Data Innovation Complementarity and Firm Growth*

Anastassia Fedyk† Roxana Mihet‡ Kumar Rishabh§ Orlando Gomes¶

February 12, 2026

Abstract

Cybercrime is expected to cost tens of trillions of dollars globally over the coming decade. This threat forces firms to invest in data security, much as external threats historically pushed states to build fiscal and military capacity. We ask whether such defensive investments can turn from necessary costs into a source of growth when integrated into firms' innovation processes. Using inventor-level data, we construct a firm-year measure of *"data-innovation complementarity"*—the extent to which data-security specialists are embedded in non-security R&D teams. Exploiting the staggered adoption of U.S. Data Breach Notification Laws as a shock to the salience and cost of data breaches, we show that firms with high ex-ante complementarity increase overall innovation and market value, while low-complementarity firms contract. The effect is specific to data-security expertise and robust to alternative constructions, controls, and placebo tests. A calibrated real-options model quantifies the importance of data-innovation complementarity for innovation-led growth.

Keywords: Organizational integration, data governance, data security, technology, innovation, patent teams, firm growth.

JEL-Codes: D8, O3, O4, G3, L1, L2, M1.

*The authors thank Tania Babina, Francesco Celentano, Fabrice Collard, Andreas Fuster, Leonardo Gambacorta, Michel Habib, Alex He, Luise Eisfeld, Gerard Hoberg, Boris Nikolov, Steven Ongena, Christian Peukert, Tarun Ramadorai, Luca Sandrini, Norman Schürhoff, Roberto Steri, Chi-Yang Tsou, Laura Veldkamp, Venky Venkateswaran, and Baozhong Yang for insightful suggestions, as well as participants at numerous conferences and seminars for useful feedback. Roxana Mihet acknowledges generous funding for this project from The Sandoz Family Foundation - Monique de Meuron Programme, as well as from UNIL's Dean's Office. All authors declare no conflicts of interest related to this project. All errors are solely the authors' responsibility.

†Anastassia Fedyk. Address: 545 Student Services Building 1900, Haas School of Business at UC Berkeley, Berkeley, CA 94720, United States. Email: fedyk@berkeley.edu.

‡Roxana Mihet. Address: Quartier UNIL-Chamberonne, Swiss Finance Institute at the University of Lausanne, Extranef 138, Lausanne, 1015 Vaud, Switzerland. Email: roxana.mihet@unil.ch.

§Kumar Rishabh. Address: Bank for International Settlements and University of Lausanne, Centralbahnplatz 2, 4051 Basel, Switzerland. Email: kumar.rishabh@unil.ch.

¶Orlando Manuel Da Costa Gomes. Address: Polytechnic Institute of Lisbon, Estr. de Benfica 529, 1549-020 Lisboa, Portugal. Email: omgomes@iscal.ipl.pt.

1 Introduction

Cybercrime is expected to cost $10.29 trillion globally by 2025, with projections reaching $16 trillion by 2029—an amount surpassing the gross domestic product of any country other than the United States and China. The cybersecurity threat forces firms to invest in protecting their data in much the same way that external threats historically pushed states to build fiscal and military capacity, with long-run consequences for development (e.g., Besley and Persson, 2009, Besley and Persson, 2011). A central organizational decision in this environment is whether to integrate data protection capabilities into the firm's broader operations or treat them as a stand-alone defensive cost. In this paper we examine how this choice shapes firms' ability to turn data security costs into innovation and growth.

We introduce a key concept of *"data innovation complementarity"*: the extent to which a firm's data-security specialists are structurally embedded in non-security R&D teams. Exploiting the staggered adoption of U.S. state-level Data Breach Notification Laws (DBNLs) as a shock to the salience and cost of data breaches, we compare otherwise similar firms with higher and lower data–security innovation complementarity across states that have and have not adopted DBNLs. We show that data security has the potential to be more than just a necessary compliance cost. While firms with low data-innovation complementarity experience DBNLs as a negative compliance cost, firms that develop data-security expertise in-house and integrate it into their broader R&D derive a benefit. These high data-innovation complementarity firms increase their general innovation output and long-run growth after DBNLs, because integrated security specialists build reuseable security-related architectures, contracts, and know-how across projects. This pattern is specific to data-security expertise and not managerial quality, firm size, or baseline patenting: analogous ''complementarity" measures constructed in other IT and non-IT domains do not yield similar results.

The dynamics we capture are illustrated by Amazon's experience: its early invest-

ments in secure payment infrastructure laid the groundwork for innovations such as the patented '1-Click' purchasing system, which in turn fueled both the company's growth and the broader e-commerce market. The Amazon case exemplifies a broader pattern in our data: firms that embed security specialists within cross-functional inventor teams are better positioned to collect high-value behavioral and transactional data that can also be used to fuel product development, personalization, and experimentation. Thus, data innovation complementarity converts a defensive capability into a source of competitive advantage. The DBNL shock serves as an impetus for these well-positioned firms to invest more in data security—investments that pay off with broader innovation gains.

We operationalize the *data–innovation complementarity* measure using inventor–level patent data from the USPTO. First, we identify *data-security specialists* as inventors whose lifetime portfolio contains a meaningful share of data-security-related patents, based on Cooperative Patent Classification (CPC) codes. Second, for each firm and year, we compute the share of data-security experts on the teams of the firm's major *non-security* patents and average this share over a rolling five-year window. This yields a firm-year measure of the extent to which security expertise has been integrated into the broader innovation process rather than siloed in a standalone function. The measure is constructed *ex ante* from pre-DBNL data and held fixed in the main analysis. We show that our results are robust to alternative measures including using a continuous index instead of a binary split, varying definitions of data-security inventors (different thresholds, minimum patent counts, lifetime vs. recent windows, broader or narrower CPC sets), using all patent teams rather than only major innovations, varying the team-size threshold, and using a citation-based notion of project importance. High-complementarity firms appear throughout the distributions of size and patent productivity, and the results are robust to controlling for lagged firm size, R&D, and past patenting. Taken together, the evidence suggests that our complementarity measure is not merely a proxy for scale, innovation intensity, or "good firms" in general.

Studying how this organizational choice shapes firms' responses to rising data-security

risk poses an empirical challenge. The recent rise of the data economy has made data security more relevant, and firms with integrated data security expertise have performed well, but these correlations are difficult to interpret: the value of data and security differs across firms and over time, and is likely endogenous to their activities. To obtain causal evidence, we need plausibly exogenous shifts in the relevance of data security that impact both high and low data-complementarity firms.

We address this challenge by exploiting quasi-experimental variation in data security relevance created by the staggered rollout of U.S. state-level Data Breach Notification Laws (DBNL). These laws, introduced between 2003 and 2018, require firms to disclose breaches involving personal data, sharply increasing the expected cost of mishandling information and the salience of data governance. Following the literature, we assign exposure based on firms' headquarters states and compare otherwise similar firms in adopting and non-adopting states, separately within the high- and low-complementarity groups.[1] The staggered adoption of the DBNLs offers substantial variation, with 27 states adopting them between 2003 and 2006 and another 18 states adopting them between 2007 and 2009. Using modern staggered difference-in-differences estimators, we examine how innovation and growth evolve around DBNL adoption for firms with different ex-ante levels of data–innovation complementarity, while absorbing firm fixed effects, calendar-year effects, and subgroup-by-year shocks, and showing clean pre-trends.

Our empirical analysis reveals three key insights. First, we document a starkly different response to data security concerns between high-data-innovation-complementarity firms and low-data-innovation-complementarity firms. Low-complementarity firms experience slight reductions in overall innovation and growth as they are forced to allocate more resources to data security. In contrast, high-complementarity firms experience an *increase* in all facets of patenting activity—including product patents—and growth in total assets and market capitalization. The effects are economically significant: for

[1] A large body of work has shown that headquarters-based assignment provides credible variation in firm exposure to legal changes (Bertrand and Mullainathan, 2003, Autor, Kerr, and Kugler, 2007, Giroud and Mueller, 2010, Kumar, Page, and Spalt, 2011, Atanassov, 2013, Hochberg and Rauh, 2013, Heider and Ljungqvist, 2015, Bena, Ortiz-Molina, and Simintzi, 2022, Atanassov, Julio, and Leng, 2024).

high-complementarity firms, DBNL adoption is associated with an average increase of 95 citation-weighted patents over the following five years, as well as a 0.45 increase in market-value-weighted patents within the high-complementarity subgroup. In contrast, low-complementarity firms experience a decline of 50 citation-weighted patents and a decrease of 0.44 in value-weighted patents over the same time period within the low-complementarity subgroup.

Second, we find that the increased innovation of high-complementarity firms translates into a long-lasting boost in growth, profitability, and market valuations, in contrast to the long-term decline experienced by low-complementarity firms. Relative to the levels implied by the control group's trajectory (i.e., low-complementarity firms not affected by DBNLs), in the first five years post-DBNL, low-complementarity firms hold roughly 16.5 percent fewer assets, have 20 percent lower market capitalization, and about 5 percent lower sales (although the sales estimate is statistically insignificant, and their return on assets increases by around 3 percentage points). By contrast, high-complementarity firms show 7.5 percent higher asset levels, 36 percent higher market capitalization, 16 percent higher sales, and 6 percentage points greater return on assets than their counterfactual levels (i.e., high-complementarity firms not affected by DBNLs).

Third, we propose a mechanism for these results. We document a high degree of cross-pollination within high-complementarity firms: the expertise developed to manage and process data also drives advancements in other general innovations. To show this, we first document that both high- and low-complementarity firms increase their cyber-risk disclosures, confirming that all firms increase attention to data security following DBNLs. However, only high-complementarity firms observe increased knowledge transfers between data-security and non-data-security innovation.

We measure knowledge spillovers in two ways: cross-disciplinary teams and self-citations. First, looking at team composition, low-complementarity firms do not increase the overall involvement of their data security experts on non-data-security-related patents, only reallocating their data security experts' outside activities towards the "ma-

jor" patents. In contrast, high-complementarity firms increase the *overall* involvement of their security experts in non-data-security-related innovation, with a particular emphasis on "major" patents that does not come at the expense of other patents. Second, looking at citation patterns, high-complementarity firms see the share of self-data-security-patent citations in their non-data-security-related patents increase by approximately 0.2 percentage points. In contrast, low-complementarity firms experience insignificant changes. Thus, the mechanisms we explore suggest that high-complementarity firms not only increase their focus on data security innovation, but also further integrate it into broader innovation activities, facilitating knowledge spillovers.

We conduct extensive robustness and placebo tests to rule out alternative explanations due to measurement, composition, timing, and treatment assignment. First, our results are invariant to alternative definitions of data-security inventors (different thresholds, minimum counts, lifetime vs. recent patenting windows, and broader or narrower CPC sets). Second, the results are unchanged when we vary the construction of the complementarity measure (using all teams rather than "major" patent teams, tightening or relaxing team-size cutoffs, and citation-weighting patents) or freeze the complementarity classification using pre-DBNL data. Third, to mitigate any composition concerns, we confirm that the estimates survive industry-by-year fixed effects and the exclusion of technology-intensive industries (where high-complementarity firms are prevalent), as well as dropping top-patenting and early-adopting states. Fourth, to verify that DBNL responses do not reflect generic specialized expertise, we reconstruct the complementarity measure in two non-security domains: "packaging" (orthogonal to data/IT) and "computer graphics" (IT-adjacent, but non-security) and show no DBNL response for these arbitrary technologies, supporting a mechanism specific to data-security know-how. Fifth, our identification is not driven by timing choices: both the estimated magnitudes and the pre-trends remain intact when we shift the treatment window from DBNL enactment to DBNL enforcement. Finally, the results are essentially unchanged when we control for lagged firm size, R&D, and past patenting, when we exclude prolific inventors, and

when we address potential SUTVA concerns by excluding very large (likely multi-state) firms, reinforcing the view that the documented heterogeneity is not driven by simple omitted-variable bias.

Taken together, the evidence supports an *organizational* mechanism: firms with sufficiently integrated data-security and innovation capabilities translate the DBNL shock into broad and durable gains in innovation and growth. While all firms increase their focus on data security following DBNLs, only high-complementarity firms are able to redeploy security experts across a wide array of innovation projects, enabling cross-domain knowledge transfer. These firms leverage their existing data-security inventors for broader innovation, resulting in increased self-citations from data-security patents to other patent classes, allowing useful data practices to travel with projects and teams to experiment at lower cost. In contrast, low-complementarity firms exhibit a narrower, compliance-driven response, allocating security expertise to only a small subset of projects, without meaningful knowledge spillovers or downstream productivity gains. As a result, for low-complementarity firms data security is a classical compliance cost, while for high-complementarity firms it actually creates an opportunity.

To interpret these findings and explore policy and counterfactual implications, we develop a parsimonious real-options model in which firms face an integration decision under uncertainty about potential data breach losses. The model generates several predictions borne out in the data: returns to integration are convex in data security relevance; innovation spillovers are stronger under internal integration than under outsourcing of data security expertise; and innovation responses are concentrated among firms with pre-existing complementarity. We validate these predictions empirically by analyzing both binary and continuous measures of complementarity. Using a continuous treatment design based on firm-year complementarity scores, we find monotonic increases in post-DBNL innovation and growth, consistent with the model. In policy analyses, we further show that policies that reduce the costs of integration, such as targeted subsidies that range from 4 percent of total firm size for medium-complementarity firms to 18 percent of total firm

size for lower-complementarity firms, can boost aggregate innovation without requiring universal compliance with security mandates. Moreover, in counterfactual simulations, we show that requiring all firms to outsource data security expertise, thereby preventing internal integration, would reduce aggregate firm value by up to a third, underscoring the innovation cost of one-size-fits-all security policies.

Literature review Our paper makes three core contributions. First, we introduce a new, scalable measure of firm-level integration between data security technologies and broader innovation capabilities, constructed from inventor-level patent data. This measure enables systematic analysis of an otherwise unobservable organizational feature and extends prior work that uses patents to study specialization, AI adoption, and creative destruction (Kakhbod, Kogan, Li, and Papanikolaou, 2024, Babina, Fedyk, He, and Hodson, 2024, Rishabh, Mihet, and Jang-Jaccard, 2024).

Second, our findings speak to the literature proposing potential mechanisms through which data technologies[2] can create value and affect long-run growth in data economies (Cong, Xie, and Zhang, 2021, Farboodi and Veldkamp, 2021): by shaping information production and improving forecasting in financial markets (Farboodi, Mihet, Philippon, and Veldkamp, 2019, Dessaint, Foucault, and Fresard, 2024), by influencing how economic agents learn from and respond to financial information (Kuhnen, 2015, Kuhnen and Miu, 2017, Das, Kuhnen, and Nagel, 2019), by generating competitive advantage through innovation and differentiation and through engagement with open-source communities (Rock, 2019, Tambe, Hitt, Rock, and Brynjolfsson, 2020, Ramadorai, Uettwiller, and Walther, 2025, Conti, Peukert, and Roche, 2025), by reducing costs and improving efficiency (Mihet and Philippon, 2019, Veldkamp, 2023), by ensuring regulatory compliance and shaping privacy regulations (Collis, Sen, and Wu, 2022, Peukert, Bechtold, Batikas, and Kretschmer, 2022, Florackis, Louca, Michaely, and Weber, 2023), and by building

[2]Broadly speaking, data technologies refer to the set of tools, systems, and capabilities that enable firms to protect data (cybersecurity, privacy-preserving computation), analyze data (machine learning, AI, analytics), manage data (data governance, data architecture, databases, cloud infrastructure), and ensure compliance (monitoring, disclosure tools, audit systems).

customer trust and loyalty or enhancing reputation and brand value (Cao, Nakhmurina, and Zhao, 2025). Our evidence aligns most closely with the competitive advantage mechanism: we show that internal integration of data-security and innovation capabilities enables firms to convert regulatory shocks into innovation-driven growth via cross-domain knowledge spillovers. We also find evidence for improved cost efficiency, as high-complementarity firms achieve greater output without proportionally higher input costs.

Third, we introduce and formalize a new mechanism, "*data innovation complementarity*," that explains how regulations affecting the relevance of data security can activate innovation spillovers when data protection is already structurally integrated into firms' innovation processes. This perspective complements prior work emphasizing the downside of cybersecurity risk, such as the effects of data breaches on reputation, and firm financing (Florackis et al., 2023), and on how stronger IT governance fosters more credible disclosures and fewer incidents (Bharadwaj, Cao, Tian, and Xue, 2025). Our findings shift the focus from risk mitigation to enabling innovation, showing that security capabilities embedded in R&D teams enhance firms' adaptive responses to external shocks. In doing so, we deepen the literature on intangible capital complementarities by emphasizing functional integration over investment intensity alone. This parallels Giroud and Mueller, 2019, who show that firms with stronger internal networks extract greater value from productivity shocks, underscoring how embedded organizational capabilities shape firm outcomes. We extend this logic to data security integration within innovation teams, demonstrating its strategic value beyond compliance.

Road-map The remainder of the paper proceeds as follows. Section 2 describes the data, introduces our novel data innovation complementarity measure, and details our empirical strategy. Section 3 presents our main empirical findings, and Section 4 discusses robustness and external validity. Section 5 introduces a real-options model to explore policy counterfactuals. Section 6 concludes.

2 Data and Empirical Strategy

In this section, we introduce the data and the methodology for constructing our novel measure of firm-level data innovation complementarity. After that, we describe the other data sets used in the paper: measures of innovation, firm financials, and operations. Finally, we detail our empirical strategy based on the staggered adoption of state-level Data Breach Notification Laws (DBNLs).

2.1 Data and Measures

Data Innovation Complementarity We introduce a novel, scalable method for measuring firm-level integration of data security with broader innovation capabilities using publicly available patent data from USPTO. Our measure captures the extent to which data security expertise is structurally embedded within a firm's broader innovation activities. Rather than relying on simple counts of data-related patents, our approach traces how firms allocate specialized data-security talent across broader innovation projects, revealing the organizational integration of data-related risk and innovation functions. This measure provides a replicable framework that can be applied over time and across industries, offering new insight into how data governance can become a meaningful driver of innovation and growth.

We construct the data innovation complementarity measure in three steps. First, we identify "data-security experts" to isolate a credible signal of specialized knowledge. A patent inventor is classified as a data-security expert if at least 10% of their lifetime patent portfolio consists of data-security patents, where we identify data-security patents using the cooperative patent classification (CPC), as described in the next subsection. The 10% threshold ensures genuine specialization, rather than incidental involvement. In Appendix A.3.7, we show that our findings are unchanged when we vary this threshold (10%, 25%, or 50%), impose minimum patent counts, use recent five-year windows instead of lifetime portfolios, or expand or restrict the underlying CPC security set.

Second, we focus on a firm's *major* non-data-security patents, defined as patents with an inventor team size above the firm's median. By construction, these patents' CPC codes are not related to data security, so the complementarity measure captures security specialists being redeployed onto *other* technological projects; it does not mechanically track the firm's level of data-security patenting. We concentrate on major patents because team size is a strong proxy for a project's strategic importance and resource intensity. However, our results are not sensitive to this definition. Appendix A.3.8 shows that our results persist when we define "major" projects based on citation impact rather than team size, and importantly when we drop the focus on "major" by considering "all" non-data-security patent teams as the baseline. For each considered non-data-security patent, we calculate the share of its inventor team comprising security experts. This step captures the deliberate allocation of scarce, high-value data-security human capital to a firm's core innovation projects.

Third, we aggregate the measure to the firm-year level. For each firm in a given year, we compute the average share of data-security experts across its major, non-data-security patents over the preceding five-year window. This rolling window is sufficiently long to reveal a consistent strategic orientation yet short enough to reflect the rapid pace of technological evolution. We classify a firm-year as having high data innovation complementarity if the average share of data-security experts in the firm's major non-data-security patents equals or exceeds the 75th percentile of the panel distribution. This classification is sticky: once a firm achieves high complementarity status, it remains a high-complementarity firm in subsequent years. The sticky classification rules out concerns that our complementarity measure could be endogenously updated after the DBNL shock (i.e., that firms might become high-complementarity because of DBNL or because prolific inventors are reassigned post-shock, so the measure would mechanically reflect treatment rather than a predetermined organizational feature). Unless otherwise noted, we fix the definition of a high data innovation complementarity firm using the 75^{th} percentile threshold throughout the main analysis. In addition to this binary split, we also

use the continuous firm-year complementarity score in a flexible treatment specification to assess the robustness and heterogeneity of our results.

Our data innovation complementarity measure captures an important firm characteristic, revealing through resource allocation how the firm embeds data expertise into its core operations. Prior evidence in other domains shows that cross-sharing of expert human capital can substantially improve firm performance due to knowledge spillovers (Fedyk, Fedyk, Hodson, and Khimich, 2023). Our paper applies this concept to data security and innovation, examining how embedded data-security expertise can drive broader innovation. Consistent with complementarity as a strategic organizational choice rather than a mechanical reflection of scale or patenting intensity, high-complementarity firms are broadly distributed across the entire range of firm size and patent-productivity levels (Figure A.5 in Appendix A.3.1). To verify that our results are specific to integrating security expertise, we construct analogous inventor-based complementarity measures in two non-security domains, "packaging" and "computer graphics", and do not observe the same results for these placebo tests (Appendix A.3.9).

Our complementarity measure varies naturally across industries, reflecting fundamental differences in business models and data integration strategies (Table 1). Data-centric service industries dominate the high-complementarity roster. Approximately 70 percent of the firms in the *Information* sector and roughly half of the firms in *Finance & Insurance* and *Professional Services* clear the 75th-percentile threshold. These sectors directly monetize user data or face significant regulatory exposure, creating strong incentives to embed security expertise into core innovation activities.

In contrast, the *Manufacturing* sector has 1,964 firms in our sample, but only 21 percent of them qualify as high-complementarity. This pattern is consistent with physical-asset-intensive business models, where security talent is typically siloed rather than integrated into mainstream R&D. Similarly, resource-heavy industries such as *Mining* and *Trade* display low rates of complementarity. Thus, our measure reflects a meaningful organizational choice, capturing the strategic integration of data-security expertise into

a firm's core innovation activities, rather than simply correlating with patent volume or firm size. Appendix Table A.6 reports further sector-level statistics on non-cybersecurity patent teams. Variation is large. In *Information*, an average major non-CS patent includes about one CS innovator (0.85). In *Healthcare*, the figure is 0.02, roughly one CS innovator for every 50 patents. One concern is that high complementarity merely reflects hiring inventors with prior data-security patents. This "mobility-dump" effect is unlikely. Our measure reflects active collaboration rather than passive IP accumulation.[3]

Table 1: Industry-Level Share of High-Complementarity (HC) Firms

NAICS Industry	**Number of Firms**	**Share of HC Firms**
Information (51)	349	71.3%
Finance & Insurance (52)	69	52.2%
Professional Services (54)	119	50.4%
Real Estate (53)	23	47.8%
Administrative & Support (56)	20	40.0%
Transportation & Warehousing (48–49)	18	33.3%
Utilities (22)	28	32.1%
Other (X)	49	26.5%
Manufacturing (31–33)	1964	21.2%
Mining & Oil/Gas Extraction (21)	38	18.4%
Retail Trade (44–45)	28	17.9%
Wholesale Trade (42)	30	16.7%
Health Care & Social Assistance (62)	21	9.5%

Legend: Firms are classified as high-complementarity (HC) if their average share of data-security specialists on major non–data-security patent teams over the prior five years exceeds the 75$^{\text{th}}$ percentile. A firm is counted as HC in this table if it met that threshold in at least half of its firm-year observations.

Finally, to assuage any concerns that results may be driven by any single industry with a large share of high complementarity firms, we conduct leave-one-out robustness tests. We sequentially exclude the *Information*, *Finance & Insurance*, and *Professional Services* sectors and rerun all key analyses. The results remain materially unchanged, confirming that complementarity captures an organizational advantage rather than industry-

[3]Year-one filings constitute only 20% of the five-year rolling average; major patents typically involve six or more inventors, diluting individual effects; and we focus on new filings under the hiring firm's name, not legacy assigned patents.

specific characteristics (Appendix A.3.2).

Innovation Measures We measure firms' innovation using their patent activity, following Kogan, Papanikolaou, Seru, and Stoffman, 2017 (KPSS dataset). Following the literature, we weight patents according to their scientific impact using forward citations (Kogan et al., 2017, Aghion, Van Reenen, and Zingales, 2013, Howell, 2017).[4] We correct for truncation bias in citations, since recent patents naturally have fewer citations, following the adjustment method proposed by Hall, Jaffe, and Trajtenberg, 2001. We also compute value-weighted patent counts following Kogan et al., 2017. Specifically, value-weighted patent counts are defined as the total market value created by patents upon award, summed over the filing year, scaled by firm assets in the filing year. Patent values measure the private economic value (as opposed to the scientific value captured by citations) derived from stock market reactions to patent awards. For robustness, we also use R&D expenditures from Compustat.

Product versus Process Innovation We classify patents by their technological focus using the claims dataset from Ganglmair, Robinson, and Seeligson, 2022. Patents primarily addressing new products or product improvements are classified as product patents. Patents centered on production methods and manufacturing processes are classified as process patents. We define a patent as product-related if at least 50 percent of its claims pertain to products; process patents follow an analogous classification.

Data-security-related Innovation We measure data-security-related innovation using the citation-weighted count of cyber-security patents filed by a firm in a given year. A patent is classified as a cyber-security patent if the United States Patent and Trademark Office (USPTO) assigns that patent at least one Cooperative Patent Classification (CPC) code associated with cyber security.[5] The time series of our aggregate cyber-

[4] More scientifically influential patents receive more citations (Hall, Jaffe, and Trajtenberg, 2005).

[5] The full list of codes appears in Table A.1; for example, CPC G06F21 denotes "Security arrangements for protecting computers, components, programs, or data against unauthorized activity."

security patent measure (computed across patents of all firms) is depicted in Figure A.1 in the Appendix. The figure indicates consistent growth in data security innovation over time. By 2022, the fraction of cyber security patents reaches 7% of all patent filings.

AI Innovation We identify AI inventions using the USPTO-based classification of Giczy, Pairolero, and Toole, 2022. This dataset assigns each U.S. patent granted since 1976 a predicted probability of belonging to one of eight AI subfields: machine learning, evolutionary computation, natural language processing, speech, vision, knowledge representation, planning and control, and AI-specific hardware.[6] A patent is classified as related to AI if its probability in any subfield exceeds 0.50. We calculate citation-weighted AI patent counts following the same methodology as the broad innovation measure described above.

Financial Data We obtain firm-level financial information from CRSP-Compustat.

Table A.2 in the Appendix summarizes the key variables used in the analysis, including outcomes and our high–data-complementarity classification. We winsorize all variables at 0.5% and 99.5% of the corresponding distribution.

2.2 Empirical Strategy

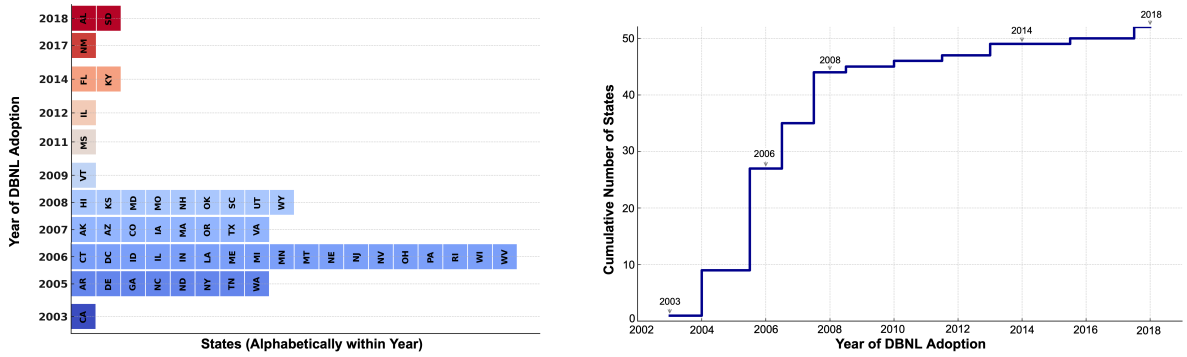
Data Breach Notification Laws (DBNLs) To estimate the causal impact of data security relevance on firm performance, we exploit the staggered introduction of Data Breach Notification Laws across U.S. states. DBNLs require organizations to notify individuals, regulators, and other stakeholders after data breaches involving unauthorized access or disclosure.[7] All 50 states enacted DBNLs between 2003 (California) and 2018

[6] Although our analysis covers the period 2000–2022, Giczy et al., 2022 identify AI-related patents even in earlier years. AI predates the currently dominant Gen-AI technology. Moreover, their classification encompasses a broad range of AI patents, including hardware-focused inventions.

[7] Data Breach Notification Laws in the United States require firms to notify affected individuals promptly (typically within 30 to 90 days) after discovering a breach. Notifications detail the nature of the breach, information compromised, and recommended protective actions. Certain states also mandate reporting breaches to state authorities or consumer agencies. Non-compliance penalties aim to ensure accountability in safeguarding personal data.

(Alabama, South Dakota). Over half of the states had introduced these laws by 2008 (Figure 1), but there was substantial variation in the exact timing: nine states introduced DBNLs in 2002–2004, another 18 states did so in 2006, and 17 states did so in 2007–2008. This variation makes DBNLs a sufficiently-powered source of quasi-random variation in firms' exposure to increased relevance of data security.[8]

Figure 1: Timeline and Cumulative Distribution of DBNL Adoption Across U.S. States.



Panel A: State-by-State Adoption

Panel B: Cumulative Adoption (CDF)

Notes: Panel A shows the state-by-state chronological adoption of Data Breach Notification Laws in the United States: blue represents earlier years, red represents later years. These laws mandate full disclosure to the public of any successful cyber-attack event on the firm. Within a given year, states are ordered alphabetically. Panel B shows the cumulative distribution of adoption.

DBNLs provide suitable exogenous variation to identify the economic effects of data security relevance. First, the staggered timing of DBNL adoption reflects state legislative processes, which are unlikely to correlate with individual firms' prior innovation or data complementarity profiles. We empirically confirm this assumption through pre-trend analyses. Second, DBNLs substantially raise firms' costs and consequences of data breaches, directly influencing firms' incentives to invest in data security (Boasiako and Keefe, 2021). Third, mandatory breach disclosures ensure accurate and consistent measurement of data breaches across states and time.

We compare innovation decisions of firms within the same data-complementarity group (high vs. high, low vs. low) headquartered in states adopting DBNLs earlier ver-

[8] A potential concern is that the relatively compressed adoption of DBNLs between 2003 and 2008 may offer limited variation, given that some outcomes, such as innovation (e.g., patenting), typically respond with a lag. We show that our design explicitly accounts for anticipatory adjustments and that the observed post-adoption divergence between high- and low-complementarity firms is long-lasting, mitigating concerns that the results are driven by short-term timing.

sus later, building on a long line of work that links staggered state laws to firm outcomes in corporate finance (e.g., Bertrand and Mullainathan, 2003, Giroud and Mueller, 2010, Heider and Ljungqvist, 2015). This comparison addresses the "forbidden comparisons" issue in staggered difference-in-differences designs (Goodman-Bacon, 2021); for identification with staggered adoption and heterogeneous effects we follow Callaway and Sant'Anna, 2021, Sun and Abraham, 2021, Borusyak, Jaravel, and Spiess, 2024. Although compliance involves operational costs, DBNLs also motivate firms to strengthen data management strategies.

A key challenge in our setting is that DBNLs formally govern the treatment of customers' personal data, and a firm headquartered in one state may already be bound by the laws of other states where it operates (leading to a SUTVA violation). Nevertheless, several considerations make headquarters-based assignment the most practical and relevant empirical strategy. First, corporate data security policies are typically centralized and set at the headquarters level, especially for large public firms (Hylender, Langlois, Pinto, and Widup, 2024). Second, firms are legally accountable to their state of incorporation and headquarters, which shapes their governance, compliance, and disclosure practices. Third, headquarters-based exposure has been the standard in prior work that exploits staggered state laws to identify causal effects (Bertrand and Mullainathan, 2003, Autor et al., 2007, Giroud and Mueller, 2010, Kumar et al., 2011, Heider and Ljungqvist, 2015, Bena et al., 2022, Atanassov et al., 2024). Finally, any spillovers across states would attenuate differences between treated and untreated firms, making our estimates conservative lower bounds. The fact that we still observe statistically significant and economically meaningful differences between high- and low-complementarity firms suggests that these concerns do not overturn our main conclusions.

Methodology To accurately identify the causal impact of DBNLs on firm outcomes, we rely on recent advances addressing key econometric challenges in staggered difference-in-differences (DID) estimation. The standard two-way fixed effects (TWFE) estimator,

historically prevalent in the literature, has recently been criticized due to biases from "forbidden comparisons" (Baker, Larcker, and Wang, 2022, Goodman-Bacon, 2021). These problematic comparisons arise when previously treated units incorrectly serve as controls for later-treated units, especially when treatment effects vary over time or across units, thus biasing the average treatment effect on the treated (ATT).

A second, unrelated issue arises from using patent counts as our outcome variable. Patent counts are highly right-skewed, and standard transformations such as $\log(1+y)$, log-linear, or inverse hyperbolic sine (IHS) produce inconsistent and biased estimates (Cohn, Liu, and Wardlaw, 2022). Negative binomial regressions address skewness but cannot incorporate firm fixed effects. Thus, we are limited to three estimators—linear, Poisson, and rate regressions—that yield unbiased and consistent fixed-effect estimates with right-skewed outcomes. Although among these three methods the linear regression approach typically exhibits lower efficiency (Cohn et al., 2022), we specifically choose linear regression because currently available estimators addressing the "forbidden comparisons" problem are developed primarily in a linear framework. Despite this efficiency trade-off, linear estimators remain unbiased and consistent.

We explicitly address the issue of "forbidden comparisons" by employing the imputation estimator introduced by Borusyak et al., 2024 (henceforth, BJS). While other recent estimators also address this issue, (e.g., Callaway and Sant'Anna, 2021, de Chaisemartin and d'Haultfœuille, 2020, Sun and Abraham, 2021) the BJS estimator is the most efficient, as it utilizes all pre-treatment observations and allows for flexible specification of the control group by choosing appropriate fixed effects (Roth, Sant'Anna, Bilinski, and Poe, 2023, Borusyak et al., 2024).

The BJS imputation estimator involves a three-step procedure:

1. Within untreated observations only, estimate firm fixed effects (α_i), calendar-year fixed effects (δ_t), and subgroup-specific calendar-year effects (δ_t^C), by ordinary least

squares (OLS) in the following model:

$$Y_{it} = \alpha_i + \delta_t + \delta_t^C \cdot HighC_{it} + \varepsilon_{it}, \tag{1}$$

where Y_{it} is the outcome for firm i in year t, and $HighC_{it}$ equals one if firm i is classified as high-complementarity in year t and zero otherwise. The inclusion of subgroup-specific calendar-year effects (δ_t^C) ensures that the constructed counterfactual outcomes (the *imputed untreated outcomes*, defined in the next step) rely exclusively on comparisons within the same subgroup. Thus, we always compare high-complementarity firms only to other high-complementarity firms, and low-complementarity firms only to other low-complementarity firms, ensuring apples-to-apples comparisons.

2. For each treated observation (firm-year combination affected by DBNL), construct the imputed untreated outcome using the estimates from step 1:

$$\hat{Y}_{it}^{(0)} = \hat{\alpha}_i + \hat{\delta}_t + \hat{\delta}_t^C \cdot HighC_{it}. \tag{2}$$

The individual treatment effect is then calculated as:

$$\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}^{(0)}. \tag{3}$$

3. Compute subgroup-specific Average Treatment on the Treated (ATT) at horizon k as a weighted average of individual treatment effects within each subgroup

$$ATT_g(k) = \sum_{i:HighC_{it}=g,\, K_{it}=k} w_{it}\, \hat{\tau}_{it}, \tag{4}$$

where $g = 1$ indicates the high-complementarity subgroup, and $g = 0$ indicates the low-complementarity subgroup. The weights w_{it} are pre-specified, non-stochastic, and depend exclusively on treatment assignment and timing, following Borusyak

et al., 2024. Specifically, within each subgroup-by-horizon cell, observations are weighted equally: $w_{it} = \frac{1}{|\{i: HighC_{it}=g,\, K_{it}=k\}|}$, where $K_{it} = t - E_i$ is the number of periods since treatment for firm i. This weighting scheme ensures that each firm-year observation within a given subgroup and horizon is given equal importance.

To assess heterogeneous responses, our estimation employs a triple-difference (i.e., DDD or Difference-in-Difference-in-Difference) approach, specifically designed to quantify DBNLs' impact on high-complementarity firms in comparison to its impact on low-complementarity firms. The triple-difference (DDD) estimate at horizon k, capturing the differential impact of DBNLs on high-complementarity firms relative to low-complementarity firms, is defined as:

$$DDD(k) = ATT_{HighC}(k) - ATT_{LowC}(k). \tag{5}$$

In all specifications, we include firm fixed effects, calendar-year fixed effects, and subgroup-by-year effects; in Section 4 we show that our results are robust to additionally including two-digit-industry-by-year fixed effects and time-varying controls for firm size, R&D, and past patenting. Moreover, in all specifications, standard errors are clustered at state level, aligning with the level of treatment variation prescribed by Abadie, Athey, Imbens, and Wooldridge, 2023.

Given the institutional context, firms likely anticipate DBNLs around the time of their passage, potentially adjusting behavior prior to the laws' effective dates. Supporting this, our textual analysis indicates that firms increasingly discuss cybersecurity risks approximately one year before the actual implementation of a DBNL (Appendix Figure A.2). Following recommendations by Borusyak et al., 2024 and Callaway and Sant'Anna, 2021, we explicitly account for this anticipation by shifting our treatment definition one year earlier. Specifically, we redefine the treatment date from the actual enforcement year to the preceding year, when the law is publicly enacted. This adjusted timing ensures that our empirical strategy accurately captures behavioral responses starting

from the credible information event. For robustness, in Appendix A.3.5, we present the main results without this anticipatory shift. Although overall ATT results remain robust, anticipation effects are evident as some outcomes exhibit significant responses in the year preceding enforcement. By contrast, there are no pre-treatment effects in our main specification incorporating the anticipatory shift, supporting our adjusted timing approach.[9] This design choice also addresses the concern, noted above in our description of DBNL adoption, that the relatively compressed rollout may not leave sufficient time for some economic outcomes to adjust. By shifting the treatment definition one year earlier and verifying robustness using the actual enforcement year, we ensure that our results capture both anticipatory responses and sustained post-adoption effects. The sustained divergence between high- and low-complementarity firms alleviates concerns that our findings are an artifact of short-term timing.

Separately, we conduct a diagnostic pre-trends test in the difference between high-complementarity and low-complementarity firms by following the recommended practice from Borusyak et al., 2024 and estimating an OLS regression on untreated observations only, including lead indicators interacted with the subgroup status (i.e., high-complementarity and low-complementarity firms) and using the same fixed effects and standard-error clustering structure as above. We obtain the subgroup-specific pre-treatment coefficients and a joint Wald test statistic.

In our baseline triple-difference specification, we include firm fixed effects, calendar-year fixed effects, and subgroup-by-year fixed effects, but we deliberately exclude time-varying firm-level covariates. Identification therefore relies strictly on parallel trends in untreated potential outcomes within subgroup-year cells, in line with Borusyak et al., 2024, who recommend conditioning only on covariates that are predetermined with re-

[9] To implement the heterogeneous treatment analysis described above, we use the `project` option in the `did_imputation` Stata package by Borusyak et al., 2024, employing the fixed-effects structure outlined previously. The *project* option performs a post-estimation regression of individual treatment effects ($\hat{\tau}_{it}$) on subgroup indicators (high-complementarity vs. low-complementarity firms). When needed, instead of DDD estimates, we separately report subgroup-specific ATTs, following the same procedure with the `hetby` option rather than `project`, estimating pre-trends individually rather than as differences. The package allows for correction of the anticipatory effects through the option `shift`.

spect to treatment. By contrast, variables such as firm size, profitability, or R&D are themselves potential outcomes of the DBNL shock; conditioning on them risks introducing "bad control" bias in the sense of Roth et al., 2023. We therefore omit these variables from the primary specification and instead treat them as outcomes of interest. In additional robustness tests reported in Appendix A.3.1, we also estimate specifications that add one-year-lagged log assets and patent filings as controls.

3 Empirical Results

In this section, we explore the impact of the Data Breach Notification Laws (DBNL) on firm innovation activities, as proxied by patenting activity, profitability, and market power. The results show different patterns for firms with low versus high data complementarity, whereby the DBNLs impose a cost on the former but create an opportunity for the latter. We end by discussing the mechanism behind this result: expertise spillovers from data-security-related to non-data-security-related innovation.

3.1 Results: Innovation Responses

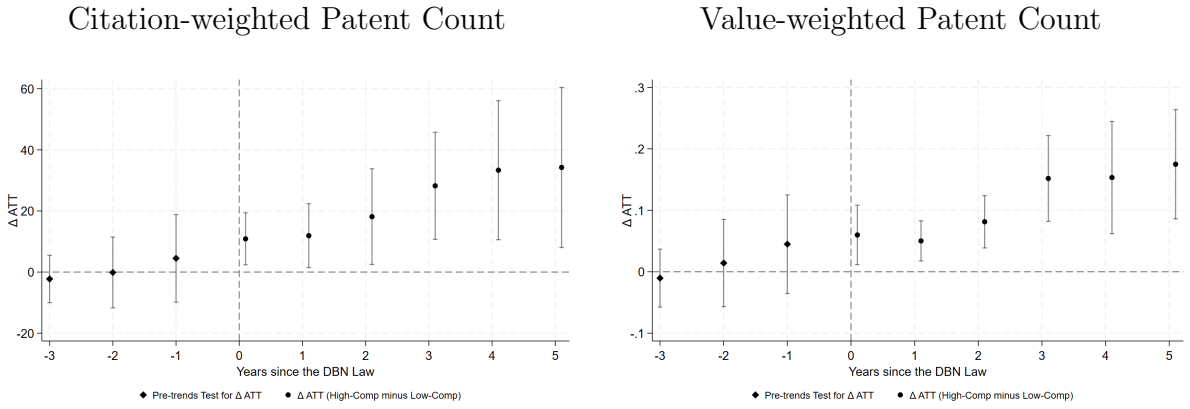
General Innovation Data Breach Notification Laws impact firms by increasing their compliance costs and the strategic importance of data governance. This is likely to impose short-term frictions, leading to a negative effect on firm performance. At the same time, the laws incentivize firms to invest in capabilities—such as cybersecurity integration—that support safer, more scalable data use. This has the potential to generate positive spillover effects for firms' innovation. Ultimately, whether DBNLs increase or decrease firms' innovation depends on whether firms treat security as a constraint or as a complementary input to data-driven innovation. This, in turn, depends on each individual firm's ex-ante organizational structure.

We examine the impact of DBNLs on firm innovation using two measures: citation-weighted patent counts, a common proxy for innovation quality and impact, and value-weighted patent counts following Kogan et al., 2017, reflecting the market valuation of

innovation. For each firm, we examine the time period from three years prior to the year of DBNL adoption in its headquarters state to five years after the announcement.

Figure 2 presents the results for the two measures of innovation using the DDD estimator. The left panel shows estimates for citation-weighted patent counts, while the right panel examines patent value. In both panels, we observe no significant pre-trends prior to DBNL adoption, supporting the parallel-trends assumption. After DBNL adoption, high-complementarity firms significantly increase their innovation relative to low-complementarity firms. These results remain robust even when we exclude industries that are most data-intensive (e.g., *Information*, *Finance & Insurance*, *Professional Services*), confirming that the effects are not concentrated in specific industries (Appendix A.3.2). We also show that our findings remain robust when including industry-by-year fixed effects and tightening identification further by combining subgroup-by-year and industry-by-year fixed effects (Appendix A.3.2).

Figure 2: Innovation: Triple Difference Estimates



Legend: This figure plots triple-difference (DDD) estimates from BJS regressions around the staggered adoption of DBNLs across U.S. states. Year "0" denotes one year before a DBNL takes effect, explicitly accounting for potential anticipation effects. We find no pre-trends. All regressions include firm, year, and year-by-subgroup fixed effects, ensuring treatment versus control comparisons remain within subgroup. Standard errors are clustered at the state level. Vertical bars show 95% confidence intervals.

This divergence becomes economically meaningful over time. By the third year, high-complementarity firms file roughly 27 additional citation-weighted patents relative to the low-complementarity group (this corresponds to approximately half the mean of

the outcome variable). Over the first five years following DBNL, high-complementarity firms see approximately 95 citation-weighted patents. Similarly, by the third year, high-complementarity firms produce 0.15 more financially valuable patents compared to their low-complementarity counterparts (for a median firm worth $933m, this corresponds to filing patents that, on average, create an addition $\$933m \times 0.15 = \$140m$ lifetime value). This pattern is consistent with high-complementarity firms deriving long-term benefits from the stronger internal innovation capabilities enabled by the increased investment in data security.

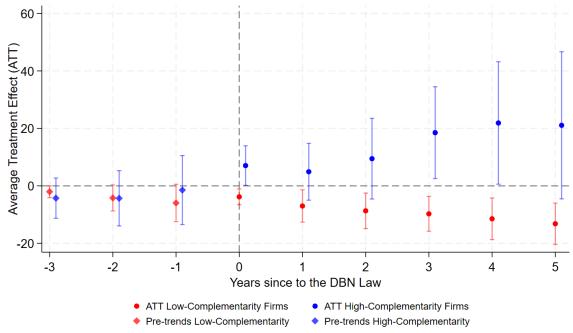
Additional analyses presented in Appendix Figure A.3 confirm similar differential responses across other innovation dimensions, including AI-related patents, product patents, and process patents. The same figure also presents the triple-difference estimate for firm R&D expenditures (log-transformed), again showing insignificant pre-trends. Post-DBNL, the difference in subgroup-specific ATT estimates between high-complementarity and low-complementarity firms significantly widens, indicating that firms with complementary innovation capabilities actively increase their resource allocation to innovation inputs, whereas firms lacking such complementarities either reduce or fail to sustain their pre-DBNL levels of R&D investment and various forms of innovation.

To understand the underlying drivers of the differential innovation response documented above, Figure 3 provides separate ATT estimates for high- and low-complementarity firms. These ATT estimates are derived from within-group comparisons (high vs. high, low vs. low). Both citation-weighted patents (left panel) and patent values (right panel) show no significant pre-treatment trends in either group separately. Post-DBNL, innovation among high-complementarity firms significantly increases, while low-complementarity firms experience a notable decline. This result highlights that DBNLs indeed impose a compliance cost on the low-complementarity firms. For high-complementarity firms, however, DBNLs activate a positive feedback loop, increasing the relative value of complementary data-security capabilities and benefiting the firms that possess them.

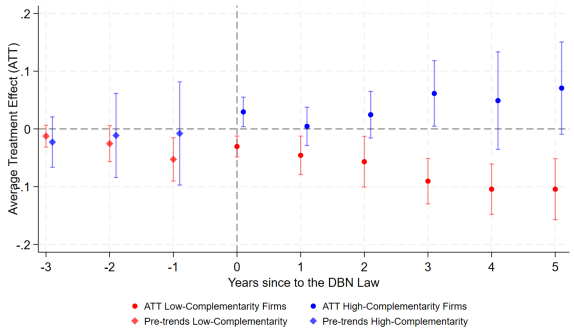
Overall, these results provide robust evidence that DBNLs activate innovation syn-

Figure 3: Innovation: DID Estimates by Complementarity Subgroup

Citation-weighted Patent Count



Value-weighted Patent Count



Legend: This figure plots DID estimates separately for high- and low-complementarity firms from BJS regressions around the staggered adoption of DBNLs across U.S. states. Year "0" denotes one year before a DBNL takes effect, explicitly accounting for potential anticipation effects. We find no pre-trends. All regressions include firm, year, and year-by-subgroup fixed effects, ensuring that treatment versus control comparisons remain within subgroup. Standard errors are clustered at the state level. Vertical bars show 95% confidence intervals.

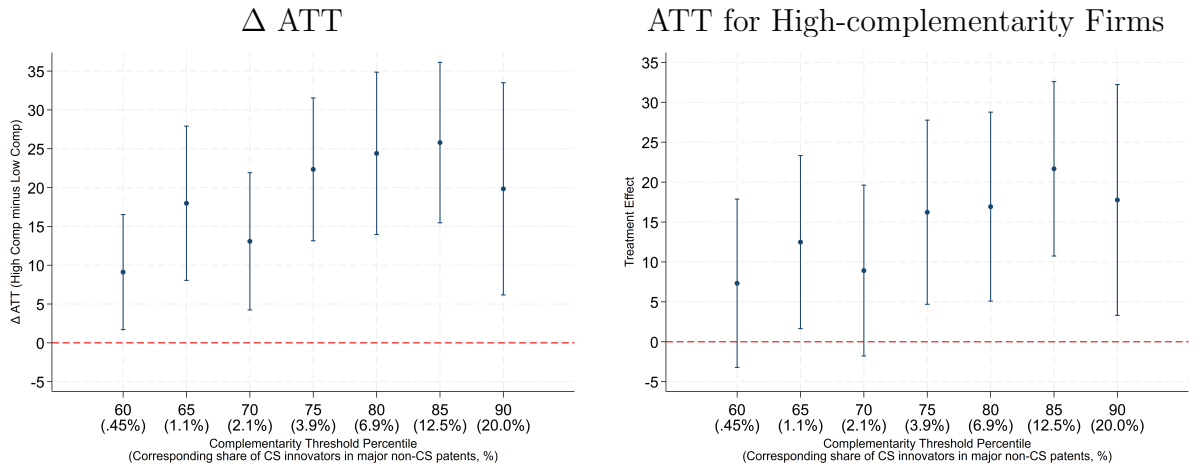
ergies specific to high-complementarity firms, reinforcing the economic significance of internal data-security capabilities in driving broader long-term benefits from firm innovation activities.

Innovation by Complementarity Threshold We examine whether the effects of increased data security relevance vary continuously with firms' degree of data innovation complementarity. For this analysis, we estimate a flexible specification allowing for heterogeneous treatment effects across the complementarity threshold distribution.

Figure 4 plots the differential innovation response to DBNL adoption, as we vary the exact threshold used to classify high- versus low-complementarity firms. We consider thresholds from the 60th percentile (which would classify 45% of firms as high-complementarity) to the 90th percentile (which would classify 20% of the firms as high-complementarity). The left panel shows the difference in the average treatment effect (in terms of citation-weighted patents) between high- and low-complementarity firms. The right panel shows DID estimates for high-complementarity firms only.

As can be seen from the left panel of Figure 4, high-complementarity firms consis-

Figure 4: Innovation Response to DBNL by Complementarity Threshold



Legend: This figure plots triple difference (DDD) estimates [left] and difference-in-differences (DID) estimates in citation-weighted patent count for high-complementarity firms [right] around the staggered adoption of DBNLs across U.S. states. The threshold for defining high-complementarity firms varies from 60% to 90%, as defined in the main text. Year "0" denotes one year before DBNL adoption. All regressions include firm, year, and year-by-subgroup fixed effects. Standard errors are clustered at the state level. Vertical bars show 95% confidence intervals.

tently outperform low-complementarity firms by approximately 10–25 additional citation-weighted patents per year across all complementarity thresholds. That said, the effect increases at higher complementarity thresholds, consistent with the firms at the top of the complementarity distribution being especially strong at harnessing data security efforts for complementary innovation. In the right panel, we focus just on high-complementarity firms to assess the complementarity threshold at which innovation starts to significantly increase. Below the 75% complementarity threshold, treatment effects on high-complementarity firms remain insignificant. The effects become positive and statistically significant above this threshold, reaching approximately 15–20 additional citation-weighted patents per year. These patterns are consistent with the idea that internal spillovers between security and innovation are necessary to unlock gains from regulatory shocks.

The innovation gap between high- and low-complementarity firms suggests that while DBNLs create compliance burdens for all firms, firms with sufficient complementarity can translate the regulatory shocks into gains via innovation spillover. This is especially

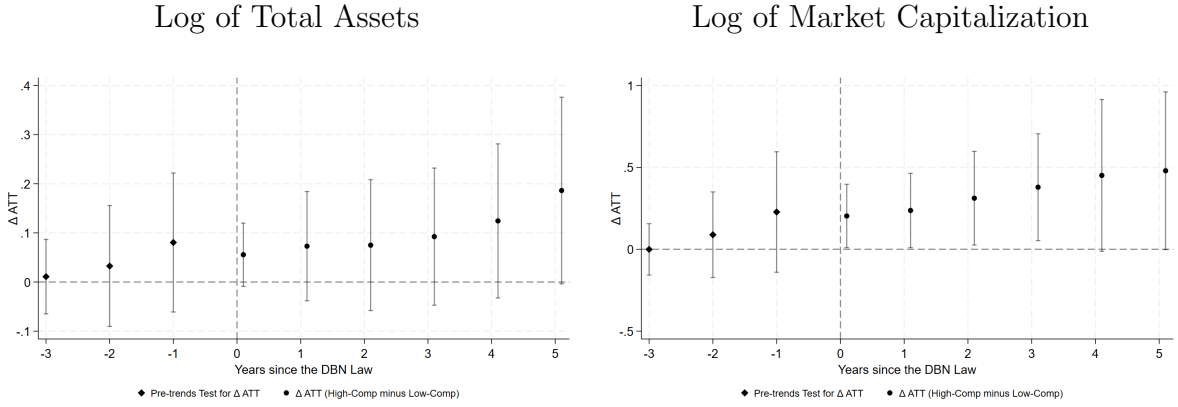
pronounced at higher complementarity thresholds, which isolate firms with greater integration capabilities. In all subsequent empirical analyses, we present results for the complementarity threshold of 75%. The results hold and often become even stronger for higher complementarity thresholds (e.g., 85%).

3.2 Results: Firm Growth Responses

We now examine whether the differential innovation responses of high- and low-complementarity firms to DBNLs translates into tangible differences in firm operational performance. We consider several measures of operational performance: total assets, market capitalization, return on assets, and sales growth.

Figure 5 presents triple-difference estimates of changes in high- versus low-complementarity firms' total assets (left panel) and market capitalization (right panel) upon DBNL enforcement. Both panels provide clean identification: pre-treatment coefficients are statistically zero, validating no significant pre-trends.

Figure 5: Firm Growth: Triple Difference (DDD) Estimates



Legend: This figure plots triple-difference estimates from BJS regressions around the staggered adoption of DBNLs across U.S. states. Year "0" denotes one year before the DBNL takes effect, explicitly accounting for potential anticipation effects. We find no pre-trends. All regressions include firm, year, and year-by-subgroup fixed effects, ensuring treatment versus control comparisons remain within-subgroup. Standard errors are clustered at state level. Vertical bars show 95% confidence intervals.

For firm size (total assets), growth differences emerge gradually. The gap appears but is statistically insignificant in the first three years after enforcement, then expands steadily

to approximately 0.19 log-points (around 21 percent) by year five. This gradual expansion suggests tangible balance-sheet growth enabled by security-innovation complementarity. Thus, DBNLs do not merely stimulate innovation; they trigger a sustained accumulation of real productive resources in firms positioned to leverage data protection advances in their broader innovation efforts These results remain robust when we sequentially exclude each major industry (*Information*, *Finance & Insurance*, *Professional Services*, and others), confirming that the effects are not concentrated in a handful of data-intensive sectors (Appendix A.3.2).

Market capitalization responses are larger and more rapid than the effects on assets. By the second year after DBNL enforcement, the gap in market capitalization responses between high- and low-complementarity firms reaches about 0.25 log-points, doubling to approximately 0.50 log-points (around 65 percent) by year five. The faster and larger market valuation increase compared to the change in firms' total assets reflects investor expectations: financial markets swiftly anticipate and capitalize the future payoff streams from enhanced data capabilities, well before these assets appear on firms' balance sheets. However, it is worth remembering that these effects capture the *difference* in the responses of high- versus low-complementarity firms, which may also reflect the costs borne by the low-complementarity firms.

The first two columns of Table A.8 examine separately (i) the effect on low-complementarity firms and (ii) the differential effect on high- versus low-complementarity firms. These results show that low-complementarity firms experience significant negative operational effects from DBNL enforcement: -0.18 log-points in assets and -0.22 log-points in market capitalization. Adding together the ATT on low-complementarity firms and the difference in ATT between high- and low-complementarity firms, high-complementarity firms experience an absolute *increase* of 0.07 log-points in total assets and 0.31 log-points in market capitalization.

Thus, investors seem to assign substantial valuation premia to firms integrating data-related expertise into their non-data-related innovation processes, when data relevance

increases with the passage of DBNL. The effects on market valuations are substantially larger than those on asset growth, highlights the important role of intangible investments, such as proprietary datasets, algorithms, or specialized expertise, which are not fully captured by standard accounting measures. The significant valuation effects also imply that investors lower their required returns, potentially reflecting reduced perceived risks as robust data security lowers anticipated breach-related costs. In terms of operational performance, the last two columns in Appendix Table A.8 show that high-complementarity firms achieve higher return on assets and stronger sales growth than low-complementarity firms following DBNL adoption.

Taken together, our results suggest that DBNLs may have effects beyond imposing compliance costs. While low-complementarity firms do suffer from the additional costs imposed by the laws, for firms with high data-complementarity, DBNLs appear to be associated with greater growth and innovation, consistent with the idea that integrating data specialists into broader innovation teams can augment competitive advantage. The observed divergence in innovation, profitability, and valuation outcomes is consistent with a reinforcing pattern, where investments in data-related capabilities support broader innovation and firm performance over time.

3.3 Mechanism: Knowledge Spillovers

The divergent responses of high- and low-complementarity firms' innovation, growth, and valuation outcomes to DBNLs are consistent with a mechanism whereby high-complementarity firms leverage their increased cybersecurity-related innovation in response to heightened data relevance for other, non-cybersecurity-related product creation. We analyze this mechanism directly by examining how firms integrate data-related talent into broader R&D operations following DBNL adoption.

Strategic Capability versus Compliance Table 2 directly examines the integration of cybersecurity expertise in non-cybersecurity-related innovation. We begin by

documenting that both high- and low-complementarity firms expand their cyber-risk disclosures by approximately +0.0105*** on a 0–1 scale post-DBNL (Column 1), indicating similar compliance-oriented attention to cybersecurity threats between the two groups of firms. However, the corresponding gains in innovation and growth materialize only for high-complementarity firms, suggesting that DBNLs act as a common nudge, but only certain firms are organizationally positioned to capitalize on it.

Table 2: Mechanisms: Organizational Integration and Innovation Spillovers

	Cybersecurity	Share of CS Experts in		Share of Self CS Citations	Citation-weighted
	risk disclosure	Major Non-CS patents	All Non-CS patents	In Non-CS patents	CS patent count
	(1)	(2)	(3)	(4)	(5)
Δ ATT Post	-0.0017	0.0036	0.0174*	0.0017***	1.0845
	(0.0106)	(0.0095)	(0.0103)	(0.0004)	(0.8937)
ATT Low Comp Post	0.0105***	0.0010*	0.0015	0.0000*	0.0756
	(0.003)	(0.0006)	(0.0012)	(0)	(0.833)
Pre-trends Test					
Δ ATT Pre (-1)	0.0003	-0.0046	0.0026	0.0003	-1.3665
	(0.0046)	(0.0135)	(0.013)	(0.0007)	(1.5183)
Δ ATT Pre (-2)	0.0022	-0.0057	-0.0012	0.0005	-0.8188
	(0.0046)	(0.0102)	(0.0091)	(0.0005)	(1.1097)
Δ ATT Pre (-3)	0.0031	-0.0007	-0.0035	0.0006	-0.494
	(0.0037)	(0.0068)	(0.005)	(0.0005)	(0.617)
ATT Low Comp Pre (-1)	0.0022	0.0011	0.0039	0.0002*	-0.0111
	(0.0027)	(0.0009)	(0.0026)	(0.0001)	(0.1906)
ATT Low Comp Pre (-2)	0.0022	0.0009	0.003	0.0001	-0.0852
	(0.0023)	(0.0008)	(0.0022)	(0.0001)	(0.1624)
ATT Low Comp Pre (-3)	0.0005	0.0002	0.0013	0.0001	-0.1675
	(0.0011)	(0.0005)	(0.0014)	(0.0001)	(0.1152)
N Observations	10,335	17,179	17,592	13,546	17,905

Legend: This table presents triple-difference (DDD) estimates from BJS regressions around the staggered adoption of DBNLs across U.S. states. The DDD estimator compares subgroup-specific ATT estimates for high-complementarity and low-complementarity firms in the post-DBNL period (high-complementarity minus low-complementarity). The table also presents ATT estimates for the low-complementarity group. In all regressions (including pre-trends diagnostics), we include firm, year, and year-by-subgroup fixed effects. Standard errors are clustered at the state level. See Table A.1 for details on the variables. $^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Inventor Deployment Patterns The strategic difference lies in how firms deploy data-related (i.e., cybersecurity) talent. As can be seen from Columns 2 and 3 of Table 2, low-complementarity firms respond to DBNLs by concentrating data security experts on a small set of "major" non-cybersecurity patents, increasing the expert share on these projects modestly (+0.10 percentage points, significant at the 10% level; Column 2) at the expense of other non-cybersecurity patents—with the total change in the presence of cybersecurity experts on non-cybersecurity patent teams remaining flat for these firms. By contrast, high-complementarity firms increase the share of cybersecurity inventors broadly across their entire patent portfolio, with a significant and economically sizable increase of +1.74 percentage points (Column 3).

These patterns are not an artifact of labeling only large teams as "major" patenting. Repeating the analysis under alternative definitions of major patenting (e.g., using citation-based measures of patent importance) and including all team sizes yields comparable differential effects (Appendix Table A.18).

Overall, low-complementarity firms have modest cybersecurity spillovers that concentrate in a handful of projects, whereas high-complementarity firms diffuse it broadly. Since the classification of high- versus low-complementarity firms is based on pre-DBNL patterns and remains fixed post-DBNL, the observed innovation shifts represent new margins of organizational adaptation rather than mechanical reassignment.[10]

Knowledge Transfer and Spillovers Broader deployment of data-security-related talent leads to more effective internal knowledge flows. Column 4 in Table 2 shows that only high-complementarity firms experience a statistically significant increase in the share of self-citations of cybersecurity patents in non-cybersecurity patents ($+0.17^{***}$ percentage points, compared to $+0.00^{*}$ for low-complementarity firms). These knowledge spillovers highlight a key finding: innovation gains stem not from more cybersecurity patents per se, but from the reuse of security expertise across technology domains.

Appendix Figure A.4 visually illustrates this divergence in citation behavior, dynamically. Low-complementarity firms see a flat line in self-citations of cybersecurity patents in non-cybersecurity patents. In contrast, high-complementarity firms see an immediate increase in the first few years, which grows larger starting around three years after DBNL enforcement.

Innovation Reallocation and Economic Upside While both firm types experience moderate increases in cybersecurity patenting in Table 2, Column 5 (ATT of roughly

[10]The initial complementarity classification is based solely on the share of cybersecurity inventors on major patents before DBNL enactment. Post-DBNL, this original metric remains stable for high-complementarity firms and increases only modestly for low-complementarity firms, confirming the absence of classification bias. The emergence of a new margin—deploying cybersecurity talent across the broader patent portfolio—is therefore an active organizational adjustment uniquely present for high-complementarity firms.

+1 for citation-weighted patents, though imprecise), only high-complementarity firms translate this into broader innovation and performance gains. This distinction suggests that the economic benefits of DBNLs hinge not on compliance alone, but on firms' ability to reallocate existing talent in ways that create firm-wide innovation synergies.

Organizational Mechanism The results in Table 2 shed light on the organizational mechanism behind data–innovation complementarity. Our findings suggest that the key organizational distinction between high- and low-complementarity firms lies in the breadth and consistency of integration of cybersecurity expertise in non-cybersecurity endeavors. High-complementarity firms systematically embed data-security experts across a wide array of major and non-major innovation projects, as reflected in a significant increase in the share of security specialists on non-security patents and in the rate of self-citations from data-security to other patent domains. In contrast, low-complementarity firms tend to allocate their security experts narrowly to a few flagship projects, with little evidence of broader knowledge spillovers.

This pattern indicates that high-complementarity firms possess organizational routines and team structures that facilitate cross-functional collaboration between security and broad innovation activities *at scale*, enabling them to redeploy talent and integrate expertise more effectively and more broadly across the firms' operations in response to shocks such as DBNL. By contrast, low-complementarity firms appear to treat security expertise as siloed and project-specific, limiting their ability to translate compliance-driven investments into firm-wide innovation gains.

4 Robustness and External Validity

In this section, we explicitly address three sets of concerns that naturally arise in our setting: (i) whether data-innovation complementarity proxies for other firm characteristics such as size or managerial quality, (ii) whether our results are sensitive to measurement choices (thresholds, "major" patents, CPC definitions), and (iii) whether

the documented effects are specific to data-security expertise rather than generic specialization.

Specifically, we rule out seven broad classes of alternative explanations. First, we ensure that our results are not a simple reflection of differences in firm size, resources, or prior innovation intensity. Second, we confirm that the results are not driven by industry composition or geographic clustering, including the possibility of industry-specific year-to-year shocks. The effects are also not concentrated in a handful of states or early adopters. Third, we confirm that the results are not an artifact of our treatment assignment at the headquarters level rather than reflecting firms' actual consumer segment exposure to DBNLs. Fourth, we address the concern that our findings arise from endogenous anticipation of regulation, rather than from exogenous shocks. Fifth, we confirm that the results are not specific to our definition of a "data-security expert" or driven only by prolific inventors. Sixth, we show that the results are not driven by our choice of "major" patents or by star-inventor concentration.

Lastly, a matched falsification test that reconstructs the complementarity measure in non-security domains ("packaging" and "computer graphics") yields null or attenuated DBNL responses, supporting a data-security–specific mechanism rather than generic specialized expertise.

First, to address the concern that high-complementarity firms might simply be larger, more innovative, or better resourced, we explicitly control for firm size, past patenting activity, and R&D expenditures in both our main and robustness specifications (Appendix *A*.3.1). The differential impact of DBNLs on innovation and growth remains statistically and economically significant even within firms of similar size and innovation intensity. When we add one-year-lagged log assets and patent filings as controls (Appendix A.3.1), the estimated DBNL effects on innovation and growth remain essentially unchanged in magnitude and significance, indicating that baseline our results are not driven by the exclusion of these covariates. This supports our interpretation that the observed spillovers stem from the organizational integration of data-security and innovation capabilities,

rather than from scale, ability, or resource advantages.

Second, we conduct leave-one-industry-out tests across all NAICS sectors, drop top-innovating states, and redefine the event year using the exact enforcement date to show robustness to these alternative specifications (Appendices *A*.3.2-*A*.3.5). Excluding even the most data-intensive industries (e.g., *Information*, *Finance & Insurance*, *Professional Services*), or data-light sectors (e.g., *Utilities* and *Mining & Oil/Gas*) does not overturn our results, confirming that our mechanism operates broadly across the economy. As an additional check, our results remain stable when controlling directly for industry-by-year shocks, or when including both subgroup-by-year and industry-by-year fixed effects simultaneously (Appendix A.3.3).

Dropping top-innovation states, which are also states that adopted DBNLs early (e.g., California and Texas), demonstrates that the results are not driven by a particular state and hence possess broad external validity (Appendix *A*.3.4). Similarly, redefining the event year using the exact enforcement date (Appendix *A*.3.5) reveals anticipatory responses consistent with our identification strategy and confirms the timing choice in our empirical design.

Third, we acknowledge that our identification strategy assigns treatment based on each firm's headquarters state, while DBNLs formally regulate customers' personal data regardless of firm location. This raises the possibility that firms could be bound by laws in states beyond their headquarters. However, headquarters-based assignment is standard in the literature exploiting staggered state laws (e.g., Bertrand and Mullainathan, 2003, Autor et al., 2007, Kumar et al., 2011). Moreover, it reflects the centralized nature of corporate data-security policies (Hylender et al., 2024). That said, we also directly address the concern that some firms may already be exposed to DBNLs in states other than the state of their headquarters by replicating our results after excluding very large multi-state corporations with significant consumer sales across states (Appendix A.3.6). Our results remain statistically and economically significant, indicating that the main findings are not unduly polluted by headquarters assignment.

Fourth, to address the concern that high-complementarity firms might have deliberately built the organizational feature of data-innovation complementarity in anticipation of DBNL adoption, rather than it being an exogenous, pre-existing characteristic, we note that our complementarity measure is constructed using firms' organizational structure over the *five years preceding* DBNL adoption and is held fixed throughout the analysis. Thus, the classification predates the regulatory shock and cannot mechanically reflect post-DBNL behavior. Moreover, we find no evidence of differential pre-trends in innovation, growth, or self-citation outcomes between high- and low-complementarity firms prior to DBNL adoption (Figures 2–5, and Table 2), supporting the interpretation that complementarity is an ex-ante firm characteristic, not an endogenous response to expected regulation.

Fifth, we perform checks to rule out that our heterogeneous effects mechanically capture prolific inventors or are sensitive to how we identify "data-security inventors" (Appendix A.3.7). In the main analysis, a patent inventor is classified as a data-security expert if at least 10% of their lifetime patent portfolio consists of data-security patents. We verify robustness of our results to alternative definitions (e.g., varying lifetime-share thresholds for security patents between 10%, 25%, and 50%, imposing minimum patent counts, using recent five-year vs. lifetime portfolios, and expanding/restricting the CPC code set). Appendix Table A.16 reports descriptive statistics, and Appendix Table A.17 re-estimates our triple-difference specification under each definition. Across all variants, the post-DBNL high-minus-low estimates for citation-weighted patents remain economically large and statistically significant; the results for value-weighted patents, assets, and ROA are similarly robust. Pre-trend diagnostics remain clean. These findings indicate that our results are not an artifact of how we label data-security inventors or, by extension, how firms are classified as having high or low data-innovation complementarity.

Sixth, our main analysis defines "major" projects as those with above-median team size. We show that the results are not driven by this choice or by star-inventor concentration (Appendix A.3.8). The results remain similar using alternative definitions (top

10% and top 90% team-size cutoffs), examining citation-based importance, including all teams, or fixing the high-complementarity status at the pre-DBNL value. Across specifications, the post-DBNL change in the average treatment effect in citation-weighted patents remains economically large ($\approx$ 18-41 additional citation-weighted patents p.a.), with positive and usually significant value-weighted patents and growth outcomes (Appendix Table A.18). Taken together with inventor-quality controls and top-inventor exclusions, these analyses rule out alternative explanations involving selection on project scale and visibility, prolific inventors, or post-treatment reclassification.

Lastly, as a falsification (placebo) test, we mirror the complementarity construction in two non-security technologies: "packaging" (CPC B65B/C/D, all groups and subgroups) and "computer graphics" (G06T/G09G, all groups and subgroups), and re-estimate the baseline DDD (Appendix A.3.9, Table A.19). If the results were driven by arbitrary technological specialization, these domains would load on DBNLs as well; they do not. "Packaging" shows no DBNL response, and "computer graphics" only attenuated, statistically weak effects, indicating that the effect is specific to complementarities in data-security know-how rather than generic specialized expertise.

5 A Real-Option Model of Data-Innovation Complementarity

Our empirical findings reveal substantial heterogeneity in how firms respond to exogenous increases in the perceived importance of data security depending on their organizational structure. High-complementarity firms significantly increase broad innovation and growth, as opposed to their low-complementarity counterparts which exhibit muted or even negative responses.

To interpret these patterns, we develop a parsimonious real-options model of quasi-irreversible organizational investment. The model formalizes our core insight: that the returns to data-security investments are amplified when they are structurally integrated into

broader innovation activities. Modeling this integration as a strategic, irreversible investment decision allows us to explain persistent gaps in firm responses, even under *common* regulatory triggers, such as DBNLs. Our framework explains why high-complementarity firms invest earlier and derive positive spillovers, while low-complementarity firms delay integration and fall behind. Our framework also generates testable predictions: convex returns to complementarity, asymmetric adoption timing, and the absence of spillovers under outsourcing, all of which are consistent with our empirical findings.

We use the theoretical framework to simulate counterfactual scenarios which cannot be directly identified in the data. We quantify the welfare costs of restricting firms to outsourcing and highlight the policy value of targeted subsidies that lower integration costs. The model shows that integrated data protection (high data-innovation complementarity) is a meaningful driver of innovation-led growth.

Setup We model a firm's decision to invest in high data-innovation complementarity (e.g., the integration decision) as a real-option, in the sense of Bar-Ilan and Strange, 1996. This occurs in the presence of uncertain, firm-specific spillovers from complementarity integration. Data security relevance increases exogenously due to regulation (e.g., DBNLs), and firms choose when to invest in integration, which is irreversible and costly.[11] Once embedded, integration efforts are rarely undone, as security becomes part of the firm's innovation infrastructure, and its value extends beyond compliance, enabling innovation spillovers.

Firms differ in their degree of data-innovation complementarity, $c \in [0, 1]$. Complementarity is a parameter reflecting the degree of alignment in infrastructure, processes, and talent between the firm's data security team and the firm's broader R&D activity and governs the strength of innovation spillovers once integration occurs. This structure generates heterogeneity in investment timing and post-shock outcomes.

[11] We assume that organizational integration of data security into broader innovation teams is not a marginal adjustment, but rather involves fundamental reconfiguration of workflows, team structures, and strategic priorities. For instance, JPMorgan Chase reportedly committed over $500M annually to embed security into IT workflows after a breach, illustrating the scale and irreversibility of such investments.

Although higher values of c capture greater potential returns from tailoring innovation to future data security relevance shocks, our empirical analysis suggests diminishing marginal returns to complementarity. We model this via a concave option-value function $\phi(c) = c^{\alpha}$ for $\alpha \in (0, 1)$. For simplicity, we assume $\alpha = \frac{1}{2}$. This modeling choice captures organizational frictions: even high-complementarity firms face capacity constraints or diminishing learning benefits from complementarity, consistent with real-options and organizational complementarity settings (e.g., Milgrom and Roberts, 1990, Bar-Ilan and Strange, 1996).

Consider a firm facing an exogenous shock to the perceived importance of data security, denoted by θ, modeled as a random variable:

$$\theta \sim \mathcal{N}(\mu, \sigma^2), \quad \text{with } \mu > 0,\ \sigma > 0.$$

This shock increases the perceived value of data protection and raises both the costs of inaction and the potential benefits from investing in data capabilities.

The firm has the option to undertake an irreversible investment in data capabilities at cost I. If the firm invests, it receives two types of benefits:

1. An *innovation spillover benefit* proportional to the firm's complementarity level, $\phi(c)$. The return to complementarity is $\gamma\phi(c)\theta$, where $\gamma > 0$ is the patent value (this means, that for every \$1M in firm size, each unit of $\phi(c)\theta$ leads to γ citation-weighted patents' worth of value).
2. A *fixed reputational or compliance benefit* denoted δ_0.

Furthermore, without investing, the firm faces a potential cost of inaction: an expected loss $\lambda\theta$, with $\lambda > 0$ representing an ex-ante cost of data exposure (e.g., direct breach-related losses and indirect trust erosion losses).

Value Functions The firm decides whether to invest immediately or wait. The waiting option is subject to a discount factor $\rho \in (0, 1)$, which captures the risk of delay, such as

missed early innovation, the decay in data value over time, or regulatory uncertainty.

If the firm invests immediately, its value is:

$$V_{\text{now}}(c) = -I + \mathbb{E}[\gamma\phi(c)\theta + \delta_0 - \lambda\theta] = -I + (\gamma\phi(c) - \lambda)\mu + \delta_0. \tag{6}$$

If the firm delays investment, it retains the option to invest later after observing θ. The value of waiting is:

$$V_{\text{wait}}(c) = \rho \cdot \mathbb{E}\left[\max\{-I + (\gamma\phi(c) - \lambda)\theta + \delta_0,\ 0\}\right]. \tag{7}$$

In our model, complementarity reduces the option value of waiting by raising the marginal value of immediate investment when θ increases. High-complementarity firms therefore exercise the "invest now" option earlier, building data capabilities sooner. Consistent with this channel, we show empirically that high ex-ante complementarity predicts faster innovation and superior post-shock performance. This is consistent with Nikolov, Schmid, and Steri, 2021: organizational flexibility conditions the payoff to acting under uncertainty; here, complementarity plays that role.

Threshold Complementarity and Optimal Timing We define the threshold level of complementarity $\bar{c}$ as the value at which the firm is indifferent between investing immediately and waiting:

$$V_{\text{now}}(\bar{c}) = V_{\text{wait}}(\bar{c}). \tag{8}$$

For given primitives $(\mu, \sigma, \gamma, \lambda, \delta_0, \rho, I)$ this equation has a unique solution $\bar{c} \in (0, 1)$ under mild regularity conditions.[12] If $c < \bar{c}$, the firm optimally delays integration; if $c > \bar{c}$, the firm invests immediately.

Real–options theory offers two seemingly conflicting prescriptions: greater uncertainty can *delay* action by magnifying the option value of waiting (Bernanke, 1983, Mc-

[12] In our calibration, $V_{\text{now}}(c)$ and $V_{\text{wait}}(c)$ are continuous in c, with $V_{\text{now}}(0) < V_{\text{wait}}(0)$, $V_{\text{now}}(1) > V_{\text{wait}}(1)$, and $V_{\text{now}}(c)$ growing faster than $V_{\text{wait}}(c)$ in c, implying a unique crossing.

Donald and Siegel, 1986), yet it can also *trigger* earlier investment when the upside of acting dominates (Bar-Ilan and Strange, 1996, Dou, 2017). Our model reconciles these views through the lens of *data–innovation complementarity*. When complementarity c is low, the marginal innovation gain $A(c) = \gamma\sqrt{c} - \lambda$ is near zero or negative, so the firm faces little certain benefit today while preserving a valuable real-option; waiting is optimal. As c rises, $A(c)$ turns positive and increases quickly, converting a security relevance shock into an immediate, firm-wide productivity boost. Beyond the threshold $c = \bar{c}$, that sure gain outweighs the convex option value of delay, so the firm invests before the resolution of uncertainty and profits from the full distribution of future shocks. Thus, uncertainty still has value, but its influence flips sign depending on the firm's internal complementarity: precaution dominates below the threshold $\bar{c}$; preemption dominates above this threshold, aligning the divergent findings in the literature.

5.1 Quantitative Calibration

We calibrate the model using empirical estimates from our staggered difference-in-differences design. All parameter values, summarized in Table C.1 in Appendix *C.1*, are chosen to align the model's predictions with the observed innovation responses to DBNL adoption among high-complementarity firms.

We calibrate $\gamma = 0.4714$ to match the observed increase in innovation value post-DBNL for high-complementarity firms (i.e., firms with $\bar{c} = 0.75$), which amounts to a total of 95 additional patents in the five years post-DBNL (per Figure 4, right panel). This assumes an average patent value of \$4.01M (per Table A.5, Column 8) and a median high-complementarity firm size of \$933M (per Table A.4).

$$\text{Benefit}_{\text{innovation}} = \gamma * \sqrt{c} * \theta * \text{Median firm size} \tag{9}$$

$$\$380.95\text{M} = \gamma * \sqrt{0.75} * 1 * 933 \implies \gamma = 0.4714$$

This parameter represents the monetary return in citation-weighted patent value,

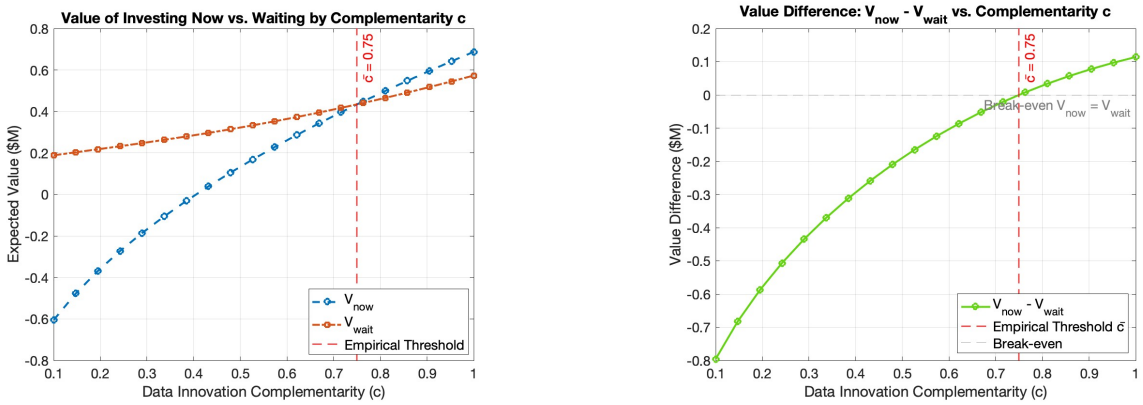
expressed in millions of dollars per unit of $\sqrt{c} \cdot \theta$ and per \$1M in firm assets, over a five-year horizon. It ensures that the model replicates the empirically observed increase of approximately 95 patents (worth \$380.95M) for the median firm, which corresponds to 0.4082 ($\$380.95M/\$933M = 0.4082$) patent-equivalent value units per \$1M in assets.

We calibrate the inaction cost parameter to $\lambda = 3$, corresponding to an expected loss of \$3M from non-integration. This value reflects the ex-ante cost of data exposure, incorporating both direct breach-related losses and indirect trust erosion and is consistent with industry estimates for firms without recent breach history (IBM Security, 2025) and with regulatory penalties (i.e., expected fines use enforcement statistics; GDPR fine is approximately EURO 2.36M). We set the reputational or compliance-related benefit of integration to $\delta_0 = 1.87$, aligning with empirical studies showing that firms with stronger IT governance face fewer repeat breaches and reduced regulatory penalties (e.g., Bharadwaj et al., 2025). At this level of benefit, only firms with sufficiently high complementarity invest, while firms with low complementarity derive limited value from integration. Finally, we set the discount factor to $\rho = 0.8$, which corresponds to an annual discount rate of approximately 4% over the five-year horizon used in our calibration.

We then numerically solve for I^* such that the value of investing equals the value of waiting at the threshold we observe empirically. Figure 6 presents the implied values of investing and waiting and implied investment levels for different thresholds of complementarity. The complementarity value $\bar{c} = 0.75$ in the model is set to be exactly the empirical point at which innovation responses to DBNLs become consistently positive in our continuous treatment empirical analysis (Figure 4, right panel). At this threshold, the implied investment cost is $I^* = 0.0723$ in model units, which corresponds to approximately \$67.45M for the median firm in our sample (with total assets of \$933M).

The calibration exercise captures the model's core mechanism: that internal data-protection capabilities become economically valuable only when they are sufficiently integrated with innovation activity. Importantly, the $\bar{c} = 0.75$ threshold captures a diverse set of firms. As shown in Appendix A.3.1, the 25% of the sample that exceeds this

Figure 6: Calibrated model



(a) Value of Investing vs Waiting

(b) Investment

Legend: The left panel reflects the value of investing now (in blue) and the value of waiting (in red) by complementarity level c. The right panel reflects the difference (in green) between investing now and waiting to invest. All values are per \$1M of firm assets over a five-year investment horizon. Parameters: $\mu = 1$; $\sigma^2 = 0.5$; $\lambda = 3$; $\gamma = 0.4714$; an average patent value $= 4.01$; $\delta_0 = 1.87$; $\rho = 0.8$; $\bar{c} = 0.75$. Table C.1 provides further details on the parameters.

threshold shows meaningful representation across industries, firm sizes, and productivity levels. While high-complementarity firms are more prevalent in data-heavy industries such as *Information* and *Finance & Insurance*, they are also present in *Manufacturing* and other sectors, reflecting diverse business models. Similarly, high-complementarity firms span the entire distribution of firm size and patent productivity, consistent with our interpretation of complementarity as a strategic organizational choice, rather than a mechanical correlate of scale. Policies or organizational investments that lower the costs of data innovation integration—such as shared infrastructure, cross-training, or managerial guidance—could help more firms surpass the threshold and unlock the innovation gains documented in our analysis.

On magnitudes, the calibrated model generates economically meaningful innovation gains for ex-ante high-complementarity firms without changing firm types. Using $\gamma = 0.4714$ and $\mathbb{E}[\theta] = 1$, the model-implied innovation yield per \$1M in assets is

$$\underbrace{0.4082}_{=\gamma\sqrt{0.75\mathbb{E}[\theta]}} \quad \text{at } c = 0.75 \qquad \text{and} \qquad \underbrace{0.4472}_{=\gamma\sqrt{0.90\mathbb{E}[\theta]}} \quad \text{at } c = 0.90.$$

At the empirical cutoff $\bar{c} = 0.75$, this maps for the median high-c firm (assets $\approx$ \$933M) into \$380.95M of citation-weighted patent value over five years, i.e., about 95 patents at \$4.01M each (Table A.5), consistent with Figure 4 (right panel).

Model validity To translate model-implied innovation into profitability, we apply a conservative elasticity of 0.07 percentage points of ROA per \$100M of citation-weighted patent value, following Hall et al., 2005. The model's \$380.95M prediction for the median high-c firm therefore implies a five-year ROA increase of 0.2665 ($0.07 \times \frac{380.95}{100}$) percentage points, or 0.0533 percentage points per year, which closely matches the empirically observed increase of 0.0618 percentage points per year for high-c firms (Table A.8). Table C.2 presents the targeted (used in the calibration) and the untargeted (validating) moments. We emphasize that ROA is not targeted in calibration; it serves as an overidentifying validation of the model's quantitative implications for firm performance.

5.1.1 Welfare Loss if Firms Are Restricted to Outsourcing

The model highlights that firms with high data innovation complementarity benefit from internal integration not only by avoiding data-related risks, but also from cross-domain innovation spillovers. By contrast, firms with low complementarity typically outsource data security, forgoing the associated innovation benefits.

We model outsourcing as an alternative organizational choice that provides the *same* protection against regulatory and operational losses as internal integration (the inaction cost $\lambda\mu$ is mitigated and the reputational/compliance benefit δ_0 is preserved), but without enabling innovation spillovers. In other words, relative to internal integration, outsourcing removes the $\gamma\sqrt{c}$ term while leaving the risk- and compliance-related components unchanged. This assumption allows us to isolate the welfare effects of shutting off the complementarity channel, holding other margins constant.

Proposition 1 *Firms with $c > \bar{c}$ incur a strictly positive welfare loss from being restricted to outsourcing data security, relative to internal integration. The welfare loss*

$\Delta V(c)$ is increasing in c, γ, and μ, and decreasing in the investment cost gap $(I - I^{ext})$.

***Proof* 1** *See Appendix B.1.*

This proposition formalizes the insight that internal investment unlocks innovation spillovers that outsourcing does not. When c is higher, the foregone innovation benefit is larger. In other words, when data innovation and security are organizationally decoupled (outsourced), high-c firms suffer the largest counterfactual losses.

Quantitative Welfare Loss We microfound the outsourcing cost I^{ext} using estimates from industry studies on managed security service providers (MSSPs). According to multiple industry reports,[13] mid- to large-size firms that outsource cybersecurity to third-party providers typically incur annual costs between \$1.5M and \$3M, depending on coverage scope and complexity. We adopt a midpoint estimate of \$2M per year, or \$10M over a five-year horizon. Scaling this value by the median firm size in our sample (\$933M in total assets), we obtain:

$$I^{\text{ext}} = \frac{10}{933} \approx 0.0107 \tag{10}$$

This assumption reflects the total cost of outsourcing cybersecurity functions, including monitoring, compliance, incident response, and threat management. By expressing I^{ext} in model units (i.e., per \$1M in assets), we ensure consistency with the scaling used for other parameters such as I, γ, λ, and δ_0.

Using the calibrated parameters ($c = 0.75$, $\mu = 1$, $\gamma = 0.4714$, $\lambda = 3.0$, $\delta_0 = 1.87$, $I = 0.0723$, and $I^{\text{ext}} = 0.0107$), the welfare loss from being restricted to outsourcing is

$$\Delta V(c) = V^{\text{internal}}(c) - V^{\text{external}} = -(I - I^{\text{ext}}) + \gamma\mu\sqrt{c} = 0.3466. \tag{11}$$

This implies a welfare loss of 0.3466 model units per \$1M in firm assets, or equivalently approximately 35% of a firm's total assets. For the median firm in our sample

[13] See, for example IBM Security, 2025, Cybrwise, 2024, and VC3, 2024.

(assets = \$933M), this translates to $0.3466 \times 933 \approx \$323.3\text{M}$. This magnitude is economically plausible and consistent with prior research linking patent output to firm value, particularly in data-intensive and innovation-driven sectors. Our empirical estimates assume a five-year horizon, during which shutting down the innovation spillovers would make the firm forgo the creation of approximately 95 additional citation-weighted patents, each valued at \$4M (Kogan et al., 2017), which is exactly equivalent to \$380M per firm, of the same order as the model-implied value gain. This scenario reflects the erosion of broader innovation-driven firm value from restricting high-complementarity firms to external data protection solutions, highlighting the critical role of internal capabilities in translating regulatory compliance into strategic advantage.

5.1.2 Minimum Policy Subsidy to Induce Early Investment

We now use the model to quantify the minimum subsidy required to induce early investment in internal data integration for firms with relatively low complementarity. This exercise informs potential policy interventions that could accelerate investment in integration capabilities among firms that would otherwise delay.

Specifically, we consider a firm with $c < \bar{c}$ that would prefer to wait in the absence of support. Let $S(c)$ denote the smallest subsidy that ensures the firm is indifferent between immediate investment and delay. This threshold satisfies:

$$V_{\text{now}}(c) + S \geq V_{\text{wait}}(c), \tag{12}$$

where $V_{\text{now}}(c)$ is the expected value of investing immediately and $V_{\text{wait}}(c)$ is the expected value of deferring the decision. The subsidy $S(c)$ thus represents the minimum intervention required to close the option value gap between immediate and delayed action.

Proposition 2 *For every firm with complementarity level $c < \bar{c}$, there exists a minimum subsidy $S(c) \geq 0$ such that the firm is indifferent between investing immediately and waiting. The subsidy can be written as $S(c) = V_{wait}(c) - V_{now}(c)$. $S(c)$ is strictly decreasing*

in c and in the innovation elasticity γ, and strictly increasing in the discount factor ρ.

***Proof* 2** *See Appendix B.2.*

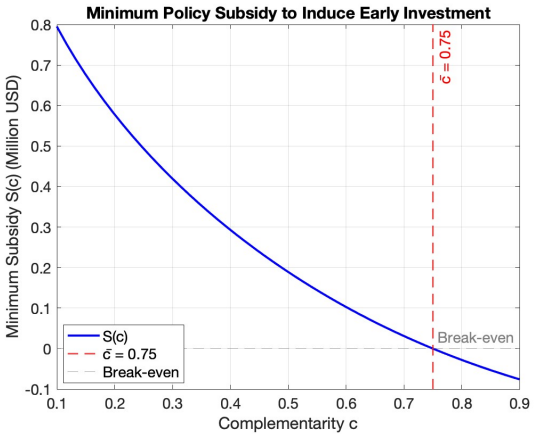
Quantitative Subsidy Policy We use our Monte Carlo simulation framework to estimate the minimum policy subsidy required to induce immediate internal investment for firms with sub-threshold data-innovation complementarity. Specifically, we consider a firm with $c = 0.5$ and simulate shocks $\theta \sim \mathcal{N}(\mu, \sigma^2)$, where we set $\mu = 1$, $\sigma = 0.5$, and use a calibrated investment cost of $I = 0.0723$ that equates V_{now} and V_{wait} at the threshold $\bar{c} = 0.75$. The simulation uses $N = 10^5$ draws and computes the expected value of investing now and waiting, given c.

At $c = 0.5$, the value of immediate investment is $V_{\text{now}}(0.5) = 0.1442$, and the expected value of waiting is $V_{\text{wait}}(0.5) = 0.3255$ (per \$1M). The implied minimum policy subsidy needed to induce immediate investment is $S(0.5) = V_{\text{wait}}(0.5) - V_{\text{now}}(0.5) = 0.1889$. For a median firm in our sample with \$933M in assets, this translates to a subsidy of approximately $0.1889 \times 933 = 176.3$M USD. While the implied subsidy of \$176.3M corresponds to approximately 18.9% of the median firm's asset base (\$933M), the actual percentage may differ for firms with lower complementarity. These firms are often smaller, implying a potentially lower subsidy in level terms, but a higher subsidy as a share of firm value. Thus, fixed integration costs may pose a disproportionately larger barrier for structurally under-prepared firms, reinforcing the case for targeted support.

We repeat this calculation for a range of $c \in [0.1, 0.9]$ and plot the resulting subsidy function $S(c)$ per \$1M in Figure 7. The subsidy declines monotonically in c and crosses zero as c reaches $\bar{c}$, by construction. Figure 7 illustrates the policy leverage available in nudging low-complementarity firms over the investment threshold. Appendix Figure C.5 presents the simulated policy subsidy across the two-dimensional space capturing a range of both data-innovation complementarity and investment cost levels.

This exercise illustrates the potential effectiveness of targeted subsidies in altering the investment thresholds of low-complementarity firms. By incentivizing data innovation

Figure 7: Minimum Subsidy For Range of Complementarity Levels (per $1M)



Legend: Simulated minimum policy subsidy $S(c)$ required to induce immediate internal investment, plotted across a range of data-innovation complementarity levels $c \in [0.1, 0.9]$. Subsidy values are expressed in model units (per $1M in assets) and are strictly decreasing in c, consistent with the model's prediction that firms with greater structural integration require less external support. The vertical line marks the threshold $\bar{c} = 0.75$, above which firms invest without subsidy. Table C.1 provides further details on the parameters.

complementarity, such policy interventions can unlock the innovation spillovers that high-complementarity firms already enjoy, reducing heterogeneity in innovation outcomes and fostering a more inclusive data economy.

Finally, our results speak to policy design. While data protection laws impose compliance costs, they can also stimulate innovation for firms with appropriate organizational structures. However, such spillovers are unevenly distributed across firms. While high-complementarity firms benefit from an increase in the perceived importance of data security, other firms fall behind. Our model quantifies the costs of this divergence and provides guidance for policy interventions that support innovation without reinforcing concentration.

6 Conclusion

We identify and quantify a novel firm capability, *data-innovation complementarity*, the degree to which data-related capabilities, particularly in data security, are structurally integrated with the firm's broader innovation activities. We show that this capability is

both measurable from publicly-available data and economically meaningful: firms with strong ex-ante data-innovation complementarity turn regulatory pressure into competitive advantage, while firms lacking such integration are left behind.

Using a new firm-level measure based on inventor-team overlap in data security and general patenting, constructed solely from publicly available data, we document that high-complementarity firms respond to the staggered adoption of U.S. Data Breach Notification Laws (DBNLs) by significantly increasing both innovation output and firm value. These effects are large, causal, and concentrated among firms with pre-existing organizational integration. By contrast, low-complementarity firms exhibit muted or even negative responses. Our empirical strategy leverages modern difference-in-differences methods that account for recent concerns about staggered adoption designs.

To interpret these findings, we develop a parsimonious real-options model, which we use for policy analysis and counterfactual simulations. High-complementarity firms invest earlier, internalize spillovers, and convert compliance pressure into innovation-fueled growth. Counterfactual simulations show that restricting firms to outsourcing data protection eliminates internal spillovers and reduces firm value by roughly one-third. Targeted subsidies can close the innovation gap by inducing earlier investment among low-complementarity firms. These findings inform both firm-level strategic decisions and the design of data governance policies.

More broadly, data protection laws can catalyze innovation when firms are organizationally prepared, but such preparedness is unevenly distributed across firms. This points to a new source of inequality and divergence in performance between firms in the modern economy. Data governance regulations can amplify asymmetries, entrenching the advantage of firms that are ex-ante organizationally prepared. Our findings inform ongoing policy debates on how to design data governance frameworks that protect consumers, while fostering firm innovation and industry competition. Future work could examine how such complementarities evolve with AI adoption, shape demand for cross-functional technical talent, and play out under different global regulatory regimes.

References

Alberto Abadie, Susan Athey, Guido Imbens, and Jeffrey Wooldridge. When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1):1–35, 2023.

Philippe Aghion, John Van Reenen, and Luigi Zingales. Innovation and institutional ownership. *American Economic Review*, 103(1):277–304, 2013.

Pat Akey, Stefan Lewellen, Inessa Liskovich, and Christoph Schiller. Hacking corporate reputations. *Review of Finance (forthcoming)*, 2025.

Julian Atanassov. Do hostile takeovers stifle innovation? evidence from antitakeover legislation and corporate patenting. *The Journal of Finance*, 68(3):1097–1131, 2013.

Julian Atanassov, Brandon Julio, and Tiecheng Leng. The bright side of political uncertainty: The case of r&d. *The Review of Financial Studies*, 37(10):2937–2970, 05 2024.

David Autor, William Kerr, and Adriana Kugler. Does employment protection reduce productivity? evidence from us states. *The Economic Journal*, 117(521):F189–F217, 2007.

Tania Babina, Anastassia Fedyk, Alex He, and James Hodson. Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151:103745, 2024. ISSN 0304-405X.

Andrew C. Baker, David F. Larcker, and Charles C.Y. Wang. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395, 2022. doi: 10.1016/j.jfineco.2022.01.

Avner Bar-Ilan and William C. Strange. Investment lags. *The American Economic Review*, 86 (3):610–622, 1996. ISSN 00028282.

Jan Bena, Hernán Ortiz-Molina, and Elena Simintzi. Shielding firm value: Employment protection and process innovation. *Journal of Financial Economics*, 146(2):637–664, 2022.

Ben S. Bernanke. Irreversibility, uncertainty, and cyclical investment. *The Quarterly Journal of Economics*, 98(1):85–106, 1983.

Marianne Bertrand and Sendhil Mullainathan. Enjoying the quiet life? corporate governance and managerial preferences. *Journal of Political Economy*, 111(5):1043–1075, 2003.

Timothy Besley and Torsten Persson. The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*, 99(4):1218–44, September 2009.

Timothy Besley and Torsten Persson. Fragile states and development policy. *Journal of the European Economic Association*, 9(3):371–398, 2011.

Anandhi Bharadwaj, Sean Cao, Jing Tian, and Ling Xue. Credibility of it disclosure narratives: An examination in the context of data breaches. Technical report, SSRN WP 5127642, 2025.

Kwabena A. Boasiako and Michael O'Connor Keefe. Data breaches and corporate liquidity management. *European Financial Management*, 27(3):528–551, 2021.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285, 2024.

Brantly Callaway and Pedro H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.

Sean Cao, Anya Nakhmurina, and Tianchen Zhao. Data breaches, debt costs, and public service provision. Mimeo, 2025.

Jonathan B Cohn, Zack Liu, and Malcolm I Wardlaw. Count (and count-like) data in finance. *Journal of Financial Economics*, 146(2):529–551, 2022.

Avinash Collis, Ananya Sen, and Joy Wu. Demand for privacy from data brokers. Working Papers 22-09, NET Institute, Sep 2022.

Lin William Cong, Danxia Xie, and Longtian Zhang. Knowledge accumulation, privacy, and growth in a data economy. *Management Science*, 67(10):6480–6492, 2021.

Annamaria Conti, Christian Peukert, and Maria Roche. Beefing it up for your investor? engagement with open source communities, innovation, and startup funding: Evidence from github. *Organization Science*, 36(4):1551–73, 2025.

Cybrwise. Understanding the cost of outsourcing cybersecurity services for smbs. Link, 2024. Accessed: August 2025.

Sreyoshi Das, Camelia Kuhnen, and Stefan Nagel. Socioeconomic status and macroeconomic expectations. *The Review of Financial Studies*, 33(1):395–432, 03 2019.

Clément de Chaisemartin and Xavier d'Haultfœuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020.

Olivier Dessaint, Thierry Foucault, and Laurent Fresard. Does alternative data improve financial forecasting? the horizon effect. *The Journal of Finance*, 79(3):2237–2287, 2024.

Winston Wei Dou. Embrace or fear uncertainty: Growth options, limited risk sharing, and asset prices. Working Paper, The Wharton School, 2017.

Maryam Farboodi and Laura Veldkamp. A growth model of the data economy. Technical Report 28427, 2021.

Maryam Farboodi, Roxana Mihet, Thomas Philippon, and Laura Veldkamp. Big data and firm dynamics. *AER Papers and Proceedings*, 109:38–42, May 2019.

Anastassia Fedyk, Tatiana Fedyk, James Hodson, and Natalya V Khimich. Do consulting services affect audit quality? evidence from the workforce. *Working paper*, 2023.

Chris Florackis, Christodoulos Louca, Roni Michaely, and Michael Weber. Cybersecurity risk. *The Review of Financial Studies*, 36(1):351–407, 2023.

Bernhard Ganglmair, W Keith Robinson, and Michael Seeligson. The rise of process claims: Evidence from a century of us patents. *Available at SSRN 4069994*, 2022.

Alexander Giczy, Nicholas Pairolero, and Andrew Toole. Identifying artificial intelligence (ai) invention: A novel ai patent dataset. *The Journal of Technology Transfer*, 47(2):476–505, 2022.

Xavier Giroud and Holger Mueller. Does corporate governance matter in competitive industries? *Journal of Financial Economics*, 95(3):312–331, 2010. ISSN 0304-405X.

Xavier Giroud and Holger Mueller. Firms' internal networks and local economic shocks. *American Economic Review*, 109(10):3617–49, October 2019.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.

Bronwyn H Hall, Adam B Jaffe, and Manuel Trajtenberg. The nber patent citation data file: Lessons, insights and methodological tools, 2001.

Bronwyn H Hall, Adam Jaffe, and Manuel Trajtenberg. Market value and patent citations. *RAND Journal of economics*, pages 16–38, 2005.

Florian Heider and Alexander Ljungqvist. As certain as debt and taxes: Estimating the tax sensitivity of leverage from state tax changes. *Journal of Financial Economics*, 118(3):684–712, 2015.

Yael Hochberg and Joshua Rauh. Local overweighting and underperformance: Evidence from limited partner private equity investments. *Review of Financial Studies*, 26(2):403–451, 2013.

Sabrina T Howell. Financing innovation: Evidence from r&d grants. *American Economic Review*, 107(4):1136–64, 2017.

David Hylender, Philippe Langlois, Alex Pinto, and Suzanne Widup. 2024 data breach investigations report (dbir). Technical report, Verizon Business, 2024.

IBM Security. Cost of a data breach report 2025. Link, 2025. Accessed: August 2025.

Ali Kakhbod, Leonid Kogan, Peiyao Li, and Dimitris Papanikolaou. Measuring creative destruction. Research paper 5008685, SSRN, 2024.

Leonid Kogan, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2):665–712, 2017.

Camelia Kuhnen. Asymmetric learning from financial information. *The Journal of Finance*, 70 (5):2029–2062, 2015.

Camelia Kuhnen and Andrei Miu. Socioeconomic status and learning from financial information. *Journal of Financial Economics*, 124(2):349–372, 2017.

Alok Kumar, Jeremy Page, and Oliver Spalt. Religious beliefs, gambling attitudes, and financial market outcomes. *Journal of Financial Economics*, 102(3):671–708, 2011.

Robert McDonald and Daniel Siegel. The value of waiting to invest. *The Quarterly Journal of Economics*, 101(4):707–728, 1986.

Roxana Mihet and Thomas Philippon. The economics of big data and artificial intelligence. *International Finance Review*, pages 29–43, 2019.

Paul Milgrom and John Roberts. The economics of modern manufacturing: Technology, strategy, and organization. *The American Economic Review*, 80(3):511–528, 1990.

Boris Nikolov, Lukas Schmid, and Roberto Steri. The sources of financing constraints. *Journal of Financial Economics*, 139(2):478–501, 2021.

Christian Peukert, Stefan Bechtold, Michail Batikas, and Tobias Kretschmer. Regulatory spillovers and data governance: Evidence from the gdpr. *Marketing Science*, 41(4):746–768, 07 2022.

Tarun Ramadorai, Antoine Uettwiller, and Ansgar Walther. Privacy policies and consumer data extraction: Evidence from u.s. firms. *Review of Finance (forthcoming)*, 2025.

Kumar Rishabh, Roxana Mihet, and Julian Jang-Jaccard. Cyberrisk and ai firms. Research paper no. 25-39, SFI, 2024.

Daniel Rock. Engineering value: The returns to technological talent and investments in artificial intelligence. Technical report, SSRN, 2019. WP 3427412.

Jonathan Roth, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, 2023.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021.

Prasanna Tambe, Lorin Hitt, Daniel Rock, and Erik Brynjolfsson. Digital capital and superstar firms. WP 28285, NBER, 2020.

VC3. How much do managed cybersecurity services cost? Link, 2024. Accessed: August 2025.

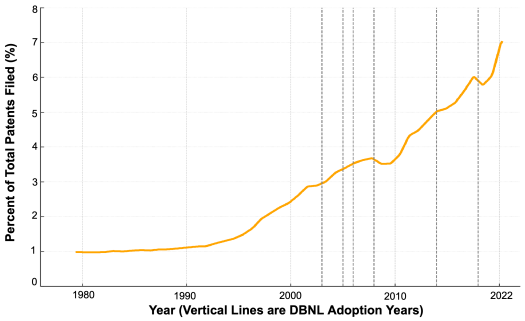
Laura Veldkamp. Valuing data as an asset*. *Review of Finance*, 27(5):1545–1562, 01 2023.

Internet Appendix

Appendix A Additional Figures and Tables

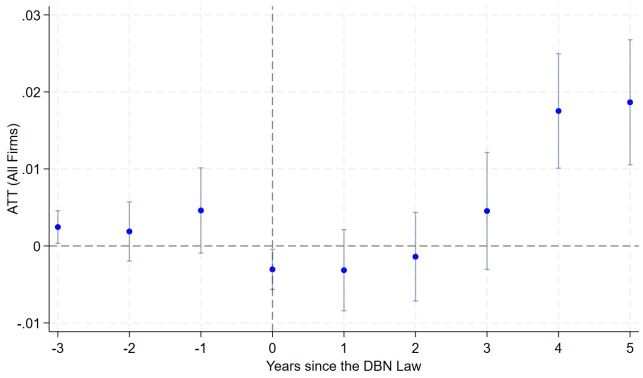
A.1 Additional Figures

Figure A.1: Share of data security (i.e., cyber security) patents among all patents filed



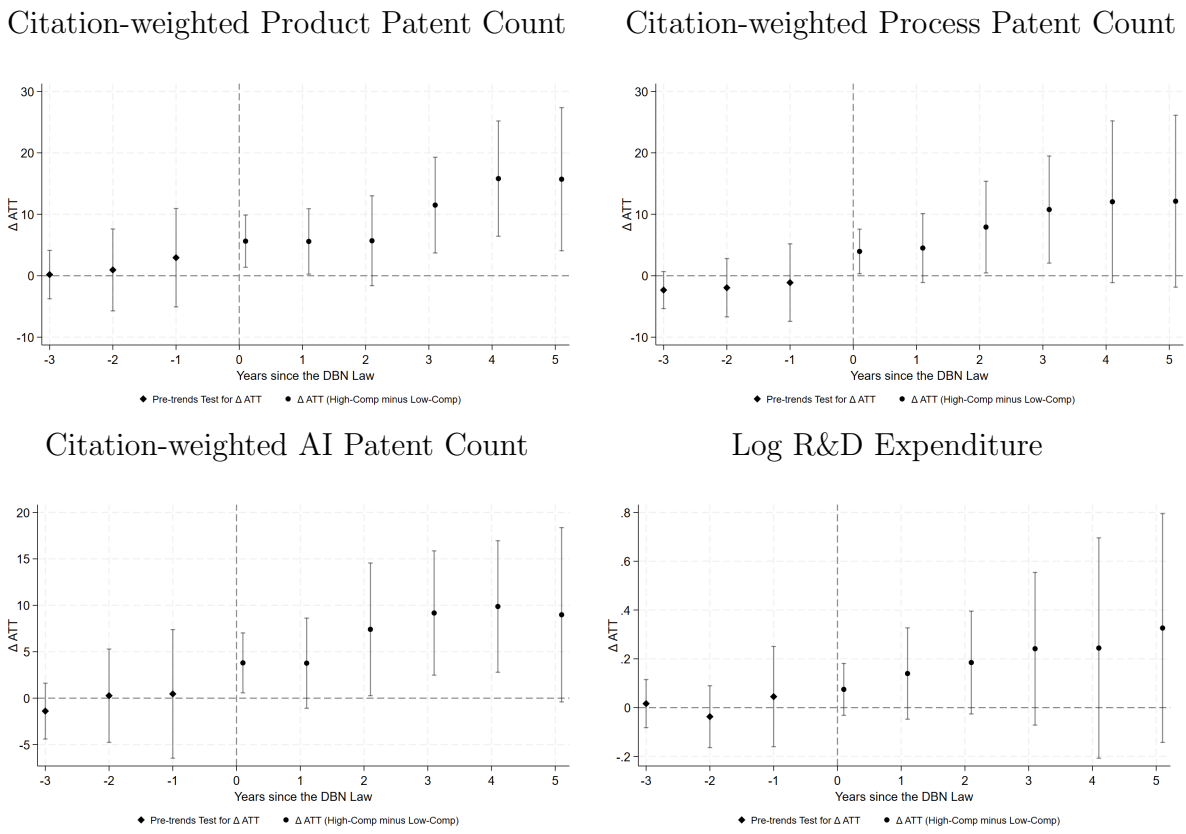
Legend: This figure shows the proportion of data-related patents (i.e., cyber security patents) out of all patents filed in a given year. A cyber security patent is defined as a patent to which the USPTO assigns at least one CPC code pertaining to cyber security. The dashed vertical grey lines represent the years of DBNL adoption.

Figure A.2: Response in Cyber Risk Score Without DBNL Anticipation Adjustment: Full Sample DID Estimates



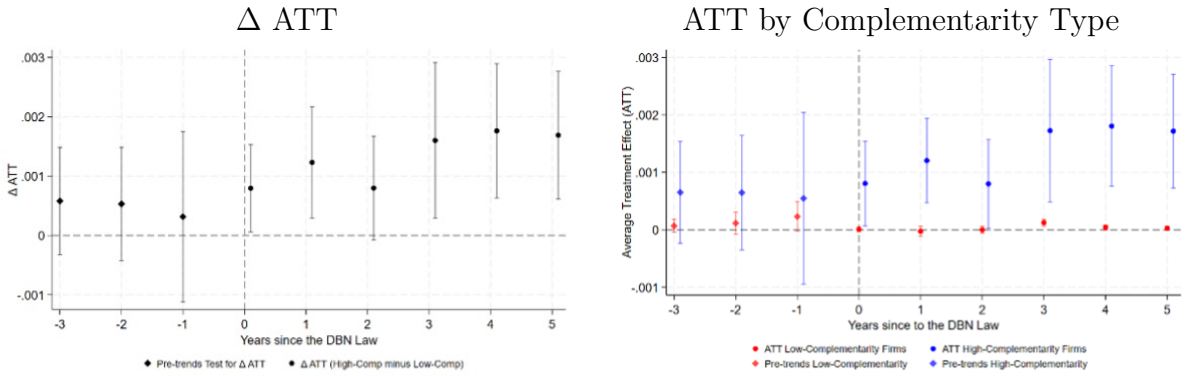
Legend: This figure tests whether treated firms pay increased attention to data risks just before the implementation of the DBNLs, in order to identify any anticipatory response to DBNL implementation. The figure plots DID estimates of changes in the cyber risk score based on the methodology in Florackis et al. [2023] for the full sample of firms around the staggered adoption of Data Breach Notification Laws (DBNLs) across U.S. states. Year "0" denotes the year a DBNL takes effect. Pre-trend diagnostics follow BJS, estimated on untreated observations only. All regressions (including the diagnostics) include firm, year, and year-by-subgroup fixed effects. Standard errors are clustered at the state level. Vertical bars show 90% confidence intervals.

Figure A.3: Innovation: Triple Difference Estimates



Legend: This figure plots triple-difference (DDD) estimates from BJS regressions around the staggered adoption of DBNLs across U.S. states. The DDD estimator compares subgroup-specific ATT estimates for high-complementarity and low-complementarity firms at each horizon. Firms are classified as high-complementarity if their average share of data-security specialists on major non-data-security patent teams over the prior five years exceeds the 75th percentile (the classification is sticky thereafter). Year "0" denotes one year before the DBNL takes effect, explicitly accounting for potential anticipation effects. Pre-trend diagnostics follow BJS, estimated using untreated observations only, with subgroup-specific lead indicators. In all regressions (including pre-trend diagnostics), we include firm, year, and year-by-subgroup fixed effects, ensuring that treatment versus control comparisons remain within-subgroup. Standard errors are clustered at the state level. Patents are classified as product (process) patents if 50 percent or more of their claims are product (process) claims, following Ganglmair et al. [2022]. The AI patent classification is based on the AIPD dataset by USPTO. See Table A.1 for details on the variables. Vertical error bars show 95% confidence intervals.

Figure A.4: Self-CS-citations in non-CS patents



Legend: This figure plots the effects on self-citations of CS (cybersecurity) patents in non-CS patents. The right panel plots the difference-in-differences (DID) effects around the staggered adoption of DBNLs across U.S. states, for high-complementarity firms in blue and for low-complementarity firms in red. The left panel plots the triple difference (DDD) estimates for the differential effect on high- versus low-complementarity firms. Year "0" denotes one year before DBNL implementation. All regressions include firm, year, and year-by-subgroup fixed effects. Standard errors are clustered at the state level. Vertical bars show 95% confidence intervals.

A.2 Additional Tables

Table A.1: Variable Descriptions

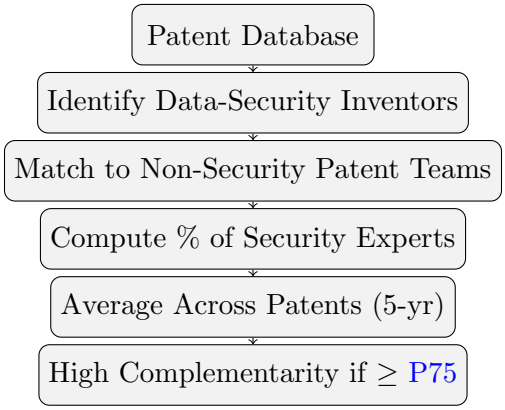
Variable	Description
Data Risk Score	Cyber-risk score for US-listed firms following Florackis et al. [2023]. Calculated through textual analysis of 10-K filings, comparing disclosed cyber risks with prior filings of breached firms. Higher scores indicate greater cybersecurity vulnerability.
Citation-weighted Patent Count	Annual patent filings weighted by forward citations and adjusted for truncation bias following Hall et al. [2001]. Captures scientific impact of patents.
Value-weighted Patent Count	Total market value created by the patents awarded to the firm in each filing year, scaled by firm assets. Derived from stock-market reactions to patent awards (KPSS dataset).
Citation-weighted Product Patent Count	Citation-weighted count of patents primarily classified as product patents per Ganglmair et al. [2022].
Citation-weighted Process Patent Count	Citation-weighted count of patents primarily classified as process patents per Ganglmair et al. [2022].
Citation-weighted AI Patent Count	Citation-weighted count of patents identified as AI inventions following the classification in Giczy et al. [2022]. A patent is classified as AI-related if any AI-subfield probability exceeds 0.50.
Citation-weighted Cybersecurity Patent Count	Citation-weighted count of patents classified under USPTO Cooperative Patent Classification (CPC) codes related to cybersecurity (G06F21 / "Security arrangements for protecting computers, components, programs, or data against unauthorized activity," H04L9 / "Cryptographic mechanisms or arrangements for secret or secure communication," H04L63 / "Network security arrangements," etc.).
"Major" Patents	Patents whose inventor team size exceeds the firm's median patent team size over a rolling five-year window.
Log R&D Expenditure	Natural logarithm of research and development expenditures (Compustat item *xrd*).
Log Total Assets	Natural logarithm of total firm assets (Compustat item *at*).
Log Market Capitalization	Natural logarithm of firm market equity, calculated as stock price (*prcc_f*) times shares outstanding (*csho*).
Log Sales	Natural logarithm of total sales revenue (Compustat item *sale*).
Return on Assets (ROA)	Operating income before depreciation scaled by total assets ($oibdp/at$).
Share of CS Experts in Major Non-CS Patents	Average fraction of cybersecurity specialists working on the firm's major non-cybersecurity patents. Major patents are defined as those with inventor team sizes above the firm's median.
Share of CS Experts in All Non-CS Patents	Average fraction of cybersecurity specialists working on all non-cybersecurity patents filed by the firm.
Share of Self CS Patent Citations	Fraction of internal citations of cybersecurity patents by the firm's non-cybersecurity patents. This measure captures internal knowledge transfers between cybersecurity and the rest of the firm's innovation activities.
High Complementarity (Yes = 1)	Dummy variable indicating firms whose average share of cybersecurity experts on major non-cybersecurity patents over the prior five years is at or above the 75th percentile across all firms (once achieved, the designation is sticky thereafter).

Table A.2: Descriptive Statistics

Variable	N	Mean	SD	p10	p25	p50	p75	p90
I. Innovation (Total Patent Count)								
Citation-weighted	26699	57.79	162.30	0.00	0.00	5.07	27.81	127.60
Value-weighted	26699	0.18	0.36	0.00	0.00	0.04	0.18	0.52
Log (R&D)	22633	3.51	1.85	1.25	2.33	3.47	4.62	5.94
II. Citation-Weighted Patent Counts								
Product	26699	32.69	92.60	0.00	0.00	2.22	15.32	73.43
Process	26699	16.48	50.40	0.00	0.00	1.00	6.90	32.69
AI	26699	14.87	50.13	0.00	0.00	0.00	3.49	27.81
Cybersecurity	26699	2.52	10.82	0.00	0.00	0.00	0.00	2.25
III. Financial Characteristics								
Log (Total Assets)	26699	6.55	2.35	3.63	4.79	6.36	8.11	9.82
Log (Market Cap)	26687	6.84	2.28	3.92	5.20	6.73	8.41	9.99
Log (Sales)	26171	6.03	2.67	2.72	4.30	6.14	7.88	9.45
Return on Assets	26644	-0.01	0.31	-0.37	-0.04	0.09	0.15	0.21
IV. Cyber Risk–Related Variables								
Share of CS Experts in Major Non-CS Patents	25326	0.06	0.13	0.00	0.00	0.00	0.04	0.20
Share of CS Experts in All Non-CS Patents	26010	0.06	0.14	0.00	0.00	0.00	0.05	0.22
Share of Self CS Patent Cites in Non-CS Patents	19256	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Cyber Risk Score	15814	0.14	0.16	0.00	0.00	0.07	0.25	0.42
V. Complementarity Measure (Dummy)								
High Complementarity	26699	0.38	0.48	0.00	0.00	0.00	1.00	1.00

Legend: This table presents descriptive statistics for the sample of firm-year observations used in our analysis. Percentiles refer to the 10^{th} through the 90^{th} percentile. High complementarity is defined as firm-years above the 75th percentile of data-security specialist integration. See Table A.1 for the full variable definitions.

Table A.3: Flowchart of Data Innovation Complementarity Index Construction



*Note: Since this classification is based on inventor team structure, it captures real integration of expertise—leading to a more precise measure than R&D spending or patent counts alone.

Table A.4: Median and Mean Firm Sizes (Descriptive Statistics)

	N Observations	Mean (USD Mil)		Median (USD Mil)	
		Total Assets	Market Cap	Total Assets	Market Cap
Sample	(1)	(2)	(3)	(4)	(5)
Full sample	26687	10172.81	9309.5	580.02	839.86
High-Complementarity sample	10727	15336.7	12253.2	933.1	1369.4
NAICS Sector (2-digit Code)					
Admin & Support (56)	174	4664.62	10106.25	2097.47	3472.54
Finance & Insurance (52)	825	101021.53	35467.25	51675	14591.04
Healthcare (62)	173	3664.57	4453.57	113.69	184.92
Information (51)	3149	12632.09	13776.88	658.45	1179.7

Legend: All statistics represent per-firm means (columns 2–3) and medians (columns 4–5) for all firm-years in our analysis. The data come from WRDS CRSP-Compustat.

Table A.5: Median and Mean Patent Values (Descriptive Statistics)

	Mean (USD Mil)				Median (USD Mil)			
Sample	Nominal (1)	Real (1982 Dollars) (2)	Nom per cit-wtd (3)	Real per cit-wtd (4)	Nominal (5)	Real (1982 Dollars) (6)	Nom per cit-wtd (7)	Real per cit-wtd (8)
Full sample	29.22	11.86	22.34	8.91	8.48	3.59	5.48	2.33
High-complementarity sample	39.85	15.76	30.47	11.83	14.29	6.10	9.40	4.01
NAICS Sector (2-digit Code)								
Admin & Support (56)	47.30	18.08	33.09	12.41	14.28	6.90	6.88	3.06
Finance & Insurance (52)	172.65	66.47	136.43	51.58	100.57	38.49	69.73	26.73
Healthcare (62)	16.87	6.63	13.91	5.41	3.33	1.35	2.53	1.00
Information (51)	35.52	14.03	27.72	10.74	9.81	4.29	6.32	2.69
Manufacturing (31-33)	24.38	10.08	18.64	7.57	7.42	3.15	4.67	2.01
Mining & Oil/Gas (21)	40.58	17.25	29.53	12.23	22.30	9.37	15.33	6.26
Other (X)	30.43	13.25	21.16	9.11	15.99	7.08	9.31	3.95

Legend: The statistics represent per-patent means (columns 1–4) and medians (columns 5–8) across all patents, computed over the baseline sample of firm-years in our analysis. The per-patent values come originally from Kogan et al. [2017].

Table A.6: Industry-Level Statistics on Non-Cybersecurity Patent Teams

NAICS Sector	N Firms	N Observations	Non-CS Patents		Major Non-CS Patents	
(2-digit Code)			Avg Patent Team Size	Avg Num of CS Innovators	Avg Patent Team Size	Avg Num of CS Innovators
	(1)	(2)	(3)	(4)	(5)	(6)
Information (51)	349	3152	2.88	0.57	4.49	0.85
Finance & Insurance (52)	69	825	3.13	0.37	4.88	0.56
Prof Services (54)	119	961	2.96	0.47	4.45	0.68
Real Estate (53)	23	251	2.67	0.25	4.22	0.37
Admin & Support (56)	20	174	3.44	0.53	5.20	0.57
Transport & Warehouse (48-49)	18	153	2.84	0.14	4.74	0.28
Utilities (22)	28	275	2.74	0.06	4.21	0.11
Other (X)	49	216	3.22	0.19	4.86	0.20
Manufacturing (31-33)	1964	19472	3.00	0.09	4.69	0.14
Mining & Oil/Gas (21)	38	430	2.66	0.05	4.26	0.07
Retail (44-45)	28	279	2.78	0.22	4.51	0.30
Wholesale (42)	30	214	2.71	0.12	4.15	0.11
Healthcare (62)	21	173	3.31	0.01	5.23	0.02

These average statistics represent per-patent means for non-cybersecurity (non-CS) patents, computed over the baseline sample of firm-years in our analysis. Average Patent Team Size is the mean number of inventors per non-CS patent. Average Number of CS Innovators indicates the mean number of cybersecurity-specialized inventors on each non-CS patent. The Average Patent Team Size for Major Patents and the Average Number of CS Innovators in Major Patents are measured analogously, but over "major" non-CS patents—defined as those patents whose inventor team size exceeds the firm's median patent team size over a rolling five-year window. CS Innovators are inventors whose cybersecurity-related patents constitute at least 10% of their lifetime patent portfolio.

Table A.7: Differential Innovation Responses to DBNLs: Estimates from Triple-Difference Regressions and ATT for Low-Complementarity Firms

	Citation-weighted Patent Count (Filed)				Value-weighted Count	
	All Patents	Product Patents	Process Patents	AI Patents	All Patents	Log (R&D Expenditure)
	(1)	(2)	(3)	(4)	(5)	(6)
Δ ATT Post	28.8767***	13.0391***	11.7046**	9.6115***	0.1756***	0.3439**
	(8.3142)	(3.525)	(4.6102)	(3.0408)	(0.0338)	(0.1718)
ATT Low Comp Post	-9.7311***	-5.3898***	-3.6971***	-1.0336*	-0.0870***	-0.2757***
	(2.7051)	(1.5899)	(1.2571)	(0.6122)	(0.0182)	(0.0769)
Pre-trends Test						
Δ ATT Pre (-1)	4.496	2.9417	-1.1145	0.4597	0.0449	0.045
	(7.3183)	(4.0849)	(3.2177)	(3.532)	(0.0409)	(0.1048)
Δ ATT Pre (-2)	-0.1215	0.9506	-1.9524	0.27	0.0142	-0.037
	(5.9068)	(3.4002)	(2.4185)	(2.9615)	(0.0362)	(0.0646)
Δ ATT Pre (-3)	-2.2465	0.1874	-2.3264	-1.3894	-0.0103	0.0163
	(3.9558)	(2.0091)	(1.5362)	(1.5346)	(0.0241)	(0.0502)
ATT Low Comp Pre (-1)	-5.9523*	-3.1755	-0.3927	0.2923	-0.0527***	0.0171
	(3.3173)	(2.1807)	(1.0655)	(0.7973)	(0.0192)	(0.059)
ATT Low Comp Pre (-2)	-4.2015*	-1.7881	-0.1792	-0.1802	-0.0255	-0.0015
	(2.3093)	(1.7142)	(0.5804)	(0.5127)	(0.0159)	(0.0367)
ATT Low Comp Pre (-3)	-2.03*	-0.8486	0.238	0.0923	-0.0125	-0.0147
	(1.0472)	(0.8357)	(0.406)	(0.357)	(0.0096)	(0.0231)
N Observations	17,905	17,905	17,905	17,905	17,905	15,159

$^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$. This table presents triple-difference (DDD) estimates from BJS regressions around the staggered adoption of Data Breach Notification Laws (DBNLs) across U.S. states. The DDD estimator compares subgroup-specific ATT estimates for high-complementarity and low-complementarity firms in the post-DBNL period (i.e., the estimated effect for high-complementarity firms minus the estimated effect for low-complementarity firms). The table also presents the estimated ATT for low-complementarity firms. Firms are classified as high-complementarity if their average share of data-security specialists on major non-data-security patent teams over the prior five years exceeds the 75th percentile (this classification is sticky thereafter). DBNL year is taken to be one year before the DBNL takes effect, explicitly accounting for potential anticipation effects. Pre-trend diagnostics follow BJS, estimated using only untreated observations with subgroup-specific leads and their differences. In all regressions (including pre-trend diagnostics), we include firm fixed effects, year fixed effects, and year-by-subgroup fixed effects, ensuring that treatment versus control comparisons remain within subgroup. Standard errors are clustered at the state level. Vertical error bars show 95% confidence intervals. See Table A.1 for details on the variables.

Table A.8: Differential Growth and Profitability Responses to DBNLs: Estimates from Triple-Difference Regressions and ATT for Low-Complementarity Firms

	Log (Total Assets)	Log (Market Capitalization)	Return on Assets	Log (Total Sales)
	(1)	(2)	(3)	(4)
Δ ATT Post	0.2524***	0.5330***	0.0311**	0.2018*
	(0.076)	(0.2039)	(0.0121)	(0.1052)
ATT Low Comp Post	-0.1798***	-0.2226***	0.0307***	-0.0505
	(0.0251)	(0.0372)	(0.0065)	(0.0457)
Pre-trends Test				
Δ ATT Pre (-1)	0.0805	0.2275	0.0147	0.0129
	(0.0721)	(0.1876)	(0.0284)	(0.1126)
Δ ATT Pre (-2)	0.0325	0.0887	-0.0028	-0.0659
	(0.0627)	(0.1334)	(0.0205)	(0.0781)
Δ ATT Pre (-3)	0.011	-0.0006	-0.0246*	-0.0948**
	(0.0386)	(0.08)	(0.0135)	(0.0433)
ATT Low Comp Pre (-1)	0.0209	-0.0302	0.0104	0.0244
	(0.041)	(0.087)	(0.0192)	(0.083)
ATT Low Comp Pre (-2)	0.0357	0.0533	0.0186	0.0458
	(0.0345)	(0.0671)	(0.0147)	(0.0607)
ATT Low Comp Pre (-3)	0.0374*	0.0809*	0.0269**	0.0728**
	(0.0202)	(0.0422)	(0.0106)	(0.036)
N Observations	17,905	17,894	17,819	17,661

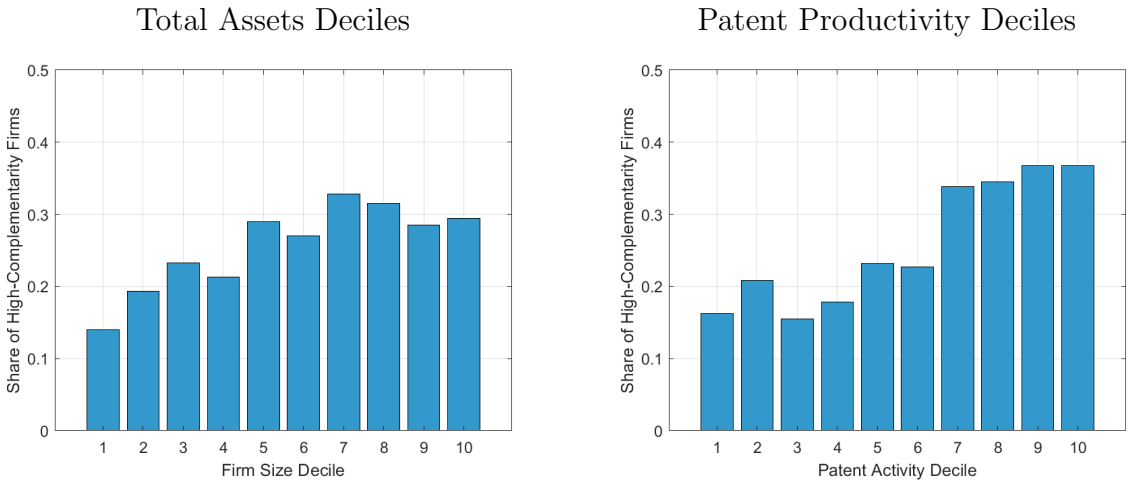
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. This table presents triple-difference (DDD) estimates from BJS regressions around the staggered adoption of Data Breach Notification Laws (DBNLs) across U.S. states. The DDD estimator compares subgroup-specific ATT estimates for high-complementarity and low-complementarity firms in the post-DBNL period (i.e., the estimated effect for high-complementarity firms minus the estimated effect for low-complementarity firms). The table also presents the estimated ATT for low-complementarity firms. Firms are classified as high-complementarity if their average share of data-security specialists on major non-data-security patent teams over the prior five years exceeds the 75th percentile (this classification is sticky thereafter). DBNL year is taken to be one year before the DBNL takes effect, explicitly accounting for potential anticipation effects. Pre-trend diagnostics follow BJS, estimated using only untreated observations with subgroup-specific leads and their differences. In all regressions (including pre-trend diagnostics), we include firm fixed effects, year fixed effects, and year-by-subgroup fixed effects, ensuring that treatment versus control comparisons remain within subgroup. Standard errors are clustered at the state level. See Table A.1 for details on the variables.

A.3 Robustness

A.3.1 Robustness: Adding Firm Size and Patenting Productivity as Controls

Figure A.5 demonstrates that high-complementarity firms are spread throughout the size and patent-productivity distributions. In the left panel, the fraction of high-complementarity firms rises smoothly from roughly 15 percent in the bottom decile based on firm assets to about 32 percent in the top three deciles. In the right panel, the share increases from 17 percent to 36 percent across deciles based on patent activity. Because every decile contains both high- and low-complementarity firms, the complementarity measure is correlated with—but clearly not reducible to—firm scale or raw inventive output. This visual evidence supports our claim that the complementarity measure captures a strategic organizational choice rather than mere capacity.

Figure A.5: Share of High-Complementarity Firms by Firm Size and Innovation Productivity Deciles



Legend: This figure plots the share of firms with high data-innovation complementarity across deciles of firm characteristics: (i) average total assets (Left panel) and (ii) average patenting activity (Right panel). Firms are assigned to deciles based on the average value of their log total assets and citation-weighted patents filed per year, computed over the pre-treatment period. Firms are classified as high-complementarity if their average share of data-security specialists on major non-data-security patent teams over the prior five years exceeds the 75th percentile of the full distribution; this status, once attained, remains sticky over time. The plotted share in each decile reflects the fraction of firms for which this status is triggered at any point during the pre-treatment period.

Table A.9 re-estimates the triple-difference regressions with one-year-lagged log assets and patent filings as additional controls. We introduce these lagged variables to explicitly account for any potential confounds arising from variations in firm scale or historical inventive intensity. By controlling for predetermined firm characteristics measured in the year prior, we ensure that observed differences in post-DBNL responses between high- and low-complementarity firms are

attributable specifically to their organizational integration of data-security expertise rather than residual differences in size or patent productivity. Because these lags enter the regression as explanatory variables, the model takes on a dynamic structure; the estimated coefficients thus reflect incremental effects relative to firms' recent size and innovation activities. Encouragingly, the results remain economically substantial, statistically significant, and close to our baseline estimates, while pre-trend tests continue to show no systematic differences. This confirms that although the lagged structure slightly attenuates short-run responses, it leaves the core finding intact and reinforces that the differential reaction to DBNLs is driven primarily by organizational complementarity.

Table A.9: Differential Innovation, Growth, and Profitability Responses to DBNL: Robustness with Controls

	Patent Count (All Patents)		Growth Variables		Profitability
	Citation weighted	Value Weighted	Log (Total Assets)	Log (Market Capitalization)	Return on Assets
	(1)	(2)	(3)	(4)	(5)
Δ ATT Post	13.6839***	0.1687***	0.1321***	0.4610**	0.0278**
	(3.8805)	(0.0299)	(0.0401)	(0.1854)	(0.0122)
ATT Low Comp Post	-6.2483***	-0.0874***	-0.0741***	-0.1130**	0.0377***
	(1.608)	(0.0184)	(0.0147)	(0.0352)	(0.0066)
Pre-trends Test					
Δ ATT Pre (-1)	-1.1652	0.0571	0.0114	0.2409	0.0004
	(5.9512)	(0.0411)	(0.0575)	(0.2003)	(0.0242)
Δ ATT Pre (-2)	1.03	0.0158	0.0007	0.098	-0.02
	(4.471)	(0.0331)	(0.041)	(0.1358)	(0.0164)
Δ ATT Pre (-3)	-2.2538	-0.0042	-0.0103	-0.0173	-0.0393**
	(3.721)	(0.0233)	(0.0303)	(0.091)	(0.0148)
ATT Low Comp Pre (-1)	-4.5452	-0.043*	0.0154	-0.0227	0.017
	(2.8332)	(0.0253)	(0.0329)	(0.0885)	(0.0188)
ATT Low Comp Pre (-2)	-2.3915	-0.0134	0.0115	0.0187	0.0174
	(1.9927)	(0.0211)	(0.0267)	(0.0592)	(0.0142)
ATT Low Comp Pre (-3)	-1.9898	-0.0089	0.0303	0.0731*	0.0251**
	(1.4065)	(0.0142)	(0.0189)	(0.0413)	(0.0093)
Size and Patenting Controls	Yes	Yes	Yes	Yes	Yes
N Observations	15,821	15,821	15,821	15,811	15,748

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. This table presents triple-difference (DDD) estimates from BJS regressions around the staggered adoption of Data Breach Notification Laws (DBNLs) across U.S. states. The DDD estimator compares subgroup-specific ATT estimates for high-complementarity and low-complementarity firms in the post-DBNL period (i.e., the estimated effect for high-complementarity minus the estimated effect for low-complementarity firms). The table also presents the estimated ATT for low-complementarity firms. Firms are classified as high-complementarity if their average share of data-security specialists on major non-data-security patent teams over the prior five years exceeds the 75th percentile (this classification remains sticky thereafter). The regressions in this table include lagged controls for size (log of total assets) and past patenting (lagged count of citation-weighted patents filed). DBNL year is taken to be one year before the DBNL takes effect, explicitly accounting for potential anticipation effects. Pre-trend diagnostics follow BJS, estimated using only untreated observations with subgroup-specific leads and their differences. In all regressions (including pre-trend diagnostics), we include firm fixed effects, year fixed effects, and year-by-subgroup fixed effects, ensuring treatment versus control comparisons remain within subgroup. Standard errors are clustered at the state level. See Table A.1 for details on the variables.

A.3.2 Robustness: Dropping High-Complementarity Industries

We conduct industry-level robustness tests to ensure our baseline results are not driven by any single sector. Table A.10 presents DDD estimates when dropping firms from one industry at a time. Each row represents a separate regression excluding the specified industry. The results demonstrate consistency with our baseline estimates of 28.88 patents for the DDD effect and -9.73 patents for low-complementarity firms.

Across most industry exclusions, the DDD estimate (Δ ATT Post) remains statistically significant and economically meaningful, ranging from 27.00 to 38.74 additional citation-weighted patents per year. The treatment effect for low-complementarity firms (ATT Low Comp Post) consistently remains negative across all specifications, ranging from -7.82 to -10.22 patents. Manufacturing presents the sole exception: when we exclude this sector, the DDD estimate drops to 13.95 patents and loses statistical significance, while the low-complementarity effect becomes larger at -13.81 patents.

Table A.10: Main Results With Leave-one-industry-out DID Regressions

Dropped Industry	Dep Var: Citation-weighted Patent Count		Dep Var: Log (Total Assets)		
(NAICS 2 digits)	Δ ATT Post	ATT Low Comp Post	Δ ATT Post	ATT Low Comp Post	N Observations
	(1)	(2)	(3)	(4)	(5)
Mining & Oil/Gas (21)	28.89***	-9.23**	0.26***	-0.17**	17593
	(8.46)	(3.14)	(0.08)	(0.07)	
Utilities (22)	29.38***	-9.91**	0.26***	-0.18**	17761
	(8.46)	(2.72)	(0.08)	(0.03)	
Manufacturing (31-33)	13.95	-13.81*	0.3***	-0.34***	3969
	(11.85)	(7.21)	(0.08)	(0.07)	
Wholesale (42)	28.99***	-9.82***	0.25***	-0.18**	17761
	(8.35)	(2.72)	(0.08)	(0.03)	
Retail (44-45)	27***	-9.97***	0.25***	-0.18**	17740
	(8.39)	(2.79)	(0.08)	(0.03)	
Transport & Warehouse (48-49)	28.83***	-9.76***	0.26***	-0.18**	17792
	(8.29)	(2.71)	(0.08)	(0.03)	
Information (51)	38.74***	-9.51***	0.19***	-0.16***	15935
	(9.22)	(2.8)	(0.07)	(0.03)	
Finance & Insurance (52)	28.59***	-10.22***	0.26***	-0.18**	17400
	(8.64)	(2.76)	(0.08)	(0.03)	
Real Estate (53)	28.95***	-9.76**	0.26***	-0.18**	17781
	(8.28)	(2.73)	(0.07)	(0.02)	
Prof Services (54)	27.49***	-9.9***	0.25***	-0.17**	17211
	(8.26)	(2.78)	(0.08)	(0.02)	
Admin & Support (56)	29.23***	-9.91***	0.26***	-0.19***	17798
	(8.43)	(2.69)	(0.08)	(0.03)	
Healthcare (62)	29.11***	-9.77***	0.23***	-0.15***	17772
	(8.65)	(2.71)	(0.08)	(0.03)	
Other (X)	27.16***	-7.82***	0.26***	-0.18***	17676
	(8.28)	(2.15)	(0.08)	(0.03)	

$^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$. This table presents triple-difference (DDD) estimates from BJS regressions around the staggered adoption of Data Breach Notification Laws (DBNLs) across U.S. states. We run separate regressions dropping firms from one industry at a time to test whether baseline results are driven by any particular sector. Each row represents a single regression with the specified industry excluded. The DDD estimator (ΔATT) compares treatment effects between high-complementarity and low-complementarity firms in the post-DBNL period. The table also presents ATT estimates for low-complementarity firms. Firms are classified as high-complementarity if their average share of data-security specialists on major non-data-security patent teams over the prior five years exceeds the 75th percentile. DBNL year is one year before the law takes effect, accounting for anticipation effects. All regressions include firm fixed effects, year fixed effects, and year-by-subgroup fixed effects, ensuring treatment versus control comparisons remain within subgroup. Standard errors are clustered at the state level. See Table A.1 for variable definitions.

The manufacturing result warrants attention but does not undermine our core findings. Excluding manufacturing leaves only 3,969 observations compared to 15,000+ for other exclusions, creating a substantial loss of statistical power. More importantly, the economic magnitude re-

mains positive at 13.95 patents, suggesting the complementarity mechanism operates even when manufacturing firms are removed from the sample, albeit with greater noise due to the smaller sample size.

Excluding the most complementarity-intensive sectors supports our mechanism rather than threatening it. When we drop Information (NAICS 51), Finance & Insurance (NAICS 52), or Professional Services (NAICS 54), the DDD estimates remain 38.74, 28.59, and 27.49 patents, respectively. Although these sectors contain the highest concentration of data-innovation complementarity, yet removing them does not undermine our results. This confirms that our findings reflect a broader economic phenomenon rather than sector-specific dynamics confined to obviously data-intensive industries.

The firm growth results mirror the patent patterns with one notable difference. DDD estimates for log total assets range from 0.19 to 0.30, clustering around our baseline estimate of 0.25. Unlike the patent results, excluding manufacturing does not weaken the firm growth effect: the DDD estimate remains strong at 0.30 percentage points and statistically significant. Low-complementarity firms consistently experience negative growth effects between -0.15 and -0.34 percentage points across all specifications. These patterns confirm that complementarity differences drive firm responses to data protection incentives across diverse industrial contexts, with the growth effects proving particularly robust to industry composition.

A.3.3 Robustness: Industry-level Cycles

Our main empirical specification already tightly controls for subgroup-specific national trajectories: firm fixed effects remove all stable firm-level differences, year fixed effects absorb aggregate shocks, and subgroup-by-year fixed effects allow high- and low-complementarity firms to follow separate national trajectories. This ensures that any nationwide shifts disproportionately benefiting high-complementarity firms do not contaminate our estimates, and each treated observation's counterfactual relies solely on untreated firms with the same complementarity classification observed in the same year. Nonetheless, one may still worry about industry-specific year-to-year shocks. High-complementarity firms are disproportionately represented in technology-intensive industries that may experience distinct annual fluctuations independent of complementarity status.

To directly address this, we augment our specification with two-digit-industry-by-year fixed effects. Panel A of Table A.11 adds these industry-year fixed effects, constraining comparisons to firms within the same industry-year. Panel B further tightens identification by including both subgroup-by-year and industry-by-year fixed effects, so that imputed untreated outcomes are drawn only from firms that share the same complementarity classification, industry, and calendar year. The results remain reassuringly stable across specifications: citation-weighted patents continue to rise strongly for high-complementarity firms, and pre-trends remain flat. This confirms that our findings are not an artifact of industry-level cycles, but reflect organizational complementarities in innovation.

Similarly, Panel A of Table A.12 adds these industry-year fixed effects, constraining comparisons to firms within the same industry-year. Panel B further tightens identification by including both subgroup-by-year and industry-by-year fixed effects, so that imputed untreated outcomes are drawn only from firms that share the same complementarity classification, industry, and calendar year. The results remain reassuringly stable across specifications: growth remains strong for high-complementarity firms, and pre-trends remain flat, confirming that sectoral or regional shocks do not explain our findings.

Table A.11: Differential Innovation (Patent Count (All Patents)) Responses to DBNLs: Robustness to Differential Industry Trends Effects

	Panel A: with Ind-by-Yr FE		Panel B: and Group-by-Yr FE	
	Citation weighted (1)	Value Weighted (2)	Citation weighted (1)	Value Weighted (2)
Δ ATT Post	14.6953***	0.0172	27.2519***	0.1369**
	(4.6413)	(0.0172)	(10.6141)	(0.0423)
ATT Low Comp Post	-2.3293	-0.0305	-7.7657***	-0.0848**
	(4.3423)	(0.0272)	(2.6448)	(0.0244)
Δ ATT Pre (-1)	4.496	0.0449	4.496	0.0449
	(7.3183)	(0.0409)	(5.9068)	(0.0362)
Δ ATT Pre (-2)	-0.1215	0.0142	-0.1215	0.0142
	(5.9068)	(0.0362)	(5.9068)	(0.0362)
Δ ATT Pre (-3)	-2.2465	-0.0103	-2.2465	-0.0103
	(3.9558)	(0.0241)	(3.9558)	(0.0241)
ATT Low Comp Pre (-1)	-2.8931	-0.0233	-2.2562	-0.0395
	(5.6733)	(0.0276)	(4.1845)	(0.0248)
ATT Low Comp Pre (-2)	-1.4207	-0.0058	-1.3035	-0.0153
	(3.984)	(0.0205)	(2.9966)	(0.019)
ATT Low Comp Pre (-3)	-0.6584	-0.0005	-1.1324	-0.008
	(1.9869)	(0.0109)	(1.2499)	(0.0104)
N Observations	14,805	14,805	14,805	14,805
Firm FE and Year FE	Yes	Yes	Yes	Yes
Group-by-Year FE	No	No	Yes	Yes
Industry-by-Year FE	Yes	Yes	Yes	Yes

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. Industry classifications are based on two-digit NAICS codes. Group designations indicate High- or Low-complementarity types. For regression specifications, see Table A.7. For variable definitions, see Table A.1.

Table A.12: Differential Growth and Profitability Responses to DBNLs: Robustness to Differential Industry Trends Effects

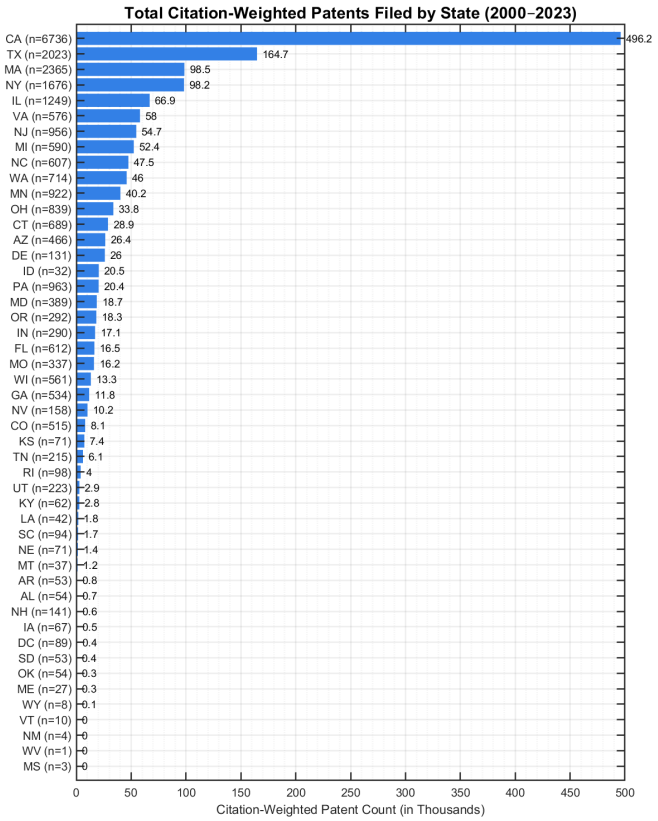
	Panel A: with Ind-by-Year FE			Panel B: and Group-by-Year FE		
	Ln (Total Assets)	Ln (Market Cap)	ROA	Ln (Total Assets)	Ln (Market Cap)	ROA
	(1)	(2)	(3)	(1)	(2)	(3)
Δ ATT Post	0.0570**	0.1279	0.0265***	0.2006***	0.4780**	0.0269
	(0.0274)	(0.1147)	(0.0067)	(0.0643)	(0.2101)	(0.0185)
ATT Low Comp Post	-0.0554**	-0.1452***	0.0154	-0.1558***	-0.2957***	0.0174*
	(0.0246)	(0.0403)	(0.0078)	(0.0229)	(0.0423)	(0.0102)
Δ ATT Pre (-1)	0.0805	0.2275	0.0147	0.0805	0.2275	0.0147
	(0.0721)	(0.1876)	(0.0294)	(0.0627)	(0.1334)	(0.0205)
Δ ATT Pre (-2)	0.0325	0.0887	-0.0028	0.0325	0.0887	-0.0028
	(0.0627)	(0.1334)	(0.0205)	(0.0627)	(0.1334)	(0.0205)
Δ ATT Pre (-3)	0.011	-0.0006	-0.0246*	0.011	-0.0006	-0.0246*
	(0.0386)	(0.08)	(0.0135)	(0.0386)	(0.08)	(0.0135)
ATT Low Comp Pre (-1)	0.0754**	0.0809	0.0223	0.0287	0.0109	0.0189
	(0.0346)	(0.0566)	(0.014)	(0.0985)	(0.0782)	(0.0388)
ATT Low Comp Pre (-2)	0.0697**	0.1069**	0.0233**	0.0349	0.0681	0.0225
	(0.0277)	(0.041)	(0.0115)	(0.0339)	(0.0603)	(0.0153)
ATT Low Comp Pre (-3)	0.0584***	0.1078***	0.0296**	0.036*	0.0803***	0.0303**
	(0.0183)	(0.0334)	(0.0077)	(0.0207)	(0.039)	(0.0099)
N Observations	14,805	14,794	14,741	14,805	14,794	14,741
Firm FE and Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Group-by-Year FE	No	No	No	Yes	Yes	Yes
Industry-by-Year FE	Yes	Yes	Yes	Yes	Yes	Yes

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. Industry classifications are based on two-digit NAICS codes. Group designations indicate High- or Low-complementarity types. For regression specifications, see Table A.7. For variable definitions, see Table A.1.

A.3.4 Robustness: State-Level Innovation and the Timing of DBNLs

Table A.13 confirms that no single "super-innovator" state is driving our results. California alone accounts for nearly half a million citation-weighted patents since 2000 (Figure A.6), yet excluding every CA-headquartered firm still yields a robust high-minus-low complementarity gap of approximately +24.4 patents and an asset-growth differential of roughly +0.26 log-points (about 25 percent). Similarly, dropping any other top-10 patent state (e.g., Texas, Massachusetts, New York, Illinois) preserves substantial patent differentials ranging from +24 to +30, with asset-growth gaps consistently falling between +0.19 and +0.27. Thus, our key results are geographically broad-based and driven solely by Silicon Valley or other highly productive clusters.

Figure A.6: US States by Patenting Activity



Legend: This figure plots total citation-weighted patent counts filed by firms headquartered in different states. Citation-weighted patent count is defined as one plus the patent's citation stock normalized by filing-year averages. "n" denotes the number of observations in each state, consisting of firms in our sample for which the complementarity status could be determined.

The observed pattern is systematic: high-complementarity firms consistently accelerate innovation and growth, while low-complementarity firms decelerate, regardless of geographic location. Every leave-one-state-out regression confirms the negative and statistically significant

ATT for low-complementarity firms (about –10 patents and –0.16 to –0.19 log-points in asset growth). Concurrently, the differential ATT (high-minus-low) remains stable and positive. Hence, the driving factor behind divergent growth trajectories is firm-level innovation complementarity, not merely regional innovation clusters or patenting intensity.

Table A.13: Main Results When Dropping Most Patenting States With Leave-one-out DID Regressions

Dropped State	Dep Var: Citation-weighted Patent Count		Dep Var: Log (Total Assets)		
(Abbreviation)	Δ ATT Post	ATT Low Comp Post	Δ ATT Post	ATT Low Comp Post	N Observations
	(1)	(2)	(3)	(4)	(5)
California (CA)	24.42**	-9.84**	0.26***	-0.18***	14010
	(9.79)	(3.2)	(0.09)	(0.03)	
Texas (TX)	26.53***	-13.06***	0.19***	-0.16***	16358
	(10.28)	(3.17)	(0.08)	(0.03)	
Massachusetts (MA)	29.56***	-10.43***	0.27***	-0.19***	16316
	(8.47)	(2.87)	(0.08)	(0.03)	
New York (NY)	26.67***	-8.21***	0.25***	-0.18***	16811
	(8.26)	(2.4)	(0.08)	(0.03)	
Illinois (IL)	30.52***	-9.72**	0.25***	-0.17***	17009
	(8.43)	(2.76)	(0.07)	(0.03)	
Virginia (VA)	27.77***	-9.06***	0.25***	-0.18***	17533
	(8.27)	(2.44)	(0.08)	(0.03)	
New Jersey (NJ)	29.24***	-9.73***	0.25***	-0.18***	17265
	(8.37)	(2.75)	(0.08)	(0.03)	
Michigan (MI)	30.51***	-9.92***	0.25***	-0.18***	17424
	(8.22)	(2.79)	(0.08)	(0.02)	
North Carolina (NC)	28.02***	-10.32***	0.26***	-0.18***	17503
	(8.32)	(2.54)	(0.08)	(0.03)	
Washington (WA)	27.44***	-9.87***	0.24***	-0.17***	17424
	(8.34)	(2.69)	(0.08)	(0.02)	

$^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$. This table presents triple-difference (DDD) estimates from BJS regressions around the staggered adoption of Data Breach Notification Laws (DBNLs) across U.S. states. We run separate regressions dropping firms from one state at a time to test whether baseline results are driven by any particular state. Each row represents a single regression with the specified state excluded. We perform this test for each of the most productive states based on their patenting activity. The DDD estimator (ΔATT) compares treatment effects between high-complementarity and low-complementarity firms in the post-DBNL period. The table also presents ATT estimates for low-complementarity firms. Firms are classified as high-complementarity if their average share of data-security specialists on major non-data-security patent teams over the prior five years exceeds the 75th percentile. DBNL year is one year before the law takes effect, accounting for anticipation effects. All regressions include firm fixed effects, year fixed effects, and year-by-subgroup fixed effects, ensuring treatment versus control comparisons remain within subgroup. Standard errors are clustered at the state level. See Table A.1 for variable definitions.

Excluding top patenting states also removes some early DBNL adopters (e.g., California, Texas, Massachusetts), yet our key results remain robust. The staggered adoption timeline (Figure 1) ensures that identification also relies on later-adopting, less patent-intensive states. Thus, our results are robust both geographically and temporally, reinforcing that differential innovation and growth outcomes hinge on firms' innovation strategies rather than on regional innovation intensity or regulatory timing.

A.3.5 Robustness: Alternative Event Date Definition

We test the robustness of our results to redefining the event date as the actual DBNL enforcement year (rather than one year prior, as in our baseline specification). The substance of our core results remains unchanged: post-event differential ATT coefficients remain consistently positive and statistically significant for innovation, market valuation, and asset-growth measures, while low-complementarity firms exhibit continued negative or negligible responses (Table A.14).

Table A.14: Differential Response to DBNL: Estimates With Actual DBNL Year as Treatment Year

	Citation-weighted Count	Value-weighted Count	Growth Variables		Share of CS Experts in		Share of Self CS Citations
	All Patents	All Patents	Log (Total Assets)	Log (Market Cap)	Major Non-CS patents	All Non-CS patents	In Non-CS patents
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δ ATT Post	21.2767***	0.1227***	0.2794***	0.4848***	0.0044	0.0186**	0.0017**
	(8.2977)	(0.0327)	(0.0731)	(0.1756)	(0.0066)	(0.0082)	(0.0003)
ATT Low Comp Post	-8.4179**	-0.0497**	-0.1862**	-0.1614**	0.0008	0.0002	0.0000**
	(2.5338)	(0.0182)	(0.0242)	(0.0297)	(0.0006)	(0.0011)	(0)
Pre-trends Test							
Δ ATT Pre (-1)	14.4809**	0.0942***	0.1041	0.392**	0.0002	0.0084	0.0009
	(7.006)	(0.0305)	(0.0705)	(0.1996)	(0.0116)	(0.0138)	(0.0007)
Δ ATT Pre (-2)	7.0362**	0.063**	0.0513	0.2312	-0.0037	0.0045	-0.0001
	(5.1443)	(0.0242)	(0.0522)	(0.1455)	(0.0086)	(0.0099)	(0.0007)
Δ ATT Pre (-3)	1.0591	0.0318	0.0098	0.1031	-0.0045	0.0012	0.0002
	(3.7341)	(0.0229)	(0.0474)	(0.1015)	(0.0066)	(0.0065)	(0.0005)
ATT Low Comp Pre (-1)	-7.618**	-0.0577**	-0.0457	-0.245**	0.0016*	0.0044*	0.0001
	(3.7756)	(0.0226)	(0.0449)	(0.0922)	(0.0008)	(0.0023)	(0.0001)
ATT Low Comp Pre (-2)	-4.4711	-0.0442**	-0.0093	-0.0872	0.0012*	0.0032	0.0002
	(2.8918)	(0.0157)	(0.0337)	(0.0609)	(0.0007)	(0.002)	(0.0001)
ATT Low Comp Pre (-3)	-3.0463*	-0.0193	0.0091	0.013	0.0009*	0.0025	0.0001
	(1.7871)	(0.0125)	(0.0275)	(0.0529)	(0.0005)	(0.0018)	(0.0001)
N Observations	19,105	19,105	19,105	19,093	18,298	18,743	14,416

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. This table presents triple-difference (DDD) estimates from BJS regressions around the staggered adoption of Data Breach Notification Laws (DBNLs) across U.S. states. We use the year each DBNL takes effect as the treatment year, rather than one year prior as in our baseline specification. The DDD estimator compares treatment effects between high-complementarity and low-complementarity firms in the post-DBNL period. The table also presents ATT estimates for low-complementarity firms. Firms are classified as high-complementarity if their average share of data-security specialists on major non-data-security patent teams over the prior five years exceeds the 75th percentile. Pre-trend diagnostics follow BJS, estimated using only untreated observations with subgroup-specific leads and their differences. All regressions include firm fixed effects, year fixed effects, and year-by-subgroup fixed effects, ensuring that treatment versus control comparisons remain within subgroup. Standard errors are clustered at the state level. See Table A.1 for variable definitions.

However, redefining event time affects pre-trend diagnostics. Specifically, with the enforcement year as event time ($t = 0$), the differential ATT at year $t = -1$ turns positive, indicating that high-complementarity firms begin diverging exactly one year before DBNL enforcement—consistent with our earlier findings that firms enhance cyber-risk disclosure in their 10-K filings in anticipation of DBNL adoption (Figure A.2). This anticipatory adjustment validates our baseline choice of event timing: firms adjust behavior immediately upon realizing regulatory change is imminent, and economic benefits accrue from that point onward.

A.3.6 Robustness: Multi-state Exposure and SUTVA

A potential SUTVA concern is that large, multi-state firms may be partially treated regardless of headquarters assignment. Because precise state-level operating footprints are unavailable at scale, we proxy multi-state activity by firm size, assuming that the largest firms are most likely to have substantial operations across states. We therefore exclude the largest 500 or 1,000 firms from the estimation sample, ranking by total sales at two reference points (the beginning of the sample, 2000, and near the end, 2021). This design removes observations with the greatest risk of ambiguous treatment exposure while preserving identification in the remainder of the sample.

Table A.15 reports triple-difference estimates. Across all cutoffs (top 500 and top 1,000, by 2000 or 2021 sales), high-complementarity firms continue to exhibit positive and statistically significant post-DBNL responses in innovation (citation-weighted and value-weighted patent counts). Low-complementarity firms show negligible or negative responses. Growth and profitability outcomes (e.g., log assets, ROA) remain positive and statistically significant for the high-complementarity subgroup. Point estimates are modestly attenuated relative to baseline, consistent with trimming influential large firms, but the qualitative conclusions are unchanged. These results indicate that our findings are not an artifact of including very large, likely multi-state firms with potentially ambiguous treatment status.

Table A.15: Differential Innovation Responses to DBNLs: Robustness to Dropping Large/Multi-state Firms.

	Panel A: Dropping Top 500 Firms (2000 definition)				**Panel B: Dropping Top 500 Firms (2021 definition)**			
	Patent Count (All Patents)		Growth Variables	Profitability	Patent Count (All Patents)		Growth Variables	Profitability
	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ ATT Post	16.1359***	0.1169***	0.2343**	0.0243*	19.2496**	0.1545***	0.2343**	0.0289**
	(3.3964)	(0.0295)	(0.0982)	(0.0136)	(11.4023)	(0.0418)	(0.0982)	(0.0147)
Δ ATT Pre (-1)	-6.6188	-0.0376*	0.0661*	0.0162	-0.0621	-0.0242	0.0571	0.0179
	(2.8186)	(0.0227)	(0.0366)	(0.0175)	(3.89)	(0.0243)	(0.0417)	(0.0169)
Δ ATT Pre (-2)	-2.4838	-0.0185	0.0574**	0.0173	-2.16	-0.0116	0.0521*	0.019
	(2.1748)	(0.0174)	(0.0283)	(0.0132)	(2.7246)	(0.0189)	(0.0309)	(0.0125)
Δ ATT Pre (-3)	-2.2323*	-0.0139	0.0532**	0.0213**	-1.941	-0.0103	0.0459**	0.0214**
	(1.1421)	(0.0098)	(0.0214)	(0.0098)	(1.5635)	(0.01)	(0.0215)	(0.0093)
N Observations	14,756	14,756	14,756	14,673	14,972	14,972	14,972	14,902

	Panel C: Dropping Top 1000 Firms (2000 definition)				**Panel D: Dropping Top 1000 Firms (2021 definition)**			
	Patent Count (All Patents)		Growth Variables	Profitability	Patent Count (All Patents)		Growth Variables	Profitability
	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ ATT Post	9.1911*	0.1074**	0.2588***	0.0362**	24.8572***	0.1822***	0.1981**	0.0303**
	(5.0722)	(0.0312)	(0.0872)	(0.0142)	(9.7727)	(0.0427)	(0.0982)	(0.0148)
Δ ATT Pre (-1)	-0.1641	-0.0265	0.0681*	0.0166	-0.496	-0.0256	0.0536	0.0173
	(2.9794)	(0.0268)	(0.0384)	(0.0194)	(3.5735)	(0.0221)	(0.0441)	(0.018)
Δ ATT Pre (-2)	-0.6309	-0.0086	0.0586*	0.0158	-1.6632	-0.0102	0.052	0.0181
	(2.1897)	(0.0197)	(0.0302)	(0.0146)	(2.5633)	(0.0178)	(0.0326)	(0.0131)
Δ ATT Pre (-3)	-1.8356	-0.0105	0.0589**	0.0214**	-1.8633	-0.0081	0.0464**	0.0225**
	(1.159)	(0.0102)	(0.0235)	(0.0109)	(1.3491)	(0.009)	(0.0228)	(0.0098)
N Observations	12,711	12,711	12,711	12,629	13,125	13,125	13,125	13,072

$^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$. For regression specifications, see Table A.7. Top firms are defined based on total sales in the definition year. For variable definitions, see Table A.1.

A.3.7 Robustness: Alternative Definitions of *Data-Security Inventors*

This subsection examines whether our results depend on how we identify *data-security inventors*. Table A.16 summarizes the sample under several operationalizations: (i) lifetime security-share thresholds (e.g., ≥10%, ≥25%, ≥50% of an inventor's portfolio), (ii) minimum counts of security patents, (iii) recent five-year vs. lifetime portfolios, and (iv) expanded/restricted sets of CPC codes associated with cybersecurity.

Table A.17 reports the triple-difference estimates corresponding to each definition. In all cases, treated high-complementarity firms are compared to untreated high-complementarity firms (and likewise within the low group) in the same calendar year; thus, any level shifts from a particular inventor-definition difference out within subgroup-year cells. The estimated post-DBNL effects remain stable across definitions for innovation outcomes (citation- and value-weighted patents) and for growth/profitability (log assets, ROA), and pre-trends remain flat. We conclude that our heterogeneous effects are not driven by the specific inventor-definition used to construct the complementarity measure.

Table A.16: Descriptive Statistics on Patent Inventors

	Number of Inventors	Patent Portfolio Size								Share of CS Patents in Patent Portfolio (%)						
		mean	median	min	p10	p25	p75	p90	p95	Mean	Median	p10	p25	p75	p90	p95
Not data-security Inventor	3675035	5.104504	2	1	1	1	4	11	19	0%	0%	0%	0%	0%	0%	0%
Specialist data-security Inventor	162646	9.07897	4	1	1	2	9	19	32	56%	50%	14%	25%	100%	100%	100%

Specialist data-security inventors are defined as those that have at least **10**% of their lifetime patent portfolio in Cybersecurity. A patent is classified as a cyber-security patent if the United States Patent and Trademark Office (USPTO) assigns that patent at least one Cooperative Patent Classification (CPC) code associated with cyber security. The full list of CPC codes appears in Table A.1.

Table A.17: Differential Innovation Responses to DBNLs: Robustness to Alternative Definition of *Data-security Experts*

	Panel A: CS Investor based on 25% threshold				**Panel B: CS Investor based on 50% threshold**			
	Patent Count (All Patents)		Growth Variables	Profitability	Patent Count (All Patents)		Growth Variables	Profitability
	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ ATT Post	14.6738*	0.0717*	0.3136***	0.0692***	14.9241	0.0209	0.2980***	0.0344***
	(8.1444)	(0.0412)	(0.0667)	(0.0108)	(10.6966)	(0.052)	(0.0474)	(0.0064)
N Observations	17,670	17,670	17,670	17,584	17,165	17,165	17,165	17,082
	Panel C: CS Investor based on 90% threshold				**Panel D: CS Investor based on 10% threshold and Minimum 10 patents**			
	Patent Count (All Patents)		Growth Variables	Profitability	Patent Count (All Patents)		Growth Variables	Profitability
	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ ATT Post	157.7571***	-0.0194	-0.0649	-0.0318	37.6462***	0.1072**	0.1504*	0.0218**
	(19.8024)	(0.0531)	(0.0954)	(0.0232)	(11.5411)	(0.0385)	(0.0847)	(0.0099)
N Observations	16,929	16,929	16,929	16,863	17,024	17,024	17,024	16,952
	Panel E: CS Investor based on 25% threshold and Minimum 10 patents				**Panel F: CS Investor based on 50% threshold and Minimum 10 patents**			
	Patent Count (All Patents)		Growth Variables	Profitability	Patent Count (All Patents)		Growth Variables	Profitability
	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ ATT Post	31.7117*	0.1332**	0.3296***	0.0147	35.0273**	0.0365	-0.0173	-0.0431***
	(16.5882)	(0.0558)	(0.122)	(0.0126)	(13.55)	(0.0458)	(0.0951)	(0.0128)
N Observations	17,179	17,179	17,179	17,105	16,770	16,770	16,770	16,700

$^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$. This table reports triple-difference estimates under alternative definitions of cybersecurity (CS) specialists, a key component in identifying high-complementarity firms. The baseline definition (used in the main text) identifies data-security specialists as inventors whose lifetime share of data-security patents is at least 10%. Firms are classified as high-complementarity if, within their "major"non-data-security patent teams (defined as teams larger than the firm-specific median), the share of these specialists places the firm in the top quartile of the sample. Panels A through F maintain this baseline definition of major patent teams but vary the threshold for classifying data-security specialists. Panel A uses a stricter 25% data-security patent threshold, keeping the top-quartile criterion to define high-complementarity firms. Panels B (≥50%) and C (≥90%) use even stricter definitions of data-security specialists. Because fewer firms meet these higher thresholds, the 75th-percentile cutoff is zero. Thus, firms are classified as high-complementarity if they have a strictly positive share of these specialists. Panels D, E, and F add another criterion, requiring data-security specialists to have at least 10 lifetime patents. Panel D keeps the baseline 10% threshold and the top-quartile criterion. Panels E (≥25%) and F (≥50%) adopt stricter cybersecurity thresholds and, due to limited qualifying firm-years, define high-complementarity firms based on a strictly positive share of specialists. Detailed regression specifications are provided in Table A.7. Variable definitions and descriptive statistics are in Table A.1.

A.3.8 Robustness: Alternative Definitions of *"Major"* Patents

Our baseline definition of major non-security patents teams above the firm's median team size could mechanically overweight projects staffed by prolific inventors or be sensitive to how "major" is operationalized. In this section, we show that our findings survive when we do not restrict to "major" teams and when "importance" is defined by citation impact rather than team size.

To address these concerns, Appendix Table A.18 implements a full battery of alternative constructions of the high-complementarity indicator while holding the empirical design fixed. Specifically, we: (i) broaden and narrow the team-size threshold (Panels A–B: top 90% vs. top 10% by team size within firm-year); (ii) use all non-security patent teams with no "major" restriction (Panel C); (iii) fix each firm's complementarity status based on its pre-DBNL (2002) organization to rule out any look-ahead (Panel D); and (iv) replace team-size screens with citation-based importance, selecting teams whose patents exceed firm-specific thresholds of normalized forward citations (Panels E–F). This suite of tests directly targets potential concerns that our results are an artifact of prolific inventors, team-size choices, or post-event reclassification.

Table A.18: Differential Innovation Responses to DBNLs: Robustness to Alternative Definition of *Major Patent Teams*

	Panel A: High Comp Defn based on Top 90% Patents on Team Size				**Panel B: High Comp Defn based on Top 10% Patents on Team Size**			
	Patent Count (All Patents)		Growth Variables	Profitability	Patent Count (All Patents)		Growth Variables	Profitability
	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ ATT Post	27.0997***	0.1220**	0.2586***	0.0213**	22.7434***	0.0586	0.1074	0.0695***
	(7.0801)	(0.0287)	(0.0668)	(0.0095)	(7.419)	(0.049)	(0.0782)	(0.0088)
N Observations	18,766	18,766	18,766	18,676	14,728	14,728	14,728	14,651
	Panel C: High Comp Defn based on All Patent Teams				**Panel D: High Comp Defn based on Baseline but Fixed to 2002**			
	Patent Count (All Patents)		Growth Variables	Profitability	Patent Count (All Patents)		Growth Variables	Profitability
	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ ATT Post	18.5773***	0.1063**	0.0634	0.0355***	33.8928**	0.1198**	-0.1071	0.0116
	(6.0816)	(0.0254)	(0.0758)	(0.0106)	(13.6667)	(0.0528)	(0.1415)	(0.0132)
N Observations	23,841	23,841	23,841	23,713	17,296	17,296	17,296	17,218
	Panel E: High Comp Defn based on Top 75% Patent Teams on Citations				**Panel F: High Comp Defn based on Top 50% Patent Teams on Citations**			
	Patent Count (All Patents)		Growth Variables	Profitability	Patent Count (All Patents)		Growth Variables	Profitability
	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets	Citation weighted	Value Weighted	Log (Total Assets)	Return on Assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ ATT Post	40.6778***	0.1626**	0.2100**	0.0296**	18.2244**	0.1364**	0.0897	0.0231**
	(12.2755)	(0.0517)	(0.0959)	(0.0119)	(8.2814)	(0.0343)	(0.0636)	(0.0098)
N Observations	19,552	19,552	19,552	19,459	20,129	20,129	20,129	20,025

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. This table reports triple-difference estimates under six alternative definitions of *major patent teams* used to classify '-complementarity"firms. In each panel, we maintain the baseline criterion for high complementarity—firms whose share of cyber-security specialists on non-CS patent teams places them in the sample's top quartile—but vary how the patent teams are selected. For reference, the baseline specification (reported in the main text) defines patent teams as "if their inventor team size exceeds the firm-specific median. Panel A broadens this definition, classifying patent teams as major if they are larger than the firm-specific 10th percentile (largest 90 percent of teams). Panel B narrows the criterion, selecting only patent teams larger than the firm-specific 90th percentile (top 10 percent). Panel C uses all patent teams regardless of team size. Panel D maintains the baseline team-size definition (above firm-specific median) but fixes each firm's high-complementarity status at its 2002 level, the year just prior to the first DBNL. Panels E and F select patent teams based on their citation impact rather than team size. Specifically, citation impact is defined as the ratio of a patent's forward citations to the mean forward citations of patents filed in the same year, correcting for the shorter citation window for later-issued patents. Panel E selects teams working on patents above the firm-specific 25th percentile of citation impact, while Panel F selects those above the median. Figure 4 provides robustness checks using alternative thresholds for defining the high-complementarity firms. For detailed regression specifications, see Table A.7. Variable definitions and descriptive statistics appear in Table A.1.

Across all alternatives, the differential post-DBNL response for high- versus low- complementarity firms remains positive and statistically significant for citation-weighted patent output, with magnitudes that are economically large (e.g., ΔATT ranging from ≈18 to ≈41 citation-weighted patents per year across Panels A-F). Value-weighted patent counts are positive and significant in most specifications. Growth and profitability responses (log assets, ROA) are

likewise positive and robust in the majority of panels, with expected variation in precision as the definition moves away from the baseline. Using all teams (Panel C) confirms that the effect is not driven by the "major" filter; using citation-based importance (Panels E–F) shows the mechanism is not a team-size artifact; and freezing complementarity at its 2002 level (Panel D) demonstrates that results are not driven by any post-DBNL reclassification. Together, these patterns indicate that our main effect is not an artifact of prolific inventors or definitional choices about "major" projects, but reflects a broader organizational feature: the embedding of data-security specialists into non-security R&D that scales beyond a few star-led teams.

A.3.9 Robustness: Placebo Tests for Non-Security Technologies

We run a placebo test to show that the DBNL effect we document is specific to complementarities built on data security know-how rather than a generic reward to any specialised expertise. We replicate the entire construction of our high complementarity indicator in two technology domains with no conceptual link to data protection.

First, we use mechanical "packaging" knowledge defined by CPC subclasses `B65B` (packaging machinery), `B65C` (labelling equipment), and `B65D` (containers and package designs), including all CPC groups and subgroups under these subclasses.

Second, we use computer "graphics" defined by `G06T` (image data processing and generation) and `G09G` (display control circuits), including all CPC groups and subgroups. "Packaging" sits far outside the IT/data space, while "graphics" is IT adjacent yet non security; together they provide a strong falsification pair.

For each domain we identify "specialist" inventors whose personal patent portfolio contains at least 10% of patents in the relevant CPC roots. For every "non-packaging" (and "non-graphics") patent we compute the share of those specialists on the team, average this share over a rolling five-year window, and label firms in the top quartile of that distribution as high complementarity. We then re-estimate the same triple difference specification used in the baseline. If the coefficients tied to "packaging" cluster around zero and those for "graphics" remain small relative to the data-security-based complementarity estimate, the evidence points to a mechanism that is specific to data security knowledge rather than general inventor quality.

Comparability of the placebo measures. Beyond their intuitive appeal, the two placebo indicators are constructed to match the data-security measure in scope and support. As expected, "graphics" shows a moderate link to data–security ($\rho = 0.40$) because both sit in the broader IT space, whereas "packaging" is nearly orthogonal ($\rho = 0.05$). This limited overlap ensures the placebo variables are not mechanically capturing the same underlying dynamics, so any divergence in their DBNL responses is informative rather than artifactual. Reassuringly, the triple-difference regressions retain almost the similar number of observations as the main specification (about 17.9 k for data-security, 16.9 k for "packaging", and 17.8 k for "graphics"), so the absence of significant effects in the placebos cannot be attributed to a loss of statistical

power.

"Packaging" placebo. Panel A of Table A.19 shows that redefining complementarity around "packaging" specialists yields no meaningful DBNL response. The post-DBNL triple-difference coefficient on citation-weighted patent output drops to −12.3 (s.e. 6.7) versus +28.9 in the data-risk baseline, and the value-weighted count is essentially zero. Coefficients for firm size, market capitalization, and return on assets are small and never significant. Taken together, these results indicate that DBNLs reward complementarities tied to data-security know-how rather than to unrelated expertise such as "packaging".

"Graphics" placebo. Panel B of Table A.19 reruns the analysis with "graphics" specialists. The post-DBNL triple-difference coefficient on citation-weighted patents is 12.5 (s.e. 9.0)—less than half the +28.9 observed for cyber complementarity and statistically weak. High-graphics firms actually contract: total assets and market capitalization fall by 0.31 and 0.30 log points, respectively, both significant at conventional levels, while profitability and value-weighted patenting stay flat. The absence of an innovation surge and the negative growth response underline that the DBNL effect is unique to complementarities anchored in data-security know-how.

Table A.19: Differential Innovation Responses to DBNLs: Placebo Tests with Alternative Dimension of *Complementarity*

	Panel A: Complementarity Defined by Packaging Experts in Non-Packaging Patents					**Panel B: Complementarity Defined by Graphics Experts in Non-Graphics Patents**				
	Patent Count (All Patents)		Growth Variables		Profitability	Patent Count (All Patents)		Growth Variables		Profitability
	Citation weighted	Value Weighted	Log (Total Assets)	Log (Market Cap)	ROA	Citation weighted	Value Weighted	Log (Total Assets)	Log (Market Cap)	ROA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ ATT Post	-12.2842*	0.0268	-0.0262	0.0984	-0.0051	12.4974	0.0184	-0.3112***	-0.2974*	0.001
	(6.7454)	(0.0409)	(0.0776)	(0.0922)	(0.0151)	(9.0198)	(0.0417)	(0.0863)	(0.1789)	(0.0242)
ATT Low Comp Post	8.999	-0.0384	-0.0079	-0.0819	0.0255***	-2.4635	-0.011	0.0901***	0.1278**	0.0382***
	(6.712)	(0.0391)	(0.0313)	(0.0656)	(0.0091)	(3.5289)	(0.0225)	(0.0323)	(0.055)	(0.0124)
Δ ATT Pre (-1)	-9.6575	0.0038	-0.0477	0.0799	0.0014	-9.6331	0.0251	-0.0811	-0.023	-0.0175
	(4.8019)	(0.02)	(0.0496)	(0.0976)	(0.0108)	(4.0502)	(0.0315)	(0.0485)	(0.1514)	(0.0155)
Δ ATT Pre (-2)	-4.2137	0.0154	-0.056	-0.0243	0	-9.0292**	-0.0002	-0.0699	-0.0264	-0.0173
	(4.8019)	(0.02)	(0.0496)	(0.0976)	(0.0108)	(4.0502)	(0.0315)	(0.0485)	(0.1514)	(0.0155)
Δ ATT Pre (-3)	2.3674	0.0054	-0.0468	0.0354	-0.0084	-8.0077***	-0.0008	-0.073*	-0.1136	-0.0268**
	(2.971)	(0.0135)	(0.032)	(0.0534)	(0.0097)	(2.5515)	(0.0211)	(0.0404)	(0.0777)	(0.0093)
ATT Low Comp Pre (-1)	-0.5696	-0.0359	0.0655	-0.0069	0.0129	-1.3758	-0.0433**	0.0736	0.0234	0.0184
	(3.4703)	(0.0292)	(0.0585)	(0.1044)	(0.0204)	(2.6354)	(0.0205)	(0.0519)	(0.1135)	(0.0177)
ATT Low Comp Pre (-2)	-2.4969	-0.0234	0.073	0.0713	0.0156	-1.387	-0.0198	0.0726*	0.0731	0.0193
	(2.559)	(0.023)	(0.0445)	(0.0701)	(0.0138)	(2.1441)	(0.0173)	(0.0391)	(0.0755)	(0.0136)
ATT Low Comp Pre (-3)	-3.3862***	-0.0148	0.0583**	0.0593	0.0209*	-0.3464	-0.0145	0.066**	0.1017***	0.0258**
	(1.1814)	(0.0119)	(0.0287)	(0.0442)	(0.0106)	(1.3187)	(0.0105)	(0.0255)	(0.0461)	(0.0089)
N Observations	16863	16863	16863	16854	16799	17812	17812	17812	17802	17726

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. This table presents triple-difference estimates from a placebo exercise in which high complementarity is redefined around technologies unrelated to data security. The goal is to test whether the data-security feedback mechanism driving our baseline results also appears when the underlying expertise is in packaging or computer graphics. Panel A constructs complementarity around *packaging* know-how. Packaging specialists are inventors whose cumulative portfolio has at least 10% of patents with CPC codes `B65B`, `B65C`, or `B65D` (including all CPC groups and subgroups under each code). Panel B repeats the procedure for *computer graphics*, defining specialists by a 10% threshold in `G06T` (including all CPC groups and subgroups). For each domain, a firm's complementarity share is the average fraction of specialists on its non-packaging (or non-graphics) patents over the prior five years. Firms whose share places them in the top quartile of the sample distribution are labelled "high complementarity"—the same rule applied in the data-security baseline. Other details remain same as in the baseline Table A.7. Variable definitions and descriptive statistics appear in Table A.1.

Appendix B Theoretical Proofs and Derivations

B.1 Proof of Proposition 1. (Welfare Loss from Outsourcing)

Let I denote the cost of internal integration and I^{ext} the cost of outsourcing. Under internal integration, the firm enjoys both the reduction in expected loss from data exposure (captured by $\lambda\mu$) and the reputational/compliance benefit δ_0, and additionally earns innovation spillovers proportional to $\gamma\sqrt{c}$. Under outsourcing, we assume that the firm obtains the same reduction in expected loss and the same reputational benefit, but no innovation spillovers, at a possibly lower up-front cost I^{ext}.

Formally, the value under internal integration is

$$V^{\text{internal}}(c) = -I + (\gamma\sqrt{c} - \lambda)\mu + \delta_0,$$

consistent with the baseline model in Section 5. Under outsourcing, there are no innovation spillovers but the risk and compliance terms are the same, so

$$V^{\text{external}} = -I^{\text{ext}} - \lambda\mu + \delta_0.$$

Taking the difference,

$$\Delta V(c) := V^{\text{internal}}(c) - V^{\text{external}} = \big[-I + (\gamma\sqrt{c} - \lambda)\mu + \delta_0\big] - \big[-I^{\text{ext}} - \lambda\mu + \delta_0\big] = -(I - I^{\text{ext}}) + \gamma\mu\sqrt{c}.$$

Since $\gamma, \mu > 0$ and $\sqrt{c}$ is strictly increasing in c, it follows that

$$\Delta V(c) > 0 \quad \text{for all } c > \tilde{c}, \qquad \text{where} \quad \tilde{c} = \left(\frac{I - I^{\text{ext}}}{\gamma\mu}\right)^2.$$

Moreover,

$$\frac{\partial \Delta V}{\partial c} = \frac{\gamma\mu}{2\sqrt{c}} > 0, \qquad \frac{\partial \Delta V}{\partial \gamma} = \mu\sqrt{c} > 0, \qquad \frac{\partial \Delta V}{\partial \mu} = \gamma\sqrt{c} > 0, \qquad \frac{\partial \Delta V}{\partial (I - I^{\text{ext}})} = -1 < 0.$$

Hence the welfare loss $\Delta V(c)$ is increasing in c, γ, and μ and decreasing in the cost gap $(I - I^{\text{ext}})$. In our calibrated model, all high-complementarity firms with $c > \bar{c}$ suffer a strictly positive loss if restricted to outsourcing. This concludes the proof.

B.2 Proof of Proposition 2. (Subsidy)

From the indifference condition

$$V_{\text{now}}(c) + S(c) = V_{\text{wait}}(c),$$

it follows directly that

$$S(c) = V_{\text{wait}}(c) - V_{\text{now}}(c). \tag{B.1}$$

Recall that

$$V_{\text{now}}(c) = -I + (\gamma\sqrt{c} - \lambda)\mu + \delta_0,$$

and

$$V_{\text{wait}}(c) = \rho\left[\max\{-I + (\gamma\sqrt{c} - \lambda)\theta + \delta_0, 0\}\right], \quad \theta \sim \mathcal{N}(\mu, \sigma^2).$$

For notational convenience, define

$$A(c) := \gamma\sqrt{c} - \lambda, \qquad X(c, \theta) := -I + A(c)\,\theta + \delta_0.$$

Then

$$V_{\text{now}}(c) = -I + A(c)\mu + \delta_0, \qquad V_{\text{wait}}(c) = \rho\left[\max\{X(c, \theta), 0\}\right].$$

It is convenient to introduce

$$F(c; \gamma, \mu, \rho) := V_{\text{now}}(c) - V_{\text{wait}}(c),$$

so that $S(c) = -F(c; \gamma, \mu, \rho)$ by (B.1). The derivatives of S with respect to (c, γ, μ, ρ) are therefore given by $S_x = -F_x$ for $x \in \{c, \gamma, \mu, \rho\}$.

Derivative with respect to c. Using $\partial A/\partial c = \gamma/(2\sqrt{c})$ we obtain

$$\frac{\partial V_{\text{now}}}{\partial c} = \mu\,\frac{\gamma}{2\sqrt{c}}.$$

For $V_{\text{wait}}(c)$, note that

$$\frac{\partial}{\partial c}\max\{X(c, \theta), 0\} = \mathbf{1}\{X(c, \theta) \geq 0\}\,\frac{\partial X(c, \theta)}{\partial c} = \mathbf{1}\{X(c, \theta) \geq 0\}\,\theta\,\frac{\partial A}{\partial c},$$

so by dominated convergence

$$\frac{\partial V_{\text{wait}}}{\partial c} = \rho\,\frac{\gamma}{2\sqrt{c}}\left[\theta\,\mathbf{1}\{X(c, \theta) \geq 0\}\right].$$

Hence

$$F_c := \frac{\partial F}{\partial c} = \frac{\gamma}{2\sqrt{c}}\left(\mu - \rho\left[\theta\,\mathbf{1}\{X(c, \theta) \geq 0\}\right]\right).$$

Under the mild condition $\mu > \rho\,[\theta^+]$, which holds in our calibration, the bracket is strictly positive for all $c \in [0, 1]$, so $F_c > 0$ and therefore

$$\frac{\partial S}{\partial c} = -F_c < 0.$$

Thus $S(c)$ is strictly decreasing in c.

Derivative with respect to γ. Using $\partial A/\partial\gamma = \sqrt{c}$ we obtain

$$\frac{\partial V_{\text{now}}}{\partial \gamma} = \mu\sqrt{c}.$$

Similarly,

$$\frac{\partial}{\partial \gamma}\max\{X(c,\theta),0\} = \mathbf{1}\{X(c,\theta)\geq 0\}\,\theta\,\frac{\partial A}{\partial \gamma} = \mathbf{1}\{X(c,\theta)\geq 0\}\,\theta\sqrt{c},$$

so

$$\frac{\partial V_{\text{wait}}}{\partial \gamma} = \rho\sqrt{c}\,[\theta\,\mathbf{1}\{X(c,\theta)\geq 0\}]\,.$$

Hence

$$F_\gamma = \sqrt{c}\,(\mu - \rho\,[\theta\,\mathbf{1}\{X(c,\theta)\geq 0\}])\,,$$

which is strictly positive under the same condition $\mu > \rho\,[\theta^+]$. It follows that

$$\frac{\partial S}{\partial \gamma} = -F_\gamma < 0,$$

so the required subsidy $S(c)$ is strictly decreasing in γ.

Derivative with respect to ρ. Finally, $V_{\text{now}}(c)$ does not depend on ρ, while

$$\frac{\partial V_{\text{wait}}}{\partial \rho} = [\max\{X(c,\theta),0\}] > 0.$$

Hence

$$F_\rho = -\frac{\partial V_{\text{wait}}}{\partial \rho} < 0, \qquad \frac{\partial S}{\partial \rho} = -F_\rho > 0,$$

so the required subsidy is strictly increasing in ρ.

Collecting these results, we obtain that $S(c)$ is strictly decreasing in c and γ, strictly increasing in ρ. This finalizes our proof.

Appendix C Simulation

C.1 Assumed and Calibrated parameters

All codes to generate the figures in this paper will be put in a public repository.

Table C.1: Microfoundation of Calibrated Parameters (5-Year Horizon, Per $1M Firm Size)

Parameter	Value	Economic Interpretation and Microfoundation	Notes
γ	0.4714	Innovation return per $1M in assets and $4.01M per patent.	Fixed from empirical data
$\bar{c}$	0.75	Threshold complementarity level at which firms begin responding to DBNL with increased innovation.	Empirically observed cutoff
δ_0	1.87	Fixed reputational or policy-related benefit from integration (e.g., improved disclosure, regulatory credibility).	$\approx 0.2\%$ of firm value over 5 years for median firm; aligns with estimates from Akey, Lewellen, Liskovich, and Schiller [2025], Bharadwaj et al. [2025]
λ	3.0	Expected cost of inaction, including breach risk, trust erosion, and opportunity loss.	$\approx$ 2.85M for median firm (0.3% of assets); consistent with expected breach probabilities and impact studies [IBM Security, 2025, VC3, 2024, Cybrwise, 2024].
μ	1.00	Mean relevance of data security shock.	Normalized
I	0.0723	One-time cost of internal integration over 5 years; equivalent to $67.45M for median firm	Calibrated in the model. Reflects irreversible organizational investment (not operational security spend); consistent with case studies (e.g., JPMorgan, Amazon)
I^{ext}	0.0107	Corresponds to a total outsourcing cost over 5 years of $10M for median firm, $2M a year, scaled by the median firm size $I^{\text{ext}} = \frac{10}{933} \approx 0.0107$	Industry reports
ρ	0.8	Discount factor over 5 years (corresponds to 0.96 annually).	Matches 5-year horizon; implies annual discount rate of 2–4%

Notes: This table presents the key parameters used in the quantitative calibration of the model. All values are expressed per $1 million of firm assets and reflect a five-year investment horizon. γ denotes the innovation return per unit of data complementarity, derived from empirical estimates of patent elasticity and value. λ captures the expected economic cost of inaction, including breach-related risk and trust erosion. δ_0 reflects the fixed reputational or compliance-related benefit of integration. I is the present value of the irreversible investment needed for integration. The discount factor ρ captures firm patience over a five-year horizon. This parameterization aligns the model's threshold prediction with observed firm behavior around $\bar{c} = 0.75$.

C.2 Targeted, Untargeted, and Empirical Moments

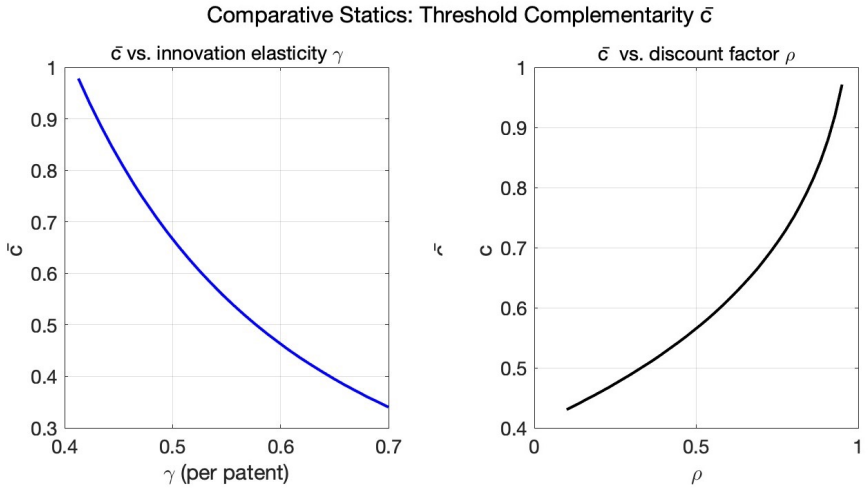
Table C.2: Calibration and Validation: Targeted vs. Untargeted Moments

Moment (Statistic)		Values	
		Data	Model
Panel A. Targeted (used in calibration)			
Implied investment (integration) cost	I^* (\$M)[a]	67.45	67.45
Innovation value	Δ patent value, \$M	380.95	380.95
Total patents	ΔP_{total} (count)	95	95
Scale parameter	γ (per Eq. (9))	0.4714	0.4714
Inaction cost	λ (\$M)	3.00	3.00
Compliance benefit	δ_0 (model units)	1.87	1.87
Discount factor	ρ	0.80	0.80
Panel B. Untargeted (over-identifying checks)			
Average ROA ($c = 0.75$)	ROA (pp)	0.0618	0.0533

Notes: [a] $I^* = 0.0723$ (model units). Dollar amount uses median high-c assets at $c = 0.75$ of \$933M: \$67.45M $= 0.0723 \times 933$. [b] Panel A lists moments used to determine $V_{wait} = V_{now}$ at a threshold level of $c = 0.75$; Panel B reports untargeted moments. "Targeted" marks objects explicitly used in calibration. High-c (low-c) indicate firms above (below) the ex-ante 75% complementarity cutoff. Post-shock windows default to $t \in \{+1, ..., +5\}$ vs. pre $\{-5, ..., -1\}$.

C.3 Comparative-statics Figure

Figure C.3: Comparative Statics of Threshold Complementarity



Legend: The left panel is the comparative static of the complementarity threshold with respect to γ (the innovation elasticity from Equation 9), and the right panel is the comparative static of the complementarity threshold with respect to ρ (the discount factor per Table C.1). Each panel varies one parameter at a time, holding the others at their calibrated baselines. The curves intersect at the benchmark $(\gamma, \rho, \bar{c}) = (0.4714, 0.8, 0.75)$. As predicted by the model, $\bar{c}$ falls in γ and rises in ρ, providing a visual sanity check. Table C.1 provides further details on the parameters.

C.4 External Validity

Holding the calibrated primitives fixed at $\gamma = 0.4714$, $\lambda = 3.0$, $\delta_0 = 1.87$, $\mu = 1$, and $\rho = 0.8$, we can ask how large the irreversible integration cost could be, for a given complementarity level c, while still making immediate investment optimal. For each $\bar{c}$, define the *break-even* integration cost $I_{\text{breakeven}}(\bar{c})$ as the unique solution to

$$V_{\text{now}}(\bar{c}; I_{\text{breakeven}}(\bar{c})) = V_{\text{wait}}(\bar{c}; I_{\text{breakeven}}(\bar{c})),$$

where V_{now} and V_{wait} are given in Section 5. By construction of our calibration, the empirical cutoff $\bar{c} = 0.75$ implies

$$I_{\text{breakeven}}(0.75) = I^* = 0.0723.$$

Keeping $(\gamma, \lambda, \delta_0, \mu, \rho)$ fixed and solving the same indifference condition for higher values of c yields

$$\begin{aligned}
\bar{c} = 0.80: &\quad I_{\text{breakeven}}(0.80) \approx 0.1402, \\
\bar{c} = 0.85: &\quad I_{\text{breakeven}}(0.85) \approx 0.2133, \\
\bar{c} = 0.90: &\quad I_{\text{breakeven}}(0.90) \approx 0.2722, \\
\bar{c} = 0.95: &\quad I_{\text{breakeven}}(0.95) \approx 0.3445.
\end{aligned}$$

Because all model variables are normalised per \$1M of assets, these figures translate for the median firm in our sample (total assets \$933M) into outlays of roughly \$67.5M, \$130.8M, \$199.0M, \$254.0M, and \$321.4M, respectively. Thus, moving from the empirical threshold $\bar{c} = 0.75$ to $\bar{c} = 0.95$ raises the break-even integration cost by a factor of almost five, even though the change in complementarity is only twenty percentage points. Thus, although $I_{\text{breakeven}}(c)$ increases roughly linearly in c over this range, the associated changes in dollar magnitudes are large: firms with very high data–innovation complementarity are willing to commit a much larger fraction of their balance sheet to internal integration than firms just at the empirical threshold.

This pattern reflects the fact that, holding the shock process fixed, higher complementarity both raises the immediate payoff from internal integration and increases the value of protecting that payoff with an irreversible organisational investment. In levels, the implied integration budgets, roughly \$130–\$320 million for highly integrated firms, are well within the range of large-scale IT and security overhauls documented in practice.[14] These magnitudes are of the same order as large-scale IT and security overhauls documented in practice, such as JPMorgan Chase's well-publicised multi-year cybersecurity programme with an annual budget of about \$500 million.[15]

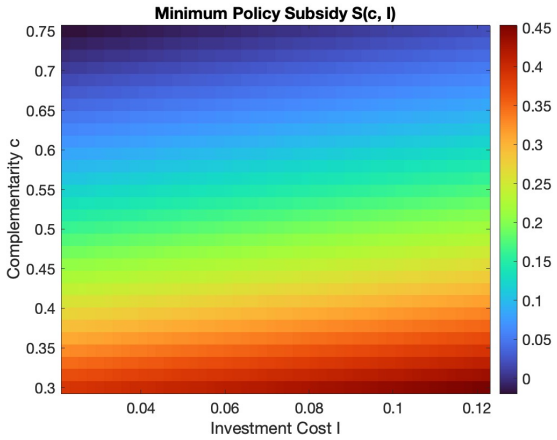
C.5 Simulated Subsidy

To assess how the required subsidy changes under different realistic assumptions, we recompute the minimum subsidy when the firm has a range of complementarity levels ($c = 0.3 - 0.75$) or when the firm faces a range of integration costs ($I = 0.0220 - 0.1220$ corresponding to \$6M - \$33M over a period of 5 years).

[14] For each c, $I_{\text{breakeven}}(c)$ is defined implicitly by $V_{\text{now}}(c; I_{\text{breakeven}}(c)) = V_{\text{wait}}(c; I_{\text{breakeven}}(c))$ and is computed numerically. These values represent the *maximum* integration costs that would still leave a firm with complementarity c indifferent between investing immediately and waiting. The baseline calibration uses $I = I^*$ at $c = \bar{c} = 0.75$; the higher numbers for $c \in \{0.80, 0.85, 0.90, 0.95\}$ are counterfactual willingness-to-pay levels, not additional cheques written on top of I^*.

[15] Press reports indicate that JPMorgan Chase committed roughly \$500 million per year to cybersecurity following a major breach, illustrating that multi-hundred-million-dollar security investments are empirically plausible for large financial institutions.

Figure C.5: Minimum Policy Subsidy for a Range of Complementarity and Investment Levels (per $1M)



Legend: Simulated minimum policy subsidy $S(c, I)$ required to induce immediate internal investment, plotted across a range of data-innovation complementarity levels $c \in [0.3, 0.75]$ and investment cost levels $I \in [0.0220, 0.1220]$ corresponding to [$6M, $33M] over a period of 5 years. Subsidy values are expressed in model units (per $1 million in assets) and are strictly decreasing in c, and strictly increasing in I, consistent with the model's prediction that firms with greater structural integration require less external support, and firms with higher costs of structural integration require more external support. The top-left square is calibrated to require a subsidy of zero for a firm of complementarity level $c = \bar{c} = 0.75$ and an investment level of $I = 0.0723$. Table C.1 provides further details on the parameters.