.

[156] *Id.*

[157] California is excluded from this analysis because of insufficient pre-treatment data.

[158] It is possible that the lack of exceptions increased the number of breach notifications and increased consumer awareness of the likelihood of identity theft and thus

Figure 19
ATT for No Harm Analysis[159]

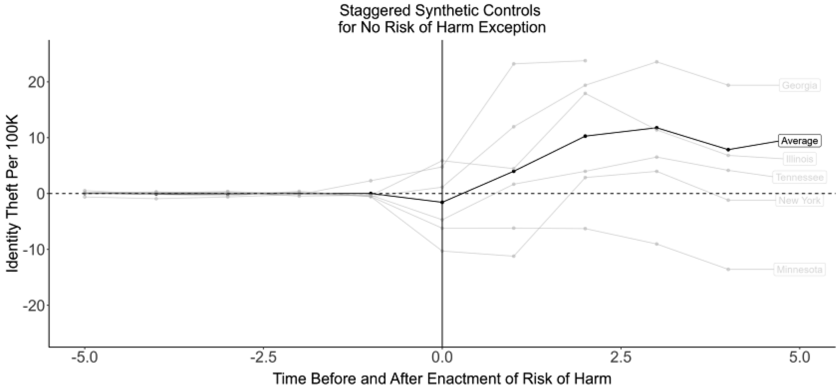
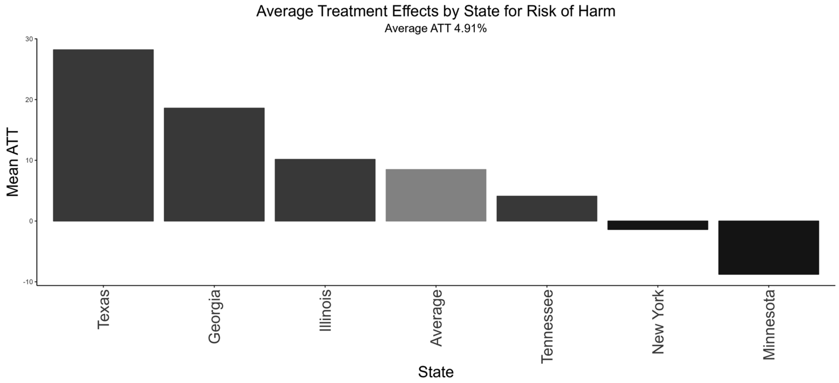


Figure 20
Treatment Effects by State for Harm Analysis[160]



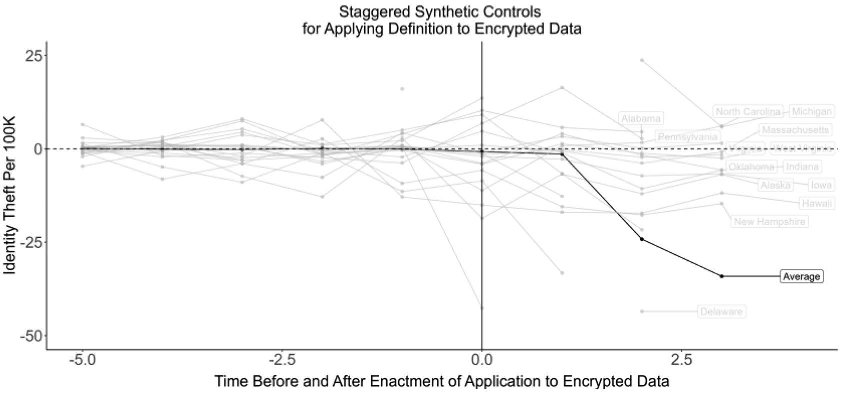
drove increased *reports* but not increased *incidents*. This explanation requires an assumption that notification increases reporting, but there is not strong evidence for this idea either in the baseline results for states' first breach notification laws in this study, or elsewhere in the literature.

[159] Synthetic control fit for providing no risk of harm exception. The black line indicates the average synthetic control and grey lines indicate the fits for each individual state. The black line closely matches 0 before treatment (enactment of the law), though with a weaker fit in the last time period before enactment. The slope upward after treatment indicates an estimated increase in identity theft reports.

[160] The Average Treatment Effect on the Treated (ATT) by state for providing a private cause of action. The positive bars indicate increases in estimated identity theft reports, negative bars indicate decreases in estimated identity theft reports, and the bar labeled "Average" indicates the average. In this case the average is an increase of 4.91% or 5.15 reports per 100,000 population.

Closing the encryption loophole tells a much clearer story. Most breach notification statutes provide a carveout for encrypted data.[161] However, many of these carveouts do not say what happens if encrypted data is stolen along with the encryption key. State laws that close this loophole usually contain language saying that breaches compromising encrypted data with the encryption key or other means of recovering the raw data must be disclosed.[162] Figures 21 and 22 show that most state laws with this kind of language see a decrease in identity-theft report rates. On average, identity-theft report rates decreased by 15.65 (11.46%) reports per 100,000 population, with some states seeing even sharper decreases.

Figure 21
ATT for Applied to Encrypted Data[163]

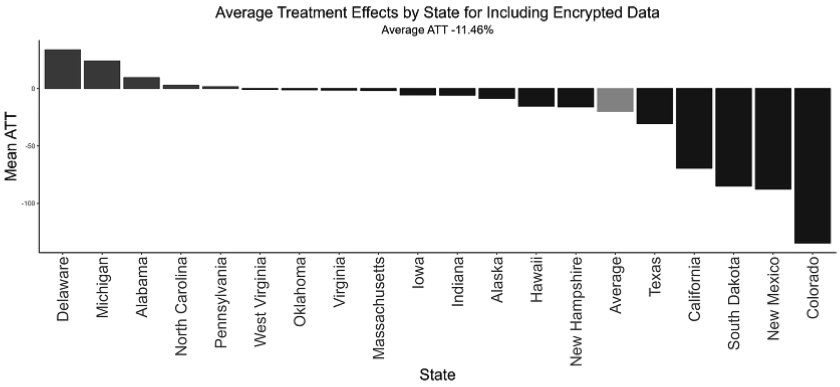


[161] GitHub, *supra* note 61.

[162] *Id.*

[163] Synthetic control fit for including encrypted data in the definition of covered data. The black line indicates the average synthetic control and grey lines indicate the fits for each individual state. The black line closely matches 0 before treatment (enactment of the law), indicating a strong fit. The slope downward after treatment indicates an estimated decrease in identity theft reports.

Figure 22
Treatment Effects by State for Applied to Encrypted Data[164]



E. Summary, Analysis, and Limitations

Table 2 summarizes the main results. The L2-imbalance columns summarize the level of global and local imbalance between pre-treatment synthetic controls and the corresponding observed units, and the average-ATT column is the estimated average increase or reduction in identity theft reports across the post-treatment time periods.[165]

[164] The Average Treatment Effect on the Treated (ATT) by state for including encrypted data in the definition of covered data. The positive bars indicate increases in estimated identity theft reports, negative bars indicate decreases in estimated identity theft reports, and the bar labeled "Average" indicates the average. In this case the average is a decrease of 11.46% or 15.65 reports per 100,000 population.

[165] L2 is also known as the Euclidean norm. In linear algebra, a norm describes the distance of a vector from its origin, and a Euclidean norm is an extension of the Pythagorean theorem to describe this distance. For the purposes of this applied context, it suffices to know that smaller values indicate shorter distances between the synthetic and observed values of a treated unit, and vice versa for larger ones.

Statutory Provision	Global L2 Imbalance	Individual L2 Imbalance	Average ATT	Percent Change	Dollars Saved[166]
Baseline Breach Notification	.113	.337	-4.5	-5.55%	$102,465,000
Attorney General Notification	.073	.420	-10.67	-9.58%	$242,955,900
Credit Reporting Agency Notification	.021	.562	-10.43	-7.71%	$237,491,100
Private Cause of Action	.026	.250	-3.66	-3.40%	$83,338,200
Applies to Encrypted Data	0.03	.247	-15.65	-11.46%	$356,350,500
No Internal Harm Analysis Exemption	.011	.060	5.15	+4.91%	**-$344,965,500**
Time Limit for Notification	.459	.448	-7.88	-9.32%	$179,427,600

Table 2
Summary of Statutory Provisions with L2 Imbalances and Average ATT

Applying breach notification to encrypted data and requiring notification to state regulators have the largest effects in decreasing identity-theft report rates. Each provision reduces identity theft reports by about 9.5 to 11%. Notification to consumer credit reporting agencies has an effect of decreasing identity-theft reports by about 7.7%, whereas the effects of a private cause of action is more muted with a reduction of around 3.4% respectively. Compelling companies to disclose even when they conclude there is a low likelihood of harm actually *increases* identity theft reports by almost 5%.

There are a number of limitations that are important be mindful of. The first are those pertaining to the staggered synthetic control method itself. Staggered synthetic control is just one option for estimating the

[166] The "dollars saved" estimate comes from multiplying the average reduction or increase in identity theft rates by the average financial loss for an identity theft case ($1,300) plus the loss in wage-hours ($28 U.S. average wage per hour multiplied by 200 hours to resolve a case). The lack of harm analysis provision dollar figure is bolded to indicate the estimated cost of the estimated increase in identity theft report rates.

effects of staggered policy adoption. Others such as two-way fixed effects, synthetic difference-in-differences, and matrix-completion methods could be appropriate as well. Staggered-adoption synthetic control also has an inherent tradeoff between global and local pre-treatment imbalance between the observed and synthetic units.[167] The algorithm searches for the optimal balance between these two in each case, but there are situations where the individual fit is poor. Apart from the staggered-adoption extension, it is also helpful to be mindful of the core assumptions of synthetic control. Excellent match between the pre-treatment observed unit and synthetic unit is absolutely necessary for inference.[168] Furthermore, many of the theoretical guarantees of synthetic control are true asymptotically. Specifically, the theoretical properties of synthetic control are shown to be true as the number of pre-treatment periods approaches infinity.[169]

The second major set of limitations pertains to the quality of the dataset. The theoretical guarantees are true with many pre-treatment periods, but the FTC data is a relatively short panel data set, only covering about twenty years. This means there are relatively few pre-treatment periods to work with.[170] Another problem with the FTC data is that it relies on aggregated reports from participating federal, state, and local law enforcement and other agencies.[171] Identity-theft crime reports suffer from the same problem as other types of crime—reported crime is only a fraction of all crime.[172] Statistical corrections to these types of systematic biases are often criticized for potentially making the problem worse, and this seems especially risky with a crime like identity

[167] Ben-Michael et al., *supra* note 19, at 374.

[168] *Id.* at 358.

[169] Eli Ben-Michael et al., *The Augmented Synthetic Control Method* 1 (Nat'l Bureau of Econ. Rsch., Working Paper No. 28885, 2021), https://www.nber.org/system/files/working_papers/w28885/w28885.pdf [https://perma.cc/HW4X-RUMK].

[170] The estimates throughout this Article exclude any states with fewer than five pre-treatment periods, such as California's 2003 law, *supra* note 44.

[171] Selection of agencies into participation can bias the overall dataset. For instance, twenty-five states' attorneys general and consumer protection agencies directly participate, making it likely that the database gets more reports from those states. The types of consumers who report to certain federal agencies, such as the CFPB, likely differ from consumers who report to local law enforcement agencies, further overrepresenting certain kinds of consumers and underrepresenting others. Police departments also typically underreport certain kinds of crimes, such as unlawful use of police force and hate crimes, and it is unknown how much of a problem this is with regards to identity theft reports.

[172] For example, the federal government uses both the Uniform Crime Reports system to track reported crime and the National Crime Victimization Survey (NCVS) to track reported and unreported crime.

theft where there are unknowns about its extent and who does or does not report it.[173]

It is also worth discussing the ways that these results support and contradict previous empirical work in this space. Romanosky, Telang, and Acquisti's original exploration of baseline data breach notification laws suggested about a 6% reduction in identity theft reports.[174] This Article suggests a similar effect. That being said, the effects of breach notification laws could have shrunk over time, and late adopters may have enjoyed smaller reductions than early adopters. Some late adopters like Texas and Florida saw *increases* in identity theft reporting after passing their laws, whereas other late adopters like Arkansas saw small reductions. This diminished effect over time may be especially pronounced if there are spillovers between states. Companies that operate in multiple states may adapt their national standards to early adopters, and thus are unaffected when late adopters enact those same standards.

The evidence presented in this Article supports previous literature suggesting that reputational market mechanisms are limited at best. Goel and Shawky study whether breach announcements affect firms stock performances and find only short-term effects.[175] Mitts and Talley similarly find evidence of insider trading prior to the disclosure of a cybersecurity incident, which could actually subsidize breaches.[176] These findings are compatible with the story that mechanisms that presuppose the market will encourage *a priori* data security investments are ineffective, whereas mechanisms that encourage better data security might be more effective.

The evidence presented does not precisely align with the two studies that are most similar to this one—the Sullivan and Maniff[177] and Greenwood and Vaaler[178] studies. Sullivan and Maniff also look at various state data breach notification provisions to assess their effectiveness. They used a regression method and treated the presence of provisions as covariates for the regression model, with the outcome

[173] *See* David Buil-Gil et al., *The Accuracy of Crime Statistics: Assessing the Impact of Police Data Bias on Geographic Crime Analysis*, 18 J. EXPERIMENTAL CRIMINOLOGY 515, 532 (2021); Hoofnagle, *supra* note 22, at 101; Michael D. Maltz, *Analysis of Missingness in UCR Crime Data* (Crim. Just. Rsch. Ctr., Paper No. 215343, 2006), https://www.ojp.gov/pdffiles1/nij/grants/215343.pdf [https://perma.cc/7SY2-RKJU]; David A. Freedman & Kenneth H. Walker, *On the Likelihood of Improving the Accuracy of the Census Through Statistical Adjustment*, *in* 40 STATISTICS AND SCIENCE: A FESTSCHRIFT FOR TERRY SPEED 197 (Darlene R. Goldstein ed., 2003).

[174] Romanosky, Telang & Acquisti, *supra* note 9, at 260.

[175] Goel & Shawky, *supra* note 9.

[176] Mitts & Tally, *supra* note 13, at 3–4.

[177] Sullivan & Maniff, *supra* note 9.

[178] Greenwood & Vaaler, *supra* note 13.

being identity theft report rates per million population.[179] They then classified states as "Better, Mixed, or Worse" in terms of identity-theft report rates and examined which provisions most frequently co-occur within each of these bins.[180] This Article asks a similar question about how particular provisions affect identity theft report rates, but it differs in the empirical approach. The main difference is that whereas Sullivan and Maniff looked at correlations between the presence of a provision and identity-theft reporting rates,[181] this Article instead frames the problem in the causal inference framework and isolates the individual effect of each provision. The two papers both find that additional notification to state regulators and credit reporting agencies, including a private right of action, reduce reported identity theft.[182] However, Sullivan and Maniff found that the risk of harm exception, encryption rules, and time limits to notify consumers *increase* identity theft,[183] whereas this Article suggests the opposite. The differences here can stem from different timeframes in the data; the Sullivan and Maniff study goes up to 2015 while this Article extends the timeframe to 2020. The other difference likely comes down to differences in choice of modeling, as the synthetic control approach yields different results compared to linear regression.

Similarly, Greenwood and Vaaler looked at the effect of data breach notification laws on the number of data breaches.[184] They used a two-way fixed effects model, which is another extension of the difference-in-differences framework, and found that breach notification laws do not decrease the number of reported data breaches.[185] They also found no effect on identity-theft report rates.[186] These results are compatible with this Article's finding that baseline breach notification laws have varied effects on reported identity theft. The main difference in these two studies is the choice of outcome variable and method: whereas Greenwood and Vaaler looked at the number of breach reports as their main outcome measure,[187] this Article uses identity theft reports. Further, staggered adoption synthetic control and two-way fixed effects estimators are two related approaches for addressing similar settings where states adopt policies at different times. This Article's different outcome measure, focus on specific provisions of breach notification

[179] Sullivan & Maniff, *supra* note 9, at 72.
[180] *Id.* at 73.
[181] *Id.*
[182] *Id.* at 76.
[183] *Id.*
[184] Greenwood & Vaaler, *supra* note 13, at 4.
[185] *Id.*
[186] *Id.* at 5.
[187] *Id.* at 4.

laws, and use of synthetic control likely drive most of the differences between the two sets of results.

IV.
POLICY IMPLICATIONS

A. Federal Data Breach Proposals

Since the first state data breach notification laws passed, scholars and policymakers have called for similar federal legislation. Federal law imposes breach notification requirements in sectors such as health,[188] education,[189] and publicly traded companies.[190] Proposals for a general data breach notification law floundered in 2011[191] and 2014,[192] and breach notification has been part of proposals for omnibus privacy laws such as the House proposal by Representatives Anna Eshoo (D-CA) and Zoe Lofgren (D-CA)[193] and Representative Frank Pallone's (D-NJ) American Data Privacy and Protection Act.[194] The latter proposal explicitly singles out data breach notification as an area that would *not* be preempted by the federal law.[195] The most recent and significant federal action in this space was the passage of the Cyber Incident Reporting for Critical Infrastructure Act (CIRA), which requires that firms

[188] *See* Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936, 2009 (1996) (codified at 42 U.S.C. §§ 1320a–7e); Health Information Technology for Economic and Clinical Health Act, Pub. L. No. 111-5, 123 Stat. 115, 226 (2009) (codified at 42 U.S.C. §§ 300jj–11); HIPAA Breach Notification Rule, 45 C.F.R. §§ 164.400–414 (2022).

[189] *See Data Security: K-12 and Higher Education*, U.S. DEP'T EDU., https://studentprivacy.ed.gov/Security [https://perma.cc/UKF6-5CQH] (last visited Oct. 6, 2022) (discussing how data breaches usually result in Family Educational Rights and Privacy Act of 1974 violations).

[190] *See* Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure, 87 Fed. Reg. 16590 (proposed Mar. 23, 2022) (to be codified at 17 C.F.R. pts. 229, 232, 239, 240, 249).

[191] *See* Press Release, Sen. Patrick Leahy, Leahy Introduces Benchmark Bill to Update Key Digit. Priv. L. (May 17, 2011), https://web.archive.org/web/20210212005126/http://www.leahy.senate.gov/press/leahy-introduces-benchmark-bill-to-update-key-digital-privacy-law&.

[192] *See* JOHN PODESTA ET AL., EXEC. OFF. PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES (2014), https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [https://perma.cc/7H6R-6JN3].

[193] *See* Press Release, Anna G. Eshoo, Cal. Congresswoman, Eshoo and Lofgren Reintroduce Sweeping Privacy Legislation (Nov. 18, 2021), https://eshoo.house.gov/media/press-releases/eshoo-and-lofgren-reintroduce-sweeping-privacy-legislation [https://perma.cc/6S8K-V7WC].

[194] JONATHAN M. GAFFNEY ET AL., CONG. RSCH. SERV., LSB10776, OVERVIEW OF THE AMERICAN DATA PRIVACY AND PROTECTION ACT, H.R. 8152 (2022), https://crsreports.congress.gov/product/pdf/LSB/LSB10776.

[195] *Id.* at 3.

maintaining "critical infrastructure" report breaches to federal regulators.[196] The theoretical justification for CIRA is different from state data breach notification laws; CIRA is mainly a national security law that imposes a tight seventy-two-hour deadline on reporting a breach after it has been discovered, in contrast with state laws that focus on consumer protection and either have a requirement that breaches are reported "without unreasonable delay" or within a time frame of forty-five to ninety days.[197]

The scholarly discourse around a federal breach notification law considers whether the current state-by-state regime creates a "patchwork" of laws that increases the costs of compliance for multistate firms.[198] Proponents of a federal law argue that such a law would harmonize confusing standards across states and bring non-adopting states up to a minimum standard.[199] Defenders of the state-by-state approach argue that state experimentation is necessary to adapt to a rapidly changing technological environment, and a federal law might preempt important innovations among the states.[200]

Who has it right? Following from their two-way fixed effects analysis of breach notification laws on breaches, Greenwood and Vaaler argued that a federal law would be more likely to deter future data breaches.[201] However, refocusing the outcome to identity-theft reports suggests that the state-by-state approach is more valuable at this point. Federal proposals that mimic the baseline requirements imposed by states is unlikely to have an effect on identity-theft report rates. All fifty states have a breach notification law, and there is little evidence that the baseline notification requirement affected identity-theft report rates. A federal law would not bring non-adopting states closer to adopting states, nor be likely to have any additional effect beyond what the state laws already accomplish. Meanwhile, a federal proposal that accomplishes what proponents advocate—removing differences across states to lower costs of compliance—could preempt further state innovation that might be effective at reducing identity theft.

Any federal data breach notification law should focus on the elements of state laws that work well and on innovations not yet proposed by the states. The focus on disclosure alone is unlikely to yield

[196] *See* Cyber Incident Reporting for Critical Infrastructure Act of 2022, Pub. L. No. 117-103, 136 Stat. 49, 1043 (2022) (codified at 6 U.S.C. § 681b).

[197] GitHub, *supra* note 61.

[198] Rachael M. Peters, Note, *So You've Been Notified, Now What? The Problem With Current Data-Breach Notification Laws*, 56 ARIZ. L. REV. 1171, 1177 (2014).

[199] *Id.* at 1176.

[200] Paul M. Schwartz, *Preemption and Privacy*, 18 YALE L.J. 902, 946 (2009).

[201] Greenwood & Vaaler, *supra* note 13, at 5.

any additional benefits. Adopting best practices from states that expand enforcement options and reduce discretion is more likely to have effects than mandatory disclosure alone. Further, if a federal proposal moves forward, it should focus on establishing *floors* for regulatory action, rather than ceilings that would preempt provisions that work. For example, consumer groups were concerned that federal proposals to implement credit freezes following the Equifax breach would preempt stronger state protections.[202] Amendments to state data breach notification laws are an important driver of policy innovation in the privacy law space, and federal action that preempts further innovation could lead to more identity theft.

There is, however, a danger to the state-by-state approach taken thus far in data breach notification law. States have led the way in enacting breach notification laws, with the few federal provisions largely mimicking state ones. In a larger context, this trend is not surprising. Scholars of federalism have noted how in recent decades states have transformed from "backwaters to major policymakers" across a range of issues such as taxation, climate change, and healthcare.[203] Within the privacy law literature, scholars have noted how states have served important roles in both drafting statutes[204] and enforcement through state attorneys general.[205] As Paul Schwartz notes, federalism should see the federal government consolidate various state laws after some experimentation.[206] However, congressional gridlock makes it hard to imagine such an effort succeeding with regularity. In the face of congressional paralysis, the importance of states as policymakers grows even more.

While breach notification is not as polarized on partisan lines as other issues, congressional apathy can lead to consumer protection being much more robust in some states than others. Critics of federalism contend that it is less a system of government encouraging innovative policymaking and more a default to decentralization in the absence of

[202] Tara Siegel Bernard, *After Equifax Breach, Credit Freeze Provision Comes at a Price*, N.Y. TIMES (Mar. 15, 2018), https://www.nytimes.com/2018/03/15/your-money/equifax-breach-credit-freezes.html [https://perma.cc/G7P9-U3Y7]. The federal proposal would have forced consumers to request credit freezes from each of the three main credit rating agencies for ten dollars. This provision would have provided rights to credit freezes nationally but would have preempted state laws that required automatic or free credit freezes.

[203] Jacob M. Grumbach, *From Backwaters to Major Policymakers: Policy Polarization in the States, 1970–2014*, 16 PERSPS. ON POL. 416 (2018).

[204] *See, e.g.*, Schwartz, *supra* note 201, at 918.

[205] *See, e.g.*, Danielle Keats Citron, *The Privacy Policymaking of State Attorneys General*, 92 NOTRE DAME L. REV. 747 (2017).

[206] Schwartz, *supra* note 201, at 941.

national norms.[207] Privacy may start to look like other issue areas such as minimum wage, taxation, and education in that where an individual lives becomes the most important determinant of their experience with these policies.[208] In such an eventuality, scholars, activists, and policymakers will need to think critically about the best path forward for policy advocacy.

B. Disentangling the Economic Theory of Data Breach Notification Laws

The economic theory of data breach notification laws has two prongs. The first prong is that breach notification laws give consumers an opportunity to ameliorate the potential harms by taking precautions to safeguard their identity. The second prong is that breach notification laws should encourage better data security practices by imposing reputational sanctions on firms that are forced to make a breach notification. Under this theory, firms should invest in strong data security measures to avoid the damaging costs of breach disclosure.[209]

Taken together, the results from this Article give clues that provisions strengthening the first prong are likely to have sizable effects, whereas encouraging data security through reputational harms alone is unlikely to work. Additional notification to third parties and imposing a time limit for when to notify consumers have sizable effects of nearly 10% reductions in estimated identity theft reports. These results suggest that consumer-side interventions, such as taking opportunities to freeze credit or enroll in identity theft protection, are at play.

In contrast, states that did not provide an exception to breach disclosure requirement for low harm scenarios see an estimated increase in reports. One justification for allowing firms to decide the likelihood of harm is that too many notices may confuse or upset consumers. Forcing firms to make disclosures for every incident could also impose needless reputational and compliance costs on them when there is little reason to think that the disclosure would better inform consumers or reduce identity theft. Does this mean that this lack of exceptions increased identity theft?

While the lack of such exceptions probably did not cause an increase directly, there is one plausible mechanism for how they could increase identity theft. States that do have a risk-of-harm exception typically require some kind of "investigation" to justify the determination

[207] *See, e.g.*, MALCOLM M. FEELEY & EDWARD RUBIN, FEDERALISM: POLITICAL IDENTITY AND TRAGIC COMPROMISE (2008).

[208] *See* Grumbach, *supra* note 204, at 416.

[209] *See* Romanosky, Telang & Acquisti, *supra* note 9, at 260.

that a breach notification is necessary.[210] In some states, these investigations must still be reported to the state Attorney General even if there is no breach notification to consumers.[211] It is possible that these investigations are effective at bolstering organizations' data security practices or that the data security infrastructure necessary to carry out an investigation also corresponds to better practices for data security more broadly. It is also possible that the possibility of avoiding a damaging disclosure by minimizing risk of harm to consumers provides enough of an incentive to invest in better data security in advance. Such a story would be compatible with what Solove and Hartzog[212] and Verstraete and Zarsky[213] characterized as the core failure of data security law: it punishes bad luck rather than encouraging good data security hygiene. Simply put, organizations that have an incentive to avoid a disclosure may invest in more robust data security practices throughout the lifecycle of their organization's data, rather than simply hope they will not be the victim of a breach.

This idea is further strengthened by the sizable estimated decrease in identity-theft reports associated with the inclusion of encryption in covered data. According to California legislators who closed the encryption loophole, they did not anticipate that the main effect would be through more disclosures to consumers. Rather, they thought the main mechanism would be prompting organizations to start using encryption.[214]

These results suggest a few policy levers that states may try going forward. One is that if consumer mitigation is the primary mechanism for reducing identity theft reports, then states should strengthen mechanisms enabling consumer intervention. For instance, mandatory reporting to credit reporting agencies usually accompanies offers to freeze a consumer's credit or enroll them in identity theft protection. Yet, only as few as 9% and as much as 30% of consumers opt into these services when offered.[215] A simple fix could be to reorient the choice architecture

[210] GitHub, *supra* note 61.

[211] *Id.*

[212] Solove & Hartzog, *supra* note 8, at 8.

[213] Verstraete & Zarsky, *supra* note 13, at 808–39.

[214] Assembly Comm. on Priv. and Consumer Protection, *supra* note 87, at 2.

[215] Sarah O'Brien, *Two Years After Huge Equifax Breach Was Revealed, Consumers Are Still Too Vulnerable to Identity Theft*, CNBC (Sept. 6 2019, 1:41 PM), https://www.cnbc.com/2019/09/06/two-years-after-equifax-breach-consumers-still-vulnerable-to-id-theft.html [https://perma.cc/K6Z6-4JYN]; Ponemon Inst. LLC, The Aftermath of a Data Breach: Consumer Sentiment 5 (2014), https://www.ponemon.org/local/upload/file/Consumer%20Study%20on%20Aftermath%20of%20a%20Breach%20FINAL%202.pdf.

to automatically enroll consumers in these services and allow them to opt out.[216]

More generally, states could consider amendments that would incentivize better data security hygiene rather than assuming that reputational sanctions will do the work. Within data breach notification laws, states could condition exceptions to disclosure requirements on adoption of certain data security measures. For instance, adopting and implementing NIST recommendations could be an incentive that states provide to organizations.[217] One key point here is that specificity might be important. The encryption provisions explicitly signal to organizations what changes they need to make. Similarly, further changes to data breach notification laws might need to peg incentives to particular practices, with the caveat that regularly updating such requirements as data security threats evolve will be necessary.

A major question that emerges from implementing provisions that promote better data security is how government should weigh the costs and benefits of such provisions. What is the cost to a firm adopting strong encryption standards? What are the benefits for both the adopting firm and its consumers? If a firm decides that this cost is not worth avoiding a breach disclosure, does this reflect a success or failure of the legal provision? These are difficult questions to answer in part because of the uncertainty involved in pricing compliance, privacy harms, and regulatory enforcement. However, they are important for understanding how strong to make data security laws. Overly harsh punishments for any failure of data security could have negative consequences for innovation, competition, and consumer choice. Lax punishments and lack of liability for weak data security practices can similarly have negative consequences for market stability, consumer protection, and national security. More research that helps quantify these aspects of data security law can help state legislatures navigate these tradeoffs and understand whether they are erring on the side of too strong or too weak legal regimes.

That said, this Article does provide one important contribution to conducting this type of cost-benefit analysis. Table 2 shows the estimated dollar amounts saved or lost by particular data breach notification law provisions. This is a rough calculation based on the average financial loss victims suffer as well as the value (based on the average

[216] Richard G. Kunkel, *Strengthening Credit Freeze Legislation in the States: Empowering Consumers to Prevent Economic Loss from Identity Theft*, 23 MIDWEST L.J. 97 (2009).

[217] *See Cybersecurity Framework,* NAT'L INST. STANDARDS & TECH., https://www.nist.gov/cyberframework [https://perma.cc/3PRZ-7LCE] (last visited Oct. 6, 2023).

U.S. wage) of hours lost resolving identity theft. While this value can certainly vary depending on the specific case of identity theft, the victim, and other factors, it provides a good benchmark for understanding the relative benefits of these laws. For example, the addition of encrypted data to the definition of covered data saved approximately $356 million nationwide. If the federal government or individual states were to adopt this provision, they could at least begin to ask whether the costs of compliance with the law outweigh this benefit.

C. The Role of Regulators

Privacy law scholars are slowly shifting away from disclosure and notice-and-consent frameworks for regulating privacy and toward conceptualizing what privacy enforcement should look like.[218] State attorneys general have already been active in this space.[219] At the federal level, the FTC is taking a leading role in using its powers to bring enforcement actions to punish privacy violations as unfair and deceptive business practices.[220] California is establishing the California Privacy Protection Agency (CPPA).[221] Breach notification represents a small slice of the types of privacy concerns that these agencies will deal with but contains lessons for them nonetheless. As arguably one of the oldest and most popular forms of privacy legislation in the United States, policymakers should look to what has worked with breach notification laws and what has not.

Agencies such as the CPPA likely play an important role in reshaping the politics of privacy federalism. The CPPA is authorized with

[218] *See, e.g.*, Matthew A. Edwards, *Empirical and Behavioral Critiques of Mandatory Disclosure: Socio-Economics and the Quest for Truth in Lending*, 14 CORNELL J.L. & PUB. POL'Y 199, 242 (2005); Ian Ayres & Alan Schwartz, *The No-Reading Problem in Consumer Contract Law*, 66 STAN. L. REV. 545, 545 (2014); Ari Ezra Waldman, *Privacy, Notice, and Design*, 21 STAN. TECH. L. REV. 74, 77–78 (2018); M. R. Calo, *Against Notice Skepticism in Privacy (and Elsewhere)*, 87 NOTRE DAME L. REV. 1027, 1027 (2013). In the context of privacy policies in particular, critics have noted that consumers rarely read notices, making the validity of their consent questionable. Even if consumers could read every privacy policy they are confronted with, the lack of meaningful choice often undermines the theoretical benefits of notice-and-consent.

[219] *See* Citron, *supra* note 206, at 748.

[220] *See* CHRIS JAY HOOFNAGLE, FEDERAL TRADE COMMISSION PRIVACY LAW AND POLICY 67, 115 (2016); Jennifer K. Wagner, *The Federal Trade Commission and Consumer Protections for Mobile Health Apps*, 48 J.L. MED. & ETHICS 103, 103 (2020); FED. TRADE COMM'N, FEDERAL TRADE COMMISSION 2020 PRIVACY AND DATA SECURITY UPDATE 3 (2021), https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-2020-privacy-data-security-update/20210524_privacy_and_data_security_annual_update.pdf; Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 583 (2011).

[221] *See California Privacy Protection Agency*, STATE OF CAL., https://cppa.ca.gov [https://perma.cc/6ZQA-6JVF] (last visited Sept. 26, 2023).

rulemaking authority under the California Privacy Rights Act (CPRA), and it is just now beginning to flex this muscle.[222] We might expect that a relatively muscular state agency dedicated to privacy protection may promulgate stronger regulations than the disclosure-based regime governing data breaches thus far. We may also see data breach laws expanded through the rulemaking process by covering new types of data, prompting agency investigations, and raising consumer awareness. For example, genetic data is now being added to data breach statutes in Illinois and California.[223] As more states create similar regulatory bodies, scholars would benefit from paying attention to how these agencies approach data breach and other areas of privacy, in addition to examining how the FTC will regulate privacy under the Biden Administration. State-level privacy regulation may prove to be an important source of regulatory innovation given that the FTC has more limited tools for regulating privacy specifically.

Although it is too early to quantitatively assess the effects of new state-level privacy regulators, one of the major lessons from this Article is that regulators should be attentive to technological details. The encryption exception illustrates this point nicely. States, reasonably, exempted organizations from disclosing breaches that involved encrypted data. Yet, it seems that some organizations interpreted this requirement in the broadest sense and may not have disclosed breaches when encrypted data was lost with the ability to decrypt it. States that tightened this exception to exclude encrypted data lost with the encryption key saw a decrease in identity theft reports. Policymakers should be vigilant and craft requirements such that organizations cannot avoid the spirit of them. Within privacy law, new consumer privacy rights such as opting in or out of tracking, data deletion rights, and a right to know what type of information is being collected should be scrutinized to see whether organizations are complying with the intent of these provisions.

The channels of enforcement also matter. Requiring disclosures to state regulators and consumer credit reporting agencies both decreased estimated identity-theft reporting rates. More work investigating the exact mechanisms of this effect would be valuable. Requiring reports to state attorneys generals and consumer protection agencies might have a deterrent effect either because firms do not want to invite an investigation into negligent cybersecurity practices or because those offices

[222] *See* Jennifer Bryant, *CPPA Board Moves CPRA Rulemaking Process Forward,* INT'L ASS'N PRIV. PROS. (June 9, 2022), https://iapp.org/news/a/cppa-board-launches-cpra-rulemaking-process/ [https://perma.cc/DJ9S-FHEZ].

[223] Genetic Information Privacy Act, 410 ILL. COMP. STAT. 513 (2022); Genetic Information Privacy Act, CAL. CIV. CODE §§ 56.18–56.186 (Deering 2023).

might be more effective at subsequently reaching consumers. Reporting to consumer credit reporting agencies may help prevent identity theft either by alerting consumers to potential risks to their credit scores after a breach or by proactively flagging suspicious behavior.[224]

These additional notification requirements may also be beneficial for bolstering the case for investments in muscular regulatory agencies. Filippo Lancieri argues that insufficient attention to information asymmetries between companies and consumers/regulators is one of the main causes for the failure of data protection law to achieve adequate enforcement.[225] Breach notification laws, and in particular provisions requiring disclosures to state regulators, are one exception within the overall data privacy law landscape. Organizations have private information about their own data security postures, and data breach notification laws might be one mechanism for helping regulators uncover relevant information. Again, there is a broader discussion around how to balance between the costs and benefits of compliance with disclosure requirements and where the responsibility should lie. But understanding the effects of these provisions on identity-theft reporting rates can help regulators think about these boundaries, rather than default to consumers needing to lead the way.

Lancieri also points out that privacy regulators are resource constrained. The California Attorney General only has a budget of $5 million to support twenty-three attorneys working on consumer protection issues, and the California Privacy Protection Agency has a budget of $10 million to enforce all of the CPPA's provisions.[226] Europe is moving to rethink its data protection law enforcement as much has been shouldered by the under-resourced Irish Data Protection Authority.[227] Arguably, many of the estimated effects of notification to state regulators in this Article represent *lower* bounds for how effective these provisions might be. Thus far, state attorneys general focus much of their data breach enforcement against high profile cases and in situations where they can pool resources in multi-state litigation.[228] Localized breaches are unlikely to be investigated or pursued the same way, yet

[224] For example, the Michigan Attorney General outlines several of the options available to consumers. Credit reporting agencies may, as a default, offer credit monitoring or fraud alert services and notify each other about such actions. *Credit Freeze; Fraud Alert; & Credit Monitoring*, MICH. DEP'T ATT'Y GEN., https://www.michigan.gov/ag/consumer-protection/consumer-alerts/consumer-alerts/credit/credit-freeze-fraud-alert-credit-monitoring-1 [https://perma.cc/VNM7-488Y] (last visited Sept. 26, 2023).

[225] Filippo Lancieri, *Narrowing Data Protection's Enforcement Gap*, 74 ME. L. REV. 15, 16 (2022).

[226] *Id.* at 56.

[227] *Id.* at 27.

[228] Press Release, Mass. Off. Atty. Gen., *supra* note 56.

these breaches can be consequential for consumer spending and access to credit.[229]

Another finding that supports existing literature is that a private cause of action can provide a useful complement to public enforcement, but it is not the most effective mechanism. Romanosky, Hoffman, and Acquisti previously found that the risk of federal litigation increased when plaintiffs could allege financial harm.[230] While a more detailed analysis of state litigation would be necessary to see how this plays out at the state level, the effects of provisions providing a private cause of action reducing identity theft suggest that private enforcement has some deterrent effect on negligent cybersecurity practices.

All of these findings point toward the idea that recent efforts to empower regulators to address privacy issues may be fruitful. While these findings are specifically about data breach notification, there are lessons for other areas of privacy law. Although giving consumers more control over how data about them is collected and processed can be beneficial, consumer choice alone may not achieve statutory aims. Consumer data exists in a broader ecosystem intermediated by insurers, data brokers, advertisers, etc., and regulation that takes these actors into account might be more effective at minimizing harms. Empowering state attorneys general and consumer protection authorities to investigate bad behavior, bring lawsuits, and inform consumers may also help. States may soon start creating California-style "privacy protection agencies" tasked with defining their own missions and priorities. Empowering these agencies through knowledge of adverse events can help get them started on the right foot.

D. Policy Evaluation and Privacy Law

Although there are deep debates about policymaking in a federal system, the state-led approach provides a good opportunity for leveraging statistical methods for policy evaluation. One of the defenses of federalism is the "laboratories of democracy" concept advanced by Justice Louis Brandeis.[231] The basic idea is that a state may "try novel social and economic experiments without risk to the rest of the country."[232] In some sense, data breach notification laws reflect this ideal. California pioneered the first version of the data breach notification law in 2003, and other states learned from its experiences and adopted similar laws.

[229] Strahilevitz & Liu, *supra* note 106.

[230] Romanosky, Hoffman & Acquisti, *supra* note 43.

[231] New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting).

[232] *Id.*

There is, however, an epistemological weakness with the laboratories of democracy concept. The metaphor is meant to invoke the notion of a scientist conducting an experiment to evaluate the effects of an intervention. State policymaking violates one of the central requirements of this kind of experimentation—there is no randomization into treatment and control groups. Randomization is a powerful tool for making credible inferences, but states select into their own policy regimes rather than be exposed to treatment or control by an experimenter. Within the social sciences, field experiments are gaining traction as a way to evaluate various policy regimes.[233] Thus far, randomized control trials have been used to study social policy areas such as education,[234] job training programs,[235] and workplace safety compliance.[236] Extending this spirit of experimentation to privacy and data security could be one way to strengthen causal claims about the efficacy of these laws.

Of course, such experimentation is frequently impossible in the social sciences, particularly when dealing with large units like states or countries. In the absence of true randomization, much of social science instead looks for "natural experiments" that approximate the experimental ideal.[237]

In the absence of randomization or quasi-randomization, is there a way to still make credible inferences about policy? Studies with observational data are often fraught with methodological problems and rely on analysts making hard-to-validate assumptions.[238] Policymakers and analysts often do not have the luxury of intervening with carefully

[233] *See generally* Donald P. Green & Dane R. Thorley, *Field Experimentation and the Study of Law and Policy*, 10 ANN. REV. L. & SOC. SCI. 53 (2014) (providing a review of field experiments in law and policy).

[234] *See, e.g.*, HANLEY CHIANG ET AL., INST. EDUC. SCI., EVALUATION OF THE TEACHER INCENTIVE FUND: FINAL REPORT ON IMPLEMENTATION AND IMPACTS OF PAY-FOR-PERFORMANCE ACROSS FOUR YEARS (2017), https://ies.ed.gov/ncee/pubs/20184004/pdf/20184004.pdf; MICHAEL PUMA ET AL., U.S. DEP'T HEALTH & HUM. SERVS., HEAD START IMPACT STUDY FINAL REPORT (2010), https://www.acf.hhs.gov/sites/default/files/documents/opre/executive_summary_final_508.pdf.

[235] *See, e.g.*, Howard S. Bloom et al., *The Benefits and Costs of JTPA Title II-A Programs: Key Findings From the National Job Training Partnership Act Study*, 32 J. HUM. RES. 549 (1997)

[236] *See, e.g.*, David I. Levine et al., *Randomized Government Safety Inspections Reduce Worker Injuries With No Detectable Job Loss*, 336 SCIENCE 907 (2012).

[237] *See* THAD DUNNING, NATURAL EXPERIMENTS IN THE SOCIAL SCIENCES: A DESIGN-BASED APPROACH 1–3 (2012) (discussing the strengths and weaknesses of various natural experiment methodologies).

[238] *See* Ryan Copus et al., *Big Data, Machine Learning, and the Credibility Revolution in Empirical Legal Studies*, *in* LAW AS DATA: COMPUTATION, TEXT, & THE FUTURE OF LEGAL ANALYSIS 21 (Michael A. Livermore & Daniel N. Rockmore eds., 2018); Joshua D. Angrist & Jörn-Steffen Pischke, *The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics*, 24 J. ECON. PERSPS. 3 (2010).

designed field experiments or waiting for good natural experiments, yet they face pressure to make decisions anyway.

Balancing the desire to make data-driven policy decisions with the realities of data paucity in many applied settings may help explain why synthetic control has become such a popular tool over the last few years. Unlike other popular methods such as regression discontinuity or instrumental-variables estimation, synthetic control does not make assumptions about the existence of quasi-randomization.[239] Identification of the causal effect instead relies on constructing a valid synthetic estimate of pre-treatment outcomes and using this synthetic estimate as the control unit to compare to a treated unit.[240] Extensions to staggered adoption settings will likely prove especially useful for applied researchers who work with observational data.

One implication of this Article is that empirical-legal-studies researchers should leverage data-driven approaches like synthetic-control methods and pay attention to research into its extensions. Empirical legal studies are often concerned with applied questions about how changes in legal regimes affect some outcome of interest. Design-based inference is not always possible in these situations, regardless of how important or consequential a policy issue may be. Synthetic control provides a powerful method for estimating the treatment effects of new policies with observational data.[241] Staggered-adoption extensions can be particularly well-suited to policies that vary at the state and local level.[242] Empirical legal studies scholars can be uniquely positioned to leverage these methods because of the tradition's emphasis on centering the nuances of law. Looking at different provisions of state laws and understanding the theoretical impacts each provision should have requires deep domain expertise in legislation, regulation, and litigation. Legal scholars can use methods like synthetic control in conjunction with their domain expertise to illuminate insights that other disciplines may not focus on. Framing policy problems in the potential-outcomes framework and using new innovations in the policy-evaluation literature can help legal scholarship realize its potential for informing real-world policy decisions.

There are some important caveats when applying this method though. Synthetic control relies on an excellent pre-treatment match between the observed and synthetic units, and this may not always be

[239] Abadie et al., *supra* note 17, at 494.
[240] *Id.*
[241] *See id.*
[242] Ben-Michael et al., *supra* note 19, at 354, 375–77.

possible.[243] In the staggered setting, there is also a tradeoff between minimizing the balance between individual treated units and minimizing the global imbalance among all treated units.[244] We see some of these limitations in practice in this Article as it is not always possible to achieve a good pre-treatment fit across all models. Ameliorating these issues is an active area of research. This approach is also just one way to frame causal problems, and other approaches might be appropriate.[245]

Turning to empirical legal studies approaches to privacy law, there are also several lessons in this Article for how to proceed moving forward. The data collection and statistical methods used in this Article can easily be adapted to study identity theft. Future extensions may look at differences across different geographies or at specific subpopulations that are most likely to submit identity theft complaints, such as older and Black Americans.[246] The methodology may also be used to study other outcomes of interest related to privacy, such as breach litigation.

However, this Article also reinforces many of the problems that have plagued empirical privacy law scholarship since the mid-2000s. In 2007, Chris Hoofnagle called for more disclosures from banks and financial institutions to provide more data to study identity theft.[247] We still have very little data of this kind. While the FTC publishes aggregated data about identity theft, researchers lack access to consumer narratives and demographic information that would provide a richer picture. There are few datasets available pertaining to firm compliance with privacy laws, though this is an active area of research.[248] To understand the broader picture of how privacy laws work, researchers need

[243] Ben-Michael et al., *supra* note 169, at 1.

[244] Ben-Michael et al., *supra* note 19, at 359–60.

[245] *See* Judea Pearl & Dana MacKenzie, The Book of Why: The New Science of Cause and Effect (2018) (introducing the Pearl causal model); *see also* Sebastian Benthall & Katherine J. Strandburg, *Agent-Based Modeling as a Legal Theory Tool,* 9 Frontiers Physics 1 (2021) (using Agent Based Modeling as an alternative to traditional law and economics approaches).

[246] *See Identity Theft Resource Center & Black Researchers Collective Research Finds ID Crime Victims in Black Communities Lose More Money Than General Population*, Identity Theft Res. Ctr. (Jan. 4, 2023), https://www.idtheftcenter.org/post/id-crime-victims-black-communities-lose-more-money-than-general-population/ [https://perma.cc/X54E-SBY5]; Marguerite DeLiema et al., *Identity Theft Among Older Adults: Risk and Protective Factors*, 4 Innovation Aging 31 (2020).

[247] Hoofnagle, *supra* note 22, at 99.

[248] *See, e.g.*, Frankenreiter, *supra* note 104; Michael Batikas et al., *European Privacy Law and Global Markets for Data* (CEPR Discussion Paper, Paper No. 14475, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstractid=3560282 [https://perma.cc/AAG2-HCBN]; Nikita Samarin et al., *Investigating the Compliance of Android App Developers With the CCPA*, IEEE-Security (2021), https://www.ieee-security.org/TC/SPW2021/ConPro/papers/samarin-conpro21.pdf.

access to more and better data and in particular, to outcomes other than identity-theft report rates.

One potential path forward for both legislative and administrative bodies is to articulate plans for introducing the best available social science techniques to the study of privacy law. Beyond just requiring the disclosure of more data that would be helpful for researchers, privacy regulators might also start thinking about introducing randomized control trials (RCTs) to the evaluation of privacy law. Federal, state, and local governments have all seen success at implementing RCTs for better understanding the effects of their programs and enforcement actions, and this spirit could be adapted to privacy law as well.[249] For instance, as federal agencies start thinking about how to conduct data security audits, running these audits as RCTs could improve their efficacy and build trust in these programs.

Conclusion

Breach notification laws will continue to be an important part of the privacy law landscape. States continue to update these laws with new provisions, thus generating more questions for scholars and policymakers to study. Previous studies, both empirical and doctrinal, typically focus on the *disclosure* aspects of these laws, a focus that misses much of the rich variation in legislative provisions. This Article contributes to the privacy law literature by analyzing these other aspects of breach notification laws and showing how innovations beyond the basic formula reduces identity theft reports.

This Article also contributes to the empirical legal studies literature by providing an example of staggered-adoption policy evaluation. Legal scholars are often concerned about the effects that laws have on particular outcomes. In the U.S. context, several issue areas such as minimum wage, education, and environmental policy are governed by state and local governments. When working with observational data, options for answering questions about different policy regimes across these sub-national units can be challenging, but new data-driven innovations in the policy evaluation literature can pave a way forward. With important caveats in mind about credible inferences from such data and methods, empirical legal studies scholars can benefit from the transparent and intuitive appeal of methods like synthetic control. This Article provides the most comprehensive and up-to-date analysis of data breach notification laws. Future work should analyze other aspects of

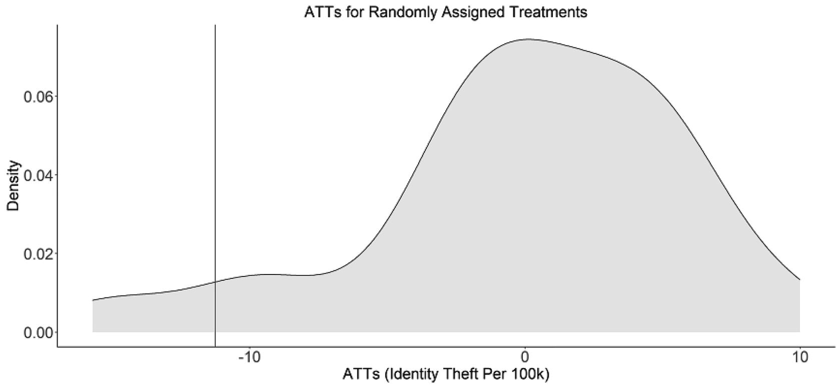
[249] *See generally* Christian R. Grose & Abby K. Wood, *Randomized Experiments by Government Institutions and American Political Development,* 185 Pub. Choice 401 (2020).

data breach notification laws, as well as new amendments that are added in coming years. This Article's study leads to a few general takeaways. The first is that state experimentation with privacy law can generate useful data for researchers to investigate what the most effective or ineffective mechanisms are. The second is that while breach notification laws do not eliminate identity theft, certain provisions can make a dent in identity-theft report rates. The third is that calls for more data will help illuminate future directions for research. The mechanisms for corporate compliance with breach notification and other privacy law are still unclear, as are the parameters of identity theft. Requiring more disclosures of these data will clarify and extend many of the findings in the Article. In general, as privacy issues are getting more attention in the regulatory sphere, looking at some of the oldest and most pervasive privacy laws in the U.S. can provide invaluable insights for moving forward.

Appendix A: Placebo and Outlier Robustness Checks

This appendix presents standard placebo checks for fitting staggered adoption synthetic controls.

The robustness check follows a similar procedure as John J. Donohue, Abhay Aneja, and Kyle D. Weber.[250] Standard synthetic control methods do not yield conventional p-values or confidence intervals. One way to address this issue is to calculate what they term a "pseudo p-value."[251] In their paper they simulate placebo treatments for states that adopt right-to-carry gun laws and calculate the proportion of placebo treatment effects whose absolute value is greater than the absolute value



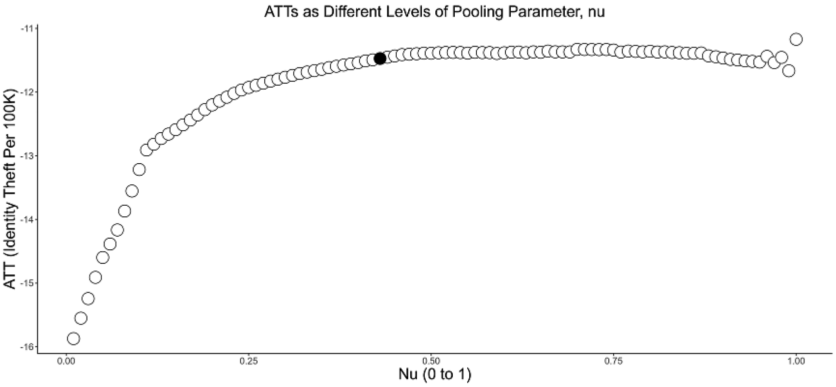
[250] Donohue et al., *supra* note 109, at 198.

[251] *Id.* at 235.

of the actual estimated treatment effect.[252] I take a similar approach and conduct 500 simulations where states are randomly assigned to treatment at random times. I then plot the distribution of estimated average treatment effects. The black line indicates the actual estimated treatment effect, which is close to the tail of the distribution that is centered around 0. This indicates that the treatment effect we see in the actual data is likely not just noise and indicates a real effect.

Appendix B: Pooling Parameter Robustness Check

Ben-Michael et al. suggest tuning the hyperparameter, υ, to see if the results depend strongly on the choice of pooling.[253] At υ = 0 this is the equivalent of fitting separate synthetic controls for each state, whereas υ = 1 is the equivalent of pooling all states together. Ideally, the choice of υ should not radically change the estimated effects and instead simply reflect the optimum on a curve. In this case, we see that choice of υ does not change the estimated treatment effect much for notification to a state regulator, particularly within the window near the optimal point of .45.



[252] *Id.*

[253] Ben-Michael et al., *supra* note 19, at 362.

Appendix C: Timeline of State Provisions

State	AG	Cause of Action	CRA	Encryption	Harm Analysis	Time Limit
AL	6/1/2018	0	6/1/2018	5/1/2018	0	6/1/2018
AK	0	0	7/1/2009	7/1/2009	0	0
AZ	8/3/2018	0	8/3/2018	0	0	8/3/2018
AR	7/23/2019	0	0	0	0	0
CA	1/1/2012	1/1/2020	0	1/1/2017	7/1/2003	0
CO	9/1/2018	0	9/1/2006	9/1/2018	0	9/1/2018
CT	10/1/2012	0	0	0	0	10/1/2021
DE	4/14/2018	0	0	4/14/2018	0	4/14/2018
DC	6/17/2020	0	7/1/2007	0	6/17/2020	0
FL	7/1/2014	0	7/1/2014	0	0	7/1/2014
GA	0	0	5/5/2005	0	7/1/2005	0
GU	0	0	0	0	0	0
HI	1/1/2007	1/1/2007	1/1/2007	1/1/2007	0	0
ID	7/1/2010	0	0	0	0	0
IL	1/1/2017	0	6/27/2006	1/1/2017	1/1/2006	0
IN	7/1/2009	0	7/1/2006	7/1/2006	0	0
IA	7/1/2014	0	0	7/1/2014	0	0
KS	0	0	1/1/2007	0	0	0
KY	0	0	7/15/2014	0	0	0
LA	1/1/2006	1/1/2006	0	0	0	8/1/2018
ME	1/31/2006	0	1/31/2006	0	0	9/19/2019
MD	1/1/2008	4/3/2007	1/1/2008	0	0	1/1/2018
MA	2/3/2008	0	0	10/31/2007	0	0
MI	0	0	7/2/2007	4/1/2011	0	0
MN	0	8/1/2008	1/1/2006	0	8/1/2008	0
MS	0	0	0	0	0	0
MO	8/28/2009	0	8/28/2009	0	0	0
MT	10/1/2015	0	3/1/2006	0	0	0
NE	7/21/2016	0	0	7/20/2016	0	0
NV	0	0	1/1/2006	0	0	0
NH	1/1/2007	1/1/2020	1/1/2007	1/1/2007	0	0
NJ	0	0	1/1/2006	0	0	0
NM	6/16/2017	0	6/16/2017	6/6/2017	0	6/16/2017
NY	12/7/2005	0	12/7/2005	0	12/7/2005	0
NC	10/1/2009	10/1/2009	12/1/2005	12/1/2005	0	0

State	AG	Cause of Action	CRA	Encryption	Harm Analysis	Time Limit
ND	8/1/2015	0	0	0	0	0
OH	0	0	2/17/2006	0	0	2/17/2006
OK	0	0	0	1/1/2008	0	0
OR	1/1/2016	0	10/1/2007	0	0	6/2/2018
PA	0	0	6/20/2006	6/20/2006	0	0
PR	1/5/2006	0	0	0	0	0
RI	6/26/2016	0	6/26/2016	0	0	6/26/2016
SC	7/1/2009	0	0	0	0	0
SD	7/1/2018	0	7/1/2018	7/1/2018	0	7/1/2018
TN	0	9/30/2019	7/1/2005	4/4/2017	7/1/2005	7/1/2016
TX	1/1/2020	0	4/1/2009	9/1/2009	9/1/2005	1/1/2020
VI	0	0	0	0	0	0
UT	0	0	0	0	0	0
VT	5/8/2012	0	1/1/2007	0	0	5/8/2012
VA	7/1/2008	1/1/2023	7/1/2008	7/1/2008	0	0
WA	7/23/2015	7/24/2005	0	0	0	3/1/2020
WV	0	0	6/6/2008	6/7/2008	0	0
WI	0	0	3/31/2006	0	0	3/31/2006
WY	0	0	0	0	0	0

Appendix D: Time Cohorts

An alternative to modeling treatment effects by state is to instead model by time cohorts. That is, identity theft report rates could be estimated by grouping states together if they passed provisions at the same time. Using "group-time" cohorts and calculating ATTs based on these groups is a similar approach to the Callaway-Sant'Anna (CS) alternative to two-way fixed effects estimators.[254] The main disadvantages of TWFE estimators are that they can introduce "bad comparisons" problems by allowing already-treated units to act as comparison groups and that they can obtain the opposite sign of the true ATT when there are dynamic treatment effects over time.[255] The CS method's major innovation was introducing the concept of "group-time" cohorts that creates comparison groups based on units that are never treated or "not-yet treated."[256] Similar to the CS approach, the multisynth approach also

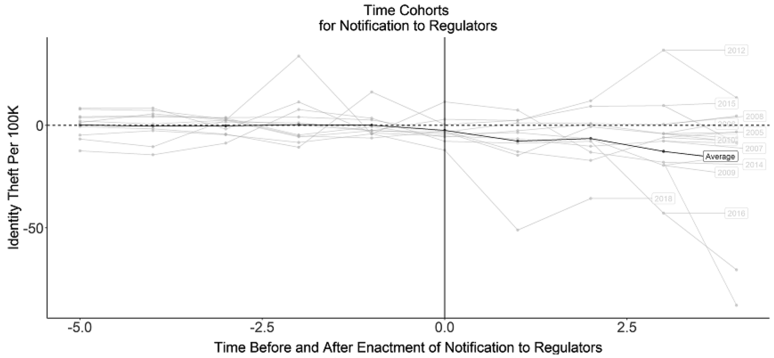
[254] Brantly Callaway & Pedro H.C. Sant'Anna, *Difference-in-Differences With Multiple Time Periods*, 225 J. ECONOMETRICS 200 (2021).

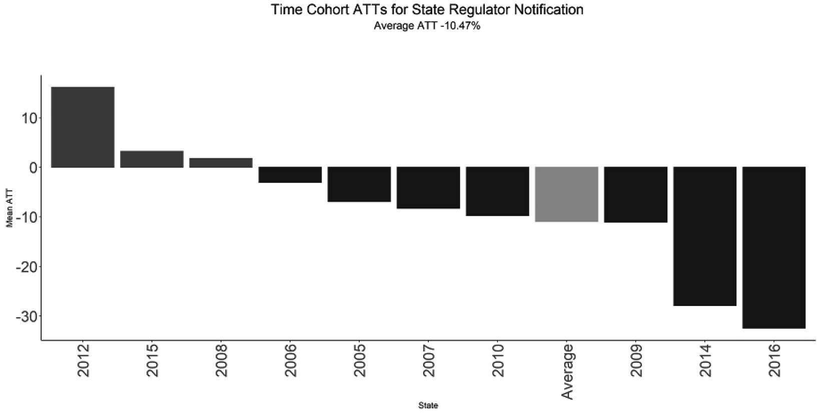
[255] Baker et al., *supra* note 126, at 371.

[256] Callaway & Sant'Anna, *supra* note 255, at 201, 206.

provides a method for estimating ATTs using group-time cohorts and "not-yet treated" units.[257]

The figures below show time cohort estimates for provisions that require additional notification to state regulators. The average ATT remains virtually unchanged (10.47% reduction with time cohorts versus 10.1% reduction with state effects), and there are similar post-treatment trends.



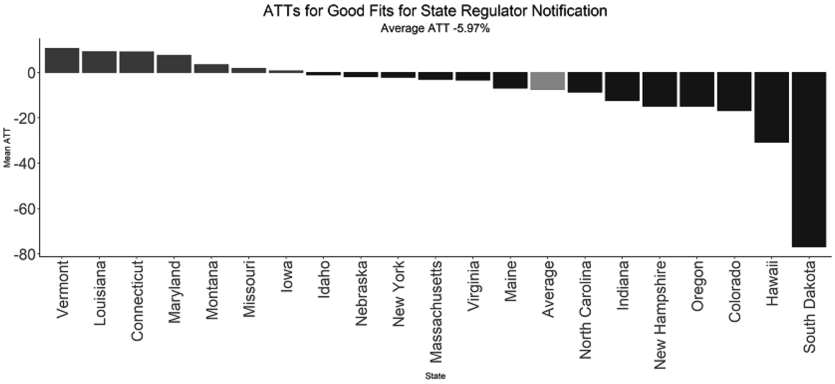


Appendix E: Bad Synthetic Control Fits and Outlier Analysis

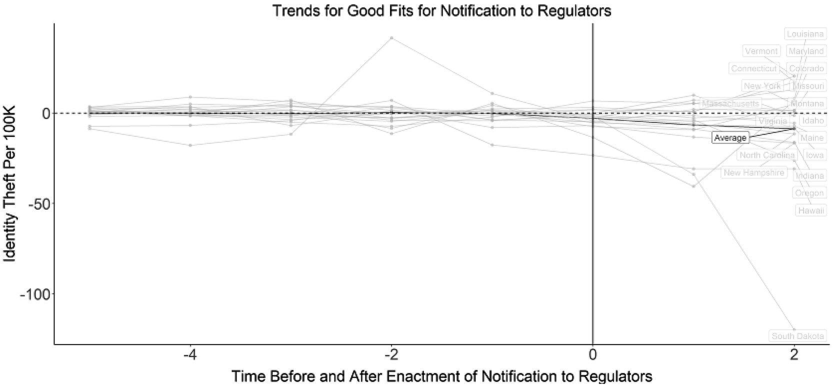
A potential problem is that the main results in this Article are driven by only a few states. For example, in Figure 12, Arizona has the largest estimated decrease in identity theft report rates by a substantial margin, with an average estimated decrease of over 100 reports compared to an average decrease of 11.25 reports across all adopting states. It is possible that Arizona truly did experience a large decrease in identity theft reports, but it is also possible that the estimate is being driven by statistical noise.

[257] Baker et al., *supra* note 126, at 374.

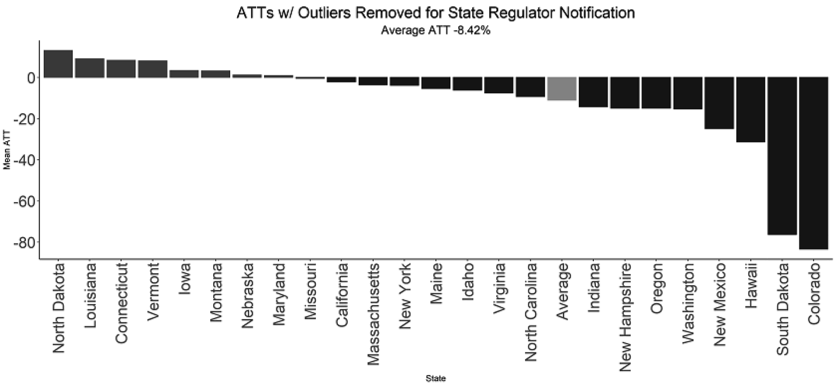
One potential cause for extreme estimates is that they are based on poorly fit synthetic controls. An individual synthetic control that is badly biased might result in extreme estimates and distort the average results as well. The figures below illustrate the average treatment effects on the treated for notification to state regulator provisions but with poor individual fits removed. Here the estimated effect does decrease (a 5.97% reduction in estimated identity theft reports without poor fits versus a 10.1% reduction in identity theft reports with all treated states), but the direction of the change is the same.

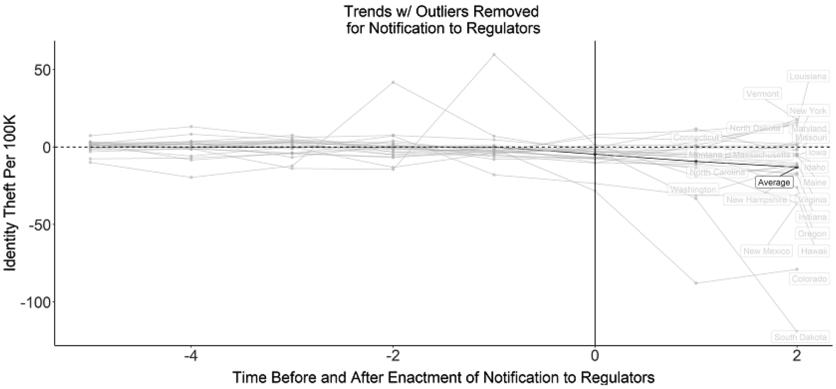

ATTs for Good Fits for State Regulator Notification

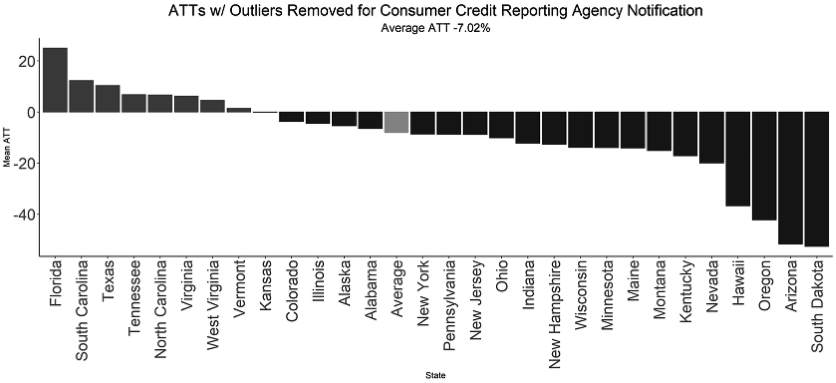
Average ATT -5.97%

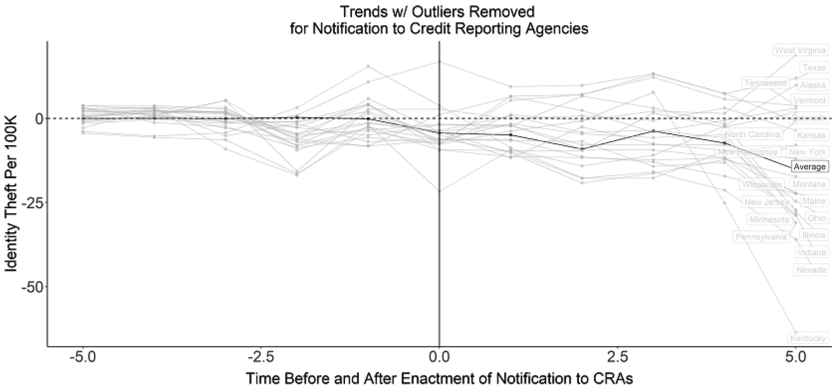

Trends for Good Fits for Notification to Regulators

Another approach might be to remove the outlier states entirely and observe whether this changes the overall result. The figures below show estimates for provisions that require notification to state regulators and credit reporting agencies but with extreme estimates removed. The treatment effect slightly decreases for regulator notification (8.42% decrease without outliers versus 10.1% decrease with outliers) and is virtually unchanged for CRA notification (7.02% without outliers versus 7.61% with outliers). In either case, we still see a decrease, indicating that the general story of the provisions decreasing reported identity theft should hold.









Appendix F: Technical Explanation of Staggered Adoption Synthetic Control

This section provides a technical review of synthetic control and its staggered adoption extension. This material is adapted from Ben-Michael et al.[258] For more details, consult sections two through four of that paper.

In the synthetic control method ("SCM"), a counterfactual outcome under control is estimated from a weighted average, the *synthetic control*. Weights are chosen to minimize the squared imbalance between the lagged outcomes for the treated unit and the weighted control, or donor units.

Ben-Michael et al. introduce a modified version of SCM that differs from Abadie et al. The original SCM formulation balances auxiliary covariates, whereas the Ben-Michael et al. SCM focuses only on lagged outcomes. Second, they add a regularization parameter, λ, that penalizes the sum of the squared weights toward uniformity. The Ben-Michael et al. SCM is:

$$\min_{\gamma_j \epsilon \Delta_j^{scm}} \frac{1}{L_j} \sum_{l=1}^{L_j} \underbrace{\left(Y_{jT_j-l} - \sum_{i-1}^{N} \gamma_{ij}\, Y_{iT_j-l} \right)^2}_{\text{objective}} + \underbrace{\lambda \sum_{i=1}^{N} \gamma_{ij}^2}_{\text{regularization}}$$

Where $\gamma_j \epsilon \Delta_j^{scm}$ has elements $\{\gamma_{ij}\}$ that satisfy $\gamma_{ij} \geq 0$ for all i, $\Sigma_i \gamma_{ij} = 1$, and $\gamma_{ij} = 0$ whenever i is not a possible donor. In other words, the vectors of weights must be non-negative and sum up to 1.

[258] Ben-Michael et al., *supra* note 19.

Given a N-vector of weights, $\hat{\gamma}_{ij}$, the SCM estimate for the treated unit j at time k, $Y_{jT_j+k}\infty$ is:

$$\hat{Y}_{jT_j+k}(\infty) \sum_{i=1}^{N} \hat{\gamma}_{ij}\, Y_{iT_j+k}$$

Ben-Michael et al. then propose the *partially pooled* SCM. This method chooses SCM weights to minimize the weighted average of the squared pooled and unit-specific pre-treatment fits:

$$\min_{\Gamma \epsilon \Delta^{scm}} \nu \left(\tilde{q}^{pool}(\Gamma)\right)^2 + (1-\nu)\left(\tilde{q}^{sep}(\Gamma)\right)^2 + \lambda||\Gamma||$$

Where the hyperparameter, $\nu\epsilon[0,1]$, and controls how much weight to place on the pooled fit relative to the separate fit of each donor unit.