Data Innovation Complementarity and Firm Growth*

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

August 27, 2025

Abstract

Data security is often viewed as a defensive cost. But in today's data-driven economy, can it also create value? We propose a novel measure of *"data innovation complementarity"*, capturing the extent to which firms have structurally integrated data-security expertise into their broader innovation. We show that such integration unlocks cross-domain knowledge spillovers, fueling firm-wide innovation and growth. Exploiting U.S. Data Breach Notification Laws as a quasi-exogenous increase in the perceived importance of data security, we find that only high-complementarity firms realize these gains. A calibrated real-options model shows that integrated data protection is a significant driver of innovation-led growth.

Keywords: Data security, data governance, patent teams, innovation, firm growth, value of data.

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

*First version: January 31, 2023. This version: August 27, 2025. This paper previously circulated under the title "Data Risk, Firm Growth, and Innovation." We thank Tania Babina, Francesco Celentano, Fabrice Collard, Andreas Fuster, Leonardo Gambacorta, 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 Venakteswaran, and Baozhong Yang for insightful suggestions, as well as participants at several 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 our own.

†Haas School of Business at UC Berkeley, United States, fedyk@berkeley.edu.

‡University of Lausanne, Switzerland, roxana.mihet@unil.ch.

§University of Lausanne, and University of Basel, Switzerland, kumar.rishabh@unil.ch.

¶ISCAL, Portugal, omgomes@iscal.ipl.pt.

1 Introduction

Data security is often viewed as a necessary cost—an insurance policy against worst-case scenarios. But in today's data-driven economy, this view is increasingly outdated. For a growing number of firms, investments in data protection are not merely defensive; they are strategic enablers of innovation and growth. Critically, robust data security can unlock access to vast quantities of sensitive customer data: on payments, preferences, and behaviors, that would otherwise be inaccessible due to privacy or trust concerns. For example, Amazon's early investments in secure payment infrastructure facilitated innovations such as the patented '1-Click' purchasing system, contributing to its growth and the broader e-commerce market.

Our paper is the first to formalize this phenomenon through what we term *"data innovation complementarity,"* the degree to which data-related capabilities, particularly in data security, are structurally integrated with a firm's broader innovation activities. When this complementarity is high, data protection becomes a foundation for competitive advantage, enabling not just compliance and risk mitigation, but also enhanced capacity for data accumulation and innovation. The Amazon case exemplifies a broader pattern in our data: firms that embed security specialists within cross-functional inventor teams are more likely to convert compliance-driven efforts into innovation-led competitive advantage. These firms are better positioned to gain customer trust, collect high-value behavioral and transactional data, and use those data to fuel product development, personalization, and firm growth. Thus, data innovation complementarity converts a defensive capability into a sustained source of competitive advantage.

We operationalize this idea by developing a new, patent-based measure of data innovation complementarity. We identify data-security specialists as engineers and inventors who allocate at least 10% of their patenting activity to data-security-related innovations, and track their participation in non-security patent teams. This team-level integration measure captures not only a firm's technical investment in data security, but also the structural embedding of that expertise into the firm's broader organizational structure.

Assessing the value of data innovation complementarity is an empirical challenge, because of potential confounds. The recent rise of the data economy has made data security more relevant, and firms with integrated data security expertise have performed well. But, is there a causal relationship between these trends? To speak to this question, we need exogenous shocks to the relevance of data security. The econometric challenge is that the value of data differs between firms and fluctuates over time, but these differences are difficult to identify empirically and likely to be endogenous to firms' activities.

We address this challenge by exploiting quasi-experimental variation in data security relevance created by the staggered adoption of U.S. state-level Data Breach Notification Laws (DBNL). These laws, introduced between 2003 and 2018, require firms to disclose data breaches involving customer information, thereby increasing the perceived stakes of data governance and protection. The staggered adoption of the DBNL offers substantial variation, with 27 states adopting them between 2003 and 2006 and another 18 states adopting them between 2007 and 2009. This variation allows us to estimate difference-in-difference effects from the DBNL on firm outcomes, separately for high data-innovation-complementarity firms and low data-innovation-complementarity firms.

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 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 using values following Kogan et al. (2017). 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.

Second, we find that the increased innovation of high–complementarity firms translates into a long-lasting boost in growth, profitability, and market valuations, while a decline for the low-complementarity firms. Relative to the levels implied by the control group's trajectory, 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 after DBNL adoption (the sales estimate is statistically insignificant). However, their return on assets increases by around 3 percentage points. High-complementarity firms show asset levels about 7.5 percent higher, market capitalization 36 percent higher, sales 16 percent higher, and return on assets roughly 6 percentage points above their counterfactual levels.

Third, we provide a mechanism for these results. Specifically, 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, with increased self-citations of own data-security-related patents, leading to a positive data feedback loop. First, we document that both high- and low-complementarity firms increase their cyber-risk disclosures. Second, there is a substantial difference in how high- versus low-complementarity firms deploy their data security (i.e., cybersecurity) talent when it comes to related non-data-security innovation. 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 most "major" patents.[1] 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. Finally, we examine the citation patterns. Low-complementarity firms experience insignificant changes in self-citations of their data-security-related patents by their non-data-security-related patents. In contrast, high-complementarity firms see the share of self-data-security-patent citations in their non-data-security-related patents increase

[1] "Major" patents are defined as patents whose inventor team size exceeds the firm's median patent team size over a rolling five-year window.

by approximately 0.2%. High-complementarity firms not only increase their focus on data security, but also increasingly integrate it into their broader innovation activities, facilitating knowledge spillovers.

Thus, the differential innovation response between high- and low-complementarity firms to the perceived increase in the importance of data security following DBNL stems from the distinct organizational capabilities of these two types of firms. High-complementarity firms respond to DBNL by redeploying security experts across a wide array of innovation projects, enabling cross-domain knowledge transfer. This effect is not due to general hiring or over-spending. Rather than expanding team size or spending more on innovation, high-complementarity firms reallocate existing inventors, resulting in increased self-citations from data-security patents to other patent classes. 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.

To interpret these findings and explore policy and counterfactual implications, we develop a real-options model where firms face an investment decision under uncertainty: whether to integrate data protection into the general innovation process or treat it as a standalone compliance function in an environment where expected data breach losses are uncertain. Integration is costly and largely irreversible, reflecting sunk investments, organizational complementarities, and reconfiguration frictions as in the standard literature (Dixit and Pindyck, 1994; Milgrom and Roberts, 1990; Bloom and Sadun, 2012). These frictions make dis-integration both economically and operationally inefficient. However, when the perceived importance of data security increases, firms with integrated structures realize innovation spillovers that increase their productivity frontier. Our model formalizes this strategic waiting game: irreversibility amplifies the value of delay, so firms with low complementarity optimally postpone integration, while high-complementarity firms act earlier and realize outsized gains when data security relevance materializes.

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 high-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 cybersecurity 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.

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 using patents to study specialization, AI adoption, and creative destruction (Bena et al., 2021; Alderucci et al., 2020; Kakhbod et al., 2024).

Second, we position our findings within five mechanisms proposed in the literature through which data technologies[2] can create value: building customer trust and loyalty or enhancing reputation and brand value (e.g., Cao et al., 2025; Bharadwaj et al., 2025; Akey et al., 2025), gaining competitive advantage through innovation and differentiation (e.g., Rock, 2019; Tambe et al., 2020; Eisfeldt et al., 2023; Rishabh et al., 2024; Ramadorai et al., 2025), reducing costs and improving efficiency (e.g., Farboodi et al., 2019; Mihet and Philippon, 2019; Mihet et al., 2025), and ensuring regulatory compliance (e.g., Florackis

[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).

et al., 2023; Jamilov et al., 2021). 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 supporting evidence consistent with 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 debt costs, reputation, and firm financing (e.g., Cao et al., 2025; Akey et al., 2025; Florackis et al., 2023; Jamilov et al., 2021), and on how stronger IT governance fosters more credible disclosures and fewer incidents (Bharadwaj et al., 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 (Eisfeldt and Papanikolaou, 2013) by emphasizing functional integration over investment intensity alone. This parallels findings by 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.

The remainder of the paper proceeds as follows. Section 2 describes the data, introduces our novel data innovation complementarity measure, and details the empirical strategy using the staggered rollout of DBNL. Section 3 presents our main empirical findings. Section 4 introduces the real-options model and explores welfare and policy counterfactuals. Section 5 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 (DBNL).

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 life-time patent portfolio consists of data-security patents, where we identify data-security patents using the CPC classification, as described in the next subsection. The 10% threshold ensures genuine specialization, rather than incidental involvement.

Second, we focus on a firm's *major* non-data-security patents, defined as those with an inventor team size above the firm's median. We concentrate on these patents because team size is a strong proxy for a project's strategic importance and resource intensity. For each of these major patents, we calculate the share of its inventor team comprising

data-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. Finally, 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. Unless otherwise noted, we fix the definition of a high-data-innovation-complementarity firm using the 75^{th} percentile threshold throughout the main analysis. However, we also examine how economic outcomes vary across alternative levels of complementarity 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 et al., 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 (see Figure A.5 in Appendix A.3.1).

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.

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^{th} percentile. A firm is counted as high-complementarity in this table if it met that threshold in at least half of its firm-year observations.

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. Table A.6 in the Appendix reports further sector-level statistics on non-cybersecurity patent teams. Variation is large: in the *Information* sector an av-

erage 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]

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, presented in Appendix A.3.2, remain materially unchanged, confirming that complementarity captures an organizational advantage rather than industry-specific artefacts.

Innovation Measures We measure firms' innovation using their patent activity. Following the literature, we weight patents according to their scientific impact using forward citations (Kogan et al., 2017; Aghion et al., 2013; Howell, 2017). More scientifically influential patents receive more citations (Hall et al., 2005). We correct for truncation bias in citations, since recent patents naturally have fewer citations, following the adjustment method proposed by Hall et al. (2001). All patent data are sourced from the publicly available dataset by Kogan et al. (2017) (KPSS dataset).

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. These patent valuation data come from the KPSS dataset.

For robustness, we also use R&D expenditures from Compustat.

[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.

Product versus Process Innovation We classify patents by their technological focus using the claims dataset from Ganglmair et al. (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.[4] The time series of our aggregate cyber-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 et al. (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. 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.

Although our analysis covers the period 2000–2022, this classification identifies AI-related patents even in earlier years. This occurs for two reasons. First, research and innovation in AI predates the currently dominant Generative AI technology and the increase in commercial AI applications over the past decade. Second, the classification encompasses a broad range of AI patents, including hardware-focused inventions.

[4]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."

Financial Data We obtain firm-level financial information from the merged CRSP-Compustat database. We calculate the following financial variables and ratios: log of total assets, log of market capitalization, and return on assets.

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.

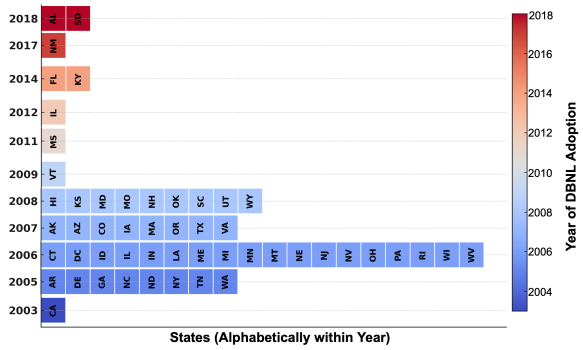
2.2 Empirical Strategy

Data Breach Notification Laws (DBNL) 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. DBNL require organizations to notify individuals, regulators, and other stakeholders after data breaches involving unauthorized access or disclosure.[5] All 50 states enacted DBNL between 2003 (California) and 2018 (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 DBNL in 2002–2004, another 18 states did so in 2006, and 17 states did so in 2007–2008. This variation makes DBNL a sufficiently-powered source of quasi-random variation in firms' exposure to increased relevance of data security.

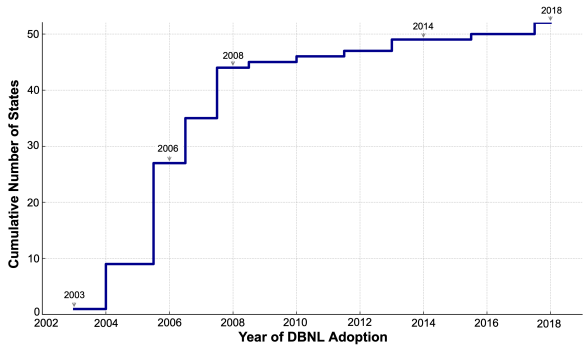
DBNL provide suitable exogenous variation to identify the economic effects of data security relevance. First, the staggered timing of DBNL adoption reflects state legislative processes, which are unlikely to correlate with individual firms' prior innovation or data complementarity profiles. We empirically confirm this assumption through pre-trend analyses. Second, DBNL substantially raise firms' costs and consequences of data breaches, directly influencing firms' incentives to invest in data security (Boasiako and Keefe, 2021; Liu and Ni, 2023; Huang and Wang, 2021). Third, mandatory breach dis-

[5]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.

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



Panel A: State-by-State Adoption



Panel B: Cumulative Adoption (CDF)

Legend: Panel A shows the state-by-state chronological adoption of Data Breach Notification Laws in the United States. 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. By 2008, over half of states had already introduced these laws.

closures ensure accurate and consistent measurement of data breaches across states and time.

We compare innovation decisions of firms headquartered in states adopting DBNL earlier versus later. This comparison addresses the "forbidden comparisons" issue common in staggered difference-in-differences designs (Goodman-Bacon, 2021). Although compliance involves operational costs (Boasiako and Keefe, 2021; Liu and Ni, 2023), DBNL also motivate firms to strengthen data management strategies. Our approach is similar to Jiang et al. (2021), who examine fintech disruptions to study how fintech regulation tests firms' innovation capacity. We study whether firms' pre-existing complementarity between data-security and other innovation domains shapes their responses to heightened data security relevance resulting from DBNL. We compare outcomes for firms in early-adopting states against those in states adopting later.

A potential concern with our identification strategy is that DBNL obligations are triggered by firms' collection or storage of personal data belonging to residents of the adopting state, whereas we assign treatment based on firms' headquarters location. This mismatch could generate spillovers that violate the Stable Unit Treatment Value Assumption (SUTVA), if firms have substantial customer bases in states with different DBNL timing. Specifically, firms headquartered in treated states, but serving customers

primarily in untreated states may face weaker exposure (false positives), while firms headquartered in untreated states, but serving customers in treated states may experience some exposure despite being classified as controls (false negatives). In practice, however, firms' data-security policies and investments are typically determined centrally at the headquarters level and apply firm-wide, particularly since most firms in our sample have customers and operations in their HQ state. Moreover, prior studies (e.g., Liu and Ni (2023); Boasiako and Keefe (2021);) similarly assign treatment based on headquarters location, consistent with organizational realities. To the extent that such spillovers exist, they attenuate the observed contrast between treated and untreated groups, implying that our estimates represent conservative lower bounds of the true effects. Finally, the statistically significant and economically meaningful differences we observe between high- and low-complementarity firms suggest that any such attenuation does not overturn our main conclusions.

Methodology To accurately identify the causal impact of DBNL 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 et al., 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 et al., 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 typi-

cally 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—such as those proposed by Callaway and Sant'Anna (2021), de Chaisemartin and d'Haultfœuille (2020), and 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 et al., 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 DBNL' 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 DBNL 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, standard errors are clustered at the state level, aligning with the level of treatment variation as prescribed by Abadie et al. (2023).

Given the institutional context, firms likely anticipate DBNL 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 (see 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.3.4, 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. These pre-treatment effects no longer emerge after incorporating the anticipatory shift, supporting our adjusted timing approach.[6]

Separately, to conduct a diagnostic pre-trends test in the difference between high-complementarity and low-complementarity firms, we follow the recommended practice from Borusyak et al. (2024) by 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.

[6]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`.

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 DBNL 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

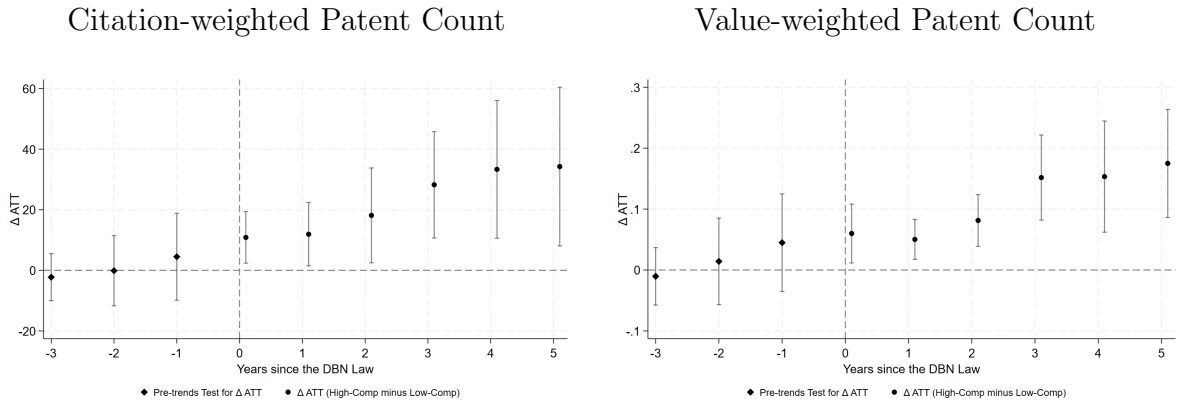
General Innovation Data Breach Notification Laws impact firms by increasing their compliance costs and the strategic importance of data governance. This is likely to impose short-term frictions, leading to a negative effect on firm performance. At the same time, the laws incentivize firms to invest in 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 DBNL 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 DBNL 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 headquarter 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.

Figure 2: Innovation: Triple Difference Estimates



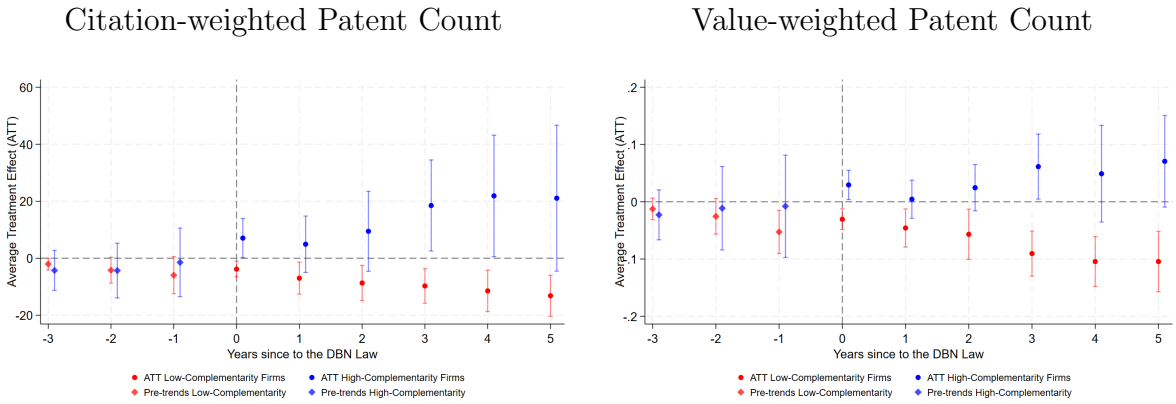
Legend: This figure plots triple-difference (DDD) estimates from BJS regressions around the staggered adoption of DBNLs across U.S. states. Year "0" denotes one year before a DBNL takes effect, explicitly accounting for potential anticipation effects. We find no pre-trends. All regressions include firm, 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 $\$933m \times 0.15 = \$140m$ additional 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.

Figure 3: Innovation: DID Estimates by Complementarity Subgroup



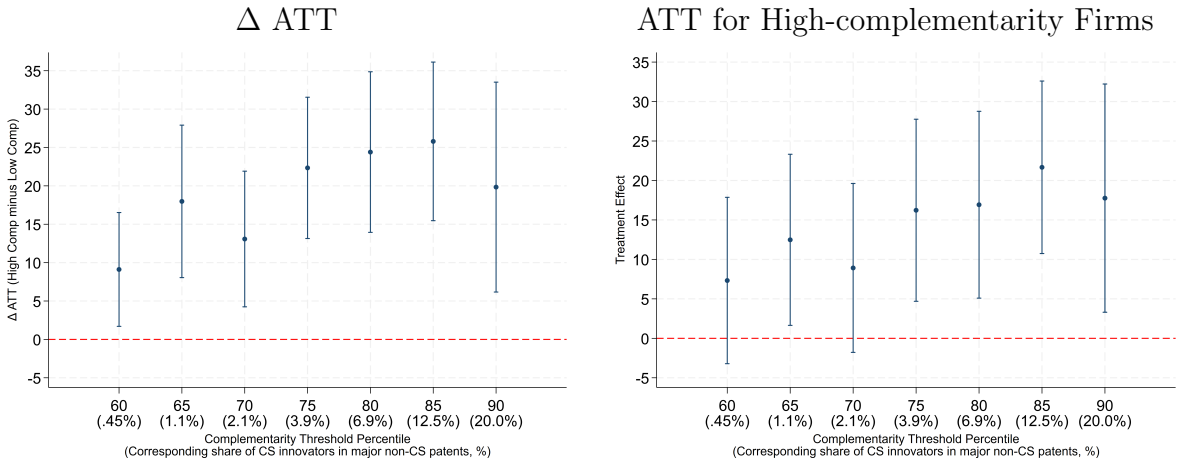
Legend: This figure plots DID estimates separately for high- and low-complementarity firms from BJS regressions around the staggered adoption of DBNLs across U.S. states. Year "0" denotes one year before a DBNL takes effect, explicitly accounting for potential anticipation effects. We find no pre-trends. 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.

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 subgroup-specific 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 DBNL indeed impose a compliance cost on the low-complementarity firms. For high-complementarity firms, however, DBNL 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 DBNL activate innovation synergies 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: 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.

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 consistently 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 DBNL 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%).

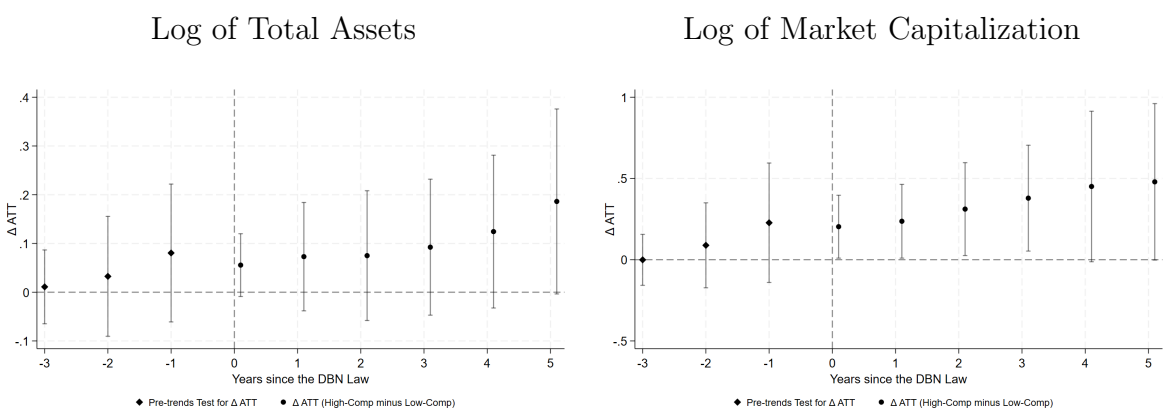
3.2 Results: Firm Growth Responses

We now examine whether the differential innovation responses of high- and low-complementarity firms to DBNL 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 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 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 the 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, DBNL 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.

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's 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 DBNL 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, DBNL 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 DBNL 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 DBNL acts as a common nudge, but only certain firms are organizationally positioned to capitalize on it.

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 DBNL 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) at the expense of other non-cybersecurity patents—with the total change in the presence of cybersecurity experts on non-cybersecurity patent team remaining flat for these firms. In contrast, high-complementarity firms increase the share of cybersecurity inventors broadly across their

Table 2: Mechanisms: Organizational Integration and Innovation Spillovers

	Cybersecurity risk disclosure (1)	Share of CS Experts in Major Non-CS patents (2)	Share of CS Experts in All Non-CS patents (3)	Share of Self CS Citations In Non-CS patents (4)	Citation-weighted CS patent count (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. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

entire patent portfolio, with a significant and economically sizable increase of $+1.74^{*}$ percentage points (Column 3).

Thus, 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.[7]

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 $+0.00^{*}$ percentage points for low-complementarity firms). These knowledge spillovers highlight a key finding: innovation gains stem not from more

[7]The initial scomplementarity 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.

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 +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 DBNL 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.

Robustness and External Validity Our results are robust to a range of alternative explanations, and we conduct targeted tests to rule out potential confounding mechanisms and establish external validity. 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 DBNL on innovation and growth remains statistically and economically significant even within firms of similar size and innovation intensity. This supports our interpretation that the observed spillovers stem from the organizational integration of data-security and innovation capabilities, rather than from scale or resource advantages.

Second, our results in Appendix A.3.2 remain unchanged when excluding any one industry (e.g., data-intensive industries such as *Information* and *Finance*, or data-light sectors such as *Utilities* and *Mining & Oil/Gas*) or dropping top-innovation states (e.g. California and Texas) — demonstrating that the results are not driven by a single sector or state and hence possess broad external validity. Similarly, redefining the event year using the exact enforcement date (Appendix A.3.3) reveals anticipatory responses consistent with our identification strategy and confirms the timing choice in our empirical design.

Finally, 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, it 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.

In all cases, high-complementarity firms exhibit stronger innovation and growth responses to DBNL. The mechanism is organizational: only firms with sufficiently integrated capabilities are able to translate regulatory shocks into broad innovation gains.

4 Data Security as a Real Option on Innovation

Our empirical findings reveal substantial heterogeneity in how firms respond to exogenous increases in data security relevance depending on their organizational structure. High-complementarity firms significantly increase innovation and growth, while their low-complementarity counterparts exhibit muted or negative responses.

To interpret these patterns and further elucidate the underlying mechanism, we develop a parsimonious real-options model of investment under uncertainty and irreversible integrating capabilities. The model formalizes our core insight: that the returns to data-security investments are amplified when they are structurally integrated with broader innovation activities. Modeling this integration as a strategic, irreversible investment allows us to explain persistent gaps in firm responses, even under common regulatory triggers, such as DBNL. In particular, the framework highlights how complementarity enhances the spillover value of security investments and alters firms' optimal timing and intensity of complementarity. The model also delivers empirically testable predictions about convexity in returns, asymmetries in adoption timing, and the effects of outsourcing—predictions borne out in the data. Finally, the model enables counterfactual simulations that quantify the cost of integration failures and the policy value of targeted support, which are difficult to isolate empirically.

4.1 Model

We model a firm's complementarity investment decision as a real option (in the sense of Bar-Ilan and Strange (1996)) in the presence of uncertain, firm-specific spillovers from complementarity integration. Data security relevance increases exogenously due to regulation (e.g., DBNL), and firms choose when to invest in integration, which is irreversible and costly. 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.[8] 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.

Importantly, our empirical evidence supports treating complementarity as persistent and predetermined at the time of DBNL adoption: firms classified as high-complementarity before the shock remain so afterwards, and low-complementarity firms do not materially increase their complementarity in response to DBNL. This observed stability over our sample horizon justifies modeling the degree of complementarity as fixed, with heterogeneity arising from pre-existing organizational structures rather than rapid post-shock adjustment.[9]

Firms differ in their underlying complementarity, which governs the strength of innovation spillovers once integration occurs. This structure generates heterogeneity in investment timing and post-shock outcomes.

[8]For instance, JPMorgan Chase reportedly committed over $500 million annually to embed security into IT workflows after a breach, illustrating the scale and irreversibility of such investments. For Amazon, its '1-Click'-ordering patent team overlapped significantly with its security engineers, highlighting how the organizational structure of innovation teams is a costly, path-dependent organizational investment that generates long-term productivity gains.

[9]In our empirical design, data innovation complementarity is measured over the five years preceding DBNL adoption and held fixed throughout the analysis, mitigating concerns that it reflects an endogenous response to anticipated regulation. Moreover, we observe no differential pre-trends in innovation or growth outcomes between high- and low-complementarity firms (see Section 3), consistent with our measure of complementarity being predetermined at the time of DBNL implementation.

Assumptions and Setup 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 $\sqrt{c} \in [0, 1]$ between data-security and product innovation. The return to complementarity is $\gamma\sqrt{c}\theta$, where $\gamma > 0$.
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 $\lambda > 0$.

Microfoundation for Diminishing Returns to Complementarity We model a firm's complementarity between data resilience and innovation as a scalar $c \in [0, 1]$, reflecting the degree of alignment between infrastructure, processes, and talent across data security and the firm's broader R&D activity. Although higher values of c capture greater potential returns from tailoring innovation to future data security relevance shocks, we assume diminishing marginal returns to complementarity, modeled via a concave option-value function $\phi(c) = c^\alpha$ for $\alpha \in (0, 1)$, such as $\phi(c) = \sqrt{c}$. This structure 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)). These functional forms are chosen without loss of generality for tractability and exposition, and alternative concave specifications would deliver

qualitatively similar insights.

As a result, the option value of waiting is lower for high-complementarity firms, making them more likely to invest early when data protection relevance rises. This mechanism links to our empirical findings: firms with high ex ante complementarity accelerate innovation and outperform following regulatory shocks. It also connects to Nikolov et al. (2021), which shows that firms' ability to respond to uncertainty is shaped by internal flexibility and organizational structure. In our setting, complementarity is the key capability that enables firms to treat regulatory compliance not as a constraint, but as a catalyst for innovation-led growth.

Value Functions The firm decides whether to invest immediately or to wait. The waiting option is subject to a discount factor $\rho \in (0, 1)$, capturing the risk of delay, such as missed early innovation, time decay in data value, or regulatory uncertainty.

If the firm invests immediately, its value is:

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

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

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

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

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

We can solve for this threshold numerically given values of $(\mu, \sigma, \gamma, \lambda, \delta_0, \rho)$. If $c < \bar{c}$, the firm optimally delays investment; if $c > \bar{c}$, the firm invests immediately.

Proposition 1 *Firms with higher data innovation complementarity c are more likely to invest early in response to data security relevance shocks. There exists a unique threshold $\bar{c}$ such that $V_{now}(c) > V_{wait}(c)$ if and only if $c > \bar{c}$. This threshold is given by:*

$$\bar{c}(\gamma, \mu, \rho) = \frac{\frac{\lambda(1-\rho)\mu + \delta_0}{\rho + \mu} + I}{\gamma} \tag{9}$$

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

Real–options theory offers two seemingly conflicting prescriptions: greater uncertainty can *delay* action by magnifying the option value of waiting (Bernanke, 1983; McDonald 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 each 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 uncertainty resolves and profits from the whole distribution of future shocks. Thus, uncertainty still has value, but its influence flips sign depending on the firm's internal complementarity: precaution dominates below $\bar{c}$; pre-emption dominates above it, neatly aligning the divergent findings in the literature.

4.2 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 and profitability 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 \$4.01 million (per Table A.5, Column 8) and a median high-complementarity firm size of \$933 million (per Table A.4).

$$\text{Benefit}_{\text{innovation}} = \gamma * \sqrt{c} * \theta * \text{Firm size} \tag{10}$$

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

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 \$1 million in firm assets, over a five-year horizon. It ensures that the model replicates the empirically observed increase of approximately 95 patents (worth \$380.95 million) for the median firm, which corresponds to 0.4082 patent-equivalent value units per \$1 million in assets.

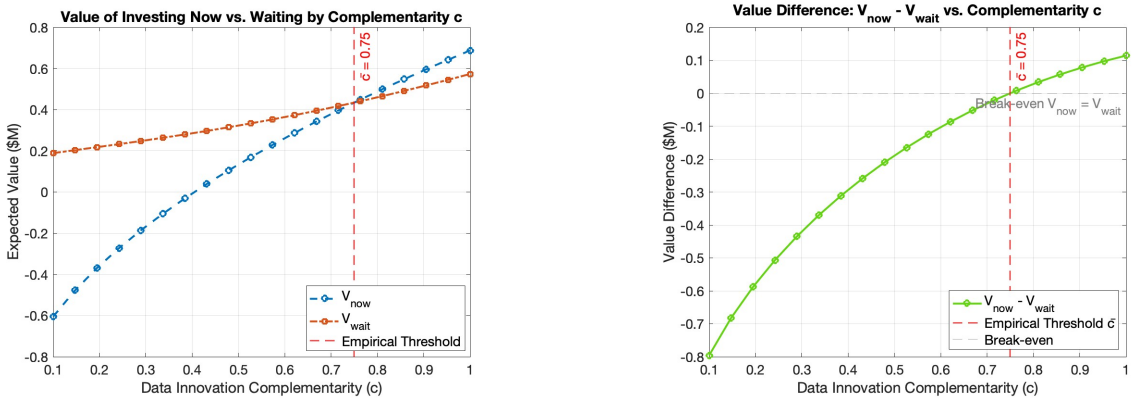
We calibrate the inaction cost parameter to $\lambda = 3$, corresponding to an expected loss of \$3 million from non-integration. This value reflects the ex ante cost of data exposure, incorporating both direct breach-related losses and indirect trust erosion, and is consistent with industry estimates for firms without recent breach history (IBM Security, 2025). 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)). This value ensures that 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, $\bar{c} = 0.75$:

$$V_{\text{now}}(\bar{c}) = V_{\text{wait}}(\bar{c}), \quad \text{for } \bar{c} = 0.75.$$

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 DBNL 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 \$67.45 million for the median firm in our sample (with total assets of \$933 million).

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.

This alignment between theory and data strengthens the interpretation that complementarity is the operative margin along which firms respond to heightened data security relevance. It also 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 is not merely a statistical artifact in the data. As shown in Appendix A.3.1, a substantial fraction of firms—about 25% of the sample by construction—exceed this threshold, with 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*, 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.

Sanity Check: Comparative Statics of Complementarity Threshold Equation (9) implies

$$\frac{\partial \bar{c}}{\partial \gamma} = -\frac{\bar{c}}{\gamma} < 0, \qquad \frac{\partial \bar{c}}{\partial \mu} = \frac{\lambda(1-\rho)}{(\rho+\mu)^2} > 0, \qquad \frac{\partial \bar{c}}{\partial \rho} = \frac{\lambda\mu + \delta_0}{(\rho+\mu)^2} > 0.$$

Corrollary B.2 derives these comparative statics in closed-form. Monte Carlo simulations in Figure C.1 confirm these monotonicities and show that the curves all intersect the calibrated point $(\mu, \rho, \bar{c}) = (1, 0.8, 0.75)$, providing a transparent sanity check that the model behaves exactly as economic intuition dictates.

Intuitively, a higher innovation elasticity γ raises the marginal return to each unit of the AI–data bundle, so the firm is willing to invest at a lower underlying complementarity; hence $\bar{c}$ falls. By contrast, increasing the mean salience of data-security shocks (μ) or shortening the effective horizon (higher ρ) raises the option value of waiting , pushing the cut-off upwards.

Quantitative Implications for Firm Growth and Innovation Our calibrated model yields a series of quantitative implications that provide a theoretically grounded benchmark consistent with the empirical patterns we observe following the implementation of Data Breach Notification Laws (DBNL). First, the model captures the value of the option to invest early: at complementarity levels above the threshold $\bar{c} = 0.75$, firms realize a net present value gain from immediate investment relative to waiting that is

increasing in the complementarity level.

Second, the model implies economically meaningful gains in innovation output. Given a calibrated innovation elasticity of $\gamma = 0.4714$ and an expected data security relevance shock $\mathbb{E}[\theta] = 1$, firms with $c = 0.75$ and $c = 0.90$ achieve an expected innovation yield of 0.4082 ($= \gamma \times \sqrt{c} \times E[\theta] = 0.4714 \times \sqrt{0.75} \times 1$) and 0.4472 ($= \gamma \times \sqrt{c} \times E[\theta] = 0.4714 \times \sqrt{0.9} \times 1$) citation-weighted patents per \$1 million in assets, respectively. For the median high-complementarity firm in our sample (assets $\approx$ \$933M), this corresponds to 380.85 and 417.24 in patent value, or approximately 95 and 104 patents, respectively, assuming an average patent value of \$4.01M, which matches our empirical findings closely.

Lastly, we translate the model-implied innovation gains into firm profitability by applying a conservative elasticity of 0.07 percentage points in return on assets (ROA) per additional citation-weighted patent, following Hall et al. (2005). Given our model's predicted innovation value of \$380.95M for the median high-complementarity firm over five years (equivalent to approximately 3.81 citation-weighted patent units), this implies an ROA increase of approximately 0.2665 percentage points over five years, or 0.0533 percentage points per year. This is closely aligned with the empirically observed ROA increase of 0.0618 percentage points for high-complementarity firms in Table *A*.8.

These quantitative implications from the model provide a theoretically grounded benchmark that aligns with our empirical findings. The magnitudes of innovation and profitability gains implied by the model are economically meaningful and consistent with observed post-DBNL firm responses.

4.3 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. In contrast, firms with low complementarity typically outsource data security, but forgo the associated innovation benefits.

We model outsourcing as an alternative organizational choice that provides equiv-

alent protection against regulatory and operational losses as internal integration, but without enabling innovation spillovers. Specifically, outsourcing eliminates the inaction cost $\lambda\mu$ and preserves the reputational or compliance-related benefit δ_0, but sets the innovation benefit to zero. That is, outsourcing delivers the same δ_0 and avoids $\lambda\mu$, but does not yield any $\gamma\sqrt{c}$ term. This assumption allows us to isolate the welfare effects of shutting off the complementarity channel, holding other margins constant.

We now examine a counterfactual scenario in which firms are restricted to outsourcing and compute the resulting the welfare loss from this constraint.

Proposition 2 *Firms with $c > \bar{c}$ incur a strictly positive welfare loss from being restricted to outsourcing cybersecurity, 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* 2** *See Appendix B.3.*

This proposition formalizes our insight: 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 reports from IBM Security (2025), Cybrwise (September 2024), and VC3 (March 2024), mid- to large-sized firms that outsource cybersecurity to third-party providers typically incur annual costs between \$1.5 million and \$ 3 million, depending on coverage scope and complexity. We adopt a midpoint estimate of \$2 million per year, or \$10 million over a five-year horizon. Scaling this value by the median firm size in our sample (\$933 million in total assets), we obtain:

$$I^{\text{ext}} = \frac{10}{933} \approx 0.0107 \tag{11}$$

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 \$1 million in assets), we ensure consistency with the scaling used for other parameters such as I, γ, λ, and δ_0.

Using our calibrated parameters—$c = 0.75$, $\mu = 1$, $\gamma = 0.4714$, $\lambda = 3.0$, $\delta_0 = 1.87$, $I = 0.0723$, and $I^{\text{ext}} = 0.0107$, we compute:

$$\Delta V(c) = -(I - I^{\text{ext}}) + \gamma\sqrt{c} = -(0.0723 - 0.0107) + 0.4082 = 0.3466 \tag{12}$$

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 (assets = \$933M), this translates to:

$$0.3466 \times 933 \approx \$323.3 \text{ million}$$

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 \$4 million (Kogan et al., 2017), which is exactly equivalent to \$380 million per firm, in line with the theoretical prediction. 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.

4.4 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:

$$V_{\text{now}}(c) + S \geq V_{\text{wait}}(c), \tag{13}$$

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 3 *For every firm with complementarity level $c < \bar{c}$, there exists a minimum subsidy $S(c)$ such that the firm is indifferent between investing immediately and waiting. The required subsidy $S(c)$ is strictly decreasing in c, γ, and μ, and increasing in ρ.*

***Proof* 3** *See Appendix B.4.*

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 for each value of c.

At $c = 0.5$, we find that the value of immediate investment is $V_{\text{now}}(0.5) = 0.1442$, and the expected value of waiting is $V_{\text{wait}}(0.5) = 0.3255$, both in millions of USD. The implied minimum policy subsidy needed to induce immediate investment is:

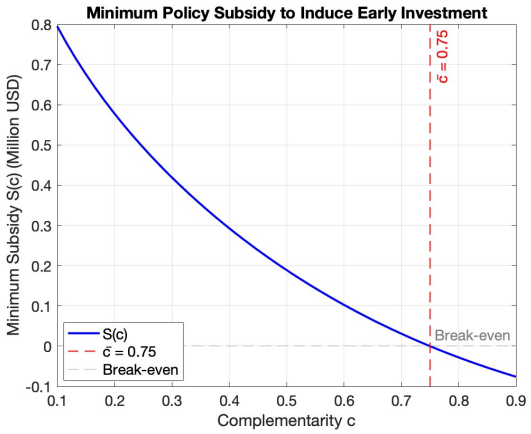
$$S(0.5) = V_{\text{wait}}(0.5) - V_{\text{now}}(0.5) = 0.1889 \tag{14}$$

These results are expressed in absolute terms for a median firm in our sample with \$933 million in assets. This translates to a subsidy of approximately:

$$0.1889 \times 933 = 176.3 \text{ million USD}.$$

While the implied subsidy of \$176.3 million corresponds to approximately 18.9% of the median firm's asset base (\$933 million), 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. This suggests that fixed integration costs may pose a disproportionately larger barrier for structurally under-prepared firms, reinforcing the case for targeted support.

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 \$1 million 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.

We repeat this calculation for a range of $c \in [0.1, 0.9]$ and plot the resulting subsidy function $S(c)$ per \$1 million in Figure 7. The subsidy declines monotonically in c and converges to zero as $c \to \bar{c}$, consistent with the model's structure. Figure 7 illustrates

the policy leverage available in nudging low-complementarity firms over the investment threshold. Figure C.2 in Appendix C.4 presents the simulated minimum policy subsidy across a range of data-innovation complementarity levels and investment cost levels.

This result illustrates the potential effectiveness of targeted subsidies in altering the investment thresholds of low-complementarity firms. By incentivizing data innovation 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 structures are unevenly distributed across firms. While high-complementarity firms benefit from an increase in the perceived importance of data security, others fall behind. Our model quantifies the costs of this divergence and provides guidance for policy interventions that support innovation without reinforcing concentration.

5 Conclusion

This paper identifies and quantifies 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 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 overlaps in data-security and general patenting constructed only from publicly-available data, we document that high-complementarity firms respond to the staggered adoption of U.S. Data Breach Notification Laws (DBNL) by significantly increasing both innovation output and firm value. These effects are large, causal, and concentrated among firms with pre-existing orga-

nizational integration. In 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 real-options model where the value of integrating data protection depends on a firm's complementarity between data-related and product innovation capabilities. High-complementarity firms invest earlier, internalize spillovers, and convert compliance pressure into innovation-led growth. Counterfactual simulations show that restricting all firms to outsource data protection—eliminating internal spillovers—reduces firm value by roughly one-third. We also show that targeted subsidies can close the innovation gap by inducing earlier investment among low-complementarity firms. Together, our results highlight that the benefits of data-security-related investment depend critically on how the investment is embedded within the firm. Outsourcing delivers protection but sacrifices spillovers; integration creates costly but persistent capabilities that transform compliance into innovation. 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. Regulations that increase the perceived importance of data security can also amplify asymmetries, entrenching the advantage of firms that are ex ante organizationally prepared. Our results 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

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

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

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

Alderucci, Dean, Lee Branstetter, Eduard Hovy, Andrew Runge, and Nikolas Zolas, "Quantifying the impact of AI on productivity and labor demand: Evidence from U.S. Census microdata," *Working paper*, 2020.

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

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

Bena, Jan, Isil Erel, Daisy Wang, and Michael S Weisbach, "Specialized investments and firms' boundaries: Evidence from textual analysis of patents," Technical Report, NBER 2021.

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

Bharadwaj, Anandhi, Sean Cao, Jing Tian, and Ling Xue, "Credibility of IT Disclosure Narratives: An Examination in the Context of Data Breaches," Technical Report, SSRN Working Paper 5127642 2025.

Bloom, Nicholas and Raffaella Sadun, "The Organization of Firms Across Countries," *The Quarterly Journal of Economics*, 2012, *127* (4), 1663–1705.

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

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

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Cao, Sean, Anya Nakhmurina, and Tianchen Zhao, "Data Breaches, Debt Costs, and Public Service Provision," Mimeo 2025.

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

Cybrwise, "Understanding the Cost of Outsourcing Cybersecurity Services for SMBs," Link September 2024. Accessed: August 2025.

de Chaisemartin, Clément and Xavier d'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Dixit, Avinash K. and Robert S. Pindyck, *Investment under Uncertainty*, Princeton University Press, 1994.

Dou, Winston Wei, "Embrace or Fear Uncertainty: Growth Options, Limited Risk Sharing, and Asset Prices," 2017. Working paper, The Wharton School, University of Pennsylvania.

Eisfeldt, Andrea L. and Dimitris Papanikolaou, "Organization Capital and the Cross-Section of Expected Returns," *The Journal of Finance*, 2013, *68* (4), 1365–1406.

_ , Gregor Schubert, Miao Ben Zhang, and Bledi Taska, "Generative AI and Firm Values," Technical Report 2023. SSRN Working Paper 4436627.

Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp, "Big Data and Firm Dynamics," *AER Papers and Proceedings*, May 2019, *109*, 38–42.

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

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

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

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

Giroud, Xavier and Holger M. Mueller, "Firms' Internal Networks and Local Economic Shocks," *American Economic Review*, October 2019, *109* (10), 3617–49.

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

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

_ , Adam Jaffe, and Manuel Trajtenberg, "Market value and patent citations," *RAND Journal of economics*, 2005, pp. 16–38.

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

Huang, Henry and Chong Wang, "Do Banks Price Firms' Data Breaches?," *The Accounting Review*, 2021, *96* (3), 261–286.

IBM Security, "Cost of a Data Breach Report 2025," Link 2025. Accessed: August 2025.

Jamilov, Rustam, Hélène Rey, and Ahmed Tahoun, "The anatomy of cyber risk," Technical Report, NBER 2021.

Jiang, Wei, Yuehua Tang, Rachel (Jiqiu) Xiao, and Vincent Yao, "Surviving the Fintech Disruption," NBER Working Paper w28668, NBER 2021.

Kakhbod, Ali, Leonid Kogan, Peiyao Li, and Dimitris Papanikolaou, "Measuring Creative Destruction," *Available at SSRN 5008685*, 2024.

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

Liu, Jinyu and Xiaoran Ni, "Ordeal by innocence in the big-data era: Intended data breach disclosure, unintended real activities manipulation," *European Financial Management*, 2023.

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

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

_ , **Kumar Rishabh, and Orlando Gomes**, "Is It AI or Data That Drives Market Power?," Research Paper 25-37, SFI April 2025. SFI Research Paper No. 25-37, Forthcoming at the Journal of Monetary Economics.

Milgrom, Paul and John Roberts, "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *The American Economic Review*, 1990, *80* (3), 511–528.

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

Ramadorai, Tarun, Antoine Uettwiller, and Ansgar Walther, "Privacy Policies and Consumer Data Extraction: Evidence From U.S. Firms," *Review of Finance (forthcoming)*, 2025.

Rishabh, Kumar, Roxana Mihet, and Julian Jang-Jaccard, "Cyberrisk and AI Firms," Research Paper 25-39, SFI October 2024. SFI Research Paper No. 25-39.

Rock, Daniel, "Engineering Value: The Returns to Technological Talent and Investments in Artificial Intelligence," Technical Report, SSRN 2019. Working Paper 3427412.

Roth, Jonathan, 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*, 2023, *235* (2), 2218–2244.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

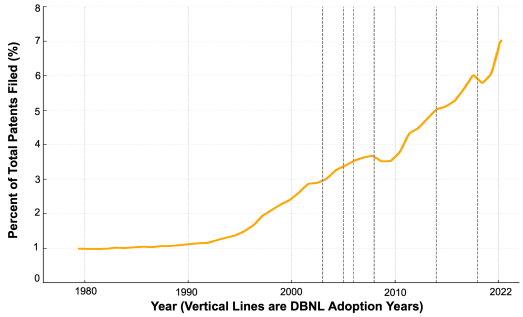
Tambe, Prasanna, Lorin Hitt, Daniel Rock, and Erik Brynjolfsson, "Digital Capital and Superstar Firms," Working Paper 28285, NBER December 2020.

VC3, "How Much Do Managed Cybersecurity Services Cost?," Link March 2024. Accessed: August 2025.

Appendix A Additional Figures and Tables

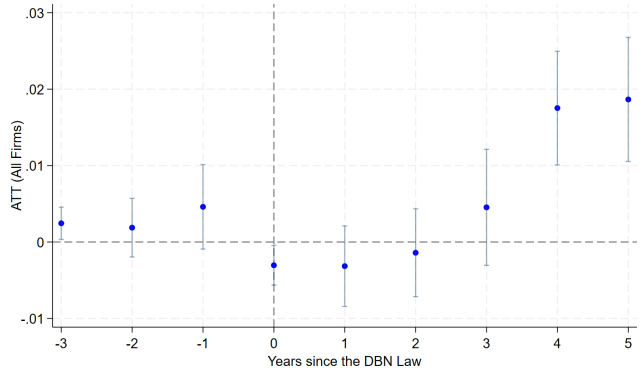
A.1 Additional Figures

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



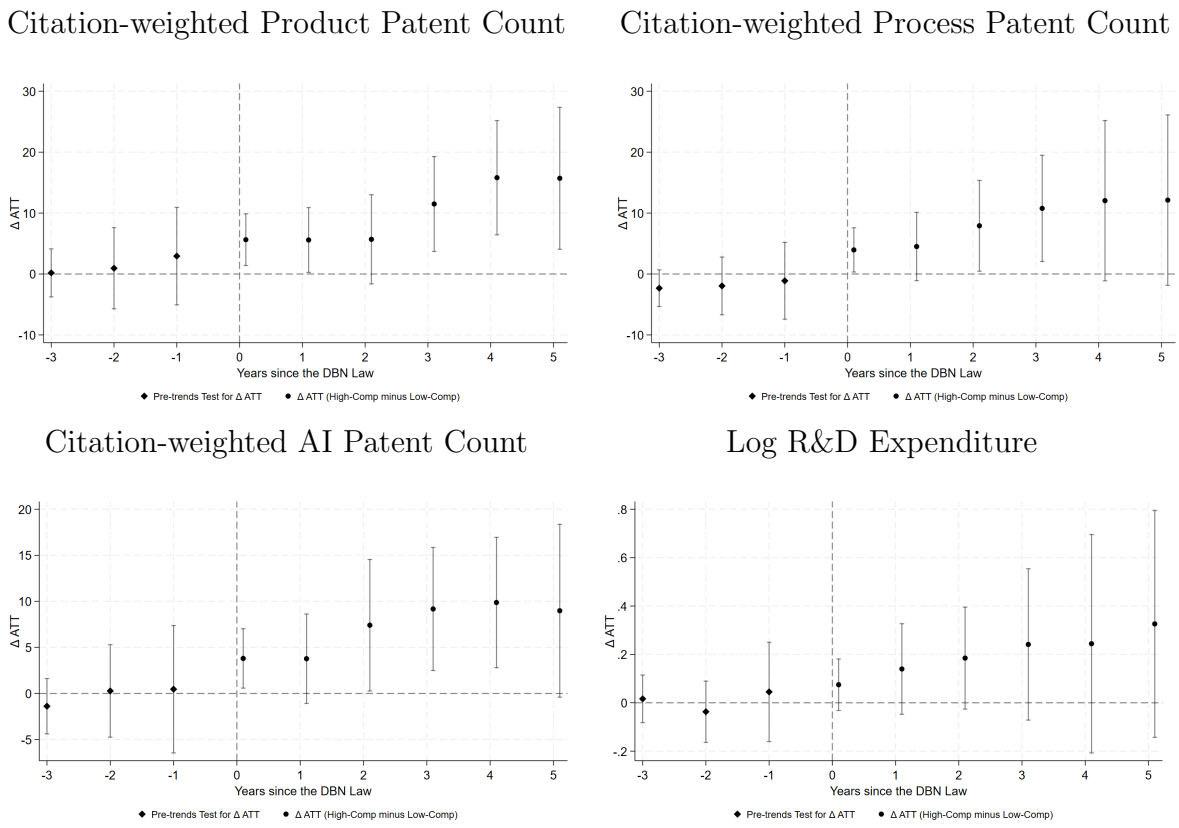
Legend: This figure shows the proportion of data-related patents (i.e., cyber security patents) out of all patents filed in a given year. A cyber security patent is defined as a patent to which the USPTO assigns at least one CPC code pertaining to cyber security. The dashed vertical grey lines represent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

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



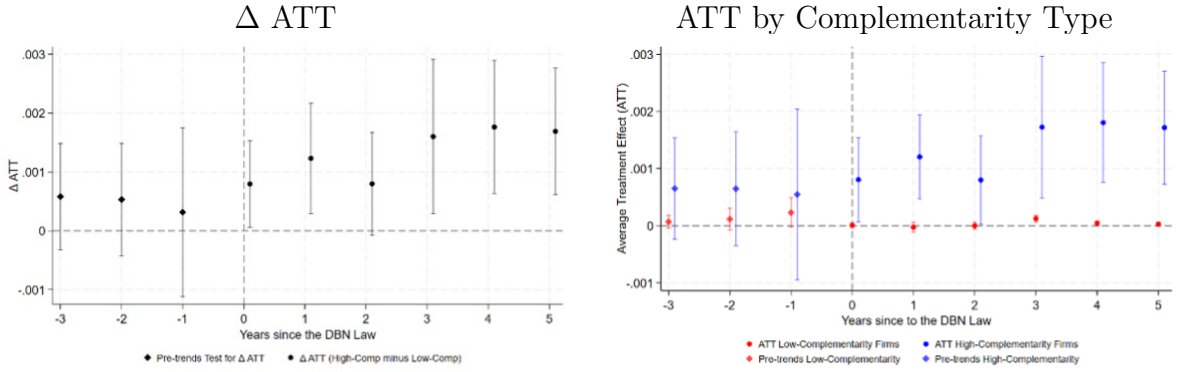
Legend: This figure tests whether treated firms start to pay attention to the data risks just before the implmentation of the DBNLs, in order to identify any anticipatory response to DBNL implementation. The figure plots DID estimates of changes in the cyber risk score based on the methodology in Florackis 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

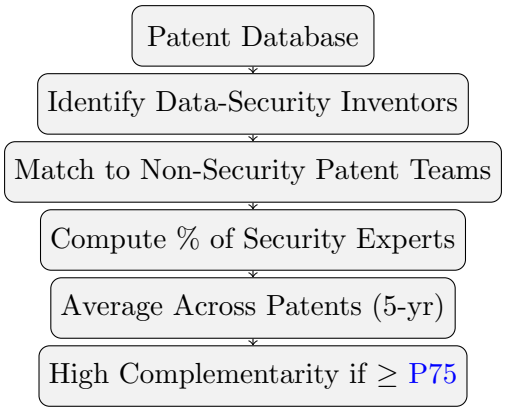
Variable	Description
Data Risk Score	Cyber-risk score for US-listed firms following Florackis et al. (2023). Calculated through textual analysis of 10-K filings, comparing disclosed cyber risks with prior filings of breached firms. Higher scores indicate greater cybersecurity vulnerability.
Citation-weighted 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)	
Sample	(1)	Total Assets (2)	Market Cap (3)	Total Assets (4)	Market Cap (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 Tripple-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 Tripple-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

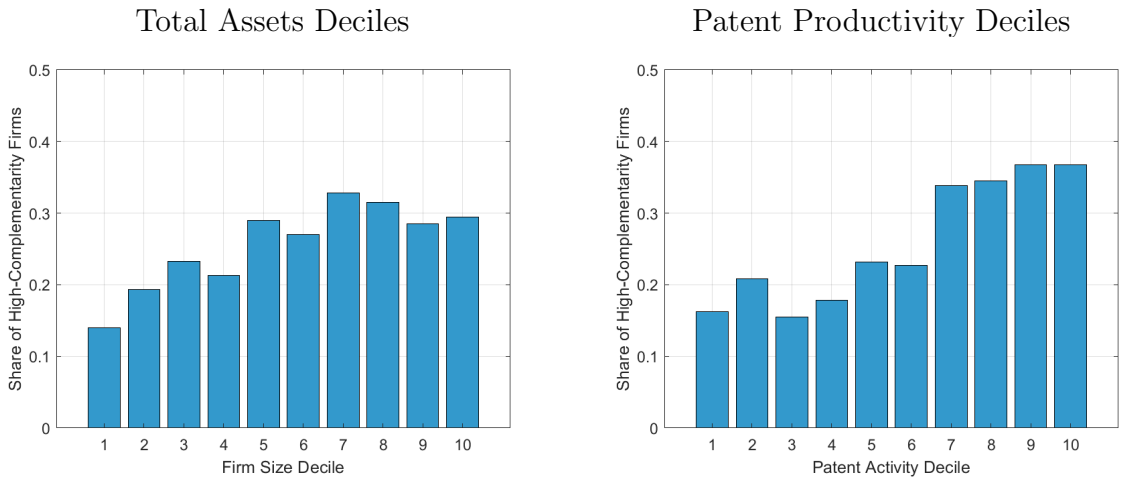
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. This table presents triple-difference (DDD) estimates from BJS regressions around the staggered adoption of Data Breach Notification Laws (DBNLs) across U.S. states. The DDD estimator compares subgroup-specific ATT estimates for high-complementarity and low-complementarity 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 with 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 re sponses, 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 Industry-Level Robustness Tests

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 remains 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.

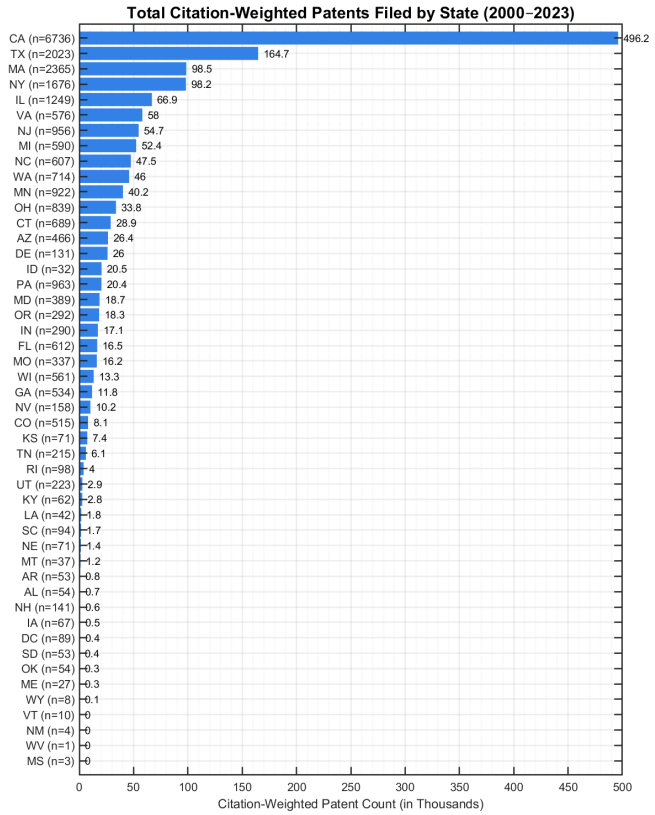
The firm growth results mirror the patent patterns with one notable difference. DDD estimates for log total assets range from 0.19 to 0.30, clustering around our baseline estimate of 0.25. Unlike the patent results, excluding manufacturing does not weaken the firm growth effect: the DDD estimate remains 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: State-Level Innovation and the Timing of DBNLs

Table A.11 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.

Excluding top patenting states also removes some early DBNL adopters (e.g., California,

Table A.11: 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.

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.3.4 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.12).

Table A.12: 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.

Appendix B Theoretical Proofs and Derivations

B.1 Proof of Proposition 1. (Threshold of Complementarity)

Define the value difference function:

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

Both $V_{\text{now}}(c)$ and $V_{\text{wait}}(c)$ are continuous and strictly increasing in c on $(0, 1]$. Moreover, both are concave in c due to the $\sqrt{c}$ specification. Since $V_{\text{wait}}(c)$ involves a maximum operator and a Gaussian expectation, it is smooth and strictly increasing but concave, as shown in the real options literature (e.g., Bar-Ilan and Strange, 1996).

At $c = 0$, we have $\sqrt{c} \to 0$, so

$$V_{\text{now}}(0) = -I + (-\lambda)\mu + \delta_0, \tag{B.2}$$

and similarly for $V_{\text{wait}}(0)$. Under reasonable parameters (e.g., moderate μ and ρ), $V_{\text{wait}}(0) > V_{\text{now}}(0)$.

At $c = 1$, we have the full value of complementarity. Since $V_{\text{now}}(1)$ increases faster than $V_{\text{wait}}(1)$ for sufficiently high μ or γ, or sufficiently low ρ, it is possible for $\Delta V(1) > 0$.

Because both functions are continuous and concave, and their difference $\Delta V(c)$ is also continuous and strictly increasing (as $V_{\text{now}}(c)$ grows faster in c than $V_{\text{wait}}(c)$ for sufficiently high c), the Intermediate Value Theorem guarantees that there exists a unique $\bar{c} \in (0, 1)$ such that Eq. (*B.1*) is zero.

$$\Delta V(\bar{c}) = 0, \tag{B.3}$$

and the sign of $\Delta V(c)$ switches exactly once. Therefore, the firm strictly prefers to invest immediately if and only if $c > \bar{c}$.

Starting from Eq. (*B.3*),

$$\gamma\,\bar{c} \;=\; \frac{\lambda\big(1-\rho\big)\mu+\delta_0}{\rho+\mu} \;+\; I, \tag{B.4}$$

one obtains the explicit solution

$$\bar{c}(\gamma,\mu,\rho) \;=\; \frac{\dfrac{\lambda(1-\rho)\mu+\delta_0}{\rho+\mu} \;+\; I}{\gamma} \tag{B.5}$$

which makes the dependence of the threshold on the innovation elasticity γ, the mean relevance of data-security shocks μ, and the discount factor ρ fully transparent.

This concludes the proof.

B.2 Corrollary 1. (Comparative Statics of Complementarity Threshold)

Totally differentiating (*B*.5) yields

$$\frac{\partial\bar{c}}{\partial\gamma} = -\frac{\bar{c}}{\gamma} < 0, \tag{B.6}$$

$$\frac{\partial\bar{c}}{\partial\mu} = \frac{\lambda(1-\rho)}{(\rho+\mu)^2} > 0, \tag{B.7}$$

$$\frac{\partial\bar{c}}{\partial\rho} = \frac{\lambda\mu+\delta_0}{(\rho+\mu)^2} > 0. \tag{B.8}$$

Equation (B.6) shows that a higher γ *lowers* the threshold. A greater return per unit of the AI–data bundle makes the fixed integration cost worthwhile at lower complementarity levels.

According to (B.7), a rise in the mean relevance of data-security shocks *raises* $\bar{c}$. When external shocks are more valuable in expectation, the option value of waiting increases and firms postpone investment until complementarity is higher.

Equation (B.8) indicates that a larger discount rate (or shorter effective horizon) also *raises* the threshold: the present value of future innovation spillovers falls, so a higher level of complementarity is required to justify the sunk cost.

These sign-definite results mirror the graphical "mental-health" check in Figure C.1: the

blue curve for γ slopes downward, whereas the red and black curves for μ and ρ slope upward, all intersecting at the calibrated point $(\mu, \rho, \bar{c}) = (1, 0.8, 0.75)$. The analytical derivatives thus validate the numerical simulation and confirm that the model responds to first-order parameter changes exactly as economic logic predicts.

B.3 Proof of Proposition 2. (Welfare Loss from Outsourcing)

Let I be the cost of internal integration and I^{ext} the cost of outsourcing. Let both regimes share the same reputational or compliance benefit, δ_0. Then:

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

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

Taking the difference:

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

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

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

The welfare loss $\Delta V(c)$ increases with c, γ, and μ, and decreases with the cost gap $(I - I^{\text{ext}})$. This concludes the proof.

B.4 Proof of Proposition 3. (Subsidy)

From the indifference condition $V_{\text{now}}(c) + S(c) = V_{\text{wait}}(c)$, it directly follows that:

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

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

$$V_{\text{wait}}(c) = \rho[\max\{-I + (\gamma\sqrt{c} - \lambda)\mu + \delta_0, 0\}] = \rho \cdot \mathbb{E}\left[\max\{-I + A\theta + \delta_0, 0\}\right], \text{ and}$$

$$A = \gamma\sqrt{c} - \lambda.$$

Because the innovation benefit $V_{\text{now}}(c)$ increases in c at a faster rate than the convex waiting value $V_{\text{wait}}(c)$, the subsidy gap $S(c)$ narrows as c increases. Moreover, since $V_{\text{now}}(c)$ is increasing in γ and μ and decreasing in ρ (which increases the value of delay), the comparative statics follow directly.

The minimum subsidy $S(c)$ satisfies the following comparative statics:

Complementarity c:

$$\frac{\partial A}{\partial c} = \frac{\gamma}{2\sqrt{c}} > 0 \quad \Rightarrow \quad \frac{\partial V_{\text{now}}}{\partial c} > 0, \quad \frac{\partial V_{\text{wait}}}{\partial c} > 0, \quad \text{but } \frac{\partial V_{\text{now}}}{\partial c} > \frac{\partial V_{\text{wait}}}{\partial c} \Rightarrow \frac{\partial S}{\partial c} < 0.$$

The subsidy is strictly decreasing in c.

Innovation Elasticity γ:

$$\frac{\partial A}{\partial \gamma} = \sqrt{c} > 0 \quad \Rightarrow \quad \frac{\partial S}{\partial \gamma} < 0.$$

Higher innovation returns reduce the required subsidy.

Shock Mean μ:

$$\frac{\partial V_{\text{now}}}{\partial \mu} = A, \quad \frac{\partial V_{\text{wait}}}{\partial \mu} > 0 \Rightarrow \frac{\partial S}{\partial \mu} < 0 \text{ (for } A < 0).$$

Stronger perceived importance of data protection reduces the need for policy intervention.

Discount Factor ρ:

$$\frac{\partial V_{\text{wait}}}{\partial \rho} = \mathbb{E}[\max\{-I + A\theta + \delta_0, 0\}] > 0, \quad \frac{\partial V_{\text{now}}}{\partial \rho} = 0 \Rightarrow \frac{\partial S}{\partial \rho} > 0.$$

More patient firms value waiting more and require higher subsidies to act early.

This concludes the 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.

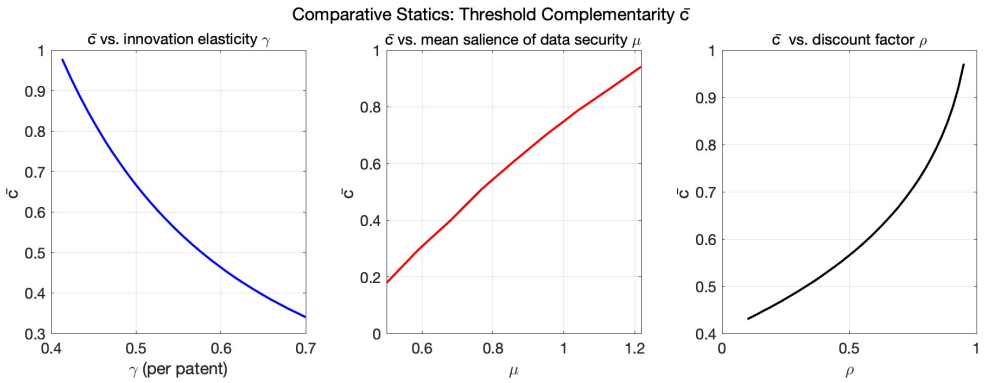
Table C.1: Microfoundation of Calibrated Parameters (5-Year Horizon, Per $1M Firm Size)

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 et al. (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, March 2024; Cybrwise, September 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 $20M 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%

Legend: 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_2 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 Comparative-statics Figure

Figure C.1: 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 10), the middle panel with respect to μ (the mean salience of data security per Table C.1), 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 $(\mu, \rho, \bar{c}) = (1, 0.8, 0.75)$. Consistent with Equation (8), $\bar{c}$ falls in γ (left) but rises in μ (center) and ρ (right), providing a visual "mental-health" check on the model's behaviour. Table C.1 provides further details on the parameters.

C.3 External Validity

Holding the calibrated primitives fixed at $\gamma = 0.4714$, $\lambda = 3.0$, $\delta_0 = 1.87$, $\mu = 1$, and $\rho = 0.8$, the implied *break-even* investment rises steeply with complementarity:

$$c = 0.75: \quad I^* = 0.0723\ ,$$
$$c = 0.80: \quad I(0.80) = 0.167\ ,$$
$$c = 0.90: \quad I(0.90) = 0.356\ .$$

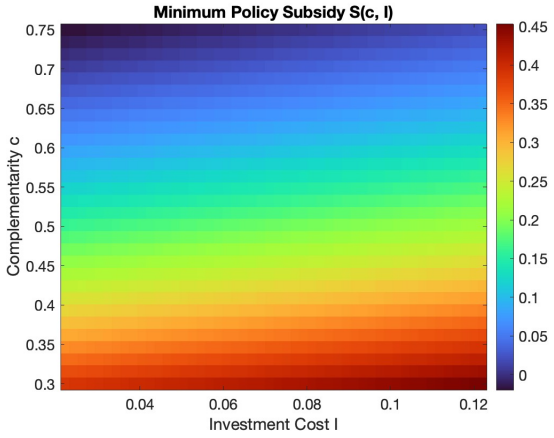
Because all model variables are normalised per \$1 million of assets, these translate into cash outlays of roughly \$67 million, \$155 million, and \$332 million, respectively, for the median firm in our sample (total assets \$933 million). The non-linear escalation reflects the convex option value in (8): once the firm is sufficiently close to full data–innovation alignment, even modest increases in c require disproportionately larger irreversible outlays to keep immediate investment competitive with the wait-and-see alternative. Importantly, the \$155–\$330 million figures are well within the scale of large-scale IT integration projects documented in practice (e.g. the \$500 million multi-year security overhaul at JPMorgan Chase), anchoring our calibration in empirically plausible magnitudes.[10] In sum, pushing complementarity from 0.75 to 0.9 more than quintuples the required up-front commitment, underscoring why only a subset of firms find it optimal to internalise data capabilities at the very top of the complementarity distribution.

C.4 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).

[10] The higher numbers reported in the comparative-statics paragraph— 0.167 ($\approx$ \$155 m) at $c = 0.80$ and 0.356 ($\approx$ \$332 m) at $c = 0.90$— are *not* additional cash outlays. They represent the *maximum* integration cost that would still leave a firm *indifferent* between investing immediately and waiting at those higher complementarity levels: $I_{\text{breakeven}}(c) = \gamma c - \frac{\lambda(1-\rho)\mu+\delta_0}{\rho+\mu}$. Because γc grows linearly in c while the opportunity-cost term is fixed, a more integrated firm is *willing* to pay more to secure the same option value; the actual cheque it writes, however, remains I^*.

Figure C.2: 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.