# Privacy Capital

Runyu Qi*

*Swiss Finance Institute*
*Università della Svizzera italiana*

This draft: August 2026

Abstract

Privacy is the governance of personal data firms hold and use in production. In this paper, I show that privacy is capital, beyond compliance obligation. I construct the first firm-year measure of privacy capital via a hybrid-LLM approach from the texts of around 90,000 privacy policies. By exploiting the 2018-2023 regulatory wave, I provide causal evidence that investment in privacy pays. A one-standard-deviation within-firm increase in privacy capital increases asset turnover by 0.06 standard deviations and ROA by 0.10 standard deviations. The mechanism is consistent with governed data being utilized more productively. Effects are more pronounced for collection-intensive, sharing-intensive, and consumer-facing firms. The market initially discounts privacy capital, consistent with concerns that privacy regulation constrains data use and growth opportunities. As its productive benefits become apparent, investors gradually correct this initial assessment and the discount dissipates. Where prior work shows that bad privacy is punished, I show that good privacy is rewarded. Findings carry implications for the ongoing regulatory debate around privacy in the AI era. Privacy regulation imposes real costs, but it can also induce productive capital formation.

*JEL classification:* E22, G14, G32, G38, K24, O33.

*Keywords:* privacy capital, privacy regulation, data economy, intangible assets, market recognition.

*Runyu Qi is a PhD candidate at the Swiss Finance Institute and USI Lugano. Contact: runyu.qi@usi.ch, Via Giuseppe Buffi 6, 6900, Lugano, Switzerland. I am deeply grateful to my advisor, Francesco Franzoni, for his support throughout. I thank Tarun Ramadorai, Antoine Uettwiller, and Ansgar Walther for sharing their privacy policy corpus. I thank Laurent Frésard, Rachel Nam, Roberto Pinto, Alberto Plazzi, Marco Zanotti, seminar participants and colleagues for helpful comments. All errors are my own. The author has no relevant financial or other conflicts of interest to disclose.

# I. INTRODUCTION

Data has become a primary input of modern firms. The most valuable companies in the world are valued largely for the data they hold (Veldkamp 2023). Among data for production, personal data plays a prominent role in how firms set prices, target customers, develop products, and compete. A large literature studies how the collection and use of personal data shape consumer welfare, advertising, competition, and innovation[1]. Meanwhile, the ways firms exploit personal data are increasingly governed by regulations. The European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) lead a global wave of statutes[2] that compel firms to adopt and disclose specific data-handling practices. Privacy is the governance of personal data firms hold and use for production. For a large and growing set of firms, privacy now shapes the data assets on which their businesses run.

Despite all this attention, a first-order question remains open. What is privacy to firms? Existing work answers only implicitly. Moreover, the answers are mostly centered on discrete events. The regulation literature treats privacy mostly as a cost of doing business[3]. Other finance literature mostly treats it as a risk, an exposure to breaches and enforcement that investors should price[4].

This paper argues that both readings miss the capital view, which emerges when privacy is seen as a routine practice beyond discrete events. A simple argument leads this new view: the practices that regulation mandates and many firms build do not just disappear post-regulation. Instead, they should accumulate into a firm-level intangible stock, which I call privacy capital. Which view holds determines where privacy shows up and how privacy regulation should be evaluated. The question has a clock. In November 2025 the European Commission proposed the Digital Omnibus, a simplification package that amends the GDPR itself and defers the AI Act's high-risk obligations. Adopted in mid-2026, it defers enforcement time for almost 2

[1]The economics of privacy and data is surveyed by Acquisti et al. (2016), Goldfarb and Tucker (2019), and Goldfarb and Que (2023).

[2]The GDPR was enacted in 2016 and came into effect in May 2018. It applies to any firm that deals with EU residents' personal data. The personal data owner's location, instead of the firm's location, determines the geographical scope. Thus, a U.S. firm is also subject to the GDPR as long as it operates in the EU. The CCPA was passed in 2018 and came into effect in 2020. It is a state-level regulation that governs firms operating in California, subject to qualifying conditions. For detailed information, see their respective official documentation.

[3]A growing body of work measures the effects of privacy regulations on operating outcomes, advertising, entry, innovation, and venture investment. See Goldfarb and Tucker (2011) and Campbell et al. (2015) for early works before the GDPR era. See Johnson (2024) for a comprehensive review of the GDPR literature through 2023. See, among others, Goldberg et al. (2024), Aridor et al. (2025), Dubé et al. (2025), Doerr et al. (2026), and Lefrere et al. (2026) for more recent works. This stream of work is discussed further in the related literature below.

[4]This stream of literature is relatively scarce. A recent work is Cheng et al. (2025). This paper is discussed further in the related literature below. Adjacent works on cybersecurity risk include Kamiya et al. (2021), Florackis et al. (2023) and Jamilov et al. (2021).

years, in a debate conducted almost entirely around the compliance cost. Over the same period roughly 20 U.S. state privacy statutes have come into force, with more enacted during 2026. The expansion of AI relies heavily on data, yet evidence on the governance of data beyond compliance cost remains scarce. This paper addresses this gap by identifying the bright side of responsible personal data governance.

I start with a series of hypotheses. Like any intangible capital[5], privacy capital is built through costly investment including personnel, data infrastructure, maintenance, and the non-pecuniary cost of collecting less data. It is built partly by regulatory compulsion, partly by choice[6], and partly by peer pressure[7].

First, if privacy practice is capital, investment should accumulate into a durable stock that persists beyond the regulatory impulse. The cost-only view, on the other hand, predicts that cross-firm differences in privacy practices compress toward the regulatory minimum once the rule binds. Firms have no reason to build beyond the mandate, since investing in privacy consumes real resources. The outcome would be a minimum-compliance equilibrium. Second, privacy capital should yield a stream of payoffs persisting beyond the discrete privacy events. Third, privacy capital is unbooked on the balance sheet, like the data asset. Its valuation depends solely on the market, and the market prices it only as investors come to recognize it. The natural hypothesis, in the spirit of Merton (1987), is that this process is slow and subject to frictions.

The results on the operating and asset-pricing sides converge on a coherent story. Firms entering the regulatory wave with less privacy capital are compelled to build more and subsequently experience larger efficiency and profitability gains, particularly where personal data matters more for production. These firms initially trade at a discount, consistent with regulation reducing perceived data-related growth opportunities, but subsequently re-rate as the productive value of privacy becomes more salient. In the cross-section of returns, firms with better privacy practices earn a premium during this transitional period, and the premium disappears as low-privacy firms catch up. Together, the evidence is more consistent with productive capital formation followed by gradual market recognition than with either a pure compliance-cost view or compensation for standing systematic privacy risk. Regulations induce capital accumulation that pays, beyond compliance burdens.

Privacy capital is the stock of data-handling practices a firm has put in place, its commitments on what data it collects, how long it keeps it, how it secures it, and what control it gives the

[5]On intangible capital and its measurement, investment, and valuation, see Corrado et al. (2009), Eisfeldt and Papanikolaou (2013), Peters and Taylor (2017), and Crouzet and Eberly (2019); on data as a productive asset, Jones and Tonetti (2020), Veldkamp (2023), and Farboodi and Veldkamp (2026).

[6]Anecdotal evidence supports this invest-beyond-regulation view even before any empirical evidence from this paper. Proton, a Swiss company providing various internet products like emails and network proxies, advertises privacy as its signature.

[7]For example, the Equifax breach in 2017 triggered a broad industry-wide upgrade in privacy provisions.

individuals whose data it holds. To make this concrete, consider a data-reliant firm that enters the wave with a thin privacy policy[8]. As the floor, the minimum standard, rises, it commits to retain less data, to delete records on request, and to give users access to and control over their data, and it builds the internal systems those commitments require. That stock of practice, costly to assemble and slow to depreciate, is what I call privacy capital. I measure the stock from firms' official privacy policies, where firms state these practices with legal effect. The privacy policy is a well-suited source for two reasons. First, it is not a marketing brochure. Under consumer-protection law a firm is bound by what its policy says, so the document is a public and enforceable description of its data-handling commitments rather than cheap talk. Second, privacy policies are far from interchangeable boilerplate. They show large time-series and cross-sectional variation in content, even within the same industry (Ramadorai et al. 2025), and it is this variation that the measure exploits.

I build the measure from the corpus of U.S. public firms' privacy policies used in Ramadorai et al. (2025). Working at the sentence level, I classify policy text into standard privacy-practice categories according to Wilson et al. (2016) via a hybrid-LLM approach and aggregate to a firm-year panel. The measure is economy-wide and reflects both what regulation compels and what firms choose, and I treat it throughout as a proxy for privacy capital. The analysis centers on the governance dimension, the commitments on data retention, deletion, user access, and choice, because these are the practices privacy regulation mandates and that define the floor. The key variable is the log number of governance sentences. The most natural concern with any text-based measure is that it captures words rather than practice. Four independent validations push against this concern, with the right timing in each case. The measure jumps for exposed firms when the European regulation takes effect, rises when industry peers suffer data breaches, tracks the staggered adoption of U.S. state privacy laws, and lines up with third-party assessments of firms' privacy practices. Empirical results relying on this measure also survive controls for policy length, mitigating the verbosity confound[9].

I discipline these predictions with a simple framework in which privacy capital accumulates against a compliance floor under rising pressure. The framework delivers the partition the identification exploits. Firms below the floor are dragged up to it while firms above it invest only when the marginal value of privacy capital pays. Privacy capital is partly firm-chosen, and it co-moves with firm characteristics. I therefore exploit the exogenous variation firms could not choose, the broad wave that raised corporate privacy practice between 2018 and 2023. The wave set a compliance floor on required practice and raised the pressure to meet it

[8]For example, a thin privacy policy may contain only boilerplate sentences. A firm might simply have listed the data it collects, without any rights granted to data owners and without any stated limits of use, prior to the privacy regulation wave.

[9]Section III takes up these validations, and the choice of a count over a share, in detail.

over the window. Firms differed in how far below the floor they began, so the least prepared firms had to build the most. Each firm's pre-wave distance below the floor, scaled by the rising pressure, is a continuous treatment intensity, and I instrument privacy capital with this dose to capture the exogenous responses. I locate the floor at the end-of-sample practice level reached by firms with the most established privacy practices[10], those operating in the European Union or headquartered in California[11]. Almost every firm begins below this standard, so the design is a continuous-dose design among treated firms.

The first stage is strong, and it has the sign the design predicts. The firms that began farthest below the floor build the most privacy capital as the wave arrives, while firms that began near the standard have little room to move. I show this catch-up pattern in Figure 1 for illustration. Sorting firms by how far they began below the floor, the least prepared close their distance fastest while the most prepared move less. The first-stage response increases monotonically with predetermined distance to the floor. The variation I ultimately use is the variation of firms compelled to build privacy capital they would not otherwise have built.

[Insert Figure 1 here]

The first hypothesis is visible in the raw data before any regression. Privacy practice persists and fans out. The middle of the distribution converges toward the floor as compelled firms catch up, while the top keeps building past the standard with no mandate left to satisfy. The evidence is inconsistent with a view in which regulation merely imposes deadweight compliance expenditure. Firms build a persistent stock and subsequently obtain operating benefits.

The direct payoff, as the second hypothesis anticipates, is efficiency. Building privacy capital raises asset turnover and subsequently ROA. A one-SD within-firm increase in privacy capital raises asset turnover by roughly 0.06 SD and subsequently ROA by roughly 0.10 SD. The efficiency gains also carry on to the next period, consistent with the persistent payoffs expected of capital. The full sample results do not show any margin gains, suggesting the ROA effects come mainly from asset turnover. The efficiency gains scale monotonically across treatment intensity. Firms who are farthest below the compliance floor build the most during the window. At the same time, these firms experience the highest efficiency boosts.

Stock prices follow the third hypothesis, with the predicted frictions. Privacy capital initially trades at a discount. Firms compelled to build the most initially exhibit lower Tobin's Q and Total Q than firms compelled to build less. Privacy regulation initially reduces firms' flexibility

[10]The floor choice, even constructed by economic motivation, may still appear arbitrary. In subsequent sections I show that the results are almost invariant to floor choices.

[11]The CCPA, as explained in previous footnotes, applies to firms operating in California subject to some conditions. However, to the best of my knowledge, there is no existing measure that could support a clean measurement of a firm's CA exposure. Thus, I rely on CA-as-headquarter as the implementable proxy.

to collect and use personal data, lowering perceived growth opportunities and generating a valuation discount for the most exposed firms. As the productive benefits of the privacy capital induced by regulation become visible, that discount subsequently reverses. The valuation loading rises sharply in the early 2020s. Recognition has traceable footprints, as institutional ownership of privacy-rich firms rises over the same window while analysts allocate more attention to data as an asset. Moreover, 2020 was the year when CCPA, the first US-state-level comprehensive privacy regulation, came into effect. A long-high-short-low privacy-capital portfolio earns a large premium before 2020 and essentially nothing after, and the premium is driven by small, thinly followed stocks, where recognition is slowest. Firms with better privacy practices in levels earn more than their low-privacy counterparts. This gap shrinks as regulations compel low-privacy firms to catch up. This evidence reads better as a transitory recognition correction than as a systematic risk. The premium accrues to good-privacy firms, the opposite sign from the natural prediction of compensation for bad-privacy exposure. In the meantime, I find little evidence that the valuation effects operate through conventional cost-of-capital channels.

I next examine the mechanism behind the efficiency gains. The natural hypothesis is data productivity. Privacy capital can impose more disciplined data architecture and reduce redundant data resources, improving the productivity of the data firms retain. This mechanism yields a discriminating cross-sectional prediction. If the gains operate through data productivity, they should be stronger where personal data matters more for production, but should not systematically scale with characteristics that carry little direct data content. Consistent with this prediction, the efficiency and profitability gains are larger for firms with greater first-party collection, greater third-party sharing, and consumer-facing businesses. The stronger operating response among consumer-facing firms echoes Crouzet and Eberly (2019), who find that productivity gains associated with intangible capital are strongest in the Consumer sector. Data-free moderators such as firm size, industry concentration, and pricing power contribute little incremental effect. The valuation patterns sharpen the interpretation. Collection-intensive and sharing-intensive firms experience larger initial discounts, consistent with greater exposure to regulation-induced losses of data-related opportunities. Consumer-facing firms, by contrast, trade at a relative premium. Their businesses rely more heavily on operational data that remains necessary for fulfilling transactions and contractual obligations, so regulation removes fewer productive opportunities, while the efficiency and trust benefits of stronger privacy practice are more directly reflected in observable business outcomes. The combination of stronger operating effects where data is more salient, null effects for data-free moderators, and opposite valuation responses across different forms of data dependence is consistent with a data-productivity channel rather than a generic firm-quality or scale effect.

This evidence alone could not distinguish it from alternative mechanisms. A general claim should be made in advance for clarification: for an omitted variable to generate the estimated effect, it must be correlated with the predetermined distance-to-floor and have a time-series evolution that loads on regulatory pressure in the same way as the treatment intensity. I run a series of discrimination and robustness tests against alternative explanations. The evidence is not consistent with the results being driven simply by technology firms, large or dominant firms, pre-existing organizational quality, or mechanical mean reversion among firms that begin the wave with weaker operating performance. The estimates remain similar within and outside the technology sector, across firm-size groups, and after allowing pre-wave firm characteristics to generate their own differential trends; the data-related heterogeneity also remains while moderators with little direct data content, including size, industry concentration, and pricing power, contribute little incremental explanatory power. The results are likewise robust to industry-by-year and state-by-year fixed effects and to tests designed to separate the regulatory wave from differential exposure to the COVID shock. Finally, the timing of the valuation re-rating does not line up naturally with an ESG-taste explanation, while I find little evidence that conventional cost-of-capital or realized-tail-risk channels account for the main results. Taken together, these tests do not eliminate every possible omitted mechanism, but they make several prominent data-free explanations difficult to reconcile with the full pattern of evidence.

The paper makes three contributions. First, it introduces privacy capital as a new corporate intangible, a persistent stock of data-governance practices that firms build through investment and that subsequently affects operating performance. This moves privacy beyond the prevailing treatment of regulations and breaches as discrete events and makes it possible to study privacy as a routine firm characteristic with productive consequences. Second, it shows that regulation can induce productive capital formation. Firms most constrained by the privacy wave are compelled to build the most privacy capital and subsequently experience the largest efficiency gains, implying that regulatory burdens can create persistent organizational assets rather than merely generate compliance expenditure. Third, it links the corporate-finance and asset-pricing consequences of that capital formation. Privacy capital is initially discounted as regulation restricts data-related opportunities, but the most affected firms subsequently re-rate as its productive value becomes recognized. The temporary return premium therefore appears more consistent with gradual recognition of an intangible asset than with compensation for standing systematic privacy risk.

**Related Literature** This paper relates to four literatures. The first is the economics of privacy and its regulation, surveyed by Acquisti et al. (2016), Goldfarb and Tucker (2019), and Goldfarb and Que (2023). A large body of work establishes how privacy and the rules that govern it shape

consumers, advertising, competition, innovation, and market structure, much of it documenting the burdens that regulation imposes (Miller and Tucker 2009; Goldfarb and Tucker 2011; Tucker 2014; Campbell et al. 2015; Jia et al. 2021; Peukert et al. 2022; Aridor et al. 2023; Johnson et al. 2023; Johnson 2024). The results in this paper do not contradictory their findings. The time periods are different, and the sample firms are different. Privacy could very reasonably affect public firms and private firms differently. What I find is complementary to existing consensus, and it adds the incremental evidence from a bright side. Within finance, a smaller literature prices data breaches and cybersecurity as downside events in the cross-section (Kamiya et al. 2021; Florackis et al. 2023; Jamilov et al. 2021). The two streams mostly observe privacy at discrete events, a statute taking effect or a breach announced. I observe it as a routine practice in a firm-year panel, which is what lets a capital emerge, and I show that the capital has first-order consequences for operations and valuation. The closest privacy-specific work is Cheng et al. (2025), which builds a firm-level privacy-risk exposure from earnings-call language and prices it as a static, persistent risk. The object and the instrument sit in tension there. Call language is a flow of attention. It moves with the news cycle, and in my tests privacy talk on calls moves inversely with the practice firms actually build. I use call language too, only where attention or managerial behavior is the object, never to measure the stock. Compensation for privacy risk should accrue to "bad-privacy" firms. In my data the premium attaches to the "good-privacy" firms, and it fades as recognition completes. The different findings between this paper and Cheng et al. (2025) are the data-privacy analogue of green returns between Pástor et al. (2022) and Bolton and Kacperczyk (2021). Results in this paper favor the realized recognition repricing explanation over a compensation for standing risk exposure explanation.

The second literature is the economics of data, which models data as a nonrival, productive input firms accumulate that shapes their growth (Begenau et al. 2018; Jones and Tonetti 2020; Abis and Veldkamp 2024; Veldkamp 2023; Farboodi et al. 2025; Farboodi and Veldkamp 2026). That work studies the quantity, the use, and the valuation of data. I study its governance, the practices that determine what a firm may collect, retain, and do with the data it holds. Privacy capital is the governance counterpart of the data stock, and I provide direct firm-level evidence that the governing practices placed on data carry consequences as real as the data themselves, reshaping firms' operations and how markets value the businesses built on data. Within this stream, Demirci et al. (2025) deserves a special place. It is the most direct supportive evidence for the mechanism this paper leans on, and the most direct bright-side evidence on the regulation itself, showing that consumers trust regulated firms more and reward them with data of better quality. My data productivity mechanism is the firm-side counterpart, privacy capital raising the efficiency of the data firms hold, and their consumer-side evidence explains why the intensive margin can win.

The third literature is the corporate finance and asset pricing of intangible capital (Corrado et al. 2009; Eisfeldt and Papanikolaou 2013; Belo et al. 2014; Gourio and Rudanko 2014; Peters and Taylor 2017; Crouzet and Eberly 2019; Eisfeldt et al. 2022). I add an intangible that none of this work considers, one built by regulation compulsion instead of voluntary investments, and I trace its payoffs and its recognition in prices. The paper also joins the stream where corporate finance and asset pricing read the same firm through one lens, with investment-based pricing recovering returns from production and financing decisions (Eisfeldt and Papanikolaou 2013; Belo et al. 2014).

The fourth literature, on the methodological side, is text as data and LLM in economics and finance (Loughran and McDonald 2011; Hoberg and Phillips 2016; Hassan et al. 2019; Gentzkow et al. 2019; Dell 2025; Ramadorai et al. 2025). I use text as the means to quantify an intangible that financial statements omit, building a firm-year measure of privacy capital from legally binding privacy policies, classified with the taxonomy of Wilson et al. (2016), and treating it throughout as a proxy for the underlying stock. Privacy policies were an understudied text in economics until the seminal work of Ramadorai et al. (2025) assembles them at scale and shows that their content varies enough to carry economics. This paper builds its capital measure on the series that work opens.

This paper carries policy implications during this era of AI expansion. The implications are forward-looking. The policy debate over data regulation, and now over artificial intelligence, is conducted almost entirely around compliance cost. Rules burden firms, and the question is how much burden firms can bear without sacrificing innovation and growth. The evidence here argues for a second entry. My evidence suggests that privacy regulation should not be viewed solely as a compliance cost that constrains data use. The privacy wave forced firms to build governance over the data they hold, and what it built behaves like capital, raising the productivity of the data operation. This matters for AI because AI potentially raises the value of exactly the governance infrastructure studied here. Models are built from firm data, and their quality rests on the provenance, consent, and organization of that data. A firm whose data is governed is a firm whose data can be used, licensed, and fed to models at scale with less legal, operational, and reputational friction. Rising governance floors for AI, like the privacy floors studied here, will incur compliance costs. But if AI deepens the productivity of well-governed data, such floors may also compel firms to build a productive asset that persists beyond compliance.

The paper proceeds as follows. Section II explains the background of privacy and regulations. Section III describes the data and the measure of privacy capital. Section IV sets out the framework and identification strategy. Section V reports the headline operating results. Section VI traces the mechanism. Section VII lists results from the asset pricing side.

Section VIII discriminates alternative explanations and performs robustness check. Section IX concludes, acknowledges limitations, and elaborates on policy implications.

# II. Institutional Background

## A. Personal data and privacy

Personal data enters modern firms as an input for production. Personal data refers to any information that relates to an identified or identifiable living person. Firms collect it and they use it to price, personalize, forecast, and operate. They may share it with vendors, advertising partners, and data brokers. They store it, secure it, and delete it subject to retention policies. Privacy, in the operational sense, is the set of practices that governs this entire procedure. It concerns what the firm collects and on what basis, what it uses the data for, with whom it shares it, how long it keeps it, how it secures it, what rights it grants the people the data describes, and how internal data infrastructure operates. These practices are costly. They require staff, systems, and hardware. This paper covers only the governance of personal data, and does not cover non-personal data.

## B. Privacy policy

A firm's privacy policy is an enforceable legal commitment. Privacy policies are not costless marketing statements. When firms make material representations about their privacy practices, failures to honor those representations can constitute deceptive practices under Section 5 of the FTC Act. Privacy-policy disclosure requirements also substantially predate GDPR and CCPA: CalOPPA has required covered commercial websites and online services collecting personal information from California consumers to post a privacy policy since 2003.

## C. Privacy regulations

The European Union's General Data Protection Regulation was adopted in April 2016 and became enforceable on May 25, 2018, replacing the 1995 Data Protection Directive. It is the reference point of modern privacy law. The GDPR regulates the full pipeline: processing requires a lawful basis (Art. 6); purposes must be specified and data minimized to them (Art. 5); individuals hold rights of access, rectification, erasure, and portability (Arts. 15–20); notices must disclose the processing in detail (Arts. 12–14); breaches must be notified within 72 hours (Art. 33); processors must be bound by contract (Art. 28); and accountability is internal

and structural processing (Art. 30), impact assessments for risky processing (Art. 35), and mandatory data protection officers for firms whose core activities involve large-scale monitoring or sensitive data (Art. 37). Fines can reach 4% of global annual revenue (Art. 83). Ireland fined Meta 1.2 billion euros in 2023 over transatlantic transfers. Article 3 applies the regulation to any firm, wherever established, that offers goods or services to people in the European Union. A U.S. retailer with European customers and a U.S. platform with European users are both in scope. Simply, this regulation applies to almost any firm that does business with EU residents.

Five weeks after GDPR enforcement, California enacted the California Consumer Privacy Act (June 2018), effective January 2020. CCPA is the first comprehensive U.S. state privacy law, granting rights to know, delete, and opt out of sale, with attorney-general enforcement and a private right of action for breaches. The 2020 ballot initiative (CPRA) extended it from January 2023 and created a dedicated regulator, the California Privacy Protection Agency. Other states like Virginia have also enacted similar regulations. As of August 2026, the number of state regulations exceeds 20. Moreover, a federal regulation is already at the discussion stage. This regulatory supports the view that privacy is salient, and its saliency is growing as AI expansion goes on.

## III. Data and the Measurement of Privacy Capital

This section explains how I build the measures from policy texts. I describe the corpus, the classification of its sentences, and the measure built from them. I then describe the sample and show the measure in the raw data. Four validations of the measure close the section. Appendices A and B document the full construction, the classification taxonomy, and the definition of every variable.

### A. The privacy-policy corpus

I draw on the same corpus[12] of U.S. public firms' privacy policies as used in Ramadorai et al. (2025), which collects dated policy versions over time and links them to firms. The corpus covers 5,463 firms and 90,273 dated policy versions. For most firms I observe a full time series of policies, allowing the construction of a rich firm-year panel. The classification of this text into privacy practice, and the privacy-capital measure built from it, are my own.

[12]For detailed documentation of sampling, crawling, and selection-bias mitigation, see the raw paper.

## B. From policy text to privacy practice

The measure is built at the sentence level. After dropping non-English text, I classify roughly 10 million policy sentences into the twelve data-practice categories of the OPP-115 taxonomy (Wilson et al. 2016)[13], a widely used, human-annotated scheme for privacy policies. I employ PrivBERT [14], a transformer-based LLM model finetuned on the Roberta model, to train a classifier. I then run inferences on the corpus sentences using the classifier. The classifying rule is intuitively simple: it measures how closely a sentence is to the seed sentences within each category. The returned value is a probability between zero and one for each category, with one meaning the sentence exactly belongs to that category. The rules for deciding the classification is three-layered. First, I retain the sentences with a probability over 0.9[15] and consider them as legitimate match. Second, I feed the ambiguous cases, which account for around 5% of the sentences to Qwen 3.5 9B, with the prompt matching the exact taxonomy. Third, I hand-craft a keyword list of privacy-native abbreviations and do a deterministic tagging. This three-layer LLM-based classification is deliberately parsimonious to balance precision and interpretability. The first step results can be near-perfectly reproducible. The classification measuring semantic distance to a human-labeled sentence set is much easier to interpret and understand than a pure black-box LLM. The second step, which is the most black-box-like step, concerns only a small proportion of the sample. The last step is deterministic, guaranteeing exact reproducibility. The details are described in full in Appendix A. A sentence may carry more than one label, though in practice almost all carry exactly one. The classification proceeds in three steps, I then set aside the categories that carry no substantive practice, like the introductory, policy-change, and contact-information sentences. These are boilerplate. I only keep the eight categories that record actual data-handling practice.

Within the substantive set, the analysis centers on the governance dimension, the firm's commitments on data retention, do-not-track, user access and deletion, and user choice and control. These are the practices privacy regulation mandates, and they define the compliance floor. Moreover, these categories are least likely to be manufactured. Each sentence is a commitment. Firms themselves have the incentives to disclose these practices as they are the "attractive" part to viewers. Over-disclosure is largely mitigated by the classification method itself. The examples of mitigation can be seen directly in the sample policy below.

[13]This paper carries the documentation of taxonomy. The seed sentences (training data for the classifier) per each category are available at https://huggingface.co/datasets/alzoubi36/opp_115.

[14]For technical details, see https://huggingface.co/mukund/privbert.

[15]This choice is arguably arbitrary, but the distribution of probabilities per each category is strongly binary at the two extreme tails. Results are nearly invariant to the cutoff probability.

## C. The measure of privacy capital

I aggregate sentences to the policy and the policy to the firm-year. A firm-year takes the most recent policy the firm published during that fiscal year, avoiding look-ahead bias. Years in which no new version is captured are left missing rather than filled forward, a conservative choice I return to below. My measure of privacy practice is the count of governance sentences in that policy, and the headline variable is its log, ln(1 + governance sentences), which I treat as a proxy for the stock of privacy capital. I count classified governance sentences rather than words, so raw length does not inflate the measure. I also prefer a count to a within-policy share. The number of distinct governance commitments is itself economically meaningful, and a share can be noisy. A firm that builds deep governance often also discloses extensive collection practice, so governance shrinks as a share of the whole even as the firm's privacy capital rises. I carry the length-adjusted residual and the share as robustness checks (Appendix A).

It helps to see one full policy as an example. I list below one of the shorter policies in the corpus in full. Perficient, a digital consulting firm, kept the policy below posted nearly unchanged from April 2016 through December 2017, which makes it the policy of record for the firm's fiscal years 2016 and 2017 and the text that fixes its pre-wave level in the design of Section IV. Sentences the pipeline classifies into the governance categories are set in bold.[16]

> *The privacy policy of Perficient, a digital consulting firm, as posted from April 2016 through December 2017, reproduced in full. Sentences classified into the governance categories are in bold.*
>
> Privacy Policy Your Privacy on Perficient.com Perficient is committed to protecting your privacy and implementing technology that gives you a rewarding and safe online experience. We want you to feel secure when visiting our site and are committed to maintaining your privacy when doing so. What Information We Gather and How We Use It Information on Perficient.com and other Perficient websites is gathered in two ways: directly and indirectly. Perficient will at times ask you to directly provide information that personally identifies you (registering for an event, downloading a file, signing up for an email list, etc.). **It is our policy to keep personal information only as long as needed to fulfill the purposes for which the information was collected.**
>
> Cookies and IP Data We use "cookies" on this site. A cookie is a piece of data stored on your device to help us improve your access to our site and identify repeat visitors. Cookies can also enable us to track and target the interests of our users to enhance the experience on our site. We also collect information indirectly through our Internet access logs. When you visit Perficient.com, your Internet Protocol (IP) address is automatically collected and is placed in our Internet access logs. IP addresses are not linked to personally identifiable information. Your IP address is used to help diagnose any technical problems, analyze site trends, administer the site, track movement within the site, and gather broad demographic information about site use that can eventually be used to make improvements that enhance your user experience on Perficient.com. Perficient does not sell, exchange, trade or give away personal information you voluntarily provide to us to any outside sources.

[16] The text is reproduced as classified. The crawl concatenates the page's section headings and footer into adjacent sentences, which is why several sentences open with a heading; I normalize spacing at a few sentence boundaries, drop one stray character, and add the paragraph breaks. Successive crawls of this policy between April 2016 and December 2017 are nearly identical.

**Your Choices When we collect information from you through a web form, you may tell us that you do not want it to be used for further marketing contact. You may also turn off cookies in your browser. Persistent cookies can be removed by following your web browser's directions.** A session cookie is temporary and disappears after you close your browser. **You can reset your web browser to refuse all cookies or to indicate when a cookie is being sent. Access to Your Information You are entitled to know whether we hold information about you and, if we do, to have access to that information and require it to be corrected if it is inaccurate.** You can do this by sending us an e-mail. **Opt-Out Information Perficient may on occasion send email communication regarding events, marketing content, its services, etc. We will always provide email recipients a clear and easy means to opt-out of receiving any future communication. We will honor unsubscribe requests from any commercial email messages within 10 days, and these unsubscribe requests will never expire.**

Site Security We take appropriate steps to provide and maintain a secure website. By visiting Perficient.com, however, you agree that the open nature of the Internet is such that data may flow over networks without security measures in place and may be accessed and used by people other than those for whom the data is intended. Your Acceptance of These Terms By using Perficient.com or any of its family of websites, you provide your consent to the Perficient Internet Privacy Policy. We reserve the right to change, modify, add, or remove portions from this policy at any time at our discretion. Please check this page periodically for changes. This site may contain links to other sites. Please be aware that we are not responsible for the content or privacy practices of such sites. We encourage our users to be aware when they leave our site and to read the privacy policy statements of any other site that collects information from you.

Safe Harbor Policy Perficient complies with the U.S.-EU Safe Harbor Framework and the U.S.-Swiss Safe Harbor Framework as set forth by the U.S. Department of Commerce regarding the collection, use, and retention of personal information from European Union member countries and Switzerland. Perficient has certified that it adheres to the Safe Harbor Privacy Principles of notice, choice, onward transfer, security, data integrity, access, and enforcement. Learn more about the Safe Harbor program If you have any questions about Perficient's Internet Privacy Policy, Safe Harbor Policy, this website, or your dealings with this website, we welcome your comments Privacy Policy © 2017 Perficient, Inc.

The bold sentences are what the measure counts. The policy pledges purpose-limited retention, grants access and correction, and spells out cookie and marketing opt-outs, eight governance sentences in all, so the firm enters the panel at $P = \ln 9 \approx 2.2$, near the sample mean in Table 1. The rest of the document is what the measure leaves out. The collection and security material belongs to other dimensions and the changes-and-links passage is boilerplate. The example also previews the adjustment the rest of the paper studies. Two weeks after the GDPR took effect, this policy was replaced by one nearly three times as long, its governance count more than tripled to twenty-seven, and the versions posted after the California wave are larger still.

## D. Sample and panel

I link the firm-year measure to Compustat for accounting data and through the CRSP-Compustat link to CRSP for market data. The sample is U.S. public firms, and I exclude financial firms and regulated utilities, sectors already governed by sector-specific privacy regimes and in which

privacy rules are different. [17] The estimation sample is 9,917 firm-years for 2,439 firms over fiscal years 2016 to 2023. The design itself uses the firms with a policy observed in calendar year 2016 or 2017, whose pre-wave level defines the treatment dose, roughly 7,400 firm-years for 1,650 firms. The tables report exact counts. For the identification tests I also build a longer panel back to 2010, which supplies the pre-period that the parallel-trends evidence in Section IV needs. The panel keeps firms that later delist, so survivorship does not drive the results. I winsorize the accounting ratios at the 1st and 99th percentiles.

The outcomes fall into four groups. Financing and liquidity are book and market leverage, cash to assets, and the current ratio. Efficiency and investment are asset turnover and capital expenditure to assets. Profitability is return on assets and operating margin. Valuation is Tobin's $Q$ and the total-$q$ measure of Peters and Taylor (2017), which puts intangibles in the denominator. Appendix B defines each variable from its Compustat and CRSP source. Section-specific data and corresponding constructions are reported in respective sections.

Table 1 reports the descriptive statistics. The sample is not limited to large-caps. Median assets are \$1.2 billion against a mean of \$8.6 billion, so the panel carries a wider distribution. The median policy records 11 governance sentences and the mean about 14. This is arguably a reasonable number and an indication of few duplications or sentence-level verbosity confounds, though not a formal justification. The log measure has a cross-firm SD of 0.99, so firms are far from a common level of practice. The financing variables sit in familiar ranges, with mean cash at 24% of assets and book leverage at 32%. Profitability is skewed, with a negative mean and a near-zero median return on assets, the loss-making tail familiar in the recent Compustat universe.

[Insert Table 1 here]

## E. The measure in the data

Figure 2 shows the measure over the sample. Panel A plots mean governance practice by year against the compliance floor, the mature standard the directly bound firms reach, which rises from about 2.0 in log points before the wave to 3.17 by its end. Average practice climbs steadily toward that floor, from 1.91 in 2016 to 2.86 in 2023. Panel B plots the cross-firm dispersion of the measure, which trends upward over the same period, about 0.0085 log points of standard deviation per year. Practice fans out as the wave proceeds instead of collapsing to a common minimum. This is a first cut against the cost-only reading. A cost-only view predicts compression toward the floor once the mandate binds, and I find the opposite. Section V returns

[17] Financial firms fall under the Gramm-Leach-Bliley Act and bank examiners, and utilities under their regulators, so privacy practice in these sectors predates and sits outside the wave I study.

to the fan-out as formal evidence.

[Insert Figure 2 here]

## F. Validating the measure

The natural concern with any text-based measure is that it captures words rather than practice. Four pieces of evidence push against this concern. Each asks whether the measure responds, with the right timing, to an event that changes real practice and that is observed from a source other than the policy text itself. First, the measure jumps for firms exposed to the European Union's regulation[18] when it takes effect, rising about 0.21 at enforcement ($t = 5.9$) with the jump concentrated in 2018 and 2019. The one pre-enforcement coefficient is statistically indistinguishable from zero and of the opposite sign. Second, it rises when a firm's industry peers[19] suffer data breaches (+0.022, $t = 4.0$), with no anticipatory lead and with the response specific to governance, as a firm tightens practice after seeing the consequences of lax practice nearby. Third, it tracks the staggered adoption of U.S. state privacy laws (+0.14, $t = 3.3$), turning on as each state's law arrives. Fourth, it lines up with two independent third-party assessments of firms' data-privacy practices, from commercial ESG providers[20], which draw on sources other than the policy text. The measure moves when real practice moves, and only then. It is also deliberately slow-moving. A firm's own breach produces no disclosure surge, so the measure behaves like a maintained stock, not crisis public relations.

Two further concerns bear on how the measure should be read. The first is classification error. This is a natural worry around automated labeling. Two things make it less troubling. The pipeline rests on a human-annotated taxonomy and combines a supervised model, a language model, and a keyword pass, so errors are unlikely to line up systematically with firm outcomes. And noise of this kind mostly attenuates the estimates toward zero, which makes the tests conservative. As a direct check, placebo regressions that use the non-substantive boilerplate sentences as the regressor produce barely any effect on the outcomes. Content-based transforms of the governance measure, a policy-length control, a length-purged residual, and a within-policy share, deliver the full set of headline results (Appendix A, Table 18).

The second concern is coverage and timing. The corpus covers listed U.S. firms with a public privacy policy, the population the regulation reaches, and because the panel keeps

[18] This number is captured by a diff-in-diff deisgn using EU sales share as the treatment intensity. The per geographical sales are reported in a highly heterogeneous way. For example, one company can report Europe + North Africa, and another firm reports only Germany alone. Thus, this is not ideal for an instrumented diff-in-diff design to carry forward. The jumps of privacy variables are robust to both using this noisy sales share measure and the `is_in_EU` dummy.

[19] Though I am not claiming a causal evidence, I use peer's breaches instead of own breaches to avoid endogeneity.

[20] MSCI and Refinitiv. In their ESG measures, they both list a sub-score of data privacy.

delisting firms, survivorship does not drive the results. For each firm-year I use the most recent policy published during the fiscal year, so the measure never uses information the market did not yet have. Years without a captured revision are left missing. I do not apply forward-fill, since filling forward would assert that nothing changed exactly when the floor was rising. The missing captures could be driven by no-update, but it could also be driven by a failed crawl. As a robustness, carrying each firm's last observed policy forward into uncaptured years (forward-fill) yields the same signs with, if anything, larger and noisier estimates.

## IV. Framework and Identification

This section states the minimal discipline the paper needs, a definition of privacy capital as a stock and a compliance floor that partitions firms into compelled and voluntary builders. I describe the continuous difference-in-differences built on distance below the floor, defend the identifying assumptions, and report the first-stage results.

### A. The stock and the floor

Firm $i$ holds a stock of privacy capital $P_{it}$, the data-governance practice described in the introduction. The stock accumulates through investment and depreciates,

$$P_{i,t+1} = (1 - \delta)\, P_{it} + I_{it}, \qquad I_{it} \geq 0, \tag{1}$$

and installing $I$ costs $I + \frac{\kappa}{2} I^2$, so building practice is costly and building it quickly is costlier. This convex cost assumption follows Hayashi (1982). Firms differ in the stock they privately want. Data-reliant firms want more privacy capital, and a growing data economy raises the desired stock over time for the firms that lean on data most.

Regulation enters as a floor. The wave of statutes, and the peer and market pressures that moved with them, set a compliance standard and raise the pressure to meet it over the sample. The binding level of practice $\underline{P}_t$ rises accordingly,

$$P_{i,t+1} \;\geq\; \underline{P}_t, \tag{2}$$

so a firm's choice is $P_{i,t+1} = \max\{\widetilde{P}_{i,t+1},\, \underline{P}_t\}$, where $\widetilde{P}$ is what it would choose on its own. The floor partitions firms, and the partition is the design.

**Proposition 1** (Floor partition)**.** *Firms whose desired stock lies below the floor are dragged to it, and locally* $\partial P_{i,t+1}/\partial \underline{P}_t = 1$*, independent of the marginal value they attach to the stock (the*

*compliance regime). Firms whose desired stock lies above the floor invest by marginal value, and their choice is locally independent of the current floor (the voluntary regime). A firm's distance below the floor therefore measures the force the floor exerts on it.*

The proposition motivates the treatment variable, each firm's pre-wave distance below the floor scaled by the rising pressure. It is a statement about who the floor moves and by how much, not an identification proof. The exogeneity, exclusion, and monotonicity of the empirical dose are assumptions I elaborate and defend in the subsections that follow. The partition also says where the two kinds of evidence live. Compelled firms supply the instrumented variation, and voluntary firms, the ones already above the standard, reveal that the stock is privately valuable when they keep building, which no mandate requires.

## B. The design

A firm chooses its privacy capital endogenously. I address the endogeneity by using variation a firm cannot choose, the broad wave that raised corporate privacy practice between 2018 and 2023. The wave set a compliance floor and the pressure to meet it rose year by year. Firms differed in how far below that standard they began. A firm that started well short of the standard had to build a great deal of privacy capital as the pressure arrived, while a firm near the standard built little. The distance a firm sat below the floor before the wave is therefore a measure of how hard the wave treats it, and that distance is predetermined.

Privacy regulation is the most visible force in the wave, but regulation alone is not sufficient, for two reasons. First, the wave was wider than the statutes. High-profile breaches and scandals, a steep rise in consumer and media attention to data, and platform and industry shifts moved privacy practice across the whole economy at the same time, and an identification strategy keyed only on the statutes would miss them. Second, treating the statutes themselves as the shock would require a clean firm-level measure of exposure to each statute, which existing data do not support, and a coarse split into treated and control firms would reintroduce the selection the design is meant to remove.

Let $s_i$ denote firm $i$'s privacy governance capital averaged over 2016 and 2017, before the wave, and let $\underline{P}$ denote the floor. The firm's dose is its censored distance below the floor, $D_i = \max(\underline{P} - s_i, 0)$, and treatment intensity is that distance scaled by the rising pressure, $T_{it} = D_i \, \Delta\underline{P}_t$. The two pieces are a level and a path. $\underline{P}$ is the level, the compliance floor, and I locate it at the end state the directly compelled firms reach. Firms with European operations or California headquarters are bound by the GDPR and the CCPA/CPRA, and by the end of the sample their practice traces out the mature compliance standard, so I set $\underline{P}$ to their 2022–2023 governance norm, 3.17 log points, up from 2.02 in 2016–2017. $\Delta\underline{P}_t$ is the path, the cumulative

rise of the binding level by year $t$, the realized compliance pressure. I measure it as the rise since 2017 of the 75th percentile of governance practice across the firms in the panel, a year-level series common to all firms. The measurement choice is arguably arbitrary but it is robust to different designs. The pressure combines statutory enforcement, which tightened over the window, with the peer and market scrutiny that moved alongside it, and the frontier of practice records their joint arrival. The floor comes from the bound firms, whose settled practice marks where the standard ends up; the pressure comes from the whole panel, which all of those forces moved. Anchoring the distance at the end state is deliberate. A firm's distance to the current year's standard would shrink with the firm's own catch-up, so it would inherit the response the design is meant to isolate. Distance to the end state is fixed by pre-wave practice and is predetermined. The dose is then, up to scale, the portion of firm $i$'s total compelled climb that the wave has demanded by year $t$. Because the standard is high, almost every firm begins below it, about 92%, so the design is not a discrete binary treated vs control setting. [21]

Because privacy capital is itself endogenous, I use the dose as an instrument for it. The two equations are

$$P_{it} = \pi\, T_{it} + \mathbf{x}'_{it}\boldsymbol{\gamma} + \alpha_i + \delta_t + u_{it}, \tag{3}$$

$$Y_{it} = \beta\, \widehat{P}_{it} + \mathbf{x}'_{it}\boldsymbol{\lambda} + \alpha_i + \delta_t + \varepsilon_{it}, \tag{4}$$

with firm fixed effects $\alpha_i$, fiscal-year fixed effects $\delta_t$, and controls $\mathbf{x}_{it}$ (log assets and book-to-market as baseline and an extended control set for robustness). This is a fuzzy difference-in-differences in the sense of de Chaisemartin and D'Haultfœuille (2018). The reduced form is the dynamic effect of the dose on the outcome, and the second-stage coefficient $\beta$ scales it into units of privacy capital, the local average treatment effect for the firms the wave compelled to build, understood in the continuous-dose sense as a dose-weighted average of responses among the firms that comply with it. I report the first stage and the design's validity here; the estimates of $\beta$ are the subject of Sections V through VII.

## C. The identifying assumption

Identification rests on parallel trends in the dose. Absent the wave, firms that began closer to and farther from the floor would have moved in parallel on the outcomes I study. The assumption does not require the pre-period level $s_i$ to be exogenous. More established and larger firms plainly began closer to the floor, and they differ from other firms in financing, investment, and

[21] The natural statutory alternative, a regression discontinuity at the CCPA's $25 million revenue threshold seems attractive. However, the existing sample does not support because the number of public firms around the cutoff is too small.

value for reasons unrelated to privacy. The firm fixed effects absorb those level differences, and identification comes from the differential movement of high- and low-dose firms after the wave.

The choice of floor is the natural place to press the design. Because the fixed effects absorb levels, the floor does not enter the estimate through its height; with firm and year effects the treatment reduces to a term in the firm's distance interacted with the wave, and the level of the floor cancels among treated firms. It enters only by setting which firms still had distance to travel when the wave arrived. Two things keep that from driving the result. First, I select the floor at the end-of-sample practice of the firms the law directly binds. This is the economic justification. Second, because the standard is high, almost every firm sits below it, so there is no clean group of already-compliant firms serving as controls. Sliding the floor across the 60th to 90th percentiles of the pre-period distribution, or removing the censoring entirely, leaves the second-stage estimates essentially unchanged (Appendix C, Table 20). An externally measured wave of privacy-policy adoption, from Refinitiv ESG, reproduces the results, so the timing is not sample-dependent. This is the statistical justification. The design therefore stands or falls on a single testable question, whether the gap between high- and low-dose firms was already moving before the wave.

## D. Parallel trends

The identifying assumption is parallel trends in the dose. Firms nearer and farther from the floor must have moved in parallel absent the wave. The financing and efficiency outcomes show no differential movement before the wave arrives. Figure 3 illustrates the identifying assumption. It plots the coefficients from interacting the distance-to-floor dose with fiscal-year dummies, the standard event study for a continuous treatment, with 2017 normalized to zero and 95 percent confidence intervals. Cash is flat with at most a faint drift, asset turnover shows a small pre-period wiggle and a clear post jump, return on assets behaves like turnover, and the margin panels, whose outcomes are dispersed ratios, are wide and individually indistinguishable from zero before the wave. The exception is capex, whose early pre-period cells are small and positive and fade before the wave; I read its post-period path with that caveat. The estimates also survive interacting pre-period firm characteristics with year, one at a time and jointly.

Raw Tobin's $Q$ is the remaining exception, its level already drifting beforehand. In the baseline panel the value of high-distance firms drifts down before the wave. The drift turns out to live in the extreme-valuation tail, not in the typical firm. The median pre-period path of a far-below-floor firm is indistinguishable from that of a near-floor firm, and in logs the drift resolves into a single old step. Relative value steps down in 2013, sits flat for the four years leading into the wave, and declines only once the wave arrives, reaching its widest point

in 2021 and narrowing after.[22] Figure 4 shows both resolutions. Total $Q$, which corrects the denominator for intangible capital, needs no repair. It steps down from early positive cells onto zero by 2013, holds there into the wave, and is the one valuation series that passes the joint pre-period test. Among firms the floor barely binds, those with equally steep pre-period declines show nothing in 2020 and 2021, so a mechanical rebound of beaten-down valuations is absent exactly when it would need to appear. I read the discount as identified once the tail is handled, and I corroborate it with the recognition dynamics of Section VII.

A second timing concern, anticipation, does not bind. The European regulation was adopted in 2016 but enforced only in 2018, and the real outcomes do not move in the 2016 and 2017 window, as one expects of compliance that is costly and slow.

[Insert Figure 3 here]

[Insert Figure 4 here]

## E. The first stage

The first stage is strong and has the sign the design predicts. The firms that began farthest below the floor build the most privacy capital as the wave arrives, while firms that began near the standard have little room to move. Table 2 reports it across specifications. In the baseline column the dose enters the privacy capital regression positively, with a coefficient near 0.44 and a first-stage $F$ of 44. A firm a log-point farther below the end-state floor closes about 0.44 log-points more practice over the wave (+0.435, t 6.60), stable from +0.415 to +0.493 across specifications. Dropping the controls, adding a richer set (profitability, R&D intensity, intangible intensity, and sales growth), interacting pre-period size and profitability with year, replacing the year effects with SIC-two-digit-industry-by-year effects, and using the uncensored distance all leave the dose strongly significant, with first-stage $F$ statistics between the 40-85[23].

The design also requires that the treatment be monotone, that no firms respond to the wave in the wrong direction. Sorting firms into terciles of their distance below the floor, the wave's effect on privacy capital is strictly ordered and same-signed, the farthest-below tercile building most (a wave loading of +0.76), the middle tercile less (+0.38), and the nearest tercile the reference, so no group defies. Figure 1 displays this as compliance catch-up, with the terciles converging toward the floor over the sample and the privacy capital built rising monotonically with distance.

[22] The step coincides with the Snowden disclosures of June 2013, the first economy-wide privacy salience shock. A level difference five years before treatment, followed by four flat years, is not a differential trend into treatment.

[23] An independent statute-exposure design agrees on relevance: instrumenting with European exposure interacted with GDPR enforcement produces a strong first stage on the measure over 2016 to 2019, though that short pre-CCPA window is underpowered for outcomes.

[Insert Table 2 here]

# V. Empirical Results

The identification section establishes that distance below the floor moves privacy capital. This section shows what the capital does. I report both the reduced form, the response of each outcome to the predetermined regulatory dose, and the corresponding 2SLS estimate, which scales that response by the first stage and therefore measures the effect of per unit of accumulated privacy capital for the firms whose practices the regulatory wave moves.

The operating results show the payoffs. Firms compelled to build privacy capital invest more, use their existing asset base more efficiently, and subsequently earn more from that asset base. These effects persist beyond the year in which the practices are built. At the same time, I find no unconditional deterioration in margins that would make privacy look only like an operating expense. The financing variables also move, Firms hold more liquidity as privacy capital accumulates. I treat this financing response as suggestive only. It is consistent with the funding needs of an intangible asset that supplies little collateral, but the cash response by itself does not distinguish non-pledgeability from other reasons a firm might build precautionary liquidity.

The baseline specification includes firm and fiscal-year fixed effects together with log assets and book-to-market. I do not attach a structural interpretation to the controls, as contemporaneous firm characteristics can themselves respond to treatment. These two controls are retained because they produce conservative estimates. Specifications with fixed effects alone and alternative control sets deliver similar or larger effects.

## A. The direct payoff

Table 3 reports the corporate footprint of privacy-capital accumulation. Each column is a separate two-stage least-squares regression of the outcome on instrumented privacy capital, with firm and fiscal-year fixed effects and the controls. The coefficient on privacy capital sits above its $t$-statistic, and the reduced-form coefficient on the dose sits below it.

The most important result is an increase in asset utilization. Instrumented privacy capital raises asset turnover by 0.101. Equivalently, A one-SD within-firm increase in privacy capital therefore raises turnover by approximately 4.7% of its sample mean. Firms compelled to build more privacy capital generate more sales relative to the assets employed in the business. This utilization gain is accompanied by higher profitability. Return on assets rises by 0.148 per unit of privacy capital, corresponding to approximately 5.8 percentage points for a one-SD within-firm

increase in the privacy-capital measure. Because the mean ROA in this sample is negative, scaling this effect by the sample mean is not economically meaningful. I therefore report the magnitude in percentage points. The later margin evidence is important for interpreting this result. Neither profit margin nor operating and gross margins increase unconditionally. The joint pattern is therefore consistent with the improvement in ROA coming primarily through more productive use of the asset base instead of a general expansion in markups or accounting margins. Capital expenditure also rises. The coefficient of 0.009 implies roughly 0.35 additional cents of capital expenditure per dollar of assets for a one-SD increase in privacy capital, about 10% of the sample mean. I do not interpret this as the expenditure that mechanically creates the privacy-capital measure. Privacy investment could take the form of personnel, organizational processes, software, and other expenditures that need not be capitalized. Instead, the role of capex response here indicates that privacy-capital accumulation does not hurt firms' investments.

Taken together, the turnover and ROA responses form the central operating result. The wave certainly requires firms to spend resources to comply. The question is whether those expenditures disappear once incurred or help create a persistent productive stock. The positive utilization response is the first evidence for the latter.

Table 3 also shows a liquidity response. Privacy capital raises cash holdings by 0.037 and the current ratio by 0.946. At a one-SD increase in privacy capital, cash rises by roughly 1.4 cents per dollar of assets, or about 6% of its sample mean. One interpretation is that accumulating an organizational intangible increases the value of internal liquidity because the stock itself supplies little conventional collateral while requiring continuing maintenance and adjustment. Consistent with this interpretation, the cash response is concentrated among firms entering the wave with relatively little tangible collateral (Table **??**). I view this evidence as suggestive only. A non-pledgeable intangible naturally generates such a liquidity footprint. But it also be can other anticipated funding needs. The cash result therefore supports the broader capital interpretation without carrying the identification of the economic mechanism.

[Insert Table 3 here]

## B. The payoff persists in the next year

A capital stock should affect production beyond the period in which it is installed. Table 4 shifts the operating outcomes forward one fiscal year while keeping privacy capital and the regulatory dose dated at $t$.

The cleanest result is asset turnover. The coefficient one year ahead is 0.110, slightly larger than the contemporaneous estimate of 0.101 and statistically significant. In economic terms, a one-standard-deviation increase in privacy capital raises next-year turnover by about

5.1% of the outcome mean. The main efficiency gain therefore does not disappear when the contemporaneous treatment year ends. This persistence is meaningful because the lead specification is mechanically conservative. the dose dated at $t$ understates the regulatory pressure prevailing when the $t + 1$ outcome is realized. Despite attenuation, the asset turnover effect survives at essentially full magnitude.

The estimates on ROA and capital expenditure remain positive one year ahead, at 0.072 and 0.004 respectively, but are not statistically distinguishable from zero. I therefore do not use them as evidence of persistence on their own. The unconditional operating and gross margins are also statistically unchanged at the lead. The economically disciplined conclusion is narrower here. Later heterogeneity tests show that profitability and margin effects become considerably more visible when personal data is economically important to the firm, but the unconditional persistence result rests on asset turnover.

This distinction helps separate a stock from a one-time adjustment. A transitory compliance episode could create contemporaneous spending or accounting effects. It is harder for that interpretation alone to explain an efficiency response that remains essentially unchanged in the following fiscal year. The evidence is therefore consistent with the practices installed during the regulatory wave continuing to affect the firm's production after the initial build.

[Insert Table 4 here]

### C. Beyond a cost-only interpretation

The results so far do not imply that privacy regulation is costless. Compliance requires real resources, and some of those costs are undoubtedly borne by firms. The relevant comparison is more limited. Is the economic effect of privacy governance limited only to these compliance costs? A cost-only benchmark has three natural implications in this setting. The required build should leave a traceable expense. Firms should have little reason to continue accumulating once the minimum standard is satisfied. The burden should be most visible among the firms forced to travel farthest to meet the new standard. The evidence does not line up well with this joint pattern.

The first implication concerns the income statement. Privacy compliance requires employees, systems, legal resources, and continuing maintenance. If these expenditures dominate the economic effect, one would expect lower margins as privacy capital accumulates. Table 5 examines operating margin, profit margin, gross margin, selling and administrative expenditure relative to sales, and R&D intensity. None moves significantly with instrumented privacy capital. Operating margin is effectively zero, profit and gross margins are positive but not statistically significant, and SG&A intensity does not increase. The absence of a margin decline should not

be read as evidence that firms incur no compliance costs. The accounting data cannot separately observe the cost of the privacy build and the benefits it generates, and the two may offset within the same firm-year. What Table 5 establishes is that compliance expenditure does not dominate the net operating response. Firms do not exhibit the broad margin compression one would expect if the regulatory wave were showing up primarily as an operating burden. Combined with the positive turnover and ROA responses, the income-statement evidence instead points to productive benefits that offset, and exceed in some cases, the costs of the build.

[Insert Table 5 here]

The second implication concerns what happens when firms reach the floor. A pure minimum-compliance equilibrium predicts convergence. Once firms have installed what regulation requires, there is no reason to continue spending on privacy practice. The data show only part of that prediction. Firms beginning below the standard do converge toward it, which is precisely the first stage of the design. But firms near the top do not stop. The most privacy-established firms reach the end-state standard early in the wave and continue adding governance practice afterward. Consequently, the middle of the distribution compresses while the upper tail continues to move outward, and overall cross-sectional dispersion rises modestly rather than collapsing. Figure 1 illustrates this.

The third implication concerns incidence. If privacy were primarily a regulatory burden, the adverse operating consequences should be most apparent among the firms forced to adjust the most, the firms entering the wave farthest below the eventual floor. Table 6 tests this directly in reduced form by dividing firms into terciles of their predetermined distance to the floor and interacting those groups with the regulatory wave. The gradient goes in the opposite direction. Relative to firms beginning nearest the floor, the farthest-below firms increase capital expenditure by 0.008, asset turnover by 0.091, and ROA by 0.188 as the wave unfolds. The middle group also increases investment and turnover, but by less and with a much smaller profitability response. The productive response therefore increases broadly with the amount of adjustment the regulation compels. At the same time, neither operating margins nor SG&A intensity deteriorates among the firms facing the largest required build. This dose pattern is economically important. The firms for which compliance should be most burdensome are also the firms that invest the most and obtain the largest subsequent gains in utilization and profitability.

The evidence in this section is not claiming compliance is free. There are certainly substantial costs induced. However, the joint patterns together suggest the cost should not be viewed as deadweight loss, and there is capital formation alongside.

[Insert Table 6 here]

# VI. The Data Productivity Mechanism

The previous section establishes that privacy capital improves firms' operating efficiency. Asset turnover rises, followed by ROA, while average margins do not fall. This section asks why. I hypothesize that privacy capital raises the productivity of firms' data. Privacy regulation restricts the breadth of data firms may collect, retain, and share, but the governance infrastructure built in response can improve the intensive margin of the data that remains. Clearer retention rules, access controls, deletion procedures, and more disciplined data architecture can reduce redundant data resources, make existing data easier to locate and use, and improve the quality of newly supplied data when stronger privacy protections increase users' willingness to provide accurate information. Privacy capital can therefore substitute better governed and more intensively used data for indiscriminate data accumulation. The net prediction is not necessarily more data, but greater productive output from the data operation the firm retains.

However, data productivity is inherently latent. Data does not appear on the balance sheet, and there is no common accounting unit that measures how productively a firm uses it. I therefore test the mechanism through its discriminating implications. The central prediction is cross-sectional. If privacy capital works through data, its operating payoff should become larger as personal data becomes more important to the firm's business. The same outcomes should not systematically appear along characteristics with little direct data content. I test these predictions using predetermined measures of firms' roles in collecting, sharing, and commercially using personal data, together with data-adjacent and data-free moderators. I then ask whether these heterogeneous payoffs persist into the following year, as the payoff from a capital stock should.

## A. Heterogeneity by what the firm does with data

The channel's sharpest test is whether the operating response varies with what the firm actually does with personal data. I estimate these interactions in reduced form, regressing each outcome on the predetermined regulatory dose and its interaction with a pre-wave moderator. The specification is saturated in the moderator's year path and its interactions with the controls, so the interaction coefficient $\delta$ captures differential exposure to the privacy wave. Continuous moderators are standardized using their 2016–2017 values; consumer-facing and technology indicators enter categorically. Because the exercise is conducted directly in reduced form, the interpretation of $\delta$ is the incremental outcome response to regulatory exposure.

Table 7 organizes the moderators into three tiers. Panel A measures the firm's actual role with personal data. The moderators are third-party sharing, first-party collection, and whether the firm is consumer facing. Collection intensity and sharing intensity are measured from

privacy policies in a similar way as the governance measure. Panel B contains characteristics adjacent to data production but not direct measures of the firm's data role, technology-sector membership and R&D intensity. Panel C contains characteristics with no intrinsic data content: industry concentration, pricing power, intangible intensity, and firm size. This partition makes the test discriminating. A generic scale, quality, or technological-sophistication explanation may generate heterogeneous operating responses, but it has no particular reason to make those responses line up systematically with the firm's collection and sharing of personal data.

Per standard deviation of pre-period sharing disclosure, sharing-intensive firms turn assets over faster (+0.044), invest more (+0.003), and gain operating and gross margin (+0.612 and +0.264) under exposure. Collection-intensive firms gain gross margin (+0.252). Consumer-facing firms take the same efficiency and investment gains (+0.098 and +0.007). No profitability interaction is significant in the role tier and no margin falls. The unconditional margin effects in the baseline result set are null, while here it flips positive. This is highly consistent with the hypothesis that the significant incremental responses cluster where privacy governance most directly changes the data operation. Furthermore, the pattern is difficult to reconcile with a pure compliance-cost channel. Where the regulated data business is most salient, profitability improves instead. The data-adjacent tier responds only in part, which is what adjacency predicts. The technology grouping, which is categorized by an industry label, shows null incremental effects on every outcome, indicating that the mechanism is not a technology-sector effect. Tech firms do handle data intensively, but the setting in this papers is relevant for personal data only. Research intensity moves profitability alone. It functions more like a scaling variable. The productivity depends heavily on the quantity and quality of data. R&D could partly proxy for the technology firms employ to benefit from their data. In the bottom tier no moderator reproduces the joint efficiency, investment, and margin pattern associated with the firm's actual data role. The evidence does not require every non-data interaction to equal zero. Its identifying content is that the economically coherent footprint is concentrated along dimensions that describe what firms actually do with personal data.

[Insert Table 7 here]

## B. Data-relevant heterogeneity is persistent

A capital mechanism should not end in the year in which the governance practices are built. I repeat the heterogeneity battery with operating outcomes measured one fiscal year ahead. This test is also useful for distinguishing persistent operating consequences from contemporaneous accounting adjustment.

Table 8 strengthens the persistence interpretation. For sharing-intensive firms, the incre-

mental asset-turnover effect still persists one year ahead (+0.044). The coefficient is almost exactly the same as its contemporaneous magnitude, while the operating and gross margin gradients strengthen to +0.772 and +0.368. For first-party collectors, the gross-margin gain persists and is joined by a +0.552 operating margin effect and greater capital expenditure. Consumer-facing firms exhibit a different dynamic. Their contemporaneous turnover advantage attenuates, while one-year-ahead ROA, capital expenditure, and operating margin increase. These patterns jointly suggest that the data-role heterogeneity does not look like a one-period accounting rearrangement. Its composition evolves, but the operating payoff persists and in several cases broadens from efficiency into profitability.

The lead results also discriminate more sharply than the contemporaneous table. Some adjacent and data-free moderators generate isolated movements in turnover, investment, or ROA, but none generates a positive margin gradient one year ahead. Persistent margin gains are confined to the direct data-role tier. The R&D–ROA interaction observed contemporaneously disappears, while the sharing and collection effects survive or strengthen. This pattern is more consistent with the privacy payoff being tied to the governed data operation than with a generic technology, intangible-capital, market-power, or scale channel.

Taken together, the two heterogeneity tests give the data-productivity interpretation its cross-sectional and dynamic content. The average effect of privacy capital operates primarily through asset utilization, but the payoff becomes broader where personal data is central to production. Sharing-intensive, collection-intensive, and consumer-facing firms exhibit stronger efficiency, investment, and eventually profitability responses. Those responses persist into the following year, while characteristics without direct data content produce only scattered effects. I therefore interpret the operating evidence as consistent with privacy capital raising the intensive-margin productivity of firms' data.

Though the heterogeneity results are hard to reconcile with general organizational quality or benefiting technological sophistication explanations, there could also be also be additional alternative explanations or confounds. I return to alternative explanations and robustness in Section VIII.

[Insert Table 8 here]

## VII. Market Recognition

Regulation compels firms to build a persistent governance stock, that stock improves asset utilization, and the operating gains concentrate where personal data matters most to production. This section asks how the stock is reflected in market prices.

The valuation problem has an asymmetry. Privacy regulation produces two effects that differ in observability. Its restriction side is immediate. Firms lose flexibility over the collection, retention, sharing, and monetization of personal data. Those lost opportunities can be priced as soon as the regulation arrives. The productive side is harder to observe. Privacy capital is unbooked, and its benefits emerge gradually through the operation of the governed data asset. The market can therefore price the restriction before it fully capitalizes the productive stock created in response. The spirit here is very similar to the regulation literature. The immediate responses in a short window tend to be negative, as the downside is most obvious.

This asymmetry organizes the evidence. Privacy capital initially carries a negative valuation loading, and that discount is largest precisely where broad collection and sharing are central to the business. Where the productive and trust benefits are comparatively important, among consumer-facing firms, the discount reverses. The valuation loading then rises sharply through the early 2020s and remains above its pre-wave level. Institutional ownership shifts toward the catch-up firms over the same period, while market attention increasingly treats data and privacy as productive economic objects. Finally, a high-minus-low privacy return premium is economically large before 2020 and essentially absent afterward. I interpret this joint pattern as the market gradually capitalizing the capital side of privacy after the initial discount.

This section uses two empirical approaches. The valuation dynamics and institutional-ownership tests reuse the distance-to-floor design and therefore exploit the regulatory variation developed in previous sections. The portfolio tests are different. Firms are sorted on lagged privacy-capital levels and I examine realized returns using standard factor models. Here the distinction is intuitive. Firms that gain the most are the heavily treated ones, and they are heavily treated because they are bad-privacy firms in levels in the beginning years.

## A. The initial discount

Panel A of Table 9 reports the aggregate valuation effect under the design of Section IV. Instrumented privacy capital carries a coefficient of 0.92 on Tobin's Q and 0.48 on the intangible-adjusted Total Q. The reduced forms have the same negative sign. Firms compelled to build more privacy capital therefore initially receive lower market valuations than firms compelled to build less. I do not interpret this level estimate as evidence that the market simply undervalues privacy capital. Regulation removes economically valuable options at the same time that it induces firms to build the new stock. A collection-intensive firm may lose flexibility over which customer information it gathers. A sharing-intensive firm may lose monetization opportunities with advertisers, vendors, or data intermediaries. An initial valuation discount is therefore economically natural. It can represent the market pricing the visible restriction

imposed on the old data business.

The recognition hypothesis concerns what happens after this initial response. If the regulatory wave only destroys data opportunities, the negative valuation effect should remain. If the wave also creates productive capital whose value becomes apparent with use, the loading of market value on privacy capital should subsequently rise. In other words, the valuation gap between good-privacy firms and bad privacy firms should shrink. Thus, the clean recognition object is not the negative level coefficient itself but the change in the valuation loading over the post-wave period.

## B. Where the discount bites

The mechanism evidence of Section VI gives the initial discount an additional evaluation direction. If the negative valuation response reflects the restriction of data opportunities, it should become more negative where the business depends most directly on broad collection and sharing. If privacy capital also creates productive and trust benefits, the response can instead attenuate or reverse where those benefits are comparatively important. Characteristics unrelated to the firm's role with data should not systematically organize the valuation effect.

Panels B through D of Table 9 apply exactly the moderator architecture used for the operating mechanism. For continuous moderators, the interaction reports how the dose response changes per standard deviation of the moderator. For the consumer-facing indicator, the level coefficient is the response among non-consumer firms and the interaction is the incremental consumer response.

The Total-Q results line up closely with the mechanism. A one-SD increase in third-party-sharing intensity makes the valuation response approximately 0.24 Q units more negative. First-party collection makes it approximately 0.25 units more negative. These are precisely the firms for which privacy regulation most directly hurts in terms of data use. The market therefore initially places the largest discount where the opportunity cost of privacy is economically largest.

Consumer-facing firms provide the more revealing case. Among non-consumer firms, the Total-Q dose effect is approximately 0.29. The consumer-facing interaction is +0.53, implying an approximately +0.24 valuation response for consumer-facing firms. and the sign reverses. This is not inconsistent with consumer firms being data intensive. Their data dependence is different. Much of the customer information used to complete transactions and fulfill contractual obligations remains operationally necessary, so privacy regulation removes fewer of the productive uses on which the core business depends. At the same time, these are firms for which consumer trust can improve the continuing data relationship. The productive and trust

side of privacy capital can consequently outweigh the lost-opportunity side.

This distinction is central to the economic interpretation. Data dependence is not one scalar exposure. Firms that monetize the breadth of collection and sharing face larger restriction costs, whereas firms whose relationship with customers depends on continuing operational data can benefit disproportionately from better governance of that relationship. The opposite valuation signs therefore provide a sharper mechanism test beyond generic data intensity claims.

The placebo structure strengthens this reading. Under Total Q, neither technology membership nor R&D intensity significantly changes the valuation effect, and neither do concentration, pricing power, conventional intangible intensity, or firm size. Some of these characteristics matter elsewhere for firms' operations or financing. For valuation, on the other hand, the effect is specifically organized by what the firm does with personal data.

The comparison between Tobin's Q and Total Q is also informative. Raw Tobin's Q omits much conventional intangible capital from its denominator and loads strongly on R&D intensity, while the collection and sharing interactions are weak. Once estimated intangible capital is incorporated in Total Q, the R&D interaction disappears and the predicted data-role interactions emerge. The consumer-facing premium is the exception that survives under both measures. I treat Total Q as the cleaner outcome for the cross-sectional mechanism test, while Tobin's Q remains useful for tracing the within-firm re-rating dynamics.

[Insert Table 9 here]

## C. The re-rating

Table 10 contains the core recognition result. The full-sample coefficients on Qs are aggregate results, yet what is more revealing is the dynamics of the effect. I interact privacy capital and its instrument with fiscal-year indicators and estimate how the valuation loading changes relative to the 2016–2017 reference period. The loading rises steadily after the regulatory wave begins. Relative to the pre-wave reference, the coefficient is +0.13 in 2018, +0.18 in 2019, +0.33 in 2020, and +0.50 by 2021. It remains elevated at approximately +0.41 and +0.40 in 2022 and 2023. Privacy capital becomes progressively more highly valued and retains much of that re-rating after the initial adjustment period.

[Insert Table 10 here]

Figure 5 makes the same point visually. The re-rating accelerates around 2020–2021 and does not subsequently return to its earlier loading. Importantly, the pattern also appears within technology firms. The within-technology path is weak or negative before 2020 and jumps in 2020–2021, ruling out a simple rotation by investors toward the technology sector. The market

is revaluing privacy capital within a sector in which data is already economically important. I interpret this movement as the market capitalizing a benefit that is initially difficult to observe. Regulation's restriction of collection and sharing is visible immediately and is reflected in the initial discount. The productive gain documented in previous sections has no separately booked asset value and becomes observable only as the governed operation performs. The market therefore prices the harm to the old data opportunity set first and learns the value of the new governance stock more slowly.

[Insert Figure 5 here]

This is also where the corporate-finance and asset-pricing evidence meet. Section VI establishes that the real payoff is organized by firms' roles with data. Table 9 shows that the initial valuation response is organized by those same roles, but with signs determined by whether restriction costs or productive benefits dominate. Table 10 then shows the overall loading rising with time. The recognition result is the market-side realization of the same capital formation.

### D. Recognition arrives through ownership and attention

If the changing valuation loading reflects recognition, the process should leave some trace in investor participation and in how market participants think about data. The evidence is consistent with both. Institutional ownership rises with privacy capital only in the later part of the sample. Under the regulation-driven privacy capital accumulation, a one-SD increase in privacy capital raises institutional ownership by roughly 2.5 percentage points after 2020, while the pre-2020 estimate is statistically flat. The response appears in ownership without a corresponding change in the conventional sell-side forecasts. Analyst coverage, forecast dispersion, and forecast error do not move significantly with instrumented privacy capital. This is more consistent with recognition of a diffuse, long-lived intangible than with a discrete revision in near-term earnings expectations. Investors become more willing to own privacy-rich firms without a simultaneous jump in analyst consensus cash-flow forecasts. The textual evidence provides broader market context. Figure 6 shows that sell-side discussion increasingly refers to data as an asset, attention to AI and machine learning rises strongly later in the sample, and management's privacy language becomes more likely to frame privacy as an opportunity than only as a burden. Institutional ownership of high-privacy firms also rises relative to high-privacy firms. These attention and sentiment series are taken as descriptive evidence only here. I do not attach any causal labels to them. I take them as recognition clocks only.

[Insert Figure 6 here]

## E. Return evidence

The return evidence provides a different view of the same transition. I form portfolios each June using privacy-capital information available with at least a six-month reporting lag. The portfolio is long firms with high privacy capital and short firms with low privacy capital. The sorting variable is in levels. Firms that build the most and gain the most are considered bad-privacy firms here.

Before 2020, the long-short portfolio[24] earns economically large abnormal returns. The alpha is approximately 13.3% annually under the CAPM, 10.0% under the Fama-French three-factor model, and 11.5% under the five-factor model. After 2020, the corresponding alphas are 1.8%, 0.5%, and 0.4%, none statistically distinguishable from zero. Figure 7 shows the same timing at higher frequency. The premium builds primarily through 2018–2019 and is essentially absent thereafter. Table 11 reports the numbers.

I do not interpret this portfolio as a new permanent privacy factor. Its composition argues against that reading. The pre-2020 premium is an equal-weighted phenomenon. The three-factor alpha falls from 10.0% to 1.5% when the smallest 20 percent of firms are removed and is negative under value weighting, as seen from Table 12. The return effect is therefore concentrated in smaller stocks, where information diffusion and investor recognition are naturally slower. That concentration does not mean that the underlying valuation result is a microcap phenomenon. Panel B of Table 12 removes the smallest quintile of firms from the panel regressions. The level discount remains significant and the 2021 re-rating remains strongly significant, falling from +0.50 to +0.365 but becoming more precisely estimated. The identified re-rating therefore survives among larger firms even though the realized long-short premium is concentrated at the thinly followed end of the market.

The returns corroborate the re-rating. A portfolio long firms rich in privacy capital and short firms poor in it earns a large premium before 2020, on the order of ten to thirteen percent a year across the factor models, and essentially nothing after (Table 11). That pattern, a premium paid while the asset is being recognized and gone once it has been, is the signature of a one-time recognition correction. But the effects are driven mainly by small stocks. As Table 12 shows, it is an equal-weighted phenomenon, since it does not survive dropping the smallest fifth of firms and it reverses sign when the portfolio is value-weighted. I therefore do not read it as a tradable anomaly, and the paper's recognition claim does not rest on it.

This distinction provides a useful hierarchy of evidence. The portfolio evidence corroborates the recognition pattern. The panel estimates identify a broad change in how privacy capital is valued. The return spread is the noisier price realization of that learning episode, and it appears

[24]This is not to state I have identified a new factor to the factor zoo. I actually argue against privacy being a systematic risk factor.

most clearly in the corner of the market where recognition should be slowest.

[Insert Table 11 here]

[Insert Figure 7 here]

[Insert Table 12 here]

### F. Why a standing systematic-risk interpretation does not fit

A final question is whether the valuation and return evidence reflects gradual recognition or instead compensation for systematic privacy risk. I do not claim that privacy carries no risk. Breaches, regulatory penalties, litigation, and other tail outcomes are real. The narrower question is whether routine privacy capital represents a standing priced systematic exposure whose compensation explains the observed return premium.

Several features weigh against that interpretation. First, the return premium is transitory. It is not static. High-privacy firms earn a positive spread during the early part of the regulatory wave, but the post-2020 spread is economically close to zero. Second, the premium attaches to firms with better privacy practices. A natural bad-privacy-risk story would place greater adverse exposure on firms with weaker governance. Third, the return premium is concentrated in equal-weighted, lightly followed stocks and disappears under value weighting, whereas the panel re-rating itself survives removal of the smallest firms.

## VIII. Alternative Explanations and Robustness

Corporate and asset-pricing results so far converge on a coherent story under the data productivity mechanism. But coherence does not by itself establish this channel as discriminative. There could be many alternative explanations. An alternative explanation, though, has to do more than correlate with privacy practice. It must also line up with predetermined distance to the floor, move with the regulatory wave, reproduce the operating response, and explain why the heterogeneity concentrates in firms' roles with data.

I consider four classes of concerns. I first ask whether the results simply reflect the composition of the treated firms. Characteristics include technology, scale, market power, or conventional intangible intensity. I then address the more serious concern that firms beginning far below the floor are observably different and may have been on different underlying trajectories. Third, I separate the regulatory wave from contemporaneous aggregate shocks, particularly the pandemic. Finally, I show statistical robustness. I examine whether the estimates depend

materially on the construction of the treatment dose, the tails of the privacy distribution, or outcome transformations.

## A. Size, Tech, and Market Power

The first possibility is that privacy capital simply proxies for a type of firm that was already becoming more productive during the sample period. Technology firms are the most natural candidate. They are data intensive, invest heavily in intangible assets, and account for an important part of the secular growth in productivity and valuations. Table 13 does not support that interpretation. Excluding technology firms leaves the main operating and valuation effects intact. Asset turnover, ROA, and capital expenditure remain positive, while both Tobin's Q and Total Q retain negative coefficients. The turnover response is, if anything, larger outside technology. Within technology itself, ROA rises and the later valuation re-rating remains visible. Thus, the results do not arise because the privacy measure mechanically selects technology firms or because technology firms were generally becoming more productive.

The consumer-facing split produces a different pattern, and that difference is informative. Removing consumer-facing firms sharply attenuates the turnover response while leaving the broader profitability effect and deepening the valuation discount. Within consumer-facing firms, turnover and investment respond positively and the valuation discount disappears. This is exactly the distinction drawn in previous sections that effects move with the firm's economic role with data, but survive removal of the technology label.

Firm size gives a second composition test. The operating results are not produced by large incumbents. Among firms below the median pre-wave size, asset turnover rises by 0.155 and ROA by 0.554, even though large firms are entirely absent from those regressions. Among larger firms, the valuation discount remains precisely estimated on both Q measures.

The same conclusion holds for market power and conventional intangible capital. Industry concentration and the Lerner index do not systematically moderate the operating or Q effects, while collection, sharing, and consumer-facing status do. Conventional intangible intensity and R&D likewise fail to reproduce the Total-Q heterogeneity associated with firms' data roles. The privacy dose itself also contains little conventional intangible-capital variation. Research, advertising, and reported intangible intensity explain only about 2% of its within-firm variation and 5% across firms. Adding those variables and their interactions with the wave leaves the valuation effects intact.

Taken together, these tests remove a broad family of generic firm-type explanations. Privacy capital is not simply technology, firm size, market dominance, or conventional intangible investment under another name. Whatever generates the results must be considerably more

specific to the privacy build itself.

[Insert Table 13 here]

## B. Selection into treatment and mean-reversion

he more serious concern comes from the design. Distance below the privacy floor is predetermined, but it is not randomly distributed across firms. Figure 8 shows that firms beginning farther below the floor are smaller and have lower asset turnover and ROA before the wave. Book-to-market, by contrast, is approximately flat across the distance terciles. Moreover, the pre-period profitability gradient is largely absorbed by size in a multivariate regression. The clearest selection pattern is therefore that the regulatory wave disproportionately compels smaller and operationally weaker firms.

[Insert Figure 8 here]

This raises a direct concern. Perhaps smaller or less efficient firms simply had more room to improve. Table 14 puts that possibility into the regression. I construct longer-run pre-wave measures of firm size, ROA, and asset turnover using 2010–2015 data and allow each to generate its own subsequent trend. The three trends also enter jointly, so generic convergence in firm scale, profitability, and efficiency is allowed to absorb the treatment effect if that is what the distance-to-floor dose represents. The central effects survive. With all three characteristic trends included, the reduced-form coefficients remain positive for asset turnover, ROA, and capital expenditure and negative for both valuation measures. The estimates attenuate under this demanding specification, as expected, but retain their signs and substantial magnitudes. When only the size trend is raced against the dose, the asset-turnover effect actually strengthens.

The trend coefficients themselves also run against a simple small-firm catch-up explanation. Generic size-related valuation drift is positive rather than negative, significantly so for Total Q. Thus, the negative privacy valuation effect is not mechanically produced by small firms converging upward toward larger firms. More importantly, the mechanism architecture survives the same horse race. Collection and sharing continue to predict stronger margin responses and more negative Total-Q effects. Consumer-facing firms continue to exhibit a stronger turnover response and positive valuation differential. Technology remains largely null. The one earlier interaction that the characteristic trends absorb completely is the size gradient in ROA—precisely the cell that a generic convergence story would be most likely to produce.

The selection tests therefore produce a useful decomposition. Baseline differences matter for some magnitudes, and in particular explain the earlier size heterogeneity in profitability. They do not explain the core operating effects or the cross-sectional structure tied to firms' data

roles. Generic mean reversion and small-firm catch-up are therefore difficult to reconcile with the full pattern of evidence.

[Insert Table 14 here]

### C. Not the pandemic or macroeconomic shocks

The privacy wave overlaps with an unusually turbulent macroeconomic period. The principal concern is the COVID-19 pandemic, especially because the valuation re-rating becomes visible around 2020. The common pandemic shock cannot drive the estimates because fiscal-year fixed effects absorb the aggregate crash and recovery. The relevant threat is differential exposure. Firms farther below the privacy floor might have been hit harder in 2020 and subsequently rebounded more strongly. Three tests weigh against that interpretation.

First, important parts of the corporate footprint predate the pandemic. Restricting the sample to 2016–2019 still produces positive cash, investment, and ROA responses together with a negative Total-Q effect. The shorter window contains much less treatment variation, so attenuation is mechanically expected, but the results are already present before COVID. Second, removing the pandemic years does not eliminate the results. Excluding 2020 and 2021 leaves cash, the current ratio, capital expenditure, ROA, and Tobin's Q statistically significant. Asset turnover attenuates in this restricted sample, so I do not claim that every coefficient is invariant to the window. The broader operating and valuation footprint clearly does not depend on the pandemic years.

[Insert Table 15 here]

Third, Table 16 directly conditions on differential pandemic exposure. Predetermined distance to the privacy floor is nearly orthogonal to three firm-specific measures of the 2020 crash. The February–March return, maximum drawdown, and largest idiosyncratic crash have a correlation of only 0.03, 0.10, and 0.13. Adding these exposures interacted with year, separately and jointly, does not attenuate the late-window valuation response.

[Insert Table 16 here]

I next consider regulatory and industry timing. Headquarters-state-by-year fixed effects absorb every shock common to firms headquartered in the same state and year, including the staggered arrival of state privacy laws. The turnover, ROA, capital-expenditure, and valuation estimates are essentially unchanged. The results are therefore not simply the consequence of particular states moving earlier than others. Industry-by-year effects require additional

interpretation. ROA and Tobin's Q survive, and the investment response remains positive, while asset turnover is absorbed and Total Q becomes less precise. This attenuation is economically unsurprising because the mechanism itself is partly organized by business roles, most notably consumer-facing status that are strongly industry structured. Industry-by-year effects therefore remove economically relevant treatment variation together with potential industry confounds. The evidence supports the narrower conclusion that profitability and raw valuation effects are visible within industry-year cells, whereas part of the turnover response operates through cross-industry differences in firms' data roles.

### D. Robustness

The remaining checks concern whether the results are artifacts of how the privacy dose or outcomes are constructed.

The first issue is the location of the compliance floor. Robustness tests replace the baseline end-state floor with an uncensored distance measure and with economy-wide floors placed at the 90th, 75th, and 60th percentiles of the pre-wave distribution. It also replaces the internally constructed regulatory-pressure path with an external Refinitiv measure of privacy-policy diffusion. The resulting coefficients remain stable, and first-stage F-statistics range from 34 to 66. The economic conclusions therefore do not depend on the precise location of the floor or on the particular time path used to represent the regulatory wave.

They also do not depend on extreme firms in the privacy-practice distribution. Removing both the top and bottom 5% of pre-wave privacy practice, including the mass of zero-practice firms and firms nearest the privacy frontier. This discards roughly one fifth of the identifying variation. Asset turnover, ROA, and capital expenditure retain their signs and statistical support; unconditional margins remain flat; and the valuation discount becomes larger. Similar spirits carry on to outcome tail robustness, and results are still relatively stable.

Finally, the results do not appear to reflect mechanical verbosity in privacy policies. Alternative content-based transforms, policy-length controls, length-purged measures, and within-policy shares reproduce the main patterns. Carrying the last observed policy forward into uncaptured years also preserves the signs, though with noisier estimates.

### E. Complementary channels outside the data-productivity mechanism

The tests above narrow the set of plausible confounds considerably, but they do not imply that data productivity is the only economic benefit generated by privacy capital. Several additional channels could coexist with the data productivity mechanism without being part of it.

One possibility is a pure demand or reputation channel. Customers may prefer firms that

make stronger privacy commitments even when that preference does not alter the quantity or quality of information supplied to the firm. Privacy can therefore function as a product attribute or reputational asset, increasing willingness to transact or customer retention independently of any improvement in the firm's use of data. The consumer-facing valuation premium is compatible with such a channel, although the current data do not separately identify it from the operating mechanism.

A second possibility is stakeholder preference more broadly. Employees, suppliers, business partners, or institutional customers may value stronger privacy governance for contractual, reputational, or ethical reasons. Such preferences could improve contracting conditions, employee retention, or commercial relationships without operating through the productivity of the firm's data. The current firm-level outcomes are not designed to distinguish these margins.

These possibilities do not contaminate the data-productivity evidence because none is needed to generate its discriminating pattern. The data mechanism is identified from the fact that the real-side payoff strengthens systematically with firms' collection, sharing, and consumer-facing roles and does not exhibit a comparable structure along data-free characteristics. The channels in this subsection are different. They may provide additional value on top of that mechanism. The present evidence therefore supports a data productivity channel while leaving open the possibility that privacy capital also creates reputational or stakeholder-relevant benefits.

## IX. Conclusion

This paper asks whether privacy is a cost, a risk, or a capital for the firm, and let a regulatory wave answer. Between 2018 and 2023, rising privacy floors compel the least-prepared firms to build governance over the data they hold. What they build behaves like capital. Relative to otherwise-similar firms, the compelled builders turn their assets over faster and earn more per dollar of assets, with no margin falling anywhere. The effects concentrate where the firm's data business is. The market prices the asset with a lag, a discount first that re-rates as recognition arrives. The short-lived privacy premium also disappears with the re-rating. The mechanism is consistent with a coherent story that privacy capital enables more efficient and productive data use. Existing finance research asks whether bad privacy is punished. The answer here is that good privacy is rewarded.

There are certain limitations in this paper. The estimates identify firm-side effects, not social welfare. Whether consumers gain from the privacy practices induced by regulation belongs to a separate welfare ledger. They are also local to firms whose behavior is moved by the regulatory

floor, not economy-wide average effects. The sample consists of U.S. listed firms, whereas compliance burdens and entry effects may be substantially more important below the listing threshold. The design identifies the consequences of the gradual regulatory wave observed in the data and cannot establish that a substantially steeper or one-shot mandate would generate the same accumulation. Nor does it establish that the mechanism travels unchanged across legal regimes. The setting combines credible enforcement, relatively large firms, deep capital markets, and business models for which data are economically important. These boundaries limit the range of the conclusions, but they also identify the settings in which the evidence is most informative.

Nevertheless, findings in this paper still carry policy implications in many dimensions.

First, regulatory cost-benefit analysis should recognize induced capital formation. Privacy regulation imposes compliance costs, but compliance expenditure is not necessarily pure deadweight loss. When regulation forces firms to reorganize how data are collected, protected, governed, and deployed, part of that expenditure can become investment in a productive organizational asset. Among firms for which the regulatory floor binds, the subsequent gains in operating efficiency and profitability per unit of assets are economically substantial. Evaluating the regulatory wave from compliance expenditures alone therefore records the input while omitting part of the output. More generally, the relevant policy question is not only what compliance costs, but also what productive capability firms build.

Second, the evidence favors minimum standards that induce catch-up while allowing the adjustment path to reflect firm capacity. A standard concern with GDPR-style regulation is that fixed compliance costs disproportionately burden smaller firms and protect incumbents. Within the population studied here, however, the strongest regulatory exposure falls on smaller firms with initially weaker privacy practices, and these firms subsequently realize the largest gains. The privacy floor therefore operates, in part, as compelled catch-up investment. This suggests a distinction between the level of the standard and the speed of adjustment toward it. Regulators need not necessarily choose between a common governance floor and sensitivity to firm size. A common long-run standard can be paired with longer implementation periods, staged compliance requirements, temporary safe harbors, or other forms of transition relief for firms facing larger adjustment costs. Such a design preserves the capital-building force of the floor while reducing the risk that short-run financing or organizational constraints prevent smaller firms from reaching it. More generally, regulation can be differentiated along the transition path rather than by permanently exempting firms with the largest capability gap.

Third, policy evaluation should distinguish value creation from value recognition. The privacy capability documented here is largely intangible and does not appear directly on firms' balance sheets. Consistent with that opacity, the market initially discounts the firms most

affected by the regulatory restriction and re-rates them only as recognition diffuses. An initial negative market response is therefore not a sufficient statistic for whether a governance mandate ultimately destroys firm value. For regulations that induce intangible investment, evaluation should extend beyond announcement-window returns and immediate compliance expenditures to subsequent operating outcomes, organizational adjustment, and medium-run valuation. The results also motivate greater attention to the observability of governance capital. If markets have difficulty distinguishing costly compliance from productive capability building, standardized disclosure of data-governance investments or capabilities may help shorten the recognition lag, although whether such disclosure is itself welfare improving remains an open question.

Fourth, the framework provides a forward-looking way to evaluate AI governance. Artificial intelligence potentially raises the productivity of exactly the organizational capability that privacy regulation forces firms to build. Effective AI deployment depends not only on access to data but also on firms' ability to classify, govern, secure, document, and use those data in a responsible way. New AI-governance requirements will therefore generate visible compliance costs, but they may also accelerate investment in an increasingly productive intangible asset. The privacy evidence suggests that these two margins should be evaluated jointly. Where governed data become more valuable as AI technologies diffuse, regulation that raises the governance floor may simultaneously impose a current adjustment cost and expand firms' future productive capacity. The appropriate empirical question for emerging AI regulation is whether the induced organizational capabilities subsequently raise the productivity of data and complementary technologies.

## References

Abis, Simona, and Laura Veldkamp, 2024, The changing economics of knowledge production, *Review of Financial Studies* 37, 89–118.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman, 2016, The economics of privacy, *Journal of Economic Literature* 54, 442–492.

Aridor, Guy, Yeon-Koo Che, Brett Hollenbeck, Maximilian Kaiser, and Daniel McCarthy, 2025, Evaluating the impact of privacy regulation on e-commerce firms: Evidence from apple's app tracking transparency, *Management Science*.

Aridor, Guy, Yeon-Koo Che, and Tobias Salz, 2023, The effect of privacy regulation on the data industry: Empirical evidence from GDPR, *RAND Journal of Economics* 54, 695–730.

Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp, 2018, Big data in finance and the growth of large firms, *Journal of Monetary Economics* 97, 71–87.

Belo, Frederico, Xiaoji Lin, and Maria Ana Vitorino, 2014, Brand capital and firm value, *Review of Economic Dynamics* 17, 150–179.

Bolton, Patrick, and Marcin Kacperczyk, 2021, Do investors care about carbon risk?, *Journal of Financial Economics* 142, 517–549.

Campbell, James, Avi Goldfarb, and Catherine Tucker, 2015, Privacy regulation and market structure, *Journal of Economics and Management Strategy* 24, 47–73.

Cheng, Daxuan, Ding Ding, Yin Liao, and Zheyao Pan, 2025, Data privacy risk, Working Paper.

Corrado, Carol, Charles Hulten, and Daniel Sichel, 2009, Intangible capital and U.S. economic growth, *Review of Income and Wealth* 55, 661–685.

Corwin, Shane A., and Paul Schultz, 2012, A simple way to estimate bid-ask spreads from daily high and low prices, *Journal of Finance* 67, 719–760.

Crouzet, Nicolas, and Janice C. Eberly, 2019, Understanding weak capital investment: The role of market concentration and intangibles, NBER Working Paper 25869, National Bureau of Economic Research.

de Chaisemartin, Clément, and Xavier D'Haultfœuille, 2018, Fuzzy differences-in-differences, *Review of Economic Studies* 85, 999–1028.

Dell, Melissa, 2025, Deep learning for economists, *Journal of Economic Literature* 63, 5–58.

Demirci, Ozge, Ayelet Israeli, and Eva Ascarza, 2025, In privacy we trust: The effect of privacy regulations on data sharing behavior, Harvard Business School Working Paper 26-001, Harvard Business School.

Doerr, Sebastian, Leonardo Gambacorta, Luigi Guiso, and Marina Sanchez del Villar, 2026, Privacy regulation and fintech lending, *Management Science*.

Dubé, Jean-Pierre, John G Lynch, Dirk Bergemann, Mert Demirer, Avi Goldfarb, Garrett Johnson, Anja Lambrecht, Tesary Lin, Anna Tuchman, and Catherine Tucker, 2025, Frontiers: The intended and unintended consequences of privacy regulation for consumer marketing, *Marketing Science* 44, 975–984.

Eisfeldt, Andrea L., and Dimitris Papanikolaou, 2013, Organization capital and the cross-section of expected returns, *Journal of Finance* 68, 1365–1406.

Eisfeldt, Andrea L., Edward T. Kim, and Dimitris Papanikolaou, 2022, Intangible value, *Critical Finance Review* 11, 299–332.

Farboodi, Maryam, Dhruv Singal, Laura Veldkamp, and Venky Venkateswaran, 2025, Valuing financial data, *Review of Financial Studies* 38, 938–980.

Farboodi, Maryam, and Laura Veldkamp, 2026, A model of the data economy, *Review of Economic Studies*.

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

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, 2019, Text as data, *Journal of Economic Literature* 57, 535–574.

Goldberg, Samuel G, Garrett A Johnson, and Scott K Shriver, 2024, Regulating privacy online: An economic evaluation of the gdpr, *American Economic Journal: Economic Policy* 16, 325–358.

Goldfarb, Avi, and Verina F Que, 2023, The economics of digital privacy, *Annual Review of Economics* 15, 267–286.

Goldfarb, Avi, and Catherine Tucker, 2019, Digital economics, *Journal of Economic Literature* 57, 3–43.

Goldfarb, Avi, and Catherine E. Tucker, 2011, Privacy regulation and online advertising, *Management Science* 57, 57–71.

Gourio, François, and Leena Rudanko, 2014, Customer capital, *Review of Economic Studies* 81, 1102–1136.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun, 2019, Firm-level political risk: Measurement and effects, *Quarterly Journal of Economics* 134, 2135–2202.

Hayashi, Fumio, 1982, Tobin's marginal q and average q: A neoclassical interpretation, *Econometrica* 50, 213–224.

Hoberg, Gerard, and Gordon Phillips, 2016, Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124, 1423–1465.

Jamilov, Rustam, Hélène Rey, and Ahmed Tahoun, 2021, The anatomy of cyber risk, NBER Working Paper 28906, National Bureau of Economic Research, Revised December 2025.

Jia, Jian, Ginger Zhe Jin, and Liad Wagman, 2021, The short-run effects of the general data protection regulation on technology venture investment, *Marketing Science* 40, 661–684.

Johnson, Garrett A, 2024, Economic research on privacy regulation: Lessons from the GDPR and beyond, in Avi Goldfarb, and Catherine E. Tucker, eds., *The Economics of Privacy*, 97–126 (University of Chicago Press, Chicago).

Johnson, Garrett A, Scott K Shriver, and Samuel G Goldberg, 2023, Privacy and market concentration: Intended and unintended consequences of the GDPR, *Management Science* 69, 5695–5721.

Jones, Charles I., and Christopher Tonetti, 2020, Nonrivalry and the economics of data, *American Economic Review* 110, 2819–2858.

Kamiya, Shinichi, Jun-Koo Kang, Jungmin Kim, Andreas Milidonis, and René M. Stulz, 2021, Risk management, firm reputation, and the impact of successful cyberattacks on target firms, *Journal of Financial Economics* 139, 719–749.

Lefrere, Vincent, Logan Warberg, Cristobal Cheyre, Veronica Marotta, and Alessandro Acquisti, 2026, Does privacy regulation harm content providers? a longitudinal analysis of the impact of the GDPR, *Management Science* 72, 1727–1747.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *Journal of Finance* 66, 35–65.

Merton, Robert C., 1987, A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42, 483–510.

Miller, Amalia R., and Catherine Tucker, 2009, Privacy protection and technology diffusion: The case of electronic medical records, *Management Science* 55, 1077–1093.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor, 2022, Dissecting green returns, *Journal of Financial Economics* 146, 403–424.

Peters, Ryan H., and Lucian A. Taylor, 2017, Intangible capital and the investment-q relation, *Journal of Financial Economics* 123, 251–272.

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

Ramadorai, Tarun, Antoine Uettwiller, and Ansgar Walther, 2025, Privacy policies and consumer data extraction: evidence from US firms, *Review of Finance* 29, 1337–1367.

Shumway, Tyler, 1997, The delisting bias in CRSP data, *Journal of Finance* 52, 327–340.

Tucker, Catherine E., 2014, Social networks, personalized advertising, and privacy controls, *Journal of Marketing Research* 51, 546–562.

Veldkamp, Laura, 2023, Valuing data as an asset, *Review of Finance* 27, 1545–1562.

Wilson, Shomir, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N Cameron Russell, et al., 2016, The creation and analysis of a website privacy policy corpus, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1330–1340.

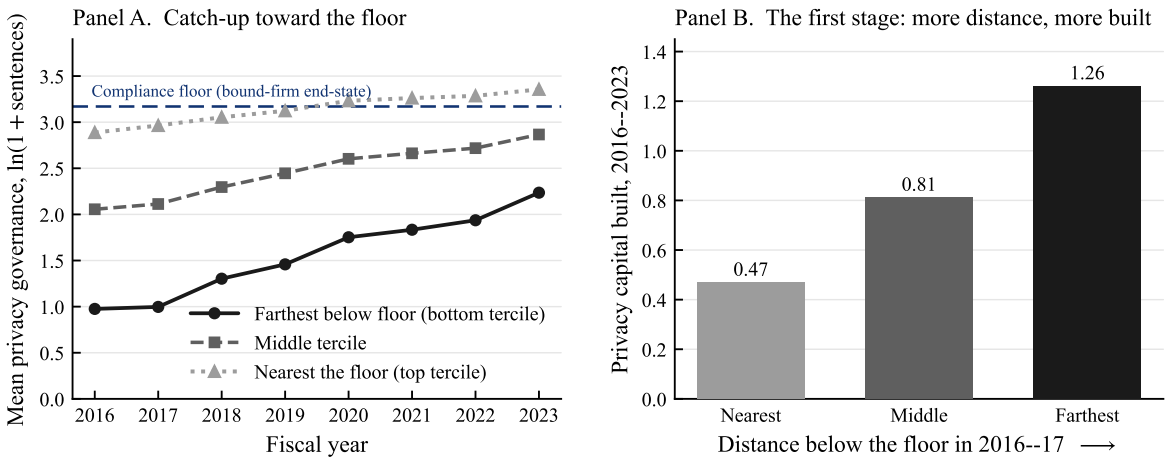



**Figure 1.** Compliance catch-up toward the floor, by distance-to-floor tercile

*Notes.* The figure shows the first stage as catch-up toward the compliance floor. Firms are sorted into terciles by how far their 2016–2017 (pre-period) privacy governance practice began below the floor, the bound-firm end-state norm, which is the 2022–2023 governance level of the firms compelled by the GDPR and the CCPA/CPRA, equal to 3.17 in ln(1 + governance sentences) and drawn as the dashed line. *Panel A* plots each tercile's mean practice by fiscal year; the firms that began farthest below the floor (bottom tercile) climb toward it fastest, while the firms that began nearest it (top tercile) rise least, so the terciles converge. *Panel B* shows the first stage directly, the privacy capital each group built between 2016 and 2023 rising monotonically with distance to the floor (nearest 0.47, middle 0.81, farthest 1.26 log points). The sample is U.S. operating firms excluding financial and utility firms, fiscal years 2016–2023.

**Table 1.** Descriptive statistics

*Notes.* The table reports summary statistics for the privacy-capital measure, the outcome variables, and the firm characteristics used in the analysis. The unit of observation is the firm-fiscal-year. The sample is U.S. operating firms, excluding financial firms (SIC 6000–6999) and utilities (SIC 4900–4949), over fiscal years 2016–2023. Privacy governance is the count of policy sentences classified into the governance categories (data retention, do-not-track, user access and deletion, and user choice and control), measured from the firm's most recent privacy policy in force at the fiscal year-end; ln(1 + governance) is the headline measure of privacy capital. All ratio variables are winsorized at the 1st and 99th percentiles and sentence counts at the 99th, matching the regressions. Variable definitions are collected in Appendix B.

| | N | Mean | SD | P25 | Median | P75 |
|---|---|---|---|---|---|---|
| *Privacy-capital measure* | | | | | | |
| Governance sentences (count) | 9,917 | 14 | 13 | 4 | 11 | 21 |
| ln(1 + governance) | 9,917 | 2.33 | 0.99 | 1.61 | 2.48 | 3.09 |
| Substantive sentences (count) | 9,917 | 61 | 47 | 26 | 50 | 86 |
| Total policy sentences (count) | 9,917 | 89 | 63 | 44 | 75 | 120 |
| Policies per firm-year | 9,917 | 3.52 | 3.88 | 1.00 | 2.00 | 4.00 |
| *Financing and liquidity* | | | | | | |
| Book leverage | 9,910 | 0.32 | 0.32 | 0.10 | 0.27 | 0.44 |
| Market leverage | 9,716 | 0.24 | 0.23 | 0.04 | 0.17 | 0.36 |
| Cash / assets | 9,910 | 0.24 | 0.26 | 0.05 | 0.13 | 0.35 |
| Current ratio | 9,696 | 2.81 | 3.18 | 1.20 | 1.82 | 3.04 |
| *Efficiency, investment, profitability* | | | | | | |
| Asset turnover | 9,909 | 0.82 | 0.68 | 0.36 | 0.66 | 1.08 |
| Capex / assets | 9,897 | 0.03 | 0.04 | 0.01 | 0.02 | 0.04 |
| Return on assets | 9,910 | -0.15 | 0.54 | -0.11 | 0.02 | 0.06 |
| Operating margin | 9,463 | -1.43 | 6.98 | -0.07 | 0.06 | 0.14 |
| *Valuation* | | | | | | |
| Tobin's $Q$ | 9,700 | 2.88 | 3.76 | 1.22 | 1.77 | 2.97 |
| Total $Q$ | 9,532 | 1.71 | 2.54 | 0.42 | 0.89 | 1.85 |
| *Firm characteristics* | | | | | | |
| Total assets ($ millions) | 9,911 | 8,637 | 20,701 | 207 | 1,221 | 5,392 |
| ln(assets) | 9,910 | 6.89 | 2.47 | 5.33 | 7.11 | 8.59 |
| Book-to-market | 9,701 | 0.40 | 0.72 | 0.14 | 0.32 | 0.62 |
| Intangibles / assets | 9,849 | 0.22 | 0.22 | 0.01 | 0.15 | 0.38 |
| R&D / assets | 9,910 | 0.09 | 0.18 | 0.00 | 0.01 | 0.10 |

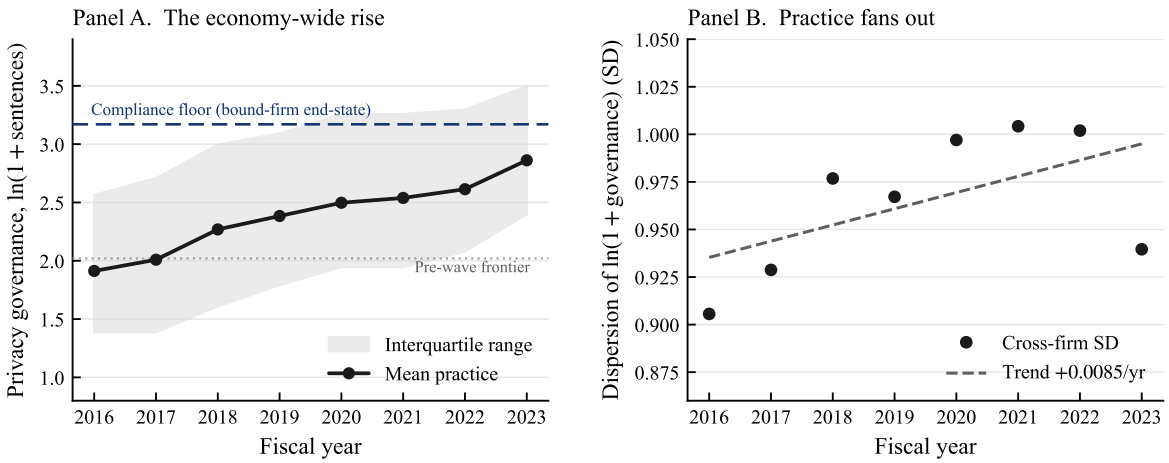



**Figure 2.** The privacy-capital measure in the data

*Notes.* The figure plots the privacy-governance measure, ln(1 + governance sentences), over fiscal years 2016–2023 for the estimation sample (U.S. operating firms excluding financial and utility firms). *Panel A* plots mean practice by year, with the shaded band the interquartile range; the dashed line is the compliance floor, the bound-firm 2022–2023 governance norm (3.17), and the dotted line is the bound-firm 2016–2017 pre-wave norm (2.02). Mean practice rises from 1.91 to 2.86 over the sample, toward the floor. *Panel B* plots the cross-firm standard deviation of the measure by year; the dashed line is its linear trend, +0.0085 per year, so the cross-section of practice fans out over the wave rather than compressing to a common level.

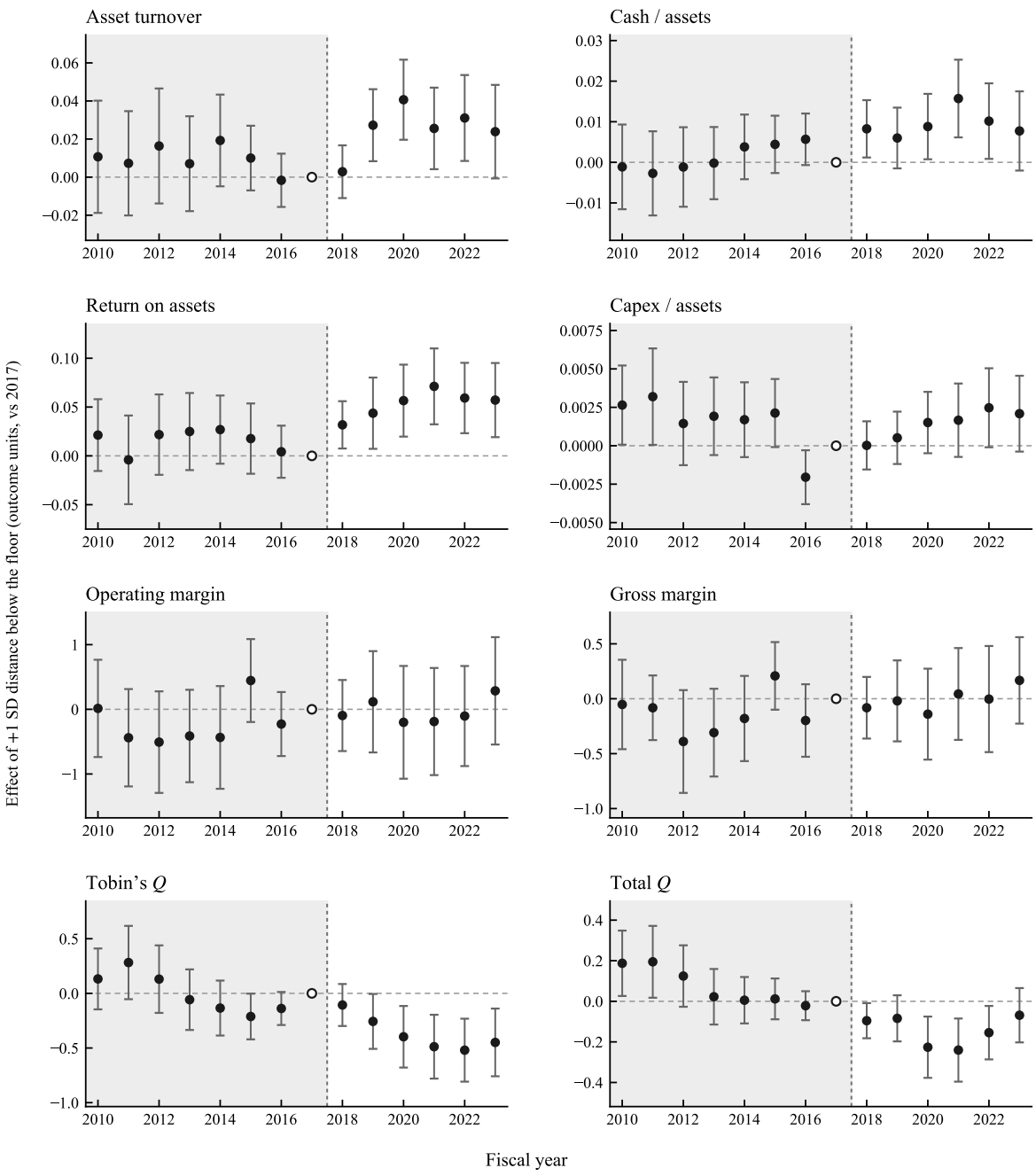



**Figure 3.** Parallel-trends event study in the distance-to-floor dose

*Notes.* Each point is the coefficient $\beta_k$ from regressing the outcome on the distance-to-floor dose interacted with fiscal-year dummies, with firm and fiscal-year fixed effects and controls (log assets, book-to-market), the dose scaled to one standard deviation and 2017 normalized to zero; bars are 95 percent confidence intervals from firm-clustered standard errors, the shaded region is the pre-period, and the dashed line marks GDPR enforcement in 2018. Cash, asset turnover, and return on assets sit on zero before the wave and move only after it; the margin panels are wide and individually indistinguishable from zero before the wave; capex shows small positive cells early that fade before the wave, the caveat under which its post-period path is read. Tobin's $Q$ shows a pre-period drift that Figure 4 traces to the extreme-valuation tail and resolves; total $Q$ (Peters and Taylor 2017) settles onto zero by 2013 and passes the joint pre-period test. U.S. operating firms excluding financial and utility firms, fiscal years 2010–2023.

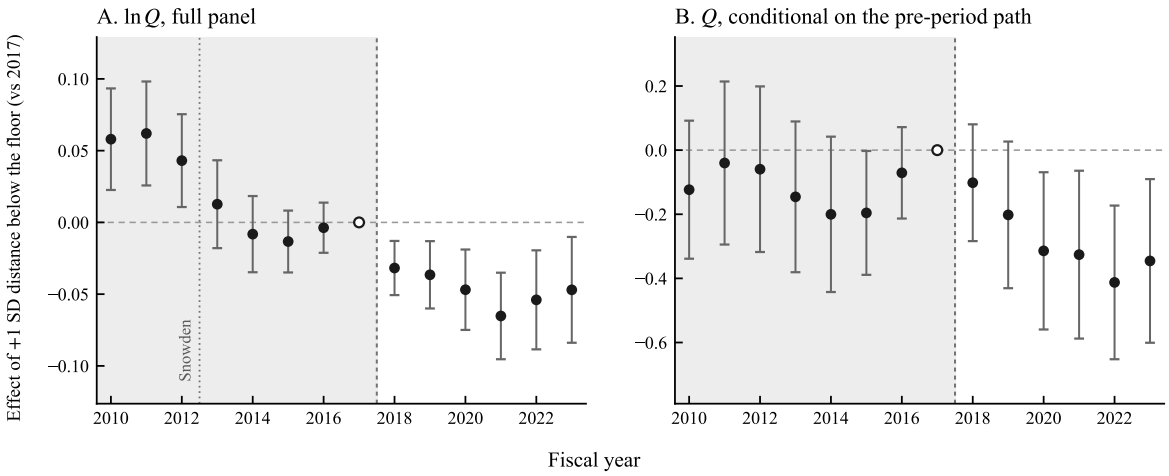



**Figure 4.** Resolving the Tobin-$Q$ pre-period

*Notes.* Both panels re-estimate the Tobin-$Q$ event study of Figure 3. Panel A takes the outcome in logs, which removes the influence of the extreme-valuation tail. The smooth pre-period drift of the levels panel resolves into a level step down at 2013, the year of the Snowden disclosures (dotted line), four flat years leading into the wave, and a post-wave discount that is widest in 2021 and narrows thereafter. Panel B keeps the outcome in levels and compares each firm with firms on the same pre-period valuation path, adding decile-of-pre-period-path by year controls, where a firm's pre-period path is the within-firm slope of its year-adjusted, winsorized $Q$ over 2011–2016. The pre-period coefficients are statistically indistinguishable from zero and the discount opens only with the wave. Randomization inference on the Panel B post-wave cells, permuting the dose across firms, gives $p = 0.010$, 0.004, 0.001, and 0.004 for 2020 through 2023 and $p = 0.002$ for their mean. Coefficients are scaled by one standard deviation of the dose; bars are 95 percent confidence intervals from standard errors clustered by firm; the shaded region is the pre-period and the dashed line marks GDPR enforcement in 2018. The sample is U.S. operating firms excluding financial and utility firms, fiscal years 2010–2023; Panel B conditions on firms with at least four pre-period observations.

**Table 2.** The first stage across specifications

*Notes.* The dependent variable is privacy governance capital, ln(1 + governance sentences), and the regressor is the distance-to-floor dose (distance below the 3.17 end-state floor interacted with the rising floor). Each column is one specification, with the dose coefficient reported above its $t$-statistic in parentheses. The size and book-to-market coefficients are reported where that control enters the column's specification and shown as "–" where it does not; book-to-market enters in levels and its row is scaled by $10^3$ for legibility. Column (1) is the baseline (size and book-to-market, firm and fiscal-year fixed effects, censored dose); (2) drops the controls; (3) adds profitability, R&D intensity, intangible intensity, and sales growth; (4) adds pre-2017 size and profitability interacted with year; (5) replaces fiscal-year effects with two-digit-SIC-by-year effects; (6) uses the uncensored distance. The dose is strongly relevant throughout, so relevance is not an artifact of the controls, the fixed effects, or the censoring; the monotonic, no-defier dose-response is shown in Figure 1. Second-stage outcome estimates and industry heterogeneity (examined by data-reliance split and interaction) are reported in the results sections. $t$-statistics use two-way (firm and fiscal year) clustered standard errors. The sample is U.S. operating firms excluding financial and utility firms, fiscal years 2010–2023. $^{*}$, $^{**}$, and $^{***}$ denote 10%, 5%, and 1%.

| | *Dependent variable: privacy governance capital* $P_{it}$ | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Distance to floor × wave | $+0.435^{***}$ | $+0.431^{***}$ | $+0.466^{***}$ | $+0.479^{***}$ | $+0.493^{***}$ | $+0.415^{***}$ |
| | (+6.60) | (+6.92) | (+7.55) | (+6.99) | (+9.09) | (+6.70) |
| Log assets | $+0.058^{***}$ | – | $+0.085^{***}$ | $+0.078^{***}$ | $+0.046^{**}$ | $+0.058^{***}$ |
| | (+2.85) | | (+3.59) | (+3.51) | (+2.23) | (+2.86) |
| Book-to-market | +0.028 | – | $+0.027^{**}$ | −0.007 | +0.016 | +0.028 |
| | (+1.64) | | (+2.04) | (−0.91) | (+1.42) | (+1.63) |
| First-stage $F$ | 44 | 48 | 57 | 49 | 83 | 45 |
| Observations | 7,435 | 7,610 | 7,014 | 7,435 | 7,435 | 7,435 |
| Firms | 1,654 | 1,675 | 1,585 | 1,654 | 1,654 | 1,654 |
| Within $R^2$ | 0.059 | 0.057 | 0.068 | 0.075 | 0.068 | 0.059 |
| Firm FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | – | Yes |
| Industry×year FE | No | No | No | No | Yes | No |
| Pre-char.×year | No | No | No | Yes | No | No |
| Controls | Size, B/M | None | Extended | Size, B/M | Size, B/M | Size, B/M |
| Dose | Censored | Censored | Censored | Censored | Censored | Uncens. |

**Table 3.** Privacy capital and the corporate footprint

*Notes.* Privacy capital raises cash and the current ratio, raises asset turnover and investment, and raises return on assets: the signature of a non-pledgeable intangible asset that is put to work, and not of a compliance cost. The endogenous regressor is privacy governance capital, $P_{it} = \ln(1 + \text{governance sentences})$, instrumented by the censored distance-to-floor dose (each firm's 2016–2017 distance below the 3.17 bound-frontier end-state floor, interacted with the rising floor). Each column is a separate two-stage least-squares regression of the column outcome on instrumented $P_{it}$, with firm and fiscal-year fixed effects and controls (log assets, book-to-market), whose estimated coefficients the table reports in full. Book-to-market enters the panel in levels and its coefficient is scaled by $10^3$ for legibility. The coefficient on $P_{it}$ is reported above its $t$-statistic in parentheses; the reduced-form row is the coefficient of the outcome on the dose itself. "Within-SD effect" scales the estimate by the within-firm standard deviation of $P_{it}$ and expresses it as a percentage of the outcome mean. Outcomes are winsorized at 1/99; $t$-statistics use two-way (firm and fiscal year) clustered standard errors. The sample is U.S. operating firms excluding financial and utility firms, fiscal years 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable* | | | | |
|---|---|---|---|---|---|
| | Cash/ assets | Current ratio | Asset turnover | Capex/ assets | Return on assets |
| Privacy capital $P_{it}$ | $+0.037^{**}$ | $+0.946^{***}$ | $+0.101^{**}$ | $+0.009^{**}$ | $+0.148^{***}$ |
| | (+2.19) | (+2.77) | (+2.06) | (+2.07) | (+3.13) |
| Log assets | −0.003 | $+1.043^{***}$ | $-0.210^{***}$ | −0.001 | $+0.316^{***}$ |
| | (−0.38) | (+4.34) | (−6.27) | (−0.38) | (+11.99) |
| Book-to-market | −0.002 | +0.012 | $-0.081^{***}$ | $+0.008^{***}$ | +0.079 |
| | (−1.21) | (+0.98) | (−30.12) | (+3.35) | (+1.23) |
| Reduced form (dose) | $+0.0162^{**}$ | $+0.4112^{***}$ | $+0.0438^{**}$ | $+0.0040^{**}$ | $+0.0643^{***}$ |
| | (+2.02) | (+2.89) | (+2.17) | (+2.50) | (+3.02) |
| Within-SD effect | +5.9% | +13.3% | +4.7% | +10.1% | +34.9% |
| First-stage $F$ | 44 | 44 | 44 | 43 | 44 |
| Observations | 7,435 | 7,266 | 7,434 | 7,424 | 7,434 |
| Firms | 1,654 | 1,625 | 1,654 | 1,654 | 1,654 |
| Firm, year FE | Yes | Yes | Yes | Yes | Yes |

**Table 4.** The operating payoff persists one year ahead

*Notes.* The operating payoff one year ahead. Each column regresses the outcome at $t + 1$, the within-firm one-year lead requiring consecutive fiscal years, on instrumented privacy capital at $t$, with the design, controls, fixed effects, winsorization, and two-way clustered standard errors of Table 3; the reduced form is the led outcome on the dose. The 2023 treatment year drops because there are no 2024 outcomes, and the panel does not support a $t + 2$ horizon. Two forces attenuate every $t + 1$ estimate mechanically: the strongest late treatment year leaves the sample, and the dose at $t$ understates the pressure in force at $t + 1$. The efficiency effect persists at full size; profitability and investment retain positive but insignificant coefficients; the margins are indistinguishable from zero unconditionally at both horizons, and their action is conditional (Table 8). "Within-SD effect" follows the convention of Table 3. The sample is U.S. operating firms excluding financial and utility firms, treatment years 2016–2022. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable, measured at $t + 1$* | | | | |
|---|---|---|---|---|---|
| | Asset turnover | Return on assets | Capex/ assets | Operating margin | Gross margin |
| Privacy capital $P_{it}$ | $+0.110^{**}$ | +0.072 | +0.004 | +0.174 | +0.353 |
| | (+2.19) | (+0.96) | (+1.03) | (+0.34) | (+1.20) |
| Log assets | $-0.077^{***}$ | +0.024 | +0.001 | −0.048 | −0.091 |
| | (−3.66) | (+0.60) | (+0.51) | (−0.16) | (−0.74) |
| Book-to-market | +0.010 | −0.050 | +0.118 | $-2.633^{**}$ | −0.083 |
| | (+1.35) | (−0.84) | (+1.43) | (−2.21) | (−0.18) |
| Reduced form (dose) | $+0.0467^{**}$ | +0.0306 | +0.0019 | +0.0819 | +0.1659 |
| | (+2.43) | (+1.02) | (+1.06) | (+0.34) | (+1.22) |
| Within-SD effect | +5.1% | +17.7% | +4.4% | +5.1% | +109.2% |
| First-stage $F$ | 37 | 37 | 37 | 45 | 45 |
| Observations | 6,763 | 6,763 | 5,226 | 5,036 | 5,036 |
| Firms | 1,635 | 1,635 | 1,470 | 1,416 | 1,416 |
| Firm, year FE | Yes | Yes | Yes | Yes | Yes |

**Table 5.** Privacy capital is beyond a compliance cost

*Notes.* If privacy were an expense, margins would fall and overhead would rise, most of all for forced compliers. Neither happens: operating, profit, and gross margins and selling and administrative intensity are statistically indistinguishable from zero, while profitability rises (Table 3). R&D intensity is a placebo. The endogenous regressor is privacy governance capital, $P_{it} = \ln(1 + \text{governance sentences})$, instrumented by the censored distance-to-floor dose (each firm's 2016–2017 distance below the 3.17 bound-frontier end-state floor, interacted with the rising floor). Each column is a separate two-stage least-squares regression of the column outcome on instrumented $P_{it}$, with firm and fiscal-year fixed effects and controls (log assets, book-to-market), whose estimated coefficients the table reports in full. Book-to-market enters the panel in levels and its coefficient is scaled by $10^3$ for legibility. The coefficient on $P_{it}$ is reported above its $t$-statistic in parentheses; the reduced-form row is the coefficient of the outcome on the dose itself. "Within-SD effect" scales the estimate by the within-firm standard deviation of $P_{it}$ and expresses it as a percentage of the outcome mean. Outcomes are winsorized at 1/99; $t$-statistics use two-way (firm and fiscal year) clustered standard errors. The sample is U.S. operating firms excluding financial and utility firms, fiscal years 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable* | | | | |
|---|---|---|---|---|---|
| | Operating margin | Profit margin | Gross margin | SG&A/ sales | R&D/ assets |
| Privacy capital $P_{it}$ | −0.008 | +0.113 | +0.359 | −0.048 | −0.012 |
| | (−0.02) | (+0.23) | (+1.31) | (−0.44) | (−0.92) |
| Log assets | +0.850** | +0.996*** | +0.265** | −0.380*** | −0.068*** |
| | (+2.53) | (+2.66) | (+2.00) | (−4.72) | (−9.67) |
| Book-to-market | −0.313*** | −0.316** | +0.048** | +0.347*** | −0.001 |
| | (−4.12) | (−2.52) | (+2.17) | (+4.33) | (−0.50) |
| Reduced form (dose) | −0.0034 | +0.0510 | +0.1623 | −0.0226 | −0.0054 |
| | (−0.02) | (+0.23) | (+1.38) | (−0.45) | (−0.98) |
| Within-SD effect | −0.2% | +2.6% | +66.9% | −2.8% | −5.1% |
| First-stage $F$ | 52 | 52 | 52 | 47 | 44 |
| Observations | 7,103 | 7,103 | 7,103 | 6,444 | 7,435 |
| Firms | 1,597 | 1,597 | 1,597 | 1,459 | 1,654 |
| Firm, year FE | Yes | Yes | Yes | Yes | Yes |

**Table 6.** Forced compliers invest: the cost incidence of the floor

*Notes.* The cost incidence of the floor. Firms are sorted into terciles of their distance below the end-state floor, and each outcome is regressed on the farthest-below and middle tercile indicators interacted with the rising floor, with the nearest-the-floor tercile as the reference, firm and fiscal-year fixed effects, and controls (log assets, book-to-market), whose coefficients the table reports; book-to-market enters in levels and its row is scaled by $10^3$ for legibility. If compliance were a pure expense, its squeeze should be most visible among the firms forced to comply, the farthest-below tercile: margins should fall and selling and administrative intensity should rise there. Instead the forced compliers raise capital expenditure and asset turnover, monotonically in distance, with no movement in margins or selling and administrative intensity. Outcomes winsorized at 1/99; $t$-statistics in parentheses use two-way (firm and fiscal year) clustered standard errors. Column samples differ with outcome availability; the design is reduced-form in the predetermined tercile, so it does not require the policy measure in every year. The sample is U.S. operating firms excluding financial and utility firms, fiscal years 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable* | | | | |
|---|---|---|---|---|---|
| | Capex/ assets | Asset turnover | Return on assets | Operating margin | SG&A/ sales |
| Farthest below × wave | $+0.008^{**}$ | $+0.091^{***}$ | $+0.188^{***}$ | +0.034 | −0.056 |
| | (+2.47) | (+2.60) | (+4.19) | (+0.06) | (−0.57) |
| Middle × wave | $+0.007^{**}$ | $+0.059^{**}$ | $+0.052^{*}$ | −0.014 | −0.128 |
| | (+2.35) | (+1.97) | (+1.68) | (−0.03) | (−1.59) |
| Log assets | −0.000 | $-0.167^{***}$ | $+0.402^{***}$ | $+0.850^{**}$ | $-0.384^{***}$ |
| | (−0.02) | (−7.81) | (+13.47) | (+2.45) | (−4.66) |
| Book-to-market | $+0.008^{***}$ | $-0.055^{***}$ | $+0.188^{**}$ | $-0.313^{***}$ | $+0.346^{***}$ |
| | (+3.50) | (−10.31) | (+2.04) | (−4.20) | (+4.42) |
| Observations | 7,424 | 11,322 | 11,322 | 7,103 | 6,444 |
| Firms | 1,654 | 1,655 | 1,655 | 1,597 | 1,459 |
| Firm, year FE | Yes | Yes | Yes | Yes | Yes |

## Table 7. Heterogeneity in the footprint by what the firm does with data

*Notes.* The real effects of the wave track the firm's role with data: the role measures of Panel A carry efficiency, investment, and margin gains with no margin falling anywhere; the data-adjacent tier of Panel B responds only in part, the technology label null on every outcome and research intensity moving profitability alone; and Panel C's one mover, a size gradient on profitability, carries no data content and, under the corrected valuation measure, does not touch the valuation response (Table 9). Each moderator contributes two rows: $\delta$, the coefficient on the dose interacted with the moderator, and the dose response $\beta$ at $M_i = 0$, the reference-group response for an indicator (the other group is $\beta + \delta$) and the mean-firm response for a standardized continuous moderator, with $\delta$ the change per one standard deviation. The design is the reduced form $Y_{it} = \beta T_{it} + \delta\,(T_{it} \times M_i) + (M\text{-saturated nuisance}) + \mathbf{x}'_{it}\gamma + \alpha_i + \delta_t + \varepsilon_{it}$, with $T_{it}$ the censored distance-to-floor dose: no instrument enters, so no exclusion restriction is invoked and there is no first stage to screen, and $\delta$ is the difference in the outcome's response to wave exposure itself, which differs from a per-unit-of-privacy-capital difference only through group first stages that are close to identical where tested. The nuisance block is $M_i$ interacted with the controls and the year indicators, with $M_i$ itself absorbed by the firm effects. Moderators are predetermined 2016–2017 firm measures and enter continuously, winsorized at 2.5/97.5 and standardized, unless definitionally categorical (consumer-facing, technology), which enter as indicators; firms missing a moderator leave that moderator's sample. The data-adjacent tier carries measures correlated with data intensity that are not statements about the firm's data role (the technology label, research intensity); the bottom tier carries moderators with no data content. Collect-and-share and first-party collection are the policy's own pre-period disclosure measures of the firm's sharing and collecting roles; consumer-facing and technology are SIC groupings. Outcomes are winsorized at 1/99; $t$-statistics use two-way (firm and fiscal year) clustered standard errors; the sample is U.S. operating firms excluding financial and utility firms, 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable* | | | | |
|---|---|---|---|---|---|
| | Asset turnover | Return on assets | Capex/assets | Operating margin | Gross margin |
| *Panel A. The firm's role with data* | | | | | |
| Third-party intensity, dose $\times M$ ($\delta$) | $+0.044^{**}$ | $-0.011$ | $+0.003^{**}$ | $+0.612^{***}$ | $+0.264^{**}$ |
| | (+2.20) | (−0.72) | (+1.99) | (+2.73) | (+2.55) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.057^{**}$ | $+0.059^{**}$ | +0.003 | +0.346 | +0.313 |
| | (+2.15) | (+2.07) | (+1.61) | (+0.88) | (+1.27) |
| First-party intensity, dose $\times M$ ($\delta$) | +0.032 | −0.004 | +0.002 | +0.386 | $+0.252^{***}$ |
| | (+1.59) | (−0.19) | (+1.26) | (+1.64) | (+2.79) |
| Dose effect $\beta$ at $M_i = 0$ | +0.029 | $+0.052^{*}$ | +0.003 | +0.477 | $+0.450^{*}$ |
| | (+1.14) | (+1.92) | (+1.46) | (+1.00) | (+1.94) |
| Consumer-facing, dose $\times M$ ($\delta$) | $+0.098^{**}$ | +0.023 | $+0.007^{*}$ | +0.443 | −0.014 |
| | (+2.04) | (+0.67) | (+1.79) | (+0.77) | (−0.05) |
| Dose effect $\beta$ at $M_i = 0$ | +0.002 | $+0.052^{**}$ | +0.001 | −0.203 | +0.119 |
| | (+0.13) | (+2.24) | (+0.79) | (−0.67) | (+0.75) |
| *Panel B. Data-adjacent placebos* | | | | | |
| Technology industry, dose $\times M$ ($\delta$) | −0.031 | +0.011 | −0.002 | +0.365 | −0.091 |
| | (−0.90) | (+0.34) | (−0.75) | (+0.80) | (−0.35) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.055^{**}$ | $+0.059^{**}$ | $+0.005^{**}$ | −0.169 | +0.179 |
| | (+2.15) | (+2.04) | (+2.44) | (−0.47) | (+0.92) |
| R&D intensity, dose $\times M$ ($\delta$) | −0.017 | $+0.090^{***}$ | −0.004 | −0.273 | +0.239 |
| | (−0.72) | (+3.15) | (−1.50) | (−0.41) | (+0.78) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.033^{*}$ | $+0.060^{***}$ | $+0.004^{**}$ | −0.160 | +0.116 |
| | (+1.71) | (+2.95) | (+2.45) | (−0.64) | (+0.91) |
| *Panel C. Moderators with no data content* | | | | | |
| Industry concentration, dose $\times M$ ($\delta$) | +0.010 | +0.020 | +0.001 | +0.034 | $-0.201^{*}$ |
| | (+0.47) | (+1.00) | (+0.46) | (+0.15) | (−1.66) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.040^{**}$ | $+0.071^{***}$ | $+0.004^{**}$ | −0.066 | +0.122 |
| | (+2.12) | (+3.20) | (+2.51) | (−0.32) | (+1.15) |
| Lerner index, dose $\times M$ ($\delta$) | −0.015 | −0.033 | −0.000 | +0.309 | −0.030 |
| | (−0.51) | (−0.90) | (−0.03) | (+1.50) | (−0.82) |
| Dose effect $\beta$ at $M_i = 0$ | +0.029 | $+0.024^{*}$ | $+0.005^{***}$ | $-0.141^{***}$ | $+0.025^{**}$ |
| | (+1.46) | (+1.93) | (+2.65) | (−2.74) | (+2.34) |
| Intangible intensity, dose $\times M$ ($\delta$) | −0.020 | −0.046 | −0.001 | +0.076 | −0.025 |
| | (−0.97) | (−1.13) | (−1.12) | (+0.33) | (−0.20) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.047^{**}$ | $+0.065^{***}$ | $+0.005^{***}$ | −0.069 | +0.140 |
| | (+2.30) | (+3.43) | (+3.21) | (−0.30) | (+1.19) |
| Firm size, dose $\times M$ ($\delta$) | −0.015 | $-0.065^{**}$ | +0.002 | +0.316 | −0.028 |
| | (−0.67) | (−2.11) | (+0.88) | (+1.24) | (−0.23) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.057^{***}$ | $+0.039^{*}$ | $+0.004^{**}$ | −0.172 | +0.111 |
| | (+2.71) | (+1.93) | (+2.18) | (−0.67) | (+0.89) |
| $M$-saturated nuisance | Yes | Yes | Yes | Yes | Yes |
| Firm, year FE | Yes | Yes | Yes | Yes | Yes |

## Table 8. Heterogeneity in the payoff one year ahead

*Notes.* The interaction battery of Table 7 re-estimated on outcomes led one year, with treatment, moderators, controls, and fixed effects at $t$: the design, moderator menu, forms, tiers, row structure, and missing-data rule are identical, and the lead construction and mechanical attenuation are those of Table 4. The role-gated margin gains strengthen at the lead: collect-and-share fires on both margins and first-party collection joins on the operating margin, while no adjacent or no-data moderator moves either margin. The collect-and-share efficiency gradient persists at its contemporaneous size, and the investment gradients of the consumer-facing and first-party roles emerge at the lead. The research-intensity profitability cell of Table 7 does not persist, consistent with a conversion scaler operating in the year of exposure. The sample is U.S. operating firms excluding financial and utility firms, treatment years 2016–2022. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable, measured at $t+1$* | | | | |
|---|---|---|---|---|---|
| | Asset turnover | Return on assets | Capex/assets | Operating margin | Gross margin |
| *Panel A. The firm's role with data* | | | | | |
| Third-party intensity, dose $\times M$ ($\delta$) | $+0.044^{**}$ | $+0.008$ | $+0.003$ | $+0.772^{***}$ | $+0.368^{***}$ |
| | (+2.09) | (+0.48) | (+1.51) | (+3.21) | (+3.52) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.044^{*}$ | $+0.054$ | $+0.000$ | $+0.020$ | $+0.289$ |
| | (+1.85) | (+1.48) | (+0.16) | (+0.05) | (+0.99) |
| First-party intensity, dose $\times M$ ($\delta$) | $+0.024$ | $+0.003$ | $+0.004^{**}$ | $+0.552^{***}$ | $+0.304^{***}$ |
| | (+1.52) | (+0.18) | (+1.99) | (+3.13) | (+3.20) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.032$ | $+0.040$ | $+0.003$ | $+0.098$ | $+0.247$ |
| | (+0.94) | (+0.74) | (+1.08) | (+0.26) | (+1.02) |
| Consumer-facing, dose $\times M$ ($\delta$) | $+0.078$ | $+0.105^{*}$ | $+0.012^{**}$ | $+0.625^{*}$ | $-0.116$ |
| | (+1.56) | (+1.88) | (+2.54) | (+1.65) | (−0.55) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.002$ | $+0.013$ | $-0.001$ | $-0.163$ | $+0.126$ |
| | (+0.13) | (+0.39) | (−0.75) | (−0.53) | (+0.77) |
| *Panel B. Data-adjacent placebos* | | | | | |
| Technology industry, dose $\times M$ ($\delta$) | $-0.085^{**}$ | $+0.046$ | $+0.001$ | $+0.492$ | $-0.194$ |
| | (−2.42) | (+0.97) | (+0.22) | (+1.19) | (−0.92) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.076^{***}$ | $+0.017$ | $+0.002$ | $-0.107$ | $+0.223$ |
| | (+3.36) | (+0.42) | (+0.75) | (−0.29) | (+1.13) |
| R&D intensity, dose $\times M$ ($\delta$) | $-0.034$ | $+0.035$ | $-0.006^{**}$ | $-0.555$ | $-0.015$ |
| | (−1.17) | (+0.66) | (−2.17) | (−0.90) | (−0.05) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.040^{**}$ | $+0.018$ | $+0.002$ | $-0.140$ | $+0.081$ |
| | (+2.13) | (+0.64) | (+0.92) | (−0.53) | (+0.58) |
| *Panel C. Moderators with no data content* | | | | | |
| Industry concentration, dose $\times M$ ($\delta$) | $+0.007$ | $+0.010$ | $+0.003$ | $-0.302$ | $-0.225$ |
| | (+0.30) | (+0.42) | (+1.62) | (−1.22) | (−1.61) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.043^{**}$ | $+0.032$ | $+0.002$ | $-0.014$ | $+0.114$ |
| | (+2.26) | (+1.09) | (+1.30) | (−0.06) | (+0.93) |
| Lerner index, dose $\times M$ ($\delta$) | $-0.033$ | $-0.060$ | $+0.006$ | $+0.010$ | $-0.024$ |
| | (−1.00) | (−0.99) | (+1.59) | (+0.05) | (−0.99) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.033$ | $+0.011$ | $+0.002$ | $-0.068$ | $+0.012$ |
| | (+1.55) | (+0.56) | (+1.24) | (−1.21) | (+1.54) |
| Intangible intensity, dose $\times M$ ($\delta$) | $-0.023$ | $-0.038^{*}$ | $-0.001$ | $-0.260$ | $-0.214$ |
| | (−1.12) | (−1.65) | (−0.52) | (−1.01) | (−1.40) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.051^{***}$ | $+0.023$ | $+0.003^{*}$ | $-0.024$ | $+0.104$ |
| | (+2.61) | (+0.78) | (+1.73) | (−0.11) | (+0.84) |
| Firm size, dose $\times M$ ($\delta$) | $-0.029$ | $-0.036$ | $+0.003^{**}$ | $+0.207$ | $-0.070$ |
| | (−1.08) | (−0.80) | (+1.99) | (+0.78) | (−0.47) |
| Dose effect $\beta$ at $M_i = 0$ | $+0.066^{***}$ | $+0.005$ | $+0.002$ | $-0.153$ | $+0.080$ |
| | (+3.41) | (+0.19) | (+0.85) | (−0.56) | (+0.59) |
| $M$-saturated nuisance | | | Yes | | |
| Firm, year FE | | | Yes | | |

**Table 9.** The valuation effect of privacy capital: the discount and where it bites

*Notes.* The valuation effect of privacy capital, in level and in cross section, on both measures; total $Q$ is the Peters and Taylor (2017) measure, which counts intangible capital in the denominator. Panel A is the unconditional effect: the two-stage least-squares coefficient on instrumented privacy capital under the design, controls, fixed effects, winsorization, and two-way clustering of Table 3, with the reduced form of the outcome on the dose beneath it; the level of the discount is the one outcome whose pre-wave path is not flat, Section IV traces the drift to the extreme-valuation tail, and the recognition claims rest on the post-wave dynamics. Panels B–D report reduced-form heterogeneity under the design, moderator menu, forms, and missing-data rule of Table 7: $\delta$ is the interaction with the dose and $\beta$ the dose response at $M_i = 0$. The two outcome columns read the measurement lesson side by side: on raw Tobin's $Q$, whose denominator omits intangible capital, the heterogeneity loads on the intangibles and the roles carry little; under total $Q$, which corrects the denominator, the three role moderators move the valuation response and every placebo is null. Book-to-market enters in levels and its Panel A row is scaled by $10^3$; $t$-statistics use two-way clustered standard errors; the sample is U.S. operating firms excluding financial and utility firms, 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | Tobin's $Q$ | | Total $Q$ | | $N$ | |
|---|---|---|---|---|---|---|
| | Dose $\times M$ ($\delta$) | Dose ($\beta$) | Dose $\times M$ ($\delta$) | Dose ($\beta$) | Tobin | Total |
| *Panel A. The unconditional valuation effect (no moderator)* | | | | | | |
| Privacy capital $P_{it}$ (2SLS) | $-0.919^{***}$ | | $-0.481^{*}$ | | 7,435 | 7,299 |
| | (−2.60) | | (−1.90) | | | |
| Reduced form (dose) | $-0.3999^{**}$ | | $-0.2086^{**}$ | | | |
| | (−2.30) | | (−2.04) | | | |
| Log assets | $-1.395^{***}$ | | +0.267 | | | |
| | (−6.08) | | (+1.13) | | | |
| Book-to-market | $-0.270^{***}$ | | $-0.080^{***}$ | | | |
| | (−5.52) | | (−3.52) | | | |
| First-stage $F$ | 44 | | 46 | | | |
| *Panel B. Heterogeneity: the firm's role with data* | | | | | | |
| Third-party intensity | +0.091 | −0.173 | $-0.238^{**}$ | −0.179 | 7,435 | 7,299 |
| | (+0.69) | (−0.67) | (−2.26) | (−1.53) | | |
| First-party intensity | −0.004 | −0.229 | $-0.250^{***}$ | −0.135 | 7,435 | 7,299 |
| | (−0.03) | (−1.00) | (−2.99) | (−1.32) | | |
| Consumer-facing | $+0.755^{**}$ | $-0.472^{**}$ | $+0.531^{*}$ | $-0.287^{**}$ | 7,435 | 7,299 |
| | (+2.15) | (−2.41) | (+1.82) | (−2.39) | | |
| *Panel C. Heterogeneity: data-adjacent placebos* | | | | | | |
| Technology industry | −0.021 | −0.343 | +0.095 | $-0.202^{**}$ | 7,435 | 7,299 |
| | (−0.08) | (−1.61) | (+0.43) | (−2.02) | | |
| R&D intensity | $-0.579^{***}$ | $-0.425^{***}$ | −0.090 | $-0.189^{*}$ | 7,435 | 7,299 |
| | (−3.24) | (−2.70) | (−0.67) | (−1.65) | | |
| *Panel D. Heterogeneity: moderators with no data content* | | | | | | |
| Industry concentration | −0.066 | $-0.419^{**}$ | +0.009 | $-0.188^{*}$ | 7,435 | 7,299 |
| | (−0.33) | (−2.55) | (+0.08) | (−1.74) | | |
| Lerner index | −0.233 | −0.167 | −0.113 | −0.182 | 6,398 | 6,336 |
| | (−1.00) | (−1.27) | (−0.47) | (−1.36) | | |
| Intangible intensity | $+0.211^{*}$ | $-0.390^{**}$ | −0.100 | $-0.185^{*}$ | 7,400 | 7,264 |
| | (+1.71) | (−2.39) | (−0.95) | (−1.73) | | |
| Firm size | +0.048 | $-0.236^{*}$ | −0.078 | −0.125 | 7,435 | 7,299 |
| | (+0.19) | (−1.73) | (−0.74) | (−1.06) | | |
| Firm, year FE | Yes | | | | | |

**Table 10.** Tobin's $Q$ re-rates upward as the asset is recognized

*Notes.* Each cell is the coefficient on privacy capital interacted with a fiscal-year indicator, from a two-stage least-squares regression of Tobin's $Q$ on privacy capital interacted with year, instrumented by the distance-to-floor dose interacted with year, with firm and fiscal-year fixed effects and controls; fiscal years 2016–2017 are the reference, so each coefficient is the valuation effect in that year relative to before the wave. Within technology, where the treated firms concentrate, the same 2020–2021 jump appears, so the re-rating is a within-sector revaluation and not a rotation into technology. The pre-wave path of $Q$ is not flat, so the claim rests on these post-wave dynamics, not on the level; cross-sectional heterogeneity is Table 9. The sample is U.S. operating firms excluding financial and utility firms, fiscal years 2016–2023; $t$-statistics use two-way clustered standard errors. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable: Tobin's Q* | | | | | |
|---|---|---|---|---|---|---|
| Privacy capital × year | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
| Full sample | $+0.13^{**}$ | $+0.18^{**}$ | $+0.33^{***}$ | $+0.50^{***}$ | $+0.41^{**}$ | $+0.40^{*}$ |
| Within technology | −0.08 | −0.28 | +0.42 | $+0.55^{*}$ | +0.21 | +0.13 |
| Observations | 7,435 (full), 2,665 (technology) | | | | | |

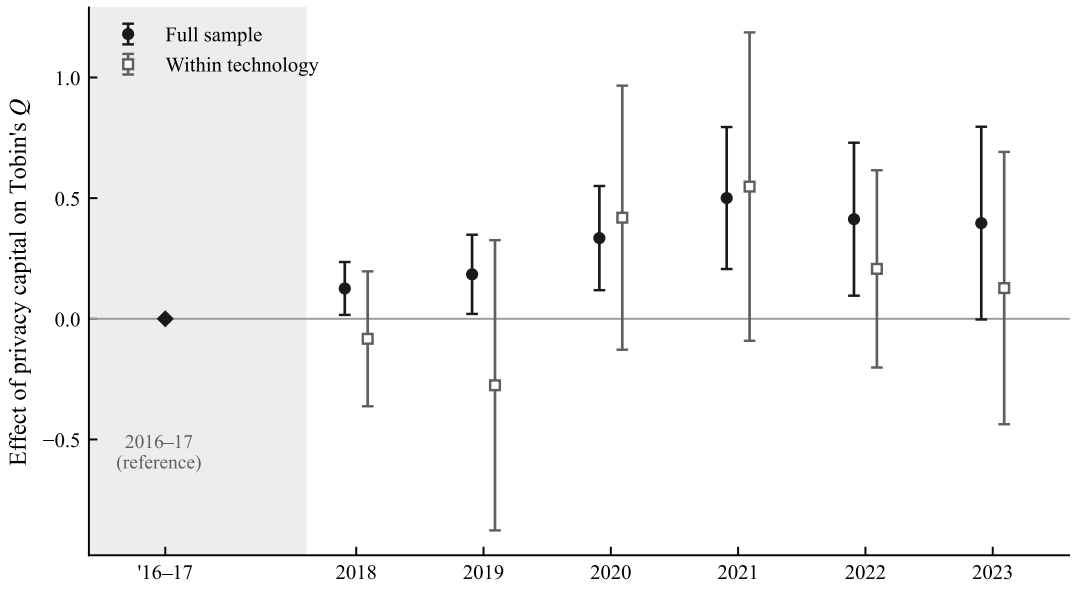



**Figure 5.** The valuation re-rating

*Notes.* The figure plots the path coefficients of Table 10: the effect of privacy capital on Tobin's $Q$ by fiscal year, relative to the 2016–2017 reference, from the year-interacted distance-to-floor two-stage least squares with firm and fiscal-year fixed effects and controls, in the full sample and within the technology sector; bars are 95 percent confidence intervals from two-way clustered standard errors. The technology path shows the same 2020–2021 jump, so the re-rating is a within-sector revaluation rather than a rotation into technology. U.S. operating firms excluding financial and utility firms, fiscal years 2016–2023.

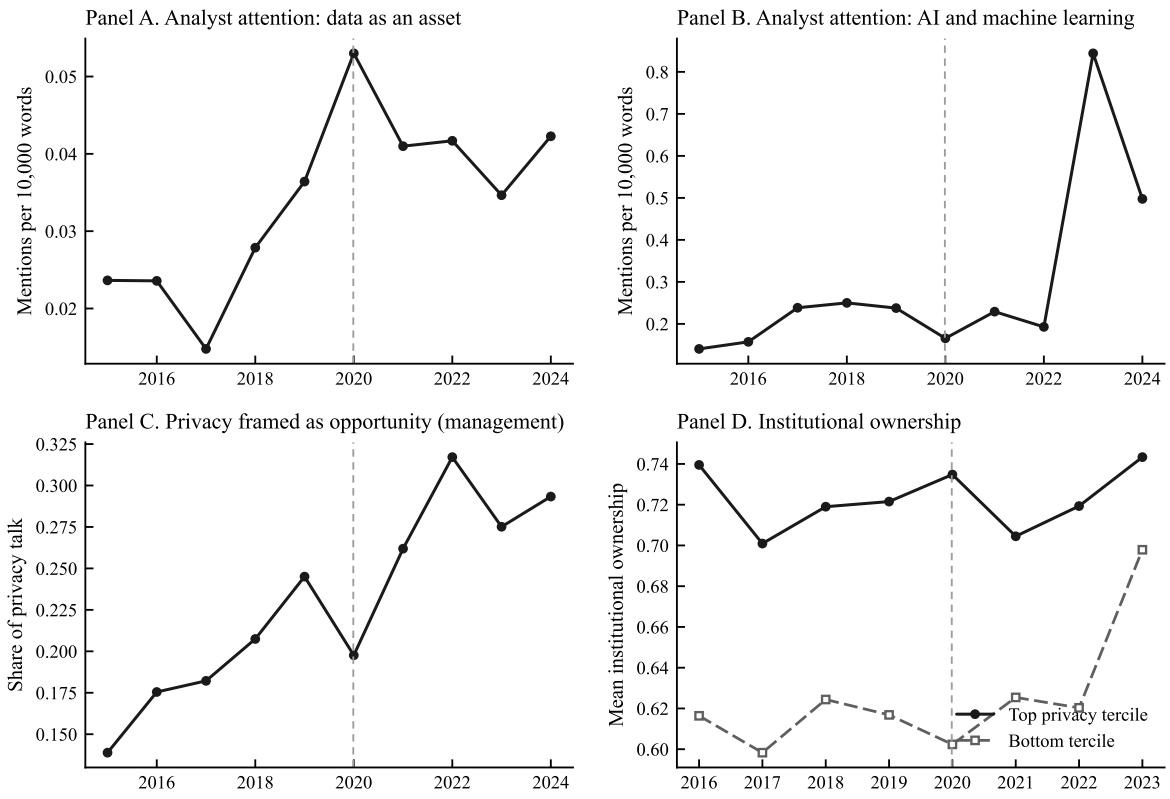



**Figure 6.** Four recognition clocks

*Notes.* *Panel A* plots the mean rate at which sell-side analysts speak about data as an asset in earnings-call question sessions, and *Panel B* the rate at which they speak about artificial intelligence and machine learning, both per ten thousand words. *Panel C* plots the share of managements' privacy talk framed as commercial opportunity rather than burden. *Panel D* plots mean institutional ownership for the top and bottom terciles of privacy capital in the estimation sample. The dashed line marks 2020. The attention and sentiment panels are economy-wide descriptive trends; the ownership panel is the raw pattern behind the instrumented post-2020 estimate in the text.

**Table 11.** The long-short privacy premium fades after 2020

*Notes.* The portfolio is long the top and short the bottom tercile of firms ranked on privacy governance capital, equal-weighted and re-formed each July from the policy in force the prior December, held July to June. Returns run through December 2024, the limit the policy corpus supports. Each row is a factor model; the columns report the annualized intercept (alpha) over 30 pre-2020 and 60 post-2020 months, and the coefficient $\delta$ on a post-2020 indicator in a pooled regression of the long-short return on the indicator and the factors. Standard errors are Newey-West with six lags; $t$-statistics in parentheses. A recognition account predicts a premium that is positive while the asset is being recognized and fades toward zero once it has been ($\delta < 0$, post-2020 $\approx 0$); a priced-risk account predicts a permanent premium ($\delta \approx 0$). The sample is U.S. operating firms excluding financial and utility firms. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | Pre-2020 $\alpha$ (%/yr) | Post-2020 $\alpha$ (%/yr) | Fade $\delta$ (%/yr) |
|---|---|---|---|
| CAPM | $+13.3^{***}$ | +1.8 | −8.4 |
| | (+3.62) | (+0.56) | (−1.54) |
| FF3 | $+10.0^{**}$ | +0.5 | −7.6 |
| | (+2.18) | (+0.15) | (−1.40) |
| FF5 | $+11.5^{**}$ | +0.4 | −8.1 |
| | (+2.13) | (+0.12) | (−1.41) |

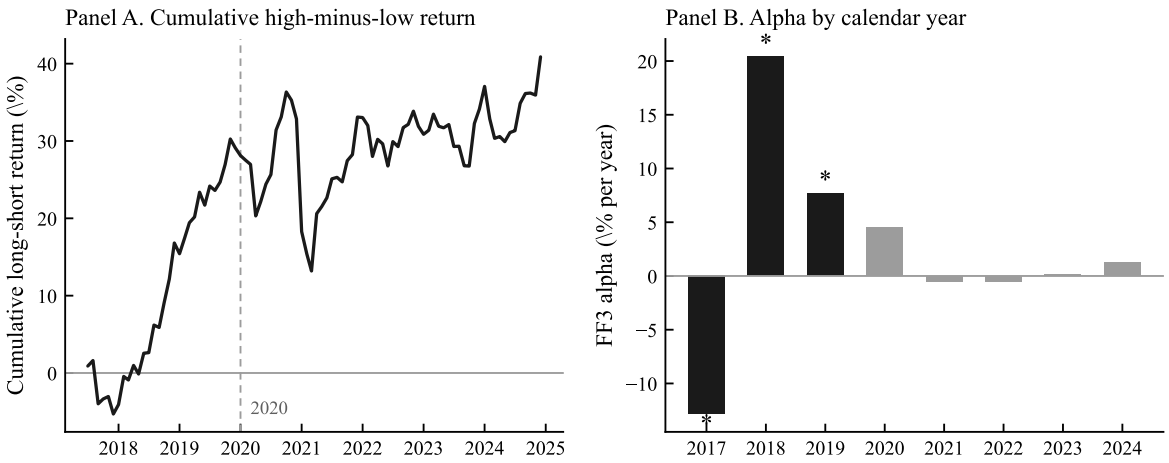



**Figure 7.** The premium builds and fades

*Notes.* *Panel A* plots the cumulative return of the high-minus-low privacy-governance portfolio (top minus bottom tercile, equal-weighted, formed each June with at least a six-month reporting lag), July 2017 through December 2024; the dashed line marks 2020. *Panel B* plots the portfolio's three-factor alpha by calendar year, from a single regression with year-specific intercepts and common factor loadings and Newey–West standard errors; asterisks mark years significant at five percent. The 2017 bar covers only the July–December formation stub, and 2024 holdings end in December under the corpus-end rule. The premium builds through 2018–2019 and is indistinguishable from zero afterward, the arc of a one-time recognition correction. The sample is U.S. operating firms excluding financial and utility firms.

**Table 12.** The premium is a small-stock phenomenon; the re-rating is not

*Notes.* *Panel A* reports the annualized three-factor alpha of the long-short privacy-governance portfolio, equal-weighted, equal-weighted after dropping the smallest twenty percent of firms at formation, and value-weighted. The premium is an equal-weighted, small-stock phenomenon: it does not survive dropping the smallest firms or value-weighting. *Panel B* reports the distance-to-floor two-stage least-squares estimate of privacy capital on the valuation level, the 2021 re-rating coefficient, book leverage, and cash, for the full sample and excluding the smallest twenty percent. These are panel estimates, not portfolio sorts, and they are not weighted by size; the valuation level, the re-rating, and the leverage effect survive microcap removal, while the cash effect weakens, so the cash build is concentrated among smaller firms, a sensitivity I disclose. The recognition claim rests on the re-rating and the footprint rather than on the portfolio premium and is unaffected. A priced systematic risk would, by contrast, be present in value-weighted returns and among large firms. Returns run through December 2024; standard errors are Newey-West (Panel A) and two-way clustered (Panel B). $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

*Panel A. Long-short alpha (annualized %), by portfolio weighting*

| | Pre-2020 | Post-2020 |
|---|---|---|
| Equal-weighted | $+10.0^{**}$ | +0.5 |
| | (+2.18) | (+0.15) |
| Excluding smallest 20% | +1.5 | −1.6 |
| | (+0.36) | (−0.64) |
| Value-weighted | −5.0 | −1.4 |
| | (−0.96) | (−0.35) |

*Panel B. Panel estimate of privacy capital, full sample vs. excluding the smallest 20%*

| *Dependent variable* | Full sample | Ex. smallest 20% |
|---|---|---|
| Tobin's $Q$ (level) | $-0.919^{***}$ | $-0.416^{**}$ |
| | (−2.60) | (−2.11) |
| Tobin's $Q$ re-rating, 2021 | $+0.500^{***}$ | $+0.365^{***}$ |
| | (+3.33) | (+4.41) |

**Table 13.** The effects follow the data roles, not the sector label, and are not driven by large firms

*Notes.* Sector, role, and size splits of the headline design. Each cell is a separate regression with the specification, controls, fixed effects, winsorization, and two-way clustered standard errors of Table 3, estimated on the subsample named in the panel; technology and consumer-facing are the indicator definitions of Table 7, and the size split is at the median of firm-level 2016–2017 mean log assets (ex-small is Panel F, so the size split carries only-panels). Two reading rules apply. First, within a subsample the fixed effects are estimated among that sample's firms alone, so a within-group estimate is the effect relative to group peers, and a within-group null does not say the group is untreated; whether an effect concentrates in a group is read from its complement, where a null after removal says the remaining sample does not carry it. Second, the splits are not symmetric objects: technology is a data-adjacent label, so the design predicts the effects survive its removal, while consumer-facing is a role, so the design predicts the role-gated effects move with it. Both predictions hold. Excluding technology, the efficiency, profitability, and valuation results stand, efficiency if anything larger. Excluding the consumer-facing role, the valuation discount deepens on both measures, the sample-split form of the sign inversion of Table 9, the turnover response moves with the role, reproducing that table's dose response at $M_i = 0$, and profitability survives, the broad gain of Table 3. In the size panels the small-firm results are self-contained, large firms absent from the regression, so the results are not driven by large firms, and the large-firm nulls are no-exposure-gradient statements, the opposite of what a regulation-entrenches-incumbents reading predicts; the identified valuation discount sits with the large firms, where $Q$ is measured precisely, while the small-firm point is larger and imprecise, a precision difference, not a loading difference (size is null as a valuation moderator in Table 9). The margins are indistinguishable from zero unconditionally in every sample; their action is conditional (Table 7). The sample is U.S. operating firms excluding financial and utility firms, 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable* | | | | | | |
|---|---|---|---|---|---|---|---|
| | Asset turnover | Return on assets | Capex/ assets | Operating margin | Gross margin | Tobin's $Q$ | Total $Q$ |
| *Panel A. Excluding technology firms* | | | | | | | |
| Privacy capital $P_{it}$ (2SLS) | +0.134* | +0.130* | +0.012* | −0.407 | +0.432 | −0.753* | −0.429* |
| | (+1.86) | (+1.95) | (+1.96) | (−0.48) | (+0.89) | (−1.96) | (−1.89) |
| Reduced form (dose) | +0.0544** | +0.0529** | +0.0050** | −0.1689 | +0.1792 | −0.3059* | −0.1733* |
| | (+2.10) | (+1.98) | (+2.44) | (−0.47) | (+0.92) | (−1.79) | (−1.86) |
| First-stage $F$ | 30 | 30 | 30 | 38 | 38 | 30 | 33 |
| Observations | 4,769 | 4,769 | 4,764 | 4,465 | 4,465 | 4,770 | 4,662 |
| *Panel B. Excluding consumer-facing firms* | | | | | | | |
| Privacy capital $P_{it}$ (2SLS) | +0.009 | +0.121** | +0.003 | −0.482 | +0.283 | −1.082*** | −0.669** |
| | (+0.22) | (+2.25) | (+0.74) | (−0.70) | (+0.72) | (−3.02) | (−2.29) |
| Reduced form (dose) | +0.0038 | +0.0493** | +0.0014 | −0.2030 | +0.1192 | −0.4428*** | −0.2715** |
| | (+0.22) | (+2.20) | (+0.79) | (−0.67) | (+0.75) | (−2.65) | (−2.37) |
| First-stage $F$ | 30 | 30 | 30 | 37 | 37 | 30 | 32 |
| Observations | 5,432 | 5,432 | 5,427 | 5,111 | 5,111 | 5,432 | 5,327 |
| *Panel C. Technology firms only* | | | | | | | |
| Privacy capital $P_{it}$ (2SLS) | +0.049 | +0.148*** | +0.005 | +0.305 | +0.075*** | −0.498 | −0.227 |
| | (+0.98) | (+2.88) | (+1.23) | (+1.10) | (+2.65) | (−0.81) | (−0.54) |
| Reduced form (dose) | +0.0245 | +0.0750*** | +0.0027 | +0.1616 | +0.0399** | −0.2515 | −0.1142 |
| | (+0.94) | (+3.16) | (+1.29) | (+1.01) | (+2.33) | (−0.78) | (−0.55) |
| First-stage $F$ | 26 | 26 | 26 | 25 | 25 | 26 | 25 |
| Observations | 2,665 | 2,665 | 2,660 | 2,638 | 2,638 | 2,665 | 2,637 |
| *Panel D. Consumer-facing firms only* | | | | | | | |
| Privacy capital $P_{it}$ (2SLS) | +0.163* | +0.127 | +0.014** | −0.048 | +0.054** | +0.920 | +0.449 |
| | (+1.94) | (+1.64) | (+2.24) | (−0.36) | (+2.09) | (+1.14) | (+0.95) |
| Reduced form (dose) | +0.0993** | +0.0770* | +0.0085** | −0.0300 | +0.0341** | +0.5596 | +0.2746 |
| | (+2.12) | (+1.78) | (+2.42) | (−0.36) | (+2.22) | (+1.18) | (+0.95) |
| First-stage $F$ | 30 | 30 | 31 | 31 | 31 | 30 | 29 |
| Observations | 2,002 | 2,002 | 1,997 | 1,992 | 1,992 | 2,003 | 1,972 |
| *Panel E. Small firms only (below-median pre-period size)* | | | | | | | |
| Privacy capital $P_{it}$ (2SLS) | +0.155* | +0.554** | +0.009 | −0.665 | +0.507 | −1.316 | −0.131 |
| | (+1.70) | (+2.52) | (+0.90) | (−0.57) | (+0.82) | (−0.57) | (−0.23) |
| Reduced form (dose) | +0.0637* | +0.2280** | +0.0038 | −0.2906 | +0.2215 | −0.5424 | −0.0534 |
| | (+1.75) | (+2.23) | (+1.02) | (−0.56) | (+0.84) | (−0.54) | (−0.23) |
| First-stage $F$ | 25 | 25 | 25 | 34 | 34 | 25 | 27 |
| Observations | 3,570 | 3,570 | 3,565 | 3,239 | 3,239 | 3,571 | 3,452 |
| *Panel F. Large firms only (above-median pre-period size)* | | | | | | | |
| Privacy capital $P_{it}$ (2SLS) | +0.068 | +0.010 | +0.008*** | +0.037 | +0.029* | −0.391** | −0.388** |
| | (+1.58) | (+0.66) | (+2.87) | (+1.57) | (+1.76) | (−2.23) | (−1.98) |
| Reduced form (dose) | +0.0325* | +0.0047 | +0.0040*** | +0.0175* | +0.0136** | −0.1861*** | −0.1838** |
| | (+1.68) | (+0.68) | (+2.63) | (+1.71) | (+1.98) | (−2.60) | (−2.16) |
| First-stage $F$ | 36 | 36 | 35 | 36 | 36 | 36 | 35 |
| Observations | 3,864 | 3,864 | 3,859 | 3,864 | 3,864 | 3,864 | 3,847 |
| Firm, year FE | | | | Yes | | | |

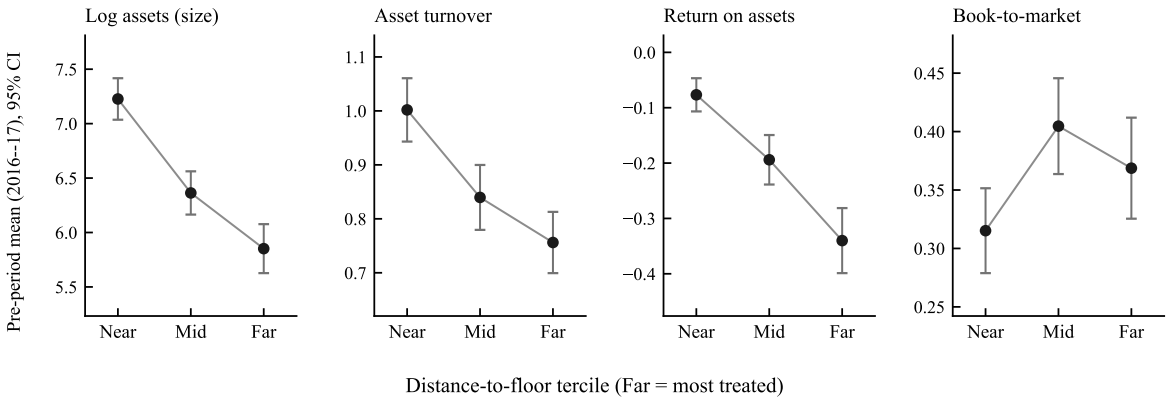



**Figure 8.** Who the wave compelled: pre-period characteristics by distance to the floor

*Notes.* Pre-period characteristics of the firms the wave compelled. Estimation-sample firms (1,675) are ranked into the terciles of Figure 1 by their pre-wave distance below the floor (558–559 firms per tercile; mean distance 0.01, 0.54, and 1.66); each panel plots the tercile means of a firm-level 2016–2017 characteristic, winsorized at the 2.5/97.5 percentiles, with 95 percent confidence intervals. By construction the far tercile begins with the thinnest privacy practice; the panels show what else those firms are. The most-compelled firms are smaller and operationally weaker at baseline, with log assets, asset turnover, and return on assets declining toward the far tercile, and they carry no value or growth tilt, with book-to-market flat across terciles. In a multivariate regression the profitability gradient is absorbed by size, so the selection statement is that the compelled are small firms, not unprofitable ones. The sample is U.S. operating firms excluding financial and utility firms.

## Table 14. Selection into treatment: the horse race

*Notes.* Selection into treatment does not carry the results: raced directly against the trends its selection could borrow, the dose survives, and the placebo trends absorb exactly one cell of the heterogeneity, the no-data size gradient on profitability. The concern is that treatment concentrates in small firms with weak baselines (Figure 8), whose generic catch-up could masquerade as the effects. The answer puts the catch-up in the regression: each pre-window characteristic—the firm's 2010–2015 mean of log assets, return on assets, and asset turnover, requiring at least three pre-window years, winsorized at 2.5/97.5 across firms and standardized—enters interacted with a linear trend, alongside the dose, under the design, controls, fixed effects, winsorization, and two-way clustering of Table 3. The pre-window mean replaces the 2016–2017 snapshot because the snapshot is nearly all transitory for profitability—its firm-level correlation with the 2010–2015 mean is 0.10, against 0.96 for size and 0.82 for turnover—so a snapshot own-level trend races noise reversion rather than the firm's standing; and the three trends enter jointly because the accounting identity ties profitability to efficiency, so their trends are one object, not three. Panel A reports the reduced-form dose coefficient on the pre-window sample: raced against the size trend—the selection margin the introduction displays—every effect survives and the efficiency response strengthens; the size trend itself enters positive, large firms drifting up, significantly so on total $Q$, so generic small-firm catch-up runs against these effects rather than producing them, and the same drift carries part of the measured total-$Q$ discount, which the full race prices at −0.167. Panels B–D re-estimate the two heterogeneity batteries with the full trend block in the regression, under the design, moderator menu, forms, and missing-data rule of Tables 7 and 9, reporting the interaction $\delta$ only; the Lerner and intangible rows lose the firms missing those moderators. The batteries' data architecture is intact—the margin structure, the efficiency interactions, profitability on research intensity, and the total-$Q$ concentration in the collecting and sharing roles all survive, with three cells shading below ten percent (the two capex interactions and the consumer-facing total-$Q$ premium, the last losing more to the pre-window sample restriction than to the race)—while the one cell the trends absorb outright is the size gradient on return on assets, which is exactly the cell a selection story could produce. Outcomes are winsorized at 1/99; $t$-statistics use two-way (firm and fiscal year) clustered standard errors; the sample is U.S. operating firms excluding financial and utility firms, 2016–2023, restricted to firms with the pre-window characteristics. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable* | | | | | | |
|---|---|---|---|---|---|---|---|
| | Asset turnover | Return on assets | Capex/ assets | Operating margin | Gross margin | Tobin's $Q$ | Total $Q$ |
| *Panel A. The dose raced against the placebo trends* | | | | | | | |
| Dose, no placebo trend | +0.0437** | +0.0551*** | +0.0043*** | +0.0987 | +0.1864* | −0.4109*** | −0.2748*** |
| | (+2.16) | (+3.19) | (+2.75) | (+0.47) | (+1.84) | (−2.95) | (−3.26) |
| Dose, size trend raced | +0.0519** | +0.0543*** | +0.0048*** | +0.0418 | +0.1569* | −0.3694*** | −0.1944** |
| | (+2.42) | (+3.20) | (+2.92) | (+0.22) | (+1.73) | (−2.76) | (−2.03) |
| Dose, all three trends raced | +0.0312* | +0.0418** | +0.0043*** | −0.0515 | +0.1139 | −0.3226** | −0.1668* |
| | (+1.65) | (+2.37) | (+2.81) | (−0.27) | (+1.30) | (−2.57) | (−1.74) |
| Observations | 6,639 | 6,639 | 6,629 | 6,433 | 6,433 | 6,640 | 6,538 |
| *Panel B. Heterogeneity under the full race: the firm's role with data* | | | | | | | |
| Collect-and-share, dose $\times M$ ($\delta$) | +0.034* | −0.006 | +0.002 | +0.502*** | +0.178** | +0.029 | −0.199** |
| | (+1.87) | (−0.44) | (+1.55) | (+2.81) | (+2.30) | (+0.28) | (−2.26) |
| First-party collection, dose $\times M$ ($\delta$) | +0.021 | +0.007 | +0.001 | +0.332** | +0.201*** | −0.051 | −0.179** |
| | (+1.06) | (+0.52) | (+0.71) | (+2.01) | (+2.81) | (−0.46) | (−2.25) |
| Consumer-facing, dose $\times M$ ($\delta$) | +0.114** | +0.028 | +0.007 | +0.167 | −0.099 | +0.563** | +0.317 |
| | (+2.34) | (+0.88) | (+1.64) | (+0.29) | (−0.36) | (+2.01) | (+1.32) |
| *Panel C. Heterogeneity under the full race: data-adjacent placebos* | | | | | | | |
| Technology industry, dose $\times M$ ($\delta$) | +0.024 | +0.025 | −0.002 | +0.360 | −0.049 | −0.183 | −0.046 |
| | (+0.73) | (+0.99) | (−0.57) | (+1.01) | (−0.25) | (−0.87) | (−0.27) |
| R&D intensity, dose $\times M$ ($\delta$) | −0.001 | +0.084*** | −0.004 | +0.296 | +0.456* | −0.495*** | −0.081 |
| | (−0.05) | (+3.19) | (−1.26) | (+0.57) | (+1.69) | (−2.62) | (−0.62) |
| *Panel D. Heterogeneity under the full race: no data content* | | | | | | | |
| Industry concentration, dose $\times M$ ($\delta$) | +0.003 | +0.023 | +0.001 | −0.097 | −0.187* | −0.013 | +0.031 |
| | (+0.14) | (+1.26) | (+0.66) | (−0.44) | (−1.67) | (−0.07) | (+0.30) |
| Lerner index, dose $\times M$ ($\delta$) | +0.013 | −0.034 | +0.002 | +0.129 | −0.032 | −0.196 | −0.079 |
| | (+0.40) | (−0.78) | (+0.30) | (+0.70) | (−0.79) | (−1.06) | (−0.62) |
| Intangible intensity, dose $\times M$ ($\delta$) | +0.006 | −0.039 | −0.001 | −0.012 | −0.034 | +0.246*** | +0.012 |
| | (+0.29) | (−1.28) | (−0.77) | (−0.05) | (−0.28) | (+2.63) | (+0.15) |
| Firm size, dose $\times M$ ($\delta$) | −0.003 | −0.045 | +0.003 | +0.322 | −0.013 | +0.065 | −0.003 |
| | (−0.14) | (−1.55) | (+1.06) | (+1.17) | (−0.11) | (+0.34) | (−0.03) |
| Placebo trend block | As labeled (Panel A); all three (Panels B–D) | | | | | | |
| Firm, year FE | Yes | | | | | | |

**Table 15.** The headline estimates and the pandemic window

*Notes.* Two-stage least-squares estimates of the headline specification (Table 3: privacy capital instrumented by the censored distance-to-floor dose, firm and fiscal-year fixed effects, controls, outcomes winsorized at 1/99, two-way clustered $t$-statistics in parentheses) re-estimated on the fiscal windows in the columns. By fiscal 2019 the floor has completed 47 percent of its rise, so the pre-pandemic column carries less than half the treatment and about half the panel; attenuation there is the design's own prediction. The final column drops fiscal years 2020 and 2021 entirely. First-stage $F$ and observation counts are reported for the cash regression; they vary little across outcomes within a window. The sample is U.S. operating firms excluding financial and utility firms. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| *Dependent variable* | Full window 2016–2023 | Pre-pandemic 2016–2019 | Pandemic excised ex 2020–21 |
|---|---|---|---|
| Book leverage | −0.049$^{*}$ | +0.010 | −0.036 |
| | (−1.67) | (+0.19) | (−1.07) |
| Cash/assets | +0.037$^{**}$ | +0.043$^{*}$ | +0.031$^{*}$ |
| | (+2.19) | (+1.73) | (+1.94) |
| Current ratio | +0.946$^{***}$ | +0.662 | +0.504$^{**}$ |
| | (+2.77) | (+1.34) | (+2.33) |
| Asset turnover | +0.101$^{**}$ | +0.092 | +0.046 |
| | (+2.06) | (+1.26) | (+1.10) |
| Capex/assets | +0.009$^{**}$ | +0.017$^{*}$ | +0.007$^{*}$ |
| | (+2.07) | (+1.88) | (+1.66) |
| ROA | +0.148$^{***}$ | +0.210$^{***}$ | +0.136$^{***}$ |
| | (+3.13) | (+2.68) | (+2.87) |
| Tobin Q | −0.919$^{***}$ | −0.233 | −0.788$^{**}$ |
| | (−2.60) | (−0.35) | (−2.44) |
| Total Q | −0.481$^{*}$ | −0.603$^{**}$ | −0.263 |
| | (−1.90) | (−2.40) | (−1.61) |
| First-stage $F$ | 44 | 26 | 30 |
| Observations | 7,435 | 4,834 | 5,779 |

**Table 16.** The late-window valuation response and pandemic-crash exposure

*Notes.* Reduced-form coefficients on the distance-to-floor dose in the 2020–2023 window, with firm and fiscal-year fixed effects and controls, progressively adding firm-level pandemic-crash exposures interacted with year dummies. Year effects absorb the common pandemic path by construction, so the rows control the remaining channel, differential crash-and-rebound. The exposures, each winsorized and standardized with higher values meaning a harder 2020 hit: the February–March 2020 fever-window return (sign reversed), the maximum drawdown of the monthly cumulative return over 2020, and the largest negative monthly market-model residual in 2020 with betas estimated over 2017–2019. Style timing adds book-to-market and log assets interacted with year. Each column block holds its sample fixed across rows: firms with continuous CRSP coverage and all three exposures, and within that the firm-years with an observed policy. The dose is essentially uncorrelated with the exposures (0.03, 0.10, 0.13). $t$-statistics use two-way (firm and fiscal year) clustered standard errors. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable: Tobin's Q* | | | |
|---|---|---|---|---|
| | CRSP-covered late panel | | Policy-observed | |
| Crash-exposure controls | Dose | $t$ | Dose | $t$ |
| Baseline | +0.437$^{*}$ | (+1.87) | +0.828$^{**}$ | (+2.41) |
| Fever-window return × year | +0.425$^{*}$ | (+1.84) | +0.799$^{**}$ | (+2.43) |
| 2020 maximum drawdown × year | +0.384$^{*}$ | (+1.73) | +0.714$^{**}$ | (+2.31) |
| Largest idiosyncratic crash × year | +0.455$^{**}$ | (+2.05) | +0.797$^{**}$ | (+2.31) |
| All three × year | +0.500$^{**}$ | (+2.23) | +0.879$^{**}$ | (+2.43) |
| All three + style timing × year | +0.542$^{**}$ | (+2.26) | +0.683$^{*}$ | (+1.74) |
| Observations | 3,802 | | 2,006 | |

# A. CONSTRUCTION OF THE PRIVACY-CAPITAL MEASURE

This appendix gives the full construction summarized in Section III. The measure is built from the corpus of dated corporate privacy policies in Ramadorai et al. (2025), 90,273 policy versions for 5,463 U.S. public firms. The unit of classification is the sentence; after dropping non-English text the corpus contains roughly 9.9 million sentences.

## A. The taxonomy

Each sentence is classified into the twelve data-practice categories of the OPP-115 taxonomy (Wilson et al. 2016), indexed by the alphabetical category ordering in Table 17. Classification is multi-label, but 96.6% of sentences carry a single label, 3.4% carry two, and fewer than 0.1% carry three, so multi-label and single-label counting are nearly equivalent. Table 17 also records the partition into substantive and boilerplate categories and the functional dimension of each substantive category.

**Table 17.** The OPP-115 classification taxonomy

*Notes.* The twelve OPP-115 data-practice categories (Wilson et al. 2016), indexed by alphabetical ordering. "Substantive" marks the eight categories that record data-handling practice; the other four are boilerplate held out of the measure and used as a verbosity placebo. "Dimension" groups the substantive categories by the kind of practice. The governance measure counts sentences in the four governance categories ($\{0, 2, 10, 11\}$). The final column is each category's share of single-label sentences in the corpus.

| Idx | OPP-115 category | Substantive | Dimension | Corpus freq. |
|---|---|---|---|---|
| 0 | Data Retention | Yes | Governance | 2.0% |
| 1 | Data Security | Yes | Security | 5.7% |
| 2 | Do Not Track | Yes | Governance | 0.4% |
| 3 | First Party Collection/Use | Yes | Collection | 26.7% |
| 4 | International & Specific Audiences | Yes | Regulatory | 5.0% |
| 5 | Introductory/Generic | No | — | 19.6% |
| 6 | Policy Change | No | — | 2.6% |
| 7 | Practice not covered | No | — | 9.7% |
| 8 | Privacy contact information | No | — | 3.5% |
| 9 | Third Party Sharing/Collection | Yes | Collection | 11.7% |
| 10 | User Access, Edit & Deletion | Yes | Governance | 3.8% |
| 11 | User Choice/Control | Yes | Governance | 9.4% |

## B. The classification pipeline

Three label sources are used in sequence. A supervised BERT-based classifier trained on the OPP-115 human annotations labels the sentences it can classify with high confidence (predicted probability above 0.90), which is 88.0% of the corpus. A large language model (Qwen 3.5 9B)

resolves the boundary sentences the supervised model leaves uncertain, a further 11.5%. A keyword rule that detects contact details assigns the remaining 0.5% to the privacy-contact category. The taxonomy is human-annotated and the pipeline relies on three independent labelers, so classification error is unlikely to align with firm outcomes, and to the extent the labels are noisy the noise is classical and attenuates estimates toward zero.

### C. The substantive partition and the functional dimensions

Write $c_{p,k}$ for the number of sentences in policy $p$ classified into category $k \in \{0, \ldots, 11\}$. The substantive categories are $\mathcal{S} = \{0, 1, 2, 3, 4, 9, 10, 11\}$ (about 64.6% of sentences) and the boilerplate categories are $\mathcal{N} = \{5, 6, 7, 8\}$. The four functional dimensions partition $\mathcal{S}$,

$$\underbrace{\{3, 9\}}_{\text{Collection}}, \qquad \underbrace{\{1\}}_{\text{Security}}, \qquad \underbrace{\{0, 2, 10, 11\}}_{\text{Governance } (\mathcal{G})}, \qquad \underbrace{\{4\}}_{\text{Regulatory}}, \tag{5}$$

so that the substantive count is the sum of the four dimension counts. The analysis uses the governance dimension $\mathcal{G} = \{0, 2, 10, 11\}$. I also carry two robustness composites, protection $\{0, 1, 2, 10, 11\}$ (governance with security) and broad governance $\{0, 2, 4, 10, 11\}$ (governance with the regulatory category).

### D. Aggregation to the firm-year and transforms

For policy $p$ the governance count is $g_p = \sum_{k \in \mathcal{G}} c_{p,k}$ and the substantive count is $\text{subst}_p = \sum_{k \in \mathcal{S}} c_{p,k}$. To form the firm-year measure, each policy is assigned to the first fiscal year that ends on or after its effective date, and a fiscal year takes the most recent policy so assigned to it; fiscal years in which no new policy version is captured are left missing and never interpolated or carried forward (carrying the last policy forward instead yields the same signs with larger, noisier estimates). Writing $g_{it}$ for the governance count of firm $i$'s policy of record in fiscal year $t$, the measure of privacy capital is

$$P_{it} = \ln(1 + g_{it}) \,. \tag{6}$$

Two alternative transforms are carried for robustness. The composition share is $\text{share}_{it} = g_{it}/\text{subst}_{it}$, and the length-adjusted residual $\tilde{g}_{it}$ is the residual from a pooled regression of $\ln(1 + g_{it})$ on $\ln(1 + n_{it})$, where $n_{it}$ is the total sentence count, which strips out any mechanical role for policy length. Table 18 carries the transforms through the design: the headline estimates survive a policy-length control, the length-purged residual, and the share transform, at similar significance throughout, so the measure's governance content, not document volume, carries

the signal. In placebo tests the boilerplate count $\sum_{k\in\mathcal{N}} c_{p,k}$ replaces $g$ as the regressor and produces no effect on the financing and efficiency outcomes; it does move the valuation level, which, read together with the transform results, says the wave lifted all disclosure rather than that the measure lacks content.

**Table 18.** Content versus volume: alternative transforms of the measure

*Notes.* Content versus volume. Each cell is the two-stage least-squares effect of the transformed privacy measure on the row outcome, instrumented by the censored end-state distance-to-floor dose, with firm and fiscal-year fixed effects and controls; design, winsorization, and two-way clustered standard errors as in Table 3. Column (1) is the baseline count, ln(1 + governance sentences). Column (2) adds the log of total policy sentences as a control, so identification uses governance content beyond document length. Column (3) replaces the endogenous variable with the length-purged residual of the count, and column (4) with the governance share of substantive sentences. First-stage $F$ is shown for the leverage sample. The sample is U.S. operating firms excluding financial and utility firms, fiscal years 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| *Dependent variable* | Baseline count | Count, length control | Length-purged residual | Governance share |
|---|---|---|---|---|
| Book leverage | $-0.049^{*}$ | $-0.100^{**}$ | $-0.088^{*}$ | $-0.499^{*}$ |
| | (−1.67) | (−1.96) | (−1.68) | (−1.67) |
| Cash/assets | $+0.037^{**}$ | $+0.072^{**}$ | $+0.067^{**}$ | $+0.378^{**}$ |
| | (+2.19) | (+2.41) | (+2.21) | (+2.17) |
| Asset turnover | $+0.101^{**}$ | $+0.168^{**}$ | $+0.180^{**}$ | $+1.024^{**}$ |
| | (+2.06) | (+2.03) | (+2.11) | (+2.08) |
| Tobin's $Q$ | $-0.919^{***}$ | $-1.619^{***}$ | $-1.645^{**}$ | $-9.337^{**}$ |
| | (−2.60) | (−2.59) | (−2.54) | (−2.50) |
| First-stage $F$ | 44 | 52 | 47 | 46 |
| Firm, year FE | Yes | Yes | Yes | Yes |

# B. VARIABLE DEFINITIONS

**Table 19.** Variable definitions

| Variable | Definition |
|---|---|
| *Privacy capital* | |
| Privacy capital $P_{it}$ | ln(1 + governance sentences) from the firm's most recent privacy policy published during the fiscal year; governance categories {0, 2, 10, 11} (Appendix A). |
| *Financing and liquidity* | |
| Book leverage | (DLC + DLTT)/AT. |
| Market leverage | (DLC + DLTT)/(DLC + DLTT + ME). |
| Cash / assets | CHE/AT. |
| Current ratio | ACT/LCT. |
| *Efficiency, investment, profitability* | |
| Asset turnover | SALE/AT. |
| Capex / assets | CAPX/AT. |
| Return on assets | IB/AT. |
| Operating margin | OIADP/REVT. |
| Profit margin | IB/REVT. |
| Gross margin | (REVT − COGS)/REVT. |
| *Valuation* | |
| Tobin's $Q$ | (ME + AT − CEQ)/AT. |
| Total $Q$ | Total $q$ of Peters and Taylor (2017): market value over the replacement cost of physical *and* intangible capital (WRDS). |
| *Controls and firm characteristics* | |
| Log assets | ln(AT). |
| Book-to-market | CEQ/ME. |
| Intangibles / assets | INTAN/AT. |
| R&D / assets | XRD/AT (set to 0 if missing). |
| Advertising / assets | XAD/AT (set to 0 if missing). |
| Sales growth | $\Delta$SALE/SALE$_{t-1}$. |
| Firm age | Fiscal year minus IPO year (IPODATE). |
| *Market and microstructure (CRSP)* | |
| Annual return | Compounded daily return, delisting-adjusted (Shumway 1997: −0.30 imputed for missing performance delistings). |
| Return volatility | Standard deviation of daily returns within the year. |
| Share turnover | Annual share volume divided by shares outstanding. |
| Bid-ask spread | Corwin and Schultz (2012) high-low estimator. |
| Amihud illiquidity | Mean of \|return\|/dollar volume within the year. |
| *Data-capital and mechanism variables (Section VI)* | |
| Forecastability (depth) | Negative absolute residual from predicting the $t+1$ fundamental (earnings, revenue growth, cash flow, or inventory) from $t$-dated firm characteristics with industry-by-year fixed effects, so higher is more predictable (Section VI). |
| Customer-data breadth | Earnings-call share of language on collecting customer data; constructed in the mechanism section. |

# C. Additional results for the operating outcome

**Table 20.** The estimates do not depend on where the floor is placed

*Notes.* Each cell is the two-stage least-squares effect of privacy governance capital, instrumented by the firm's distance below the floor interacted with the rising floor, on the column outcome. The design, controls (log assets, book-to-market), winsorization (1/99), and fixed effects (firm and fiscal year) are held fixed across rows; only the dose definition changes. The primary row places the floor at the bound-frontier end-of-sample governance norm (3.17); the economy-$p$ rows place it at the indicated percentile of the 2016–2017 cross-section; the uncensored row keeps above-floor firms (no control group). The floor rows vary the dose's cross-sectional gap; the final row varies its time path instead, keeping the primary gap and replacing the wave path with the privacy-policy adoption rate of the Refinitiv ESG universe, an external vendor's measured diffusion over a different set of firms, rescaled to the primary path's range, so the pressure path is not an object of my construction; a path mixing the privacy and cybersecurity policy flags reproduces the same four columns. $t$-statistics in parentheses use standard errors clustered two-way by firm and fiscal year. The sample is U.S. operating firms excluding financial and utility firms ($n \approx 7,435$ firm-years, 1,654 firms), fiscal years 2010–2023. $^{*}$, $^{**}$, and $^{***}$ denote significance at the 10%, 5%, and 1% levels.

| | *Dependent variable* | | | | |
|---|---|---|---|---|---|
| Dose definition | Cash/ assets | Book leverage | Asset turnover | Tobin's $Q$ | First-stage $F$ |
| End-state floor (censored, primary) | $+0.037^{**}$ | $-0.049^{*}$ | $+0.101^{**}$ | $-0.919^{***}$ | 44 |
| | (+2.19) | (−1.67) | (+2.06) | (−2.60) | |
| Uncensored distance | $+0.037^{**}$ | $-0.048^{*}$ | $+0.101^{**}$ | $-1.009^{***}$ | 45 |
| | (+2.16) | (−1.66) | (+2.02) | (−2.78) | |
| Economy floor, p90 | $+0.038^{**}$ | −0.048 | $+0.098^{**}$ | $-0.926^{***}$ | 43 |
| | (+2.23) | (−1.60) | (+2.04) | (−2.62) | |
| Economy floor, p75 | $+0.043^{**}$ | −0.049 | $+0.088^{*}$ | $-0.934^{**}$ | 39 |
| | (+2.47) | (−1.54) | (+1.86) | (−2.57) | |
| Economy floor, p60 | $+0.051^{***}$ | $-0.061^{*}$ | +0.078 | $-0.971^{***}$ | 34 |
| | (+2.76) | (−1.82) | (+1.62) | (−2.62) | |
| External pressure path (Refinitiv) | $+0.038^{**}$ | $-0.056^{*}$ | $+0.098^{**}$ | $-0.945^{**}$ | 66 |
| | (+2.21) | (−1.93) | (+2.06) | (−2.57) | |

**Table 21.** The practice-distribution trim

*Notes.* The headline does not ride the tails of the practice distribution: dropping the extreme-practice firms leaves every coefficient intact and deepens the valuation discount. Panel B removes firms whose pre-period practice score—the 2016–2017 level the dose gap is built from—lies below the 5th or above the 95th percentile of the firm cross-section, so both the weakest-practice, largest-gap firms and the near-floor frontier firms leave; the 5th percentile is exactly zero, a mass point of zero-practice firms, so the lower cut removes that mass, about seven percent of firms, and 1,476 of 1,675 firms remain. Each cell keeps the specification, controls, fixed effects, winsorization, and two-way clustered standard errors of Table 3; Panel A re-estimates the full sample for side-by-side reading. The trim removes roughly a fifth of the dose's identifying variation, so the Panel B $t$-statistics are conservative; the coefficients themselves hold or strengthen everywhere, and the valuation discount deepens on both measures. The sample is U.S. operating firms excluding financial and utility firms, 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable* | | | | | | |
|---|---|---|---|---|---|---|---|
| | Asset turnover | Return on assets | Capex/ assets | Operating margin | Gross margin | Tobin's $Q$ | Total $Q$ |
| *Panel A. Full sample* | | | | | | | |
| Privacy capital $P_{it}$ (2SLS) | $+0.101^{**}$ | $+0.148^{***}$ | $+0.009^{**}$ | $-0.008$ | $+0.359$ | $-0.919^{***}$ | $-0.481^{*}$ |
| | (+2.06) | (+3.13) | (+2.07) | (−0.02) | (+1.31) | (−2.60) | (−1.90) |
| Reduced form (dose) | $+0.0438^{**}$ | $+0.0643^{***}$ | $+0.0040^{**}$ | $-0.0034$ | $+0.1623$ | $-0.3999^{**}$ | $-0.2086^{**}$ |
| | (+2.17) | (+3.02) | (+2.50) | (−0.02) | (+1.38) | (−2.30) | (−2.04) |
| First-stage $F$ | 44 | 44 | 43 | 52 | 52 | 44 | 46 |
| Observations | 7,434 | 7,434 | 7,424 | 7,103 | 7,103 | 7,435 | 7,299 |
| *Panel B. Interior-practice sample (bottom and top 5% of pre-period practice removed)* | | | | | | | |
| Privacy capital $P_{it}$ (2SLS) | $+0.117^{*}$ | $+0.228^{**}$ | $+0.014^{**}$ | $-0.164$ | $+0.641$ | $-1.426^{**}$ | $-0.751^{**}$ |
| | (+1.70) | (+2.46) | (+2.02) | (−0.18) | (+1.27) | (−2.46) | (−2.09) |
| Reduced form (dose) | $+0.0414^{*}$ | $+0.0808^{**}$ | $+0.0052^{**}$ | $-0.0614$ | $+0.2404$ | $-0.5055^{**}$ | $-0.2685^{**}$ |
| | (+1.76) | (+2.46) | (+2.43) | (−0.18) | (+1.27) | (−2.49) | (−2.26) |
| First-stage $F$ | 37 | 37 | 37 | 43 | 43 | 37 | 40 |
| Observations | 6,546 | 6,546 | 6,536 | 6,253 | 6,253 | 6,547 | 6,439 |
| Firm, year FE | | | | Yes | | | |

**Table 22.** State-by-year and industry-by-year robustness

*Notes.* The dose effect is not state regulatory timing, and its profitability and valuation core is not an industry trend. Each cell is the reduced form of the outcome on the dose under the controls, winsorization, and two-way (firm and fiscal year) clustered standard errors of Table 3, with the year effects replaced by the interacted effects named in the panel. Panel A interacts the headquarters state with the fiscal year, absorbing every state-level shock, including the arrival of the state privacy laws; the sample requires a headquarters state, the firm-and-year baseline is re-estimated on that sample in the row above, and the estimates are essentially unchanged. Panel B interacts the Fama-French twelve-industry grouping with the fiscal year on the full sample: the profitability response and the valuation discount are within-industry-year phenomena, investment holds at ten percent, and the efficiency response is absorbed; finer grids, two-digit and division, repeat the same split, investment excepted, which holds only on the coarse grid. Total $Q$'s within-industry component sits below conventional significance, so its discount evidence rests on the cross-sectional battery of Table 9 and the post-wave dynamics. The sample is U.S. operating firms excluding financial and utility firms, 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable* | | | | | | |
|---|---|---|---|---|---|---|---|
| | Asset turnover | Return on assets | Capex/ assets | Operating margin | Gross margin | Tobin's $Q$ | Total $Q$ |
| *Panel A. Headquarters-state-by-year effects* | | | | | | | |
| Dose, firm and year effects | +0.0510$^{**}$ | +0.0641$^{***}$ | +0.0040$^{**}$ | +0.0311 | +0.2028 | −0.3586$^{**}$ | −0.2159$^{**}$ |
| | (+2.42) | (+3.11) | (+2.19) | (+0.13) | (+1.55) | (−2.30) | (−1.99) |
| Dose, firm and state × year effects | +0.0523$^{**}$ | +0.0613$^{***}$ | +0.0042$^{**}$ | +0.1060 | +0.2460$^{*}$ | −0.3608$^{**}$ | −0.2021$^{*}$ |
| | (+2.48) | (+3.06) | (+2.06) | (+0.39) | (+1.69) | (−2.19) | (−1.79) |
| Observations | 6,659 | 6,659 | 6,651 | 6,367 | 6,367 | 6,660 | 6,535 |
| *Panel B. Industry-by-year effects (Fama-French 12)* | | | | | | | |
| Dose, firm and year effects | +0.0438$^{**}$ | +0.0643$^{***}$ | +0.0040$^{**}$ | −0.0034 | +0.1623 | −0.3999$^{**}$ | −0.2086$^{**}$ |
| | (+2.17) | (+3.02) | (+2.50) | (−0.02) | (+1.38) | (−2.30) | (−2.04) |
| Dose, firm and industry × year effects | +0.0148 | +0.0583$^{***}$ | +0.0030$^{*}$ | −0.1583 | +0.0893 | −0.2963$^{**}$ | −0.1178 |
| | (+0.79) | (+2.73) | (+1.90) | (−0.75) | (+0.83) | (−2.02) | (−1.12) |
| Observations | 7,434 | 7,434 | 7,424 | 7,103 | 7,103 | 7,435 | 7,299 |
| Controls, winsorization, clustering | As in Table 3 | | | | | | |

# D. ADDITIONAL RESULTS FOR THE DATA PRODUCTIVITY CHANNEL

**Table 23.** Privacy capital and the firm's data: the dual margin

*Notes.* The data channel, as a joint test across the dimensions of privacy practice. Each row instruments one dimension of privacy capital (governance or third-party sharing) on its own exposure-based shift (the firm's 2016–2017 mean times a leave-one-out industry shift) and estimates its effect on a measure of data depth or breadth, with firm and fiscal-year fixed effects and controls (log assets, book-to-market), whose coefficients the table reports alongside; the book-to-market column is scaled by $10^3$ for legibility. Each row is a separate regression and the row label names its dependent variable. Depth is the forecastability of the firm's own fundamentals, breadth the language firms use about collecting customer data in earnings calls and 10-Ks. The effect is the two-stage least-squares coefficient with its $t$-statistic below; "WCB $p$" is the shock-level wild-cluster bootstrap $p$-value and "RW $p$" the Romano–Wolf familywise $p$-value across the battery. Depth cells are predicted positive and breadth cells negative. The sample is U.S. operating firms excluding financial and utility firms, fiscal years 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| *Instrumented dimension → dependent variable* | Effect (2SLS) | Log assets | B/M | First-stage $F$ | WCB $p$ | RW $p$ | $N$ |
|---|---|---|---|---|---|---|---|
| *Panel A. Depth, predicted +: the firm uses data more intensively* | | | | | | | |
| Governance → earnings forecastability | $+0.079^{**}$ | $+0.054^{***}$ | $-0.384$ | 48 | 0.105 | 0.136 | 4,259 |
| | (+2.09) | (+3.43) | (−0.04) | | | | |
| Third-party sharing → earnings forecastability | $+0.096^{**}$ | $+0.054^{***}$ | +1.716 | 40 | 0.097 | 0.136 | 4,259 |
| | (+2.20) | (+3.38) | (+0.19) | | | | |
| *Panel B. Breadth, predicted −: the firm collects less* | | | | | | | |
| Governance → customer-data talk (calls) | $-0.072^{***}$ | +0.008 | −2.596 | 74 | 0.019 | 0.051 | 6,209 |
| | (−2.87) | (+0.77) | (−0.71) | | | | |
| Governance → customer-data orientation (10-K) | −0.231 | +0.076 | −9.264 | 59 | 0.019 | 0.136 | 6,422 |
| | (−1.48) | (+1.35) | (−0.55) | | | | |

# E. Additional asset-pricing results

**Table 24.** The effects are invariant to return volatility

*Notes.* Each cell is the distance-to-floor two-stage least-squares effect of privacy capital on the row outcome, adding the indicated control. Total and idiosyncratic volatility are the standard deviation of monthly returns and of market-model residuals within the fiscal year. If the financing and valuation effects were compensation for priced risk, they would attenuate once volatility is controlled; instead they barely move. The sample is the firm-years with return data, U.S. operating firms excluding financial and utility firms, 2016–2023; $t$-statistics in parentheses use two-way clustered standard errors. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | Baseline | + total volatility | + idiosyncratic volatility |
|---|---|---|---|
| *Dependent variable: Tobin's Q* | | | |
| Privacy capital $P_{it}$ | −0.489** | −0.495** | −0.501** |
| | (−2.15) | (−2.12) | (−2.13) |
| Log assets | −0.462*** | −0.446*** | −0.442*** |
| | (−3.14) | (−2.98) | (−2.94) |
| Book-to-market | −19.061 | −18.066 | −17.911 |
| | (−1.41) | (−1.39) | (−1.38) |
| Observations | 5,786 | 5,786 | 5,786 |
| *Dependent variable: Book leverage* | | | |
| Privacy capital $P_{it}$ | −0.059** | −0.059** | −0.060** |
| | (−2.51) | (−2.51) | (−2.53) |
| Log assets | +0.011 | +0.013 | +0.013 |
| | (+1.05) | (+1.19) | (+1.19) |
| Book-to-market | −5.879 | −5.781 | −5.783 |
| | (−1.61) | (−1.60) | (−1.59) |
| Observations | 5,786 | 5,786 | 5,786 |

**Table 25.** Ownership and the analyst margins under the canonical dose

*Notes.* Ownership and analyst responses to instrumented privacy capital under the canonical censored distance-to-floor dose of Section IV. Institutional ownership is the year-end 13F institutional share; coverage is the log number of analysts, dispersion the standard deviation of annual EPS forecasts over price, and forecast error the absolute consensus miss over price, from the last pre-period-end IBES consensus. Each column is a two-stage least-squares regression with firm and fiscal-year fixed effects, controls (log assets, book-to-market), outcomes winsorized at 1/99, and two-way clustered standard errors; the pre- and post-2020 columns split the sample, and the post-window first stage is weaker ($F = 16$), which the text discloses. Ownership rises with privacy capital only after 2020, about two and a half percentage points per within-firm standard deviation, while none of the analyst margins moves, the pattern of a re-rating without any accompanying revision in analyst cash-flow expectations. The sample is U.S. operating firms excluding financial and utility firms, fiscal years 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| | *Dependent variable* | | | | | |
|---|---|---|---|---|---|---|
| | Institutional ownership | | | Coverage | Dispersion | Forecast error |
| | 2016–23 | Pre-2020 | Post-2020 | 2016–23 | | |
| Privacy capital $P_{it}$ | +0.027$^{*}$ | +0.032 | +0.102$^{***}$ | +0.021 | +0.000 | −0.001 |
| | (+1.76) | (+1.14) | (+2.64) | (+0.57) | (+0.24) | (−0.20) |
| Log assets | +0.120$^{***}$ | +0.133$^{***}$ | +0.057$^{***}$ | +0.253$^{***}$ | −0.002$^{*}$ | −0.009$^{***}$ |
| | (+5.24) | (+3.68) | (+5.04) | (+11.52) | (−1.76) | (−3.54) |
| Book-to-market | −2.641 | −2.981 | +25.634$^{*}$ | −11.660 | +3.175 | +0.602 |
| | (−0.31) | (−0.34) | (+1.71) | (−0.76) | (+1.08) | (+0.12) |
| Per within-SD of $P$ | +1.1pp | +1.0pp | +2.5pp | | | |
| First-stage $F$ | 68 | 42 | 16 | 59 | 71 | 69 |
| Observations | 6,115 | 3,903 | 2,212 | 5,289 | 4,876 | 5,223 |
| Firms | 1,421 | 1,387 | 840 | 1,226 | 1,150 | 1,218 |
| Firm, year FE | Yes | Yes | Yes | Yes | Yes | Yes |

**Table 26.** Privacy capital does not raise the price of risk

*Notes.* Two-stage least-squares estimates of the effect of privacy capital on the price of risk, instrumented by the distance-to-floor dose, with firm and fiscal-year fixed effects and controls. The implied cost of equity is the PEG-ratio estimate; the cost of debt is interest expense over total debt; the breach hazard is an indicator that the firm suffers a data breach in the year. A priced-risk account predicts that privacy exposure raises the cost of capital and the hazard of the events it represents; none of the three is significantly positive, and the breach estimate is if anything protective. The sample is U.S. operating firms excluding financial and utility firms, 2016–2023. $^{*}$, $^{**}$, $^{***}$ denote 10%, 5%, 1%.

| *Dependent variable* | Privacy capital | Log assets | B/M | *t* | Obs. |
|---|---|---|---|---|---|
| Implied cost of equity | +0.003 | −0.003 | +36.483$^{***}$ | (+0.33) | 3,272 |
| Own data-breach hazard | −0.022 | +0.002 | +0.002 | (−1.16) | 7,435 |