Unintended Winners of Privacy: Data Protection and Labor Share

Conggang Li[a] Peigong Li[a] Linxuan Wu[b] Yifan Zhou[c]

November 21, 2025

Abstract

We develop a simple model in which data and skilled labor are imperfect substitutes in generating firm insight; a privacy shock raises the shadow cost of data and predicts substitution toward higher-quality labor and greater R&D. Using China's PIPL, we show digital-intensive firms raise labor share via higher employee compensation, tilt hiring toward experienced and R&D roles, and expand human-capital investments. Effects are strongest in more digitally focused firms and those with better governance, and are not explained by Covid-related shifts. Revenues remain broadly stable while costs rise, leading to compressed profitability.

Keywords: Privacy, data protection, GDPR, labor share, employee

[a]School of Accounting, Shanghai Lixin University of Accounting and Finance.

[b]Economics and Management School, Wuhan University.

[c]*Corresponding author.* Department of Finance, International Business School Suzhou, Xi'an Jiaotong-Liverpool University. Email: yifan.zhou@xjtlu.edu.cn.

We thank comments from Sudipta Basu, Wei Wang, Bochen Wu, and participants from the 2025 Temple-Lixin Accounting Research Conference. Authorship is alphabetically ordered; all authors contributed equally. Conggang Li acknowledges support from the National Natural Science Foundation of China (grant number 72304203). Yifan Zhou acknowledges support from the National Natural Science Foundation of China (grant number 72403054).

Email addresses: conggangli@lixin.edu.cn (C. Li), lpg37@163.com (P. Li), 2024201050046@whu.edu.cn (L. Wu), yifan.zhou@xjtlu.edu.cn (Y. Zhou).

Declarations of Interest

None for all coauthors.

1 Introduction

The relationship between data protection regulations and firm-level outcomes has garnered significant academic attention, particularly in the wake of sweeping legislation such as the General Data Protection Regulation (GDPR) in the European Union and similar laws in other jurisdictions. Not only do these regulations impact firms' financial performance, innovation, corporate strategy, they also carry wider economic implications (Goldfarb and Tucker, 2011; Miller and Tucker, 2009).

In this paper, we shed some light on the how the implementation of the the *Personal Information Protection Law* (PIPL)—a landmark legislation that fundamentally altered the landscape of consumer data governance and privacy in China—affect firm decisions and outcomes. Unlike previous privacy regulations, which mostly lay out high-level frameworks around privacy, this law categorically prohibited and criminalized the collection and use of consumer data without the consumer's explicit consent, a commonplace practice among Chinese firms.[1] Specifically, we examine how the PIPL impacts the labor market dynamics between firms and employees of affected industries. The PIPL represent a significant regulatory shift aimed at safeguarding data security and protecting personal information amidst the growing digital economy. As Chinese firms navigate the implications of these regulations, it is crucial to understand their broader impact on corporate dynamics.

[1] We provide a detailed discussion on the uniqueness of the PIPL in comparison to related laws in Section 4.

One area of particular interest is how these privacy and data security laws influence labor economics within firms. The labor share, defined as the ratio of labor expenses to revenue, provides a lens through which to examine this impact. By imposing stricter data management and protection requirements, the PIPL may lead to increased compliance costs and necessitate greater investments in specialized personnel, potentially altering the distribution of financial resources within firms (e.g., Johnson, 2022; Klein, Manini, and Shi, 2022; Blind, Niebel, and Rammer, 2024; Demirer, Hernández, Li, and Peng, 2024; Frey and Presidente, 2024).

In this paper, we adopt a difference-in-differences strategy to examine the treatment effect of the PIPL on corporate decisions and outcomes, with an emphasis on labor share. Defining the treatment group as firms in industries classified as "core to the digital economy" by the National Statistics Bureau of China in its *Classification of Digital Economy and its Core Industries (2021)*, we find core firms' labor share increases by 1.3 to 1.4 percentage points following the introduction of data protection regulations. Moreover, we discover that this increase is driven almost exclusively by employees', rather than managers', labor share. Given that labor expense is the product of employee number and average pay (and that we find no significant treatment effect on firms' revenue), we then separately study these. Our analysis indicates that firms in digital-core industries do not significantly adjust the size of their executive team or workforce, but increase average employee compensation (but not manager compensation), by approximately two percent following the implementation of data protection regulations.

Theoretically, data can be viewed as a firm-specific intangible asset that enhances pre-

dictive accuracy, personalization, and innovation (Bharadwaj, 2000; Bharadwaj, Anandhi and El Sawy, Omar A and Pavlou, Paul A and Venkatraman, N v, 2013). When such data access is curtailed, firms must seek alternative inputs capable of restoring competitive advantage. Human capital represents the closest functional substitute, as innovation, product refinement, and algorithmic improvement ultimately depend on human expertise rather than physical capital (Aral, Brynjolfsson, and Wu, 2012), for which we do not observe a significant increase among digital-core firms post-PIPL. The ability to compensate for data loss rests primarily on skilled labor rather than machines or capital equipment (Lambrecht and Tucker, 2015). Supporting this view, technology leaders frequently assert that talent is their most essential resource. For example, at the 2025 Bloomberg Tech conference, Alphabet CEO Sundar Pichai stressed the importance of workers and his plan to expand its engineering workforce in 2026, despite its investments in AI.[2] This conceptual foundation helps explain why we observe labor-for-data substitution.

We formalize these ideas in a simple conceptual framework (Section 3). The model treats output as depending on an "insight" component produced by data and skilled labor that are imperfect substitutes. A tighter privacy regime raises the marginal cost of data, so firms optimally substitute toward skilled labor. Because labor enters in efficiency units, the adjustment loads on worker quality and pay more than on headcount, predicting higher average compensation and a tilt toward experienced and R&D roles. A dynamic extension with capability accumulation implies greater investment in human-capital for-

[2]Mint by Govind Choudhary, June 5, 2025, "Google to expand engineering workforce by 2026 amid AI investments: CEO Sundar Pichai". Retrieved from https://www.livemint.com/technology/tech-news/google-to-expand-engineering-workforce-by-2026-amid-ai-investments-ceo-sundar-pichai-11749122073860.html.

mation and retention (e.g., training, benefits, and ownership incentives) and higher R&D effort.

An intuitive way to interpret the positive treatment effect on labor share, especially in light of higher average employee compensation, is that digital-core firms are placing greater strategic value on their workforce, now that the ability to exploit consumer data for competitive advantage has been curtailed. We test this hypothesis by focusing on talent recruitment and retention. First, we analyze changes in talent recruitment strategy using job advertisement data collected from major online recruitment platforms. Following the implementation of data protection regulations, digital-core firms significantly increase the share of job postings requiring postgraduate qualifications, offering high monthly salaries (exceeding 10,000 CNY in 2018-adjusted terms), and substantial prior work experience. These patterns suggest that firms are actively seeking to attract more skilled and experienced workers, consistent with an increased reliance on human capital. More importantly, we find that digital-core firms increase hiring for R&D-related positions—such as software development, testing, and engineering—suggesting that firms substitute high-skilled labor for data as an alternative source of competitive advantage.

Second, we examine retention responses. If labor substitutes for data, we should see not only higher pay but broader profit-sharing and deeper investment in key staff. Consistent with this, we find digital-intensive firms are more likely to introduce employee stock ownership plans (ESOPs) after the PIPL, and both employee participation and subscription shares rise. We then look at employee turnover as an imperfect proxy for job satisfaction, and find a significant decline among treated firms. Finally, consistent

with expanded benefits and upskilling for employees, we observe a significant increase in fringe benefits plus training costs as a fraction of operating expenses for digital-core firms following the introduction of the PIPL. Viewed alongside our recruitment evidence, these patterns indicate a coordinated retention strategy focused on skilled workers and suggest that data protection regulations prompt firms to upgrade human capital and strengthen employee incentives, yielding gains in both compensation and job security.

To further support our hypothesis that digital-core firms are actively substituting labor for data to regain lost competitive advantage, we examine changes in firms' R&D investments. Specifically, we test whether digital-core firms expand their R&D efforts following the introduction of China's PIPL. In line with the observed rise in R&D-related hiring, we find that these firms not only increase R&D spending as a share of revenue, but also achieve higher R&D productivity, with revenue per R&D employee rising by 2.57 million CNY. These results suggest that firms are not merely complying with regulatory constraints, but are strategically reallocating resources toward high-skilled labor and innovation to rebuild capabilities previously supported by unrestricted access to consumer data. The increase in revenue per R&D worker underscores that this substitution is targeted and skill-intensive, consistent with a shift toward internal innovation as a new source of competitive advantage (Yoo et al., 2010). This interpretation aligns with Blind, Niebel, and Rammer (2024), who document that firms respond to GDPR by reorienting innovation strategies toward enhancing product offerings within the bounds of new privacy regulations. Our findings also resonate with evidence from European content providers, who responded to the GDPR through data minimization and strategic

consent mechanisms to maintain content quality (Lefrere, Warberg, Cheyre, Marotta, and Acquisti, 2025).

Additionally, we explore potential heterogeneous treatment effects on labor share to provide supporting evidence for our story. Specifically, we find the treatment effect is stronger in firms more concerned with digitalization. This finding is intuitive—firms with a stronger digitalization focus would likely have collected and utilized consumer data to a greater extent compared to those less concerned with digitalization. Therefore, restricting data usage would likely lead to a strong need to quickly substitute labor for data.

A potential concern is that these effects reflect pandemic-driven shifts in labor demand rather than a response to privacy regulation. However, multiple tests rule out Covid-19 as the primary driver. First, the divergence in labor share between treated and control firms begins in 2021, coinciding with the enactment of the PIPL, rather than in 2020, when pandemic disruptions peaked. Second, the treatment effect remains when absorbing regional pandemic shocks using province-year fixed effects. Third, in a triple-difference framework interacting treatment status with local Covid-19 severity (total confirmed cases and deaths in 2020 at the firm's headquarter province), the triple-interaction terms are insignificant. By contrast, interacting treatment with pre-PIPL digital intensity yields a positive and significant triple-interaction effect, consistent with the mechanism that firms most reliant on digital operations, not firms in regions most affected by Covid-19, drive the labor-data substitution response.

As an additional falsification exercise, we test whether our findings merely reflect broader innovation policy rather than privacy regulation. In particular, China amended its *Scientific and Technological Progress Law* beginning in 2022 to strengthen national support for R&D investment and high-tech talent. If our results were driven by this policy, we would expect firms officially recognized as high-tech enterprises to exhibit similar increases in labor share. However, when we replace the treatment indicator with a *High-Tech* status indicator, we find no significant effect, suggesting that the rise in labor share is specific to data protection regulation and not a byproduct of general innovation-promotion initiatives.

Having established that digital-core firms engage in labor-data substitution following the introduction of the PIPL, we conclude our empirical analysis by examining the effectiveness of this strategy in terms of its impact on firm performance. We find that labor is an effective substitute for consumer data when it comes to revenue generation, as digital-core firms' revenue did not significantly decrease following the introduction of data protection regulations. However, labor is more expensive than data and negatively affects the same firms' bottom line. For example, these firms' cost of sales ratio increases by almost one percentage point, or 11 percent, after the introduction of data protection regulations. Consistently, treated firms' EBIT-over-revenue also decreases. These findings echo the EU GDPR's affect on firms. For example, Frey and Presidente (2024) find the introduction of the GDPR in the EU harmed technology firms' profitability, leading to a decline in profits without affecting sales. Similarly, Demirer et al. (2024) argues that the GDPR led to increased operational costs and a re-evaluation of resource allocation,

particularly among firms heavily reliant on data.

Our study contributes to several strands of the literature. First, we add to the emerging body of research on the economic consequences of data protection regulations such as the GDPR (e.g., Johnson, 2022; Demirer et al., 2024; Frey and Presidente, 2024), by providing the first comprehensive firm-level evidence from China's PIPL. Second, while prior work has largely focused on firm performance, innovation, or market structure, we highlight a novel channel—labor share and human capital investment—through which data regulations reshape corporate resource allocation. Third, by documenting changes in both recruitment and retention practices, as well as heterogeneity by digitalization intensity and governance, we provide micro-level evidence on firms' substitution of data with skilled labor, an aspect largely overlooked in the privacy regulation literature. Finally, our analysis contributes to the literature on the determinants of labor share (e.g., Autor, Dorn, Katz, Patterson, and Van Reenen, 2020; Barkai, 2020), by identifying data protection laws as a new regulatory factor influencing labor's share of firm value.

2 Related Literature

The introduction of significant data protection regulations, such as the EU's GDPR, has reshaped firm performance, governance, and market structure. Firms face compliance frictions and strategic reallocations (Johnson, 2022); boards with higher *ex-ante* cyber risk add expertise and tighten oversight (Klein et al., 2022); website traffic and reported revenue fall under GDPR (Goldberg, Johnson, and Shriver, 2024); and vendor portfolios

initially contract and then concentrate (Johnson, Shriver, and Goldberg, 2023). In China's digital economy, curbs on personal-data use reduce engagement and transactions (Sun, Yuan, Li, Zhang, and Xu, 2024). Relatedly, Lattanzio and Taillard (2022) find lower cyber-security risk facilitates M&A with better completion terms and post-deal outcomes. In media markets, stricter privacy can induce more polarized ideological positioning when behavioral tracking is banned (Bondi, Rafieian, and Yao, 2025).

Privacy also shapes innovation. In mobile apps, GDPR depresses entry and availability (Janssen, Kesler, Kummer, and Waldfogel, 2022), yet by lowering concentration it can foster competitive innovation (Niebel, 2021). Rising cyber exposure shifts firms from trade secrets toward patents and shorter innovation cycles (Lattanzio and Ma, 2023). Among Chinese AI firms, Beraja, Yang, and Yuchtman (2023) find access to data-rich contracts substantially boosts commercial AI output.

Acquisti, Taylor, and Wagman (2016) establish the core trade-off in the economics of privacy between protecting consumers and maintaining operational efficiency. Goldfarb and Tucker (2019) argue that privacy rules raise the costs of digital activity and reshape market structure. Goldfarb and Que (2023) extends this perspective, emphasizing the need to balance privacy protection with data-driven innovation. Consistent with these themes, incentive models predict stronger investment in data protection when revenues are data-driven rather than usage-driven (Fainmesser, Galeotti, and Momot, 2023).

Our paper also relates to determinants of labor share. In the US, the decline cannot be fully explained by neoclassical substitution, implicating offshoring and supply-chain

shifts Elsby, Hobijn, and Şahin (2013). Globally, Karabarbounis and Neiman (2014) document cheaper investment goods from IT account for about half of the decline. Autor et al. (2020) argue that superstar-firm dynamics further depress labor share through markups and concentration. Institutional forces matter: deregulation weakens worker bargaining power (Ciminelli, Duval, and Furceri, 2022), lower corporate taxes favor capital-intensive firms (Kaymak and Schott, 2023), and concentration and oligopsony reduce labor share and welfare (Smith, Yagan, Zidar, and Zwick, 2022; Berger, Herkenhoff, and Mongey, 2022). Technology plays a central role via automation and capital deepening (Dao, Das, Koczan, and Lian, 2017; Autor and Salomons, 2018; Acemoglu and Restrepo, 2020; Oberfield and Raval, 2021; Bergholt, Furlanetto, and Maffei-Faccioli, 2022). At the same time, capital–labor complementarity and pure-profit dynamics imply rising profits can absorb value added even when labor share stabilizes or rises in some settings Barkai (2020); Bergholt et al. (2022).

3 Conceptual Framework

Privacy regulation limits a firm's ability to extract value from consumer data. When such rules tighten, the firm must restore the "insight" it uses for prediction, targeting, and product improvement by reallocating toward inputs that can play a similar functional role. We model insight as produced jointly by data and skilled labor. When regulation raises the shadow cost of data or constrains its use, the firm substitutes toward skilled labor in order to preserve insight. Interpreting skilled labor in efficiency units naturally

links the adjustment to higher average pay and a shift toward experienced and R&D roles. Extending the framework to a dynamic setting where capability accumulates through skilled labor explains increased investment in training, benefits, and retention policies and connects to higher R&D intensity.

Consider a firm that produces output Y at price $p > 0$ from physical capital K and an "insight" aggregate I:

$$Y = A\,K^{\beta}\,I^{1-\beta}, \qquad A > 0,\ \beta \in (0,1), \tag{1}$$

where A is total factor productivity and K is rented at rate r. The insight aggregate I is produced from two inputs: effective data $D \geq 0$ and skilled labor efficiency units $H \geq 0$. Skilled labor is paid a wage $w > 0$ per efficiency unit. Let

$$I = \left[\alpha D^{\rho} + (1-\alpha)\,(\theta H)^{\rho}\right]^{1/\rho}, \qquad \rho \leq 1,\ \alpha \in (0,1),\ \theta > 0. \tag{2}$$

Here α measures pre-regulation digital intensity (the weight on data), θ scales the effectiveness of skilled labor in producing insight, and ρ governs the elasticity of substitution $\sigma = 1/(1-\rho)$. When $\rho < 1$, data and skilled labor are imperfect substitutes.

Privacy regulation is summarized by a stringency parameter $\tau \geq 0$. The total cost of using data is $\kappa(D,\tau)$, increasing and convex in D and increasing in τ:

$$\kappa_D(D,\tau) > 0, \qquad \kappa_{DD}(D,\tau) \geq 0, \qquad \kappa_{\tau}(D,\tau) > 0.$$

A simple parametric case is $\kappa(D,\tau) = (a_0 + a_1\tau)D + \frac{b}{2}D^2$ with $a_0, a_1, b > 0$, so the marginal cost $\kappa_D = a_0 + a_1\tau + bD$ rises with both D and τ. One may also impose a feasibility cap $D \le \overline{D}(\tau)$ with $\overline{D}'(\tau) < 0$; the results below carry through.

Static profits are

$$\Pi = p\,A\,K^{\beta}\,I^{1-\beta} - w\,H - r\,K - \kappa(D,\tau), \tag{3}$$

and the firm chooses D and H (and K, which is not central for the substitution margin) to maximize Π given (1) and (2). Define the marginal value of insight,

$$M \equiv p\,A\,(1-\beta)\,K^{\beta}\,I^{-\beta}.$$

From (2),

$$\frac{\partial I}{\partial D} = \alpha\,D^{\rho-1}\,I^{1-\rho}, \qquad \frac{\partial I}{\partial H} = (1-\alpha)\,\theta^{\rho}\,H^{\rho-1}\,I^{1-\rho}.$$

The first-order conditions equate marginal benefits and marginal costs:

$$M\,\frac{\partial I}{\partial D} = \kappa_D(D,\tau), \qquad M\,\frac{\partial I}{\partial H} = w. \tag{4}$$

Eliminating the common factor $M\,I^{1-\rho}$ yields the relative condition

$$\frac{(1-\alpha)\,\theta^{\rho}\,H^{\rho-1}}{\alpha\,D^{\rho-1}} = \frac{w}{\kappa_D(D,\tau)}. \tag{5}$$

An increase in τ raises the marginal cost of data $\kappa_D(D,\tau)$ at any given D, and lowers the

right-hand side of (5). With $\rho < 1$, the left-hand side must fall to restore equality, which requires an increase in H relative to D. Hence a tightening of privacy regulation induces substitution away from data and toward skilled labor in the insight aggregator (Table 2):

$$\frac{\partial H^*}{\partial \tau} > 0, \qquad \frac{\partial D^*}{\partial \tau} < 0, \qquad \text{for } \rho < 1.$$

The strength of this response is larger when α is higher, since a more data-reliant firm experiences a larger loss on the data margin that must be offset through labor (Table 7). Capital does not undo the privacy shock in this framework because the regulation operates within $I(\cdot)$ rather than on K; increasing K scales output but does not restore the constrained insight channel (Table A3).

To connect the substitution to wage-based changes in labor share, interpret skilled labor in efficiency units. Write $H = qN$, where N is headcount and q is average worker quality (e.g., skill or experience). Suppose individual wages increase with quality, for example a worker of quality q is paid $w(q)$ with $w'(q) > 0$ and $w''(q) \geq 0$. To expand H after a privacy shock, the firm can hire more workers or upgrade the quality mix. With convex wage–quality schedules and standard recruitment frictions, the cheapest way to raise efficiency units quickly is to increase q more than N. The adjustment in $\partial H^*/\partial \tau$ therefore shows up primarily as higher average pay and a tilt toward experienced and R&D-oriented roles rather than large changes in headcount (Tables 3 and 4). This mechanism links the model to an empirical rise in labor share via average compensation, together with shifts in job requirements toward postgraduate degrees and longer prior

experience (Table 4).

A dynamic extension ties substitution to capability accumulation and R&D. Let $S(t)$ denote a capability stock (e.g., productionized code, documented know-how, model pipelines, and tacit routines) that is accumulated by skilled labor according to

$$\dot{S}(t) = \phi\, H(t) - \delta\, S(t), \qquad \phi > 0,\ \delta > 0. \tag{6}$$

Let insight depend on capability rather than contemporaneous labor,

$$I(t) = \left[\alpha D(t)^{\rho} + (1-\alpha)\,(\theta S(t))^{\rho}\right]^{1/\rho}, \tag{7}$$

and let per-period profits be

$$\pi(t) = p\, A\, K^{\beta}\, I(t)^{1-\beta} - w\, H(t) - r\, K - \kappa(D(t), \tau).$$

The firm chooses paths $D(t)$ and $H(t)$ to maximize $\int_0^{\infty} e^{-rt}\pi(t)\, dt$ subject to (6) and $S(0)$ given. An increase in τ raises the shadow cost of data and induces both an immediate rise in $H(t)$ and an investment in capability $S(t)$ so that future insight depends less on raw data. In a stationary allocation with $\dot{S} = 0$, the capability target satisfies $S^* = (\phi/\delta)H^*$. A permanent tightening of privacy regulation therefore implies a higher target S^* and a persistently higher demand for skilled labor to offset depreciation (Tables 4, 5, and 6). During the transition, revenue can be stabilized by maintaining $I(t)$, while profits compress because wages, training, and retention costs rise (Tables 5 and 10).

Taken together, the framework delivers three distinct implications. First, a tightening of privacy that raises the marginal cost of data induces substitution away from data and toward skilled labor in the insight aggregator, with stronger responses among firms that were more data-reliant ex ante; capital deepening does not offset this shock because it operates outside the insight channel. Second, because skilled labor enters in efficiency units, the adjustment loads primarily on the quality/compensation margin, rather than on headcount (Tables 3 and 4). Third, when insight draws on an accumulated capability, stricter privacy raises the value of building and preserving that capability over time, prompting greater investment in human-capital formation (Table 5) and retention and stepped-up R&D effort (Table 6), which helps stabilize revenue while increasing the wage bill in transition (Table 10).

4 Setup and Data

4.1 Institutional Background

China's legislative landscape regarding data security and privacy has evolved significantly over the last decade, culminating in the enactment of the PIPL in 2021. The PIPL was introduced to address the complexities of data management in the digital age, setting stringent guidelines for data handling and imposing severe penalties for non-compliance.

Since the booming of China's e-commerce industry, China has been steadily rolling out regulations for the digital age. However, prior regulations were not adequate in

protecting consumer data. For example, the *Cybersecurity Law*, enacted in November 2016, focused primarily on protecting the network security infrastructure.

In contrast, the PIPL takes a detailed and stringent approach by setting specific requirements for personal data processing. It emphasizes the necessity of obtaining explicit consent from data subjects (e.g., customers) before collecting or using their personal data, adhering to the principles of data minimization and purpose limitation. This law also expanded the rights of individuals, granting them extensive control over their personal data through rights to access, rectification, and deletion, thus imposing greater operational demands on businesses to align their data practices with these enhanced rights.[3]

A pivotal case that illustrates the impact of the PIPL is that of Didi Chuxing, China's largest ride-hailing platform. In July 2022, after a one-year investigation, Didi was fined a record amount of over 8 billion CNY (approximately 1.2 billion USD) for violations of cybersecurity and data protection laws, including the PIPL. The investigation revealed that Didi had engaged in the illegal collection and use of consumer personal data, including failing to obtain adequate consent from users and inadequately anonymizing personal data before use.[4] The company's CEO and chairman were also fined one million CNY

[3]Another relevant and contemporaneous legislation is China's *Data Security Law* (DSL), enacted in June 2021. While the DSL also influences how firms collect, store, and manage consumer information, its primary orientation is toward safeguarding data as a strategic resource at the national level. In contrast, the PIPL, enacted in August 2021, directly regulates the collection and processing of personal information, emphasizing consent, purpose limitation, and transparency. Accordingly, the PIPL provides the most direct and identifiable regulatory shock to firms' consumer data practices examined in this study. The distinction between the two laws is immaterial to our analysis, which concerns the broader firm-level implications of heightened data regulatory pressure. Given their temporal proximity and substantive overlap, we do not endeavor to isolate their individual effects.

[4]Reuters by Julie Zhu, Yingzhi Yang, and Kane Wu, July 22, 2022, "China fines Didi $1.2 bln but outlook clouded by app relaunch uncertainty". Retrieved from https://www.reuters.com/technology/china-fines-didi-global-12-bln-violating-data-security-laws-2022-07-21/.

(approximately 147,000 USD) each.[5] In addition to the financial penalties, the company's mobile applications was removed from China's app stores, a move that Didi said "will affect business."[6]

These enforcement actions against Didi signaled a clear message to all businesses operating in China about the government's intent on clamping down data privacy violations. With the introduction of the PIPL, companies not only face higher compliance costs, such as acquiring consumer consent and data encryption, but likely also lose some of their existing revenue channels that rely heavily on exploiting consumer data. Collectively, these may lead affected companies to explore other avenues to compensate for higher costs and potential loss of revenue.

4.2 Data

Our sample period is between 2018 to 2023, giving us three years each side of the data protection regulations. The PIPL applies to all firms without prejudice. This imposes a problem for us in testing these regulations' effect; dividing the sample into pre- and post-regulation periods does not allow us to measure the effect without intervention or account for spontaneous changes. However, one way to address this is to acknowledge that these regulations predominantly affect data-intensive firms. As a counter example, a dairy farm that sells all its products to Yili (China's largest dairy producer) is unlikely to

[5]CNN by Yong Xiong, Larry Register, and Laura He, July 21, 2021, "China fines Didi $1.2 billion for violating cybersecurity and data laws". Accessed from https://edition.cnn.com/2022/07/21/economy/china-fines-didi-data-law-violation-intl-hnk/index.html.

[6]BBC, July 5, 2021, "Didi says removal of app in China will affect business". Accessed from https://www.bbc.com/news/business-57716131.

be impacted by these data protection regulations.

Leveraging on the nature of these regulations, we divide our sample into treatment and control groups. A firm is categorized as treated if it belongs to an industry defined as "core to the digital economy" by the China's National Bureau of Statistics in its *Classification of Digital Economy and its Core Industries (2021)*, and is control otherwise. The classification divides digital economy industries into five categories: digital product manufacturing, digital product services (including digital product wholesale, retail, and rental), digital technology application (including software development, Internet-related services, and IT services), digital factor-driven (including online platforms, online finance, and data vendors), and digital efficiency improvement. The first four are defined as "core to the digital economy", covering industries which rely heavily on data elements. For example, Perfect World (*Wanmei shijie* in Chinese) belongs to the the digital technology application industry, as it needs to continuously collect user data for the development and optimization of its online games; Mango Excellent Media (aka Mango TV) belongs to the digital factor-driven industry, as it needs to collect audience rating data to understand their preferences and to determine the theme of radio and television program production. Industry data are obtained from WIND. Firms' financial data are obtained from CSMAR. Continuous variables are winsorized at the 1 and 99 percent level.

5 Empirical Results

5.1 Data Protection Regulations and Labor Share

To examine the effects of China's implementation of the PIPL in 2021 on firms, we adopt a classic two-way fixed effects difference-in-differences model.[7] Specifically, as these legislation were originally designed amid the explosive growth of China's e-commerce platforms and primarily focus on the protecting consumer privacy, we define our treatment group to be firms in industries core to the digital economy, as classified in the *Classification of Digital Economy and its Core Industries (2021)* issued by the National Bureau of Statistics of China. Formally, we adopt the following baseline model:

$$LS_{i,t} = \beta_1 Core_i Post_t + \gamma' X_{i,t-1} + \alpha_i + \delta_t + \varepsilon_{i,t}, \tag{8}$$

where i and t index firm and year, respectively. LS is the firm's labor share, defined as cash labor expense (including salaries, wages, and benefits) divided by revenue. *Core* indicates if the firm belong to a digital-core industry, as discussed above. *Post* indicates if the year is after 2021, the year in which China introduced the PIPL. X is a vector of controls, including the natural logarithm of market value of assets (*Ln(Assets)*), ROA, leverage, fraction of common shares owned by the firms' executives (*Mgmt Hold*), the proportion of independent directors (*Indep Dir*), revenue growth from $t-1$ to t, the natural logarithm of the firm's age (*Ln(Firm Age)*), and property, plant, and equipment

[7] Our setting does not have any variation in treatment timing.

normalized by total assets (*PPE/Assets*). Detailed variable definitions can be found in Table A1 in the Internet Appendix. We adopt a six-year sample period between 2018 and 2023 for symmetry (regulations were implemented in 2021) and to avoid confounding with earlier cybersecurity reforms, most notably the *Cybersecurity Law*.[8]

Before delving into the results of equation (8), we first perform a parallel trends test. Figure 1 shows that the labor share of digital-core and non-core firms are almost perfectly parallel in the years leading up to 2021, but unlike non-core firms, core firms' labor share exhibit a clear and steeper upward trend after 2021, indicating a positive treatment effect.

Having passed the parallel trends test, we move to the results of equation (8) presented in Table 2. Consistent with Figure 1, we observe that digital-core firms' labor share increases by an average of 1.3 to 1.4 percentage points (between 6.2 and 6.7 percent in relative terms) following the implementation of data protection regulations (columns 1 and 2).[9] We then separately examine whether this increase is driven by manager or employee labor share by sequentially replacing the dependent variable in equation (8) with manager labor share (*LS (Mgmt)*) and employee labor share (*LS (Ee)*). The results indicate that the positive treatment effect of data protection regulations on labor share

[8]The Cybersecurity Law represents China's first comprehensive legislation governing data processing and network security. Unlike the PIPL, however, it primarily establishes security obligations for network operators rather than explicitly prohibiting the collection and commercial use of personal data without consent. As a robustness exercise, we replicate our baseline specification around the introduction of the *Cybersecurity Law*, adjusting the sample period to 2014-2019. Consistent with our main findings, we observe a statistically significant but economically smaller increase in labor share for core firms after 2016 (Table A2). This pattern is consistent with the *Cybersecurity Law* affecting data-driven business models, but doing so less directly and less forcefully than the PIPL, which explicitly restricted firms' ability to collect and monetize personal data without consent.

[9]We conduct a placebo test in which we randomly draw firms from our sample to create a placebo treatment group, before re-estimating equation (8). We repeat this process 500 times and plot the distribution of β_1. Figure A1 in the Internet Appendix shows that our true treatment effect is to the far-right of the distribution of placebo effects.

is almost entirely driven by increases in employee labor share (Table 2 columns 3 to 6). In other words, our results indicate that employee of digital-core firms benefit from the introduction of data protection regulations in China.[10]

We also examine a dynamic version of our baseline model:

$$LS_{i,t} = \sum_{k=-3}^{2} d_k D_{i,t}^{k} + \gamma' X_{i,t-1} + \alpha_i + \delta_t + \varepsilon_{i,t}, \tag{9}$$

where i and t index firm and year, respectively. The dummy variable, $D^k, k \in \{-3, -2, 0, 1, 2\}$, equals to one for digital-core firms in years relative to 2021 (i.e., 2021 is defined as year 0) and zero otherwise. We then plot the point estimate (d_k) of each sub-period and the associated 95 percent confidence interval calculated using standard errors clustered by firm. The point estimate immediately before the event year is normalized to zero following Freyaldenhoven, Hansen, and Shapiro (2019). Notwithstanding the relative short sample period, we observe from Figure 2 a lack of obvious pre-trend and a dramatic increase in the point estimate subsequent the introduction of data protection regulations.

Figure 2 also sheds some light on a potential alternative explanation of our results. Similar to large technology companies in the US, revenues and profits of Chinese e-commerce platforms and Internet companies rose significantly in the wake of the Covid-19 pandemic and subsequent lockdowns. Consequently, it is not difficult to imagine these

[10]One may ask whether firms respond to privacy regulation by substituting capital investment (e.g., automation, physical or digital infrastructure) for data, rather than labor. We re-estimate equation (8) replacing labor-share outcomes with measures of capital expenditures (logged, normalized by revenue, and normalized by assets). The coefficient of the interaction term is statistically insignificant across specifications, suggesting that capital investment does not increase in response to the regulation (Table A3). This pattern supports our interpretation that firms substitute toward human capital, rather than physical capital, when data resources are constrained.

companies, all of which are core to the digital economy, sharing some of these gains with employees, thus resulting in higher labor costs. However, there are two reasons why this is unlikely to be driving our findings. First, our measure of labor costs is normalized by revenue. Therefore, unless these firms' increase in labor expense is more than proportionate to the increase in revenue, labor share would not increase. Second, Figure 2 shows a monotonic increase in the treatment effect as we progress in time from 2021 to 2023, by which time the profits of these Internet companies had declined significantly compared to 2021. For example, Alibaba, China's largest e-commerce company, saw its net profits decrease from 150 billion CNY in to 73 billion CNY in 2023.[11] As such, we largely rule out digital-core firms distributing their higher revenues during Covid-19 to employees as an explanation of our results.

Having documented that non-executive employees are responsible for a general increase in the labor share of digital-core firms, we drill deeper into the two components of labor share—employee number and average pay. If data protection regulations lead firms hiring more employees, it is indicative of a higher compliance burden. On the other hand, if average employee pay is increasing, it implies higher-skilled workers are becoming more valuable in the absence of the ability to exploit consumer data (e.g., for targeted advertising). This would reflect a general increase in labor costs driven by greater competition for talent. We may also infer that, in our context and to a certain extent, labor and data are substitute factors of production.

To investigate our question, we replace the dependent variable in equation (8) with

[11] Yahoo Finance, "Alibaba Group Holding Limited (BABA)". Accessed from https://finance.yahoo.com/quote/BABA/financials/.

the natural logarithm of employee number (*Ln(Ee Number)*) and the natural logarithm of average employee pay (*Ln(Avg Pay Ee)*). For completeness, we also included executive number (*Ln(Mgmt Number)*) and average executive pay (*Ln(Avg Pay Mgmt)*) as dependent variables. The results are presented in Table 3. Consistent with Table 2's finding that labor share of managers was not affected by the passage of data protection regulations, we observe no significant changes in both the number of executives or their average pay (columns 1 to 4). Moving to employee number and average pay, we find that the positive treatment effect on employee labor share is driven by higher average pay, rather than increases in employee number (columns 5 to 8). Specifically, the treatment effect on average employee pay is roughly two percent. This finding aligns with the idea that in the absence of the ability to exploit consumer data, labor' market value increase.

5.2 Talent Recruitment and Retention

In this subsection, we present additional evidence supporting the notion that the positive treatment effect reflects digital-core firms placing greater value on skilled workers. We begin by examining firms' talent recruitment activities. Specifically, we compile job advertisements posted by our sample firms during the study period from major Chinese recruitment platforms, including 51job, Liepin, Wuba, Zhilian Zhaopin, and BOSS Zhipin. For each job advert, we extract information on the offered salary (using the midpoint if a range is provided), minimum work experience requirement, and academic qualification requirement. These data are aggregated to the firm-year level to construct five measures capturing recruitment targeting skilled and experienced workers: the percentage of post-

ings requiring postgraduate degrees (*Postgrad Pct*), the percentage of postings offering salaries above 10,000 CNY per month (*High Income Pct*),[12] the percentages of postings requiring at least five and ten years of work experience (*Min 5 Pct* and *Min 10 Pct*), and the percentage of postings for R&D-related jobs, including IT research, development, engineering, testing, automation, and system design (*R&D Pct*). Each measure is normalized by the total number of job postings by the firm in the corresponding year.

We then re-estimate equation (8) using each of these five recruitment measures as dependent variables. The results, reported in Table 4, are consistent with our conjecture: digital-core firms increase recruitment along all four dimensions. For instance, the share of postings requiring postgraduate degrees rises by 0.7 percentage points following the implementation of the PIPL (column 1 and 2). Similarly, these firms increase the proportion of higher-paying roles (columns 3 and 4) and raise experience requirements (columns 5 to 8). Taken together, these findings indicate that digital-core firms place greater emphasis on attracting skilled and experienced workers, reinforcing the notion that such firms place increased value on human capital in the post-regulation period.

In addition to general indicators of skill intensity, we further examine whether digital-core firms increase recruitment specifically in R&D-related positions—roles that are critical for product innovation and technological adaptation. To do so, we identify job postings related to software and IT research, development, engineering, testing, automation, and system design, and calculate their share as a fraction of total job advertisements

[12] According to a 2019 report by the National Bureau of Statistics of China, individuals earning over 10,000 CNY per month in 2018 are considered high-income (see https://www.stats.gov.cn/sj/sjjd/202302/t20230202_1896141.html). We adopt this definition and adjust salaries using the CPI. The results are similar using unadjusted figures.

posted by each firm in a given year. We find that digital-core firms significantly increase the proportion of such R&D-related positions following the implementation of the PIPL (columns 9 and 10). This pattern is consistent with the notion that, in the face of diminished access to consumer data, firms strategically invest in high-skilled labor to support alternative sources of competitive advantage, such as proprietary technologies, improved product offerings, or internal innovation. This shift toward R&D-intensive recruitment provides direct evidence of firms substituting labor for data and further underscores the central role of human capital in adapting to privacy-centric regulatory environments.

Building on the recruitment evidence, we next examine firms' efforts to retain talent. If digital-core firms are indeed placing greater value on skilled employees in response to data protection regulations, we would expect them not only to attract more qualified candidates but also to invest in retaining them. We focus on stock ownership plans (ESOP)—a common tool for incentivizing and retaining key personnel, particularly in knowledge-intensive industries. In China, listed-company ESOPs are typically aimed at senior management and core/backbone employees, so movements in ESOP adoption and intensity are informative about incentives for key talent.[13]

Empirically, we re-estimate equation (8) using a set of ESOP-related dependent variables. Specifically, an indicator for if the firm has an employee stock ownership plan in place (*ESOP*), the natural logarithm of one plus the number non-executive employees participating in ESOPs (*Ln(ESOP Ee Num)*), and the percentage of ESOP shares subscribed

[13]For example, Shanghai Golden Bridge InfoTech's 2025 ESOP lists "core employees" as the target group. See https://static.sse.com.cn/disclosure/listedinfo/announcement/c/new/2025-02-27/603918_20250227_L3KO.pdf (*in Chinese*).

by non-executive employees (*ESOP Ee Sub (%)*). A positive treatment effect on these variables would suggest a strategic shift toward long-term employee investment. From Table 5, we observe that following the introduction of the PIPL, digital-core firms increase the probability of implementing ESOPs by 2.1 percentages points, representing a 30 percent relative increase (column 2). Simultaneously, non-executive employee participation increases, as evident by the positive treatment effect on both the number and percentage of subscriptions (columns 3 to 6), indicating a strong emphasis on talent retention.

We complement the ESOP evidence with two additional retention proxies: (i) employee turnover (*Ee Turnover*), defined as the percentage change in total headcount, and (ii) investment in employee support and upskilling, measured as human capital investment (fringe benefits plus training costs) divided by operating expenses (*HCI/OpEx*). While turnover can move for many reasons (e.g., macro conditions, restructuring) and is therefore an imperfect proxy for satisfaction, the combination of lower employee turnover and higher fringe-and-training points to stronger retention and capability building among key employees, echoing our recruitment results. In the data, digital-core firms' turnover falls by about two percentage points (a 50% relative decline; Table 5 columns 7 and 8) following PIPL, and fringe-plus-training costs rise significantly for digital-core firms post-PIPL (Table 5 columns 9 and 10). Taken together with the ESOP results, these patterns indicate that firms' post-PIPL emphasis on experienced hires is matched by deeper ownership, improved benefits, and expanded training for core/backbone staff.

5.3 R&D

Having established that digital-core firms seek to substitute their lost ability to unilaterally exploit consumer data for profits with human capital, we move to examine whether they also place a greater emphasis on R&D. The latter would seem a natural substitute for consumer data; as firms may seek to regain their lost competitive advantage from restricted data use by enhancing product quality through intensified R&D activities. To examine this, we look at the firm's R&D investment, as well as its revenue generate process's reliance on R&D. Specifically, we modify equation (8) by replacing the dependent variable with R&D expenditure normalized by revenue and revenue per R&D employee. The former allows us to examine whether the firm places a greater emphasize on R&D following the introduction of data protection regulations; the latter enables us to study the firm's reliance on R&D employees for its revenue.

The results presented in Table 6 are consistent with our conjecture that digital-core firms also rely on R&D to substitute for their lost ability to exploit consumer data. Specifically, we observe that R&D expenditure as a fraction on revenue increases by 0.5 percentage points (or 5.6 percent in relative terms) in the post-regulations period for digital-core firms (column 2). The same firms also rely more on R&D for their revenue generation, as demonstrated by an increase in revenue per R&D employee of 2.57 million CNY, representing a 28 percent increase (column 4). Note that an alternative explanation for our finding is that R&D employee number fell after the introduction of data protection regulations. However, we do not find this to be the case; R&D employee number do not

exhibit significant changes before and after the regulations' introduction.

The results in columns 3 and 4 also shed more light on the earlier finding that digital-core firms place higher values on worker following the introduction of data protection regulations. Our earlier tests treated workers are a homogeneous group. However, if firms are indeed substituting consumer data with labor, then we would expect to seek a greater emphasis on higher-skilled workers. The finding that average revenue per R&D employee significantly increases is consistent with the notion of firms placing a greater emphasis on higher-skilled workers, which in turn supports with substitution explanation. Overall, Table 6 results indicate that, in addition to labor, digital-core firms also rely on R&D investment and employees to substitute for their lost ability to unilaterally exploit consumer data for profits.

5.4 Digitalization Intensity

In this subsection, we examine a layer of heterogeneity—the firm's degree of concern with regards to digitalization (digital intensity). Digitalization broadly refers to the use of digital technologies to change a business model and/or provide new revenue and value-producing opportunities. If the underlying assumption that the introduction of data protection regulations predominantly affect digital-core firms, then we would expect the effect to be even greater among those that place a heavier emphasis on digitalization. Affirmation of this would also highlight the importance of consumer data in profit generation.

To examine this, we first adopt a simple split-sample analysis and divide our full sample into four sub-samples based on digitalization intensity. Following common practice in the literature,[14] we define *Digi Intensity* as the natural logarithm of one plus the number of digitalization-related words in the Management Discussion and Analysis section of the firm's annual report. Digitalization-related words are those related to artificial intelligence, blockchain, cloud computing, big data, and digital technology application.

We then repeat equation (8) for the top and bottom quartile of each measure. This allows us to directly observe the difference in the magnitude of β_1 in each sub-sample. However, this method does not permit us to easily calculate the statistical significance of the difference. Therefore, we also include a triple-interaction model:

$$LS_{i,t} = \beta_1 Core_i Post_t Digi\ Intensity_{i,t} + \gamma' X_{i,t-1} + \alpha_i + \delta_t + \varepsilon_{i,t}, \tag{10}$$

where i and t index firm and year, respectively. Other variables share the same definition as in equation (8). Pairwise interactions are included in X. In this model, β_1 allows us to test the statistical significance of the modifying effect.

From Table 7, we observe that the treatment effect is significantly positive among firms in the top quartile of digitalization intensity (columns 3 and 4), but insignificant among those with in the bottom quartile (columns 1 and 2). The difference in the treatment effect is statistically significant, as we may observe from the results of the triple-

[14] See, for example, Gordon, Loeb, and Sohail (2010); Berkman, Jona, Lee, and Soderstrom (2018); Lawrence, Minutti-Meza, and Vyas (2018); Florackis, Louca, Michaely, and Weber (2023); Lattanzio and Ma (2023) among others.

interaction model (columns 5 and 6). These results are expected. Firms with a stronger digitalization focus would likely have collected and utilized consumer data to a greater extent compared to those less concerned with digitalization. Therefore, restricting data usage would likely lead to a strong need to quickly substitute labor for data.

5.5 Alternative Explanations

5.5.1 Covid-19

A natural concern is that the documented increase in labor share among digital-core firms may reflect pandemic-induced changes in labor demand rather than a response to China's personal data protection regime. The Covid-19 pandemic accelerated online activity and remote service delivery, potentially increasing demand for technical labor irrespective of privacy regulation. To address this alternative explanation, we conduct several empirical exercises.

First, the timing of the break strongly supports a regulatory mechanism. Figures 1 and 2 show that labor share trends between treated and control firms are parallel through 2020 and begin to diverge only in 2021, the year the PIPL was enacted. If pandemic-related adjustments were the primary driver, we would expect the break to emerge in 2020, coinciding with the onset of large-scale lockdowns and digital activity surges.

Second, we augment our baseline specification with province-year fixed effects, absorbing all time-varying shocks at the provincial level—including infection rates, lock-

down policies, containment stringency, and reopening speed. Untabulated results indicate that the treatment effect remains positive and statistically significant, indicating that regional pandemic dynamics do not explain our results.

Third, we estimate a triple-difference model, similar to that in the previous subsection, using Covid severity—measured by the natural logarithm of total 2020 confirmed cases (*Ln(Covid Cases)*) and deaths (*Ln(Covid Deaths)*) in each firm's headquarters province—as a moderator. The results are reported in Table 8. The triple-interaction terms are statistically insignificant across specifications (columns 1 to 4). This implies that the increase in labor share is not concentrated in regions more severely affected by Covid-19, contradicting a pandemic-driven labor shift story.

For contrast, we also interact *Core* $\times$ *Post* with pre-PIPL digital intensity (measured in 2020), as introduced in the prior subsection. This is to parallel the first four columns of Table 8, so that both moderators reflect firms' pre-treatment characteristics and are not mechanically influenced by post-PIPL developments or endogenous responses to the regulation. The triple-interaction coefficient is positive and significant in columns 5 and 6, reinforcing the interpretation that firms more reliant on digital strategies and data assets, not those most exposed to Covid-19, drive the labor share response.

Taken together, these tests confirm that the documented labor-for-data substitution is not a byproduct of Covid-19 disruptions, but rather a strategic response to the loss of unilateral access to consumer data under China's privacy regime.

5.5.2 Scientific and Technological Progress Law

As an additional falsification exercise, we test whether our results could instead reflect broader pro-innovation policy changes, rather than privacy regulation. In particular, China amended its *Scientific and Technological Progress Law* in 2021 (effective from January 2022) to strengthen national support for R&D and high-technology development. If increases in labor share simply reflect policy-driven innovation incentives, we would expect a similar pattern among certified high-tech enterprises. To test this, we re-estimate our baseline specification replacing the *Core* indicator with *High-Tech*—an indicator for firms formally recognized as high-tech enterprises (for special tax treatment) as of 2021 year-end.[15] We define *High-Tech* based on the firm alone, and alternatively incorporating affiliated parent and subsidiary entities. Across all specifications, the interaction term is statistically insignificant (Table 9). This evidence indicates that our findings are not driven by general technology-promotion policies, but are specific to the constraints imposed by data protection regulation.

5.6 Firm Performance

We conclude our empirical analysis by examining the treatment effect of data protection regulations on digital-core firms' operating performance. This indirectly allows us to

[15]To qualify as a high-tech enterprise, a company must be registered in mainland China for over one year and have at least 10% of its employees working in technical roles. R&D spending over the past three years must reach 5%, 4%, and 3% of revenue for firms with less than 50 million, 50-200 million, and over 200 million CNY annual sales, respectively. At least 60% of total revenue must come from qualified high-tech products or services. The company must also hold core proprietary intellectual property, pass an innovation evaluation, and have had no major safety, quality, or environmental violations in the prior year.

make some comments on the effectiveness of the labor-data substitution. If labor is a close substitute for consumer data, then we would expect to see very little adverse impact on digital-core firms' operating performance. To do so, we again modify equation (8) by replacing the dependent variable with various measures of firm performance, namely, the natural logarithm of revenue (*Ln(Rev)*), the cost of sales ratio (*Cost of Sales/Rev*), and EBIT normalized by revenue (*EBIT/Rev*).

The results are presented in Table 10. We make two observations. First, we observe that labor can effectively substitute for consumer data when it comes to revenue generation. There is no evidence indicating a decrease in digital-core firms' revenue following the introduction of the data protection regulations (columns 1 and 2). Second, the labor-data substitution appears to adversely impact digital-core firms' bottom line—cost of sales represents a larger fraction of total revenue, while EBIT represents a smaller fraction (columns 3 to 6). Collectively, these suggest that although labor can replace the role of consumer data to a large extent in terms of revenue generation, it is much more expensive that consumer data and thus negatively affects treated firms' bottom line.

6 Conclusion

In 2021, China introduced a landmark legislation aimed at protecting consumer data—the *Personal Information Protection Law*. Unlike prior regulations that lay out high-level frameworks, the PIPL explicitly prohibited the collection and usage of consumer data without consent. We therefore examine whether and how the introduction of the PIPL

affected digital-core firms' operations, with a focus on labor share. We hypothesize that because digital-core firms likely relied relatively heavily on consumer data in its daily operations (e.g., for targeted advertising), they are more likely to see an urgent need to substitute labor for data.

Our results are consistent with this hypothesis. We find that digital-core firms' labor share significantly increases following the introduction of data protection regulations. Moreover, we discover that this increase is driven almost exclusively by an increase in employees' average pay, rather than executive compensation, executive team size, or workforce size. We therefore conjecture that this is due to digital-core firms placing a large emphasis and value on high-skill workers. Exploiting major recruitment platforms' job advertisement data, we confirm on conjecture and find digital-core firms significantly increase recruitment efforts aimed at skilled labor following China's introduction of data regulations. This includes a higher share of job postings requiring advanced degrees, prior experience, high salaries, and a rise in the proportion of R&D-related positions. In addition to intensified recruitment, digital-core firms strengthen retention post-PIPL: they are more likely to adopt ESOPs with higher employee participation and subscription shares, they raise fringe benefits and training costs, and they experience lower employee turnover—together indicating a coordinated emphasis on retaining skilled, experienced workers.

Finally, we provide additional evidence that digital-core firms respond to data protection regulations by investing in innovation. In line with the observed increase in R&D-related hiring, we find that these firms raise their R&D spending as a share of

revenue and become more productive in their use of R&D labor. Collectively, these results suggest that digital-core firms are not simply reacting to regulatory constraints, but are strategically substituting high-skilled labor for data to rebuild lost competitive advantage. The combined evidence from recruitment patterns, retention strategies, and innovation investment supports a broader interpretation: in the face of reduced access to consumer data, firms pivot toward human capital and internal capabilities—especially those tied to R&D—as a new basis for value creation.

These empirical patterns line up with the implications of our conceptual framework. In the model, tighter privacy raises the shadow cost of data and induces substitution toward skilled labor within the "insight" technology; because labor enters in efficiency units, the adjustment operates primarily through worker quality and pay rather than headcount, yielding more postings for experienced and R&D roles and higher average compensation. In the dynamic extension, capability accumulation further implies stronger retention and human-capital formation (e.g., ESOP participation, benefits and training) alongside higher R&D effort.

Additionally, we find that the treatment effect on labor share is strong among treated firms with higher digitalization intensity. We attribute this finding to *ex ante* heavier reliance on consumer data. We find local Covid-19 severity and the 2022 amendment to the *Scientific and Technological Progress Law* had no significant relation to firm labor share. Finally, we find that although labor-data substitution is effective in maintaining revenue levels, it has a negative impact on digital-core firms' bottom line (higher cost of sales and lower EBIT), likely due to higher costs associated with high-skilled workers as compared

to consumer data.

Overall, our study shows that data protection regulations trigger meaningful shifts in firms' internal resource allocation, particularly among those most reliant on digital strategies. Rather than passively bearing the cost of lost data access, digital-core firms actively invest in high-skilled labor and innovation—substituting data with human capital as a new engine of competitive advantage. These adjustments not only benefit workers through higher pay, better retention, and expanded R&D roles, but also highlight the evolving role of labor in data-constrained environments. Our findings underscore the importance of consumer data as a critical input in the modern economy and raise broader questions about how firms—and the labor market—adapt when data is no longer freely exploitable.

References

Acemoglu, Daron, and Pascual Restrepo, 2020, Robots and jobs: Evidence from US labor markets, *Journal of Political Economy* 128, 2188–2244.

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

Aral, Sinan, Erik Brynjolfsson, and Lynn Wu, 2012, Three-way complementarities: Performance pay, human resource analytics, and information technology, *Management Science* 58, 913–931.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, 2020, The fall of the labor share and the rise of superstar firms, *Quarterly Journal of Economics* 135, 645–709.

Autor, David, and Anna Salomons, 2018, Is automation labor-displacing? Productivity growth, employment, and the labor share, *Working Paper,* National Bureau of Economic Research.

Barkai, Simcha, 2020, Declining labor and capital shares, *Journal of Finance* 75, 2421–2463.

Beraja, Martin, David Y Yang, and Noam Yuchtman, 2023, Data-intensive innovation and the State: Evidence from AI firms in China, *Review of Economic Studies* 90, 1701–1723.

Berger, David, Kyle Herkenhoff, and Simon Mongey, 2022, Labor market power, *American Economic Review* 112, 1147–1193.

Bergholt, Drago, Francesco Furlanetto, and Nicolò Maffei-Faccioli, 2022, The decline of the labor share: New empirical evidence, *American Economic Journal: Macroeconomics* 14, 163–198.

Berkman, Henk, Jonathan Jona, Gladys Lee, and Naomi Soderstrom, 2018, Cybersecurity awareness and market valuations, *Journal of Accounting and Public Policy* 37, 508–526.

Bharadwaj, Anandhi S, 2000, A resource-based perspective on information technology capability and firm performance: An empirical investigation, *MIS Quarterly* 169–196.

Bharadwaj, Anandhi and El Sawy, Omar A and Pavlou, Paul A and Venkatraman, N v, 2013, Digital business strategy: Toward a next generation of insights, *MIS Quarterly* 471–482.

Blind, Knut, Crispin Niebel, and Christian Rammer, 2024, The impact of the EU General Data Protection Regulation on product innovation, *Industry and Innovation* 31, 311–351.

Bondi, Tommaso, Omid Rafieian, and Yunfei Yao, 2025, Privacy and polarization: An inference-based framework, *Management Science* .

Ciminelli, Gabriele, Romain Duval, and Davide Furceri, 2022, Employment protection deregulation and labor shares in advanced economies, *Review of Economics and Statistics* 104, 1174–1190.

Dao, Mai Chi, Mitali Das, Zsoka Koczan, and Weicheng Lian, 2017, Why is labor receiving a smaller share of global income? Theory and empirical evidence, *Working Paper,* International Monetary Fund.

Demirer, Mert, Diego J Jiménez Hernández, Dean Li, and Sida Peng, 2024, Data, privacy laws and firm production: Evidence from the GDPR, *Working Paper,* National Bureau of Economic Research.

Elsby, Michael WL, Bart Hobijn, and Ayşegül Şahin, 2013, The decline of the US labor share, *Brookings Papers on Economic Activity* 2013, 1–63.

Fainmesser, Itay P, Andrea Galeotti, and Ruslan Momot, 2023, Digital privacy, *Management Science* 69, 3157–3173.

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

Frey, Carl Benedikt, and Giorgio Presidente, 2024, Privacy regulation and firm performance: Estimating the GDPR effect globally, *Economic Inquiry* 62, 1074—-1089.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M Shapiro, 2019, Pre-event trends in the panel event-study design, *American Economic Review* 109, 3307–38.

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, 2011, Search engine advertising: Channel substitution when pricing ads to context, *Management Science* 57, 458–470.

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

Gordon, Lawrence A, Martin P Loeb, and Tashfeen Sohail, 2010, Market value of voluntary disclosures concerning information security, *MIS Quarterly* 567–594.

Janssen, Rebecca, Reinhold Kesler, Michael E Kummer, and Joel Waldfogel, 2022, GDPR and the lost generation of innovative apps, *Working Paper,* National Bureau of Economic Research.

Johnson, Garrett, 2022, Economic research on privacy regulation: Lessons from the GDPR and beyond, *Working Paper,* National Bureau of Economic Research.

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.

Karabarbounis, Loukas, and Brent Neiman, 2014, The global decline of the labor share, *Quarterly Journal of Economics* 129, 61–103.

Kaymak, Barış, and Immo Schott, 2023, Corporate tax cuts and the decline in the manufacturing labor share, *Econometrica* 91, 2371–2408.

Klein, April, Raffaele Manini, and Yanting Shi, 2022, Across the pond: How US firms' boards of directors adapted to the passage of the General Data Protection Regulation, *Contemporary Accounting Research* 39, 199–233.

Lambrecht, Anja, and Catherine E Tucker, 2015, Can big data protect a firm from competition?, *Working Paper,* Available at SSRN 2705530.

Lattanzio, Gabriele, and Yue Ma, 2023, Cybersecurity risk and corporate innovation, *Journal of Corporate Finance* 82, 102445.

Lattanzio, Gabriele, and Jérôme Taillard, 2022, M&A and cybersecurity risk: Empirical evidence, *Working Paper,* Available at SSRN 4170093.

Lawrence, Alastair, Miguel Minutti-Meza, and Dushyantkumar Vyas, 2018, Is operational control risk informative of financial reporting deficiencies?, *Auditing: A Journal of Practice & Theory* 37, 139–165.

Lefrere, Vincent, Logan Warberg, Cristobal Cheyre, Veronica Marotta, and Alessandro Acquisti, 2025, Does privacy regulation harm content providers? A longitudinal analysis of the impact of the GDPR, *Management Science* .

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

Niebel, Crispin, 2021, The impact of the General Data Protection Regulation on innovation and the global political economy, *Computer Law & Security Review* 40, 105523.

Oberfield, Ezra, and Devesh Raval, 2021, Micro data and macro technology, *Econometrica* 89, 703–732.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick, 2022, The rise of pass-throughs and the decline of the labor share, *American Economic Review: Insights* 4, 323–340.

Sun, Tianshu, Zhe Yuan, Chunxiao Li, Kaifu Zhang, and Jun Xu, 2024, The value of personal data in internet commerce: A high-stakes field experiment on data regulation policy, *Management Science* 70, 2645–2660.

Yoo, Youngjin, Ola Henfridsson, and Kalle Lyytinen, 2010, Research commentary—The new organizing logic of digital innovation: An agenda for information systems research, *Information systems research* 21, 724–735.

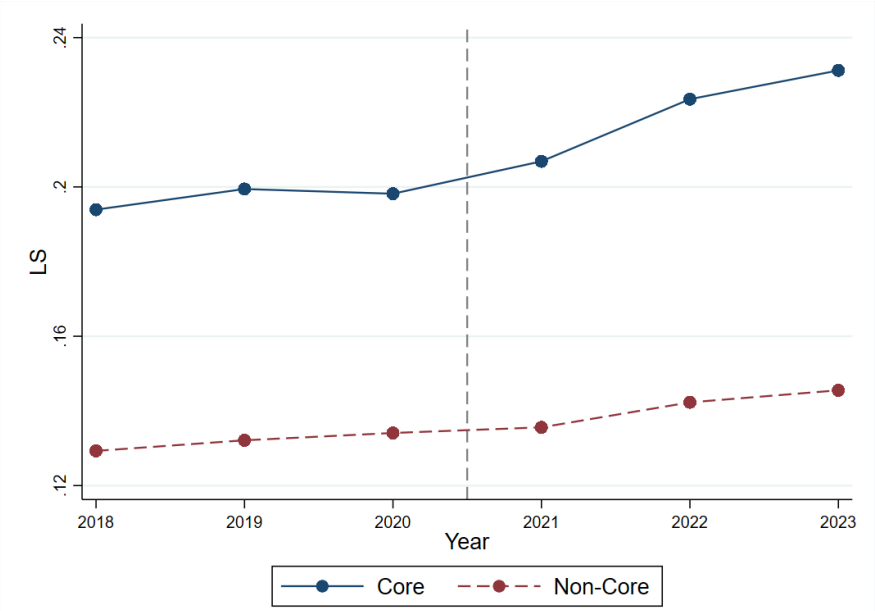


Fig. 1: Parallel trends graph. *LS* is defined as cash labor expense (including salaries, wages, and benefits) divided by revenue. *Core* is an indicator for industries core to the digital economy as classified in the *Classification of Digital Economy and its Core Industries (2021)* issued by the National Bureau of Statistics of China.

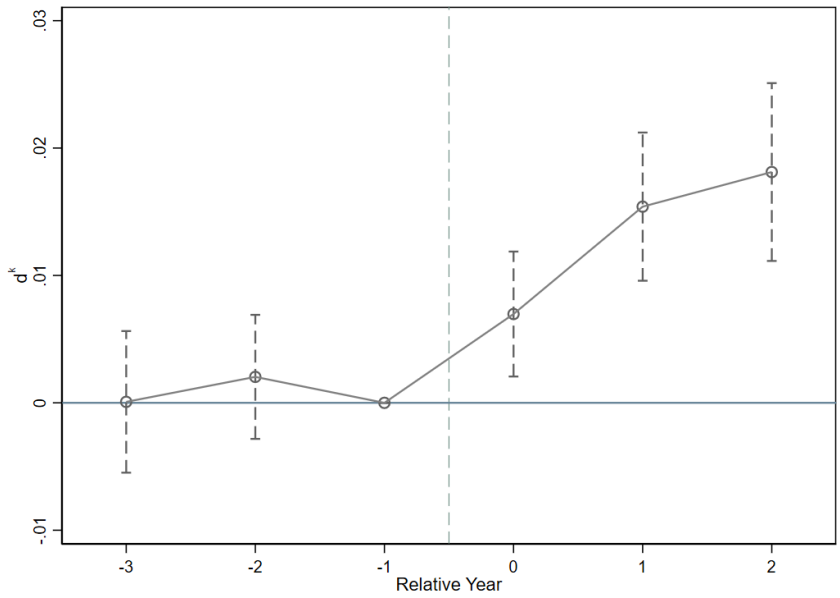


Fig. 2: Estimated dynamic difference-in-differences coefficient, d^k, of equation (9), with vertical dashed lines representing 95 percent confidence intervals. The point estimate of the year 2021 (year 0), in which the *Personal Information Protection Law* was introduced, is normalized to zero following Freyaldenhoven et al. (2019).

Table 1 Summary Statistics
Detailed variable definitions are presented in Table A1.

	Core						Non-Core					
	N	Mean	S.D.	p25	p50	p75	N	Mean	S.D.	p25	p50	p75
LS	3492	0.21	0.13	0.12	0.18	0.26	12 648	0.14	0.09	0.08	0.12	0.18
Ln(Assets)	3492	22.77	1.05	22.00	22.62	23.37	12 648	22.92	1.19	22.04	22.72	23.62
ROA	3492	0.02	0.09	0.01	0.03	0.06	12 648	0.03	0.07	0.01	0.04	0.07
Leverage	3492	0.22	0.16	0.10	0.18	0.30	12 648	0.31	0.21	0.14	0.27	0.44
Mgmt Hold	3492	0.18	0.19	0.00	0.10	0.31	12 648	0.13	0.19	0.00	0.01	0.24
Indep Dir	3492	0.39	0.05	0.33	0.38	0.43	12 648	0.38	0.05	0.33	0.36	0.43
Rev Growth	3492	0.14	0.33	−0.04	0.10	0.26	12 648	0.14	0.32	−0.02	0.10	0.25
Ln(Firm Age)	3492	2.99	0.27	2.83	3.00	3.18	12 648	3.07	0.27	2.89	3.09	3.26
PPE/Assets	3492	0.14	0.12	0.05	0.10	0.20	12 648	0.22	0.15	0.10	0.19	0.31
LS Mgmt	3489	0.01	0.01	0.00	0.00	0.01	12 633	0.00	0.00	0.00	0.00	0.00
Ls Ee	3489	0.20	0.13	0.12	0.17	0.26	12 633	0.13	0.09	0.07	0.12	0.17
Ln(Avg Pay Mgmt)	3488	13.34	0.59	12.95	13.30	13.69	12 624	13.26	0.62	12.85	13.23	13.61
Ln(Avg Pay Ee)	3489	11.99	0.45	11.67	11.95	12.27	12 631	11.83	0.43	11.53	11.78	12.08
Postgrad Pct	3315	0.03	0.05	0.00	0.01	0.03	11 186	0.03	0.06	0.00	0.00	0.03
High Income Pct	3306	0.38	0.25	0.20	0.36	0.56	11 059	0.24	0.23	0.05	0.19	0.36
Min 5 Pct	3315	0.14	0.12	0.06	0.13	0.20	11 184	0.14	0.14	0.02	0.10	0.21
Min 10 Pct	3315	0.01	0.03	0.00	0.00	0.02	11 184	0.02	0.03	0.00	0.00	0.02
R&D Pct	3315	0.49	0.22	0.35	0.50	0.64	11 186	0.32	0.24	0.13	0.29	0.47
COD	3314	3.12	5.91	0.00	0.00	3.31	12 267	4.60	7.24	0.00	0.00	7.12
Digi Intensity	3492	2.48	1.38	1.39	2.64	3.56	12 648	0.87	1.01	0.00	0.69	1.61
ESOP	3492	0.07	0.25	0.00	0.00	0.00	12 648	0.05	0.22	0.00	0.00	0.00
Ln(ESOP Ee Num)	3491	0.30	1.20	0.00	0.00	0.00	12 648	0.23	1.07	0.00	0.00	0.00
ESOP Ee Sub (%)	3491	4.76	18.61	0.00	0.00	0.00	12 648	3.51	16.17	0.00	0.00	0.00
Ee Turnover	3492	0.04	0.20	−0.06	0.02	0.12	12 644	0.04	0.19	−0.05	0.01	0.09
HCI/OpEx	2817	0.51	1.20	0.01	0.06	0.29	11 233	0.41	0.83	0.02	0.08	0.39
R&D/Rev	3385	0.09	0.07	0.04	0.07	0.12	11 276	0.04	0.04	0.02	0.04	0.05
Rev/R&D Ee	3378	9.05	26.62	2.13	3.83	7.14	11 162	33.72	83.02	5.62	10.06	22.22
Ln(Covid Cases)	3492	7.14	0.91	6.75	7.18	7.62	12 648	6.87	0.98	6.53	6.90	7.18
Ln(Covid Deaths)	3492	1.81	1.43	0.69	2.08	2.20	12 648	1.59	1.40	0.69	1.61	2.20
High-Tech (Firm)	3492	0.22	0.41	0.00	0.00	0.00	12 648	0.15	0.36	0.00	0.00	0.00
Ln(Rev)	3492	21.40	1.32	20.44	21.28	22.16	12 648	21.82	1.42	20.79	21.68	22.71
Cost of Sales/Rev	3492	0.07	0.08	0.02	0.05	0.10	12 648	0.08	0.10	0.01	0.04	0.09
EBIT/Rev	3492	0.02	0.25	0.02	0.07	0.13	12 648	0.08	0.21	0.03	0.08	0.16

Table 2 Data Protection Regulations' Effect on Labor Share of Digital-Core Industries
The dependent variables are *LS*, defined as cash labor expense (including salaries, wages, and benefits) divided by revenue (columns 1 and 2), management labor share (columns 3 and 4), and employee labor share (columns 5 and 6). *Core* is an indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China. *Post* is an indicator for years including and after 2021 (i.e., enactment of the PIPL). Standard errors are clustered by firm; *t*-statistics are in parentheses.

Dep Variable:	LS		LS (Mgmt)		LS (Ee)	
	(1)	(2)	(3)	(4)	(5)	(6)
Core × Post	0.014***	0.013***	0.000	0.000	0.014***	0.013***
	(5.05)	(4.91)	(0.92)	(0.74)	(5.14)	(4.98)
Ln(Assets)		-0.010***		-0.001***		-0.009***
		(-5.96)		(-12.71)		(-5.35)
ROA		-0.047***		-0.001**		-0.047***
		(-4.41)		(-2.11)		(-4.39)
Leverage		-0.016**		-0.004***		-0.012
		(-2.05)		(-7.71)		(-1.62)
Mgmt Hold		-0.026***		-0.001*		-0.025***
		(-2.72)		(-1.78)		(-2.70)
Indep Dir		-0.037**		-0.002**		-0.035**
		(-2.40)		(-2.39)		(-2.34)
Rev Growth		-0.021***		-0.001***		-0.020***
		(-11.41)		(-8.77)		(-11.07)
Ln(Firm Age)		-0.000		-0.002**		0.002
		(-0.00)		(-2.15)		(0.11)
PPE/Assets		0.020		-0.001		0.019
		(1.54)		(-0.99)		(1.55)
N	16140	16140	16122	16122	16122	16122
Adj R-sq	0.04	0.09	0.00	0.09	0.04	0.09
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3 Data Protection Regulations and Labor Expense

The dependent variables are the natural logarithm of the number of executives (columns 1 and 2), the natural logarithm of average manager compensation (columns 3 and 4), and the natural logarithm of the number of employees (columns 5 and 6), the natural logarithm of average employee compensation (columns 7 and 8). *Core* is an indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China. *Post* is an indicator for years including and after 2021 (i.e., enactment of the PIPL). Standard errors are clustered by firm; *t*-statistics are in parentheses.

Dep Variable:	Ln(Mgmt Number)		Ln(Avg Pay Mgmt)		Ln(Ee Number)		Ln(Avg Pay Ee)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Core × Post	-0.004	-0.004	-0.002	-0.011	-0.011	-0.006	0.020**	0.019**
	(-0.82)	(-0.78)	(-0.14)	(-0.82)	(-0.62)	(-0.42)	(2.51)	(2.39)
Ln(Assets)		0.017***		0.084***		0.357***		0.049***
		(4.99)		(8.13)		(22.39)		(7.74)
ROA		0.078***		0.343***		0.378***		-0.008
		(4.38)		(6.68)		(6.74)		(-0.22)
Leverage		0.037**		-0.078*		0.571***		0.121***
		(2.37)		(-1.67)		(8.91)		(4.13)
Mgmt Hold		-0.002		-0.222***		0.053		-0.010
		(-0.09)		(-3.68)		(0.86)		(-0.28)
Indep Dir		-0.252***		-0.104		-0.022		-0.085
		(-6.51)		(-1.07)		(-0.23)		(-1.38)
Rev Growth		0.000		0.065***		0.067***		0.022***
		(0.05)		(7.36)		(6.07)		(2.97)
Ln(Firm Age)		0.045		0.403***		0.237		0.124
		(0.95)		(2.89)		(1.64)		(1.50)
PPE/Assets		-0.018		-0.094		0.149**		0.042
		(-0.93)		(-1.64)		(2.11)		(1.08)
N	16140	16140	16112	16112	16138	16138	16120	16120
Adj R-sq	0.01	0.03	0.19	0.22	0.02	0.25	0.40	0.41
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4 Data Protection Regulations and Talent Recruitment

The dependent variables are the percentage of advertised job positions (i) requiring postgraduate qualification (columns 1 and 2), offering a CPI-adjusted monthly income of more than 10,000 CNY in 2018 yuan (columns 3 and 4), requiring at least five years of experience (columns 5 and 6), and requiring at least 10 years of experience (columns 7 and 8). *Core* is an indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China. *Post* is an indicator for years including and after 2021 (i.e., enactment of the PIPL). Standard errors are clustered by firm; *t*-statistics are in parentheses.

Dep Variable:	Postgrad Pct		High Income Pct		Min 5 Pct		Min 10 Pct		R&D Pct	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Core × Post	0.007***	0.007***	0.037***	0.033***	0.012**	0.011**	0.004***	0.004***	0.015**	0.015**
	(3.93)	(3.83)	(4.52)	(4.00)	(2.50)	(2.29)	(3.43)	(3.21)	(2.07)	(2.06)
Ln(Assets)		0.003*		0.028***		0.009**		0.001		0.007
		(1.77)		(4.52)		(2.42)		(1.38)		(1.32)
ROA		0.007		0.029		0.010		0.007		0.032
		(0.89)		(0.87)		(0.50)		(1.29)		(1.11)
Leverage		0.002		-0.059**		-0.020		-0.000		-0.042
		(0.32)		(-2.08)		(-1.12)		(-0.03)		(-1.63)
Mgmt Hold		-0.006		-0.089**		-0.029		-0.005		0.029
		(-0.71)		(-2.49)		(-1.42)		(-0.85)		(0.86)
Indep Dir		0.003		-0.015		-0.014		0.009		-0.061
		(0.15)		(-0.25)		(-0.39)		(0.99)		(-1.06)
Rev Growth		-0.003		-0.003		0.002		0.001		0.000
		(-1.61)		(-0.51)		(0.58)		(1.16)		(0.01)
Ln(Firm Age)		-0.004		0.153*		0.027		0.012		0.066
		(-0.18)		(1.91)		(0.59)		(0.96)		(0.87)
PPE/Assets		-0.013		-0.055		-0.067**		-0.011*		-0.013
		(-1.15)		(-1.30)		(-2.56)		(-1.75)		(-0.33)
N	14501	14501	14365	14365	14499	14499	14499	14499	14501	14501
Adj R-sq	0.01	0.01	0.13	0.13	0.04	0.04	0.02	0.02	0.02	0.02
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5 Data Protection Regulations and Talent Retention

The dependent variables are an indicator for if the firm has an employee stock ownership plan in that year (columns 1 and 2), the natural logarithm of one plus the number of employees participating in ESOPs (columns 3 and 4), the percentage of ESOP shares subscribed by non-executive employees (columns 5 and 6), employee turnover calculated as the percentage change in employee number from year $t-1$ to year t (columns 7 and 8), and human capital investment (employee fringe benefits plus training costs) divided by operating expenses (columns 9 and 10). *Core* is an indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China. *Post* is an indicator for years including and after 2021 (i.e., enactment of the PIPL). Standard errors are clustered by firm; *t*-statistics are in parentheses.

Dep Variable:	ESOP		Ln(ESOP Ee Num)		ESOP Ee Sub (%)		Ee Turnover		HCI/OpEx	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Core × Post	0.022**	0.021**	0.093**	0.095**	1.704**	1.660**	-0.020**	-0.021***	0.044**	0.042**
	(2.33)	(2.26)	(2.12)	(2.14)	(2.49)	(2.41)	(-2.55)	(-2.80)	(2.03)	(2.05)
Ln(Assets)		0.012*		0.051*		0.853*		-0.045***		-0.073***
		(1.83)		(1.65)		(1.78)		(-7.11)		(-4.50)
ROA		-0.034		-0.023		-0.652		0.123***		-0.012
		(-1.14)		(-0.18)		(-0.33)		(3.43)		(-0.15)
Leverage		-0.072***		-0.331***		-5.062***		-0.452***		-0.197***
		(-3.09)		(-2.84)		(-2.89)		(-15.41)		(-4.29)
Mgmt Hold		-0.016		0.007		-0.697		0.022		-0.053
		(-0.58)		(0.06)		(-0.35)		(0.68)		(-0.57)
Indep Dir		0.109*		0.498*		7.024*		0.020		-0.039
		(1.75)		(1.77)		(1.66)		(0.36)		(-0.27)
Rev Growth		0.003		0.012		0.195		0.019***		-0.102***
		(0.54)		(0.44)		(0.45)		(2.68)		(-6.25)
Ln(Firm Age)		-0.054		-0.319		-1.210		-0.013		-0.033
		(-0.64)		(-0.80)		(-0.20)		(-0.18)		(-0.17)
PPE/Assets		0.006		0.034		0.468		-0.199***		-0.048
		(0.18)		(0.22)		(0.20)		(-4.81)		(-0.29)
N	16140	16140	16130	16130	16129	16129	16136	16136	14050	14050
Adj R-sq	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.06	0.00	0.02
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6 Data Protection Regulations and R&D

The dependent variables are R&D expense normalized by revenue (columns 1 and 2) and revenue per R&D employee (columns 3 and 4). *Core* is an indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China. *Post* is an indicator for years including and after 2021 (i.e., enactment of the PIPL). Standard errors are clustered by firm; *t*-statistics are in parentheses.

Dep Variable:	R&D/Rev		Rev/R&D Ee	
	(1)	(2)	(3)	(4)
Core × Post	0.005***	0.005***	2.783**	2.574**
	(3.51)	(3.57)	(2.44)	(2.45)
Ln(Assets)		0.002**		3.390***
		(2.26)		(2.62)
ROA		0.001		-0.677
		(0.29)		(-0.17)
Leverage		-0.006*		4.989
		(-1.67)		(0.70)
Mgmt Hold		-0.002		-6.005
		(-0.33)		(-1.48)
Indep Dir		-0.007		2.338
		(-0.93)		(0.19)
Rev Growth		-0.008***		4.984***
		(-9.41)		(3.50)
Ln(Firm Age)		0.003		14.706
		(0.31)		(0.84)
PPE/Assets		-0.003		9.970
		(-0.58)		(1.16)
N	14661	14661	14540	14540
Adj R-sq	0.03	0.05	0.01	0.01
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7 Data Protection Regulations, Labor Share, and Digitalization

The dependent variable is *LS*, defined as cash labor expense (including salaries, wages, and benefits) divided by revenue. *Core* is an indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China. *Post* is an indicator for years including and after 2021 (i.e., enactment of the PIPL). *Digi Intensity* is the natural logarithm of the number of words related to digitalization in the Management Discussion & Analysis section of the firm's annual report. Columns 1 and 2 (3 and 4) restrict the sample to firms in the bottom (top) quantile of *Digi Intensity*. Columns 5 and 6 adopt a triple interaction approach to test the statistical significance of the moderator variable. Standard errors are clustered by firm; *t*-statistics are in parentheses.

Sample:	Digi Intensity (Lo)		Digi Intensity (Hi))		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Core × Post	-0.010	-0.008	0.024***	0.023***	-0.011**	-0.008*
	(-0.74)	(-0.64)	(5.27)	(5.36)	(-2.24)	(-1.67)
Core × Post × Digi Intensity					0.009***	0.008***
					(4.20)	(3.79)
Post × Digi Intensity					0.002*	0.001
					(1.89)	(1.25)
Core × Digi Intensity					-0.004**	-0.004**
					(-2.05)	(-2.19)
Digi Intensity					-0.002	-0.000
					(-1.53)	(-0.46)
Ln(Assets)		-0.001		-0.010**		-0.009***
		(-0.30)		(-2.36)		(-5.54)
ROA		-0.041		-0.038**		-0.046***
		(-1.50)		(-2.23)		(-4.28)
Leverage		0.010		-0.030*		-0.017**
		(0.84)		(-1.70)		(-2.18)
Mgmt Hold		-0.038*		-0.002		-0.025***
		(-1.77)		(-0.08)		(-2.62)
Indep Dir		-0.060**		-0.002		-0.038**
		(-2.06)		(-0.07)		(-2.53)
Rev Growth		-0.022***		-0.025***		-0.021***
		(-5.11)		(-5.58)		(-11.27)
Ln(Firm Age)		0.010		0.032		0.000
		(0.23)		(0.62)		(0.01)
PPE/Assets		0.036		0.047*		0.020
		(1.30)		(1.72)		(1.59)
N	4035	4035	4035	4035	16140	16140
Adj R-sq	0.02	0.07	0.11	0.15	0.05	0.10
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8 Covid-19 and Labor Share

The dependent variable is *LS*, defined as cash labor expense (including salaries, wages, and benefits) divided by revenue. *Core* is an indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China. *Post* is an indicator for years including and after 2021 (i.e., enactment of the PIPL). *Ln(Covid Cases (Deaths)* is natural logarithm of total Covid-19 cases (deaths) in the firm's headquarter province in 2020. *Digi Intensity (pre-PIPL)* is the natural logarithm of the number of words related to digitalization in the Management Discussion & Analysis section of the firm's annual report, fixed at each firm's 2020 value. Standard errors are clustered by firm; *t*-statistics are in parentheses.

Dep Variable:	LS					
Moderator:	Covid-19				Digitalization	
	(1)	(2)	(3)	(4)	(5)	(6)
Core × Post	0.019	0.021	0.012***	0.011***	-0.007	-0.005
	(1.11)	(1.27)	(2.84)	(2.80)	(-1.61)	(-1.17)
Core × Post × Ln(Covid Cases)	-0.001	-0.001				
	(-0.30)	(-0.52)				
Post × Ln(Covid Cases)	0.000	0.000				
	(0.10)	(0.24)				
Core × Post × Ln(Covid Deaths)			0.001	0.001		
			(0.84)	(0.73)		
Post × Ln(Covid Deaths)			0.000	0.000		
			(0.37)	(0.17)		
Core × Post × Digi Intensity (pre-PIPL)					0.008***	0.007***
					(3.91)	(3.57)
Post × Digi Intensity (pre-PIPL)					0.002	0.001
					(1.56)	(1.01)
Ln(Assets)		-0.010***		-0.010***		-0.009***
		(-5.96)		(-5.95)		(-5.56)
ROA		-0.047***		-0.047***		-0.046***
		(-4.40)		(-4.40)		(-4.28)
Leverage		-0.016**		-0.016**		-0.017**
		(-2.05)		(-2.06)		(-2.19)
Mgmt Hold		-0.026***		-0.026***		-0.025***
		(-2.72)		(-2.72)		(-2.63)
Indep Dir		-0.037**		-0.037**		-0.038**
		(-2.41)		(-2.39)		(-2.50)
Rev Growth		-0.021***		-0.021***		-0.021***
		(-11.40)		(-11.41)		(-11.31)
Ln(Firm Age)		0.000		0.000		0.001
		(0.00)		(0.01)		(0.05)
PPE/Assets		0.020		0.020		0.020
		(1.54)		(1.54)		(1.58)
N	16140	16140	16140	16140	16140	16140
Adj R-sq	0.04	0.09	0.04	0.09	0.05	0.09
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9 Scientific and Technological Progress Law's Effect on Labor Share of Digital-Core Industries
The dependent variable is *LS*, defined as cash labor expense (including salaries, wages, and benefits) divided by revenue. *High-Tech* is an indicator for firms recognized as a high-tech enterprise by the Chinese government (for special tax treatment). We define *High-Tech* based both on the firm itself only (column 1 and 2) and include parent and subsidiary firms (columns 3 and 4). *Post STPL* is an indicator for years including and after 2022 (amendments to the STPL took effect in January 2022). Standard errors are clustered by firm; *t*-statistics are in parentheses.

Dep Variable:	LS			
High-Tech Definition:	Firm Only		Firm, Parent, & Subs	
	(1)	(2)	(3)	(4)
High Tech × Post STPL	0.003	0.003	0.002	0.002
	(1.04)	(1.14)	(0.89)	(0.85)
Ln(Assets)		-0.010***		-0.010***
		(-6.16)		(-6.16)
ROA		-0.047***		-0.048***
		(-4.44)		(-4.46)
Leverage		-0.017**		-0.017**
		(-2.17)		(-2.16)
Mgmt Hold		-0.031***		-0.031***
		(-3.18)		(-3.21)
Indep Dir		-0.037**		-0.037**
		(-2.39)		(-2.39)
Rev Growth		-0.021***		-0.021***
		(-11.33)		(-11.35)
Ln(Firm Age)		0.011		0.011
		(0.48)		(0.51)
PPE/Assets		0.020		0.021
		(1.58)		(1.59)
N	16140	16140	16140	16140
Adj R-sq	0.04	0.09	0.04	0.06
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10 Data Protection Regulations and Firm Performance

The dependent variables are the natural logarithm of revenue (columns 1 and 2), cost of sales normalized by revenue (columns 3 and 4), and EBIT normalized by revenue (columns 5 and 6). *Core* is an indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China. *Post* is an indicator for years including and after 2021 (i.e., enactment of the PIPL). Standard errors are clustered by firm; *t*-statistics are in parentheses.

Dep Variable:	Ln(Rev)		Cost of Sales/Rev		EBIT/Rev	
	(1)	(2)	(3)	(4)	(5)	(6)
Core × Post	-0.033	-0.022	0.009***	0.008***	-0.018**	-0.019**
	(-1.50)	(-1.56)	(4.78)	(4.58)	(-1.99)	(-2.10)
Ln(Assets)		0.477***		-0.004***		-0.021***
		(29.17)		(-3.31)		(-3.25)
ROA		0.595***		-0.016**		-0.117**
		(9.36)		(-2.09)		(-2.37)
Leverage		0.763***		0.007		-0.117***
		(11.34)		(1.19)		(-3.92)
Mgmt Hold		0.192***		-0.015		-0.003
		(3.28)		(-1.58)		(-0.08)
Indep Dir		0.152		-0.005		0.089
		(1.58)		(-0.47)		(1.32)
Rev Growth		0.241***		-0.004***		0.053***
		(17.90)		(-3.76)		(6.78)
Ln(Firm Age)		0.382**		0.007		-0.020
		(2.52)		(0.42)		(-0.28)
PPE/Assets		0.144*		-0.007		-0.015
		(1.90)		(-0.96)		(-0.38)
N	16140	16140	16140	16140	16140	16140
Adj R-sq	0.11	0.41	0.05	0.06	0.02	0.03
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Internet Appendix

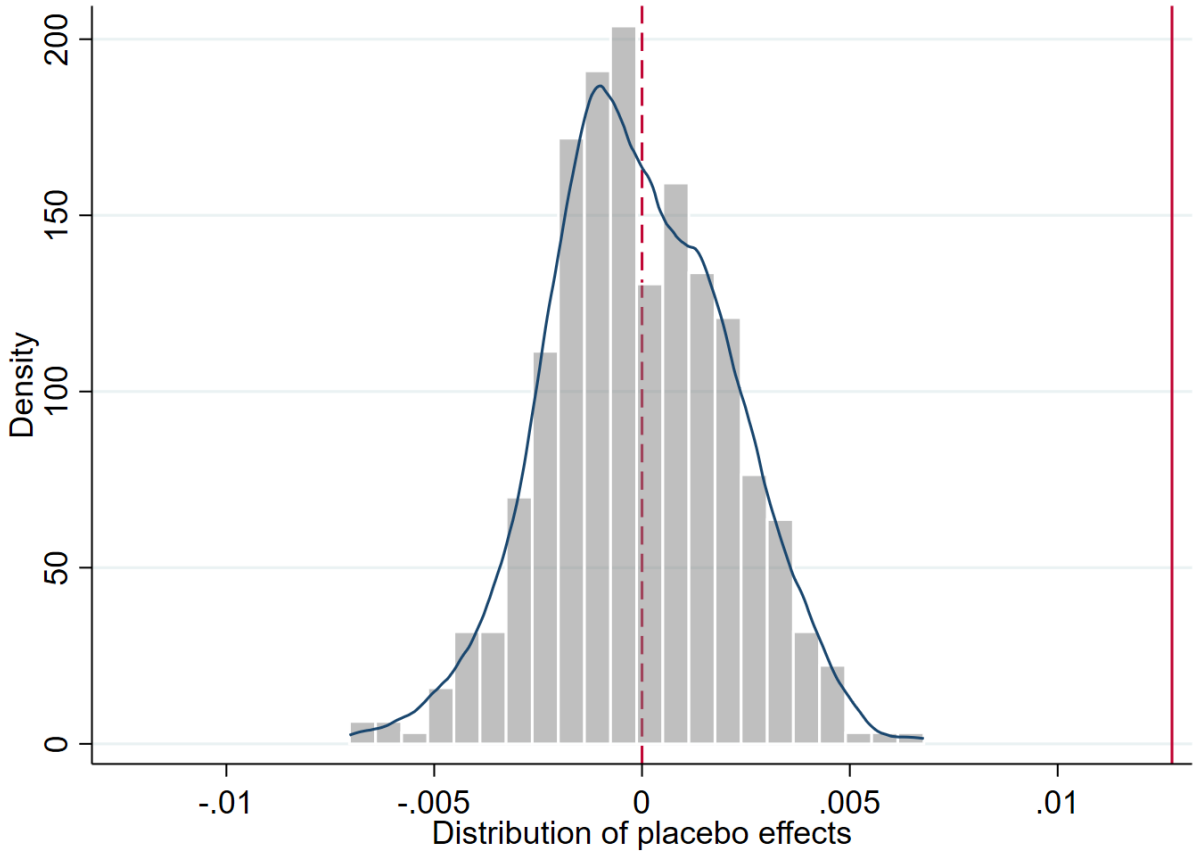


Fig. A1: Placebo treatment effect. Firms are first randomly drawn from our sample to create a placebo treatment group, before equation (8) is re-estimated. We repeat this process 500 times and plot the distribution of β_1.

Table A1 Variable Definition

Variable	Definition
Core	An indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China.
Cost of Sales/Rev	Cost of sales divided by revenue.
Digi Intensity	The natural logarithm of one plus the number of digitalization-related words in the Management Discussion and Analysis section of the annual report. Digitalization-related words are those related to artificial intelligence, blockchain, cloud computing, big data, and digital technology application.
EBIT/Rev	EBIT divided by revenue.
Ee Turnover	The percentage change in employee number from year $t-1$ to year t.
ESOP	Equals to one if the firm enacted employee stock ownership plans during the year, and zero otherwise.
ESOP Ee Sub (%)	The percentage of ESOP shares subscribed by non-executive employees.
HCI/OpEx	Human capital investment (fringe benefits plus training costs) divided by operating expenses.
High Income Pct	The number of advertised positions offering a CPI-adjusted monthly income of more than 10,000 CNY in 2018 yuan, divided by the total number of advertised job positions.
High-Tech	Equals to one if the firm is recognized as a high-tech enterprise by the Chinese government (for special tax treatment) as of 2021 year-end, and zero otherwise. To qualify as a high-tech enterprise, a company must be registered in mainland China for over one year and have at least 10% of its employees working in technical roles. R&D spending over the past three years must reach 5%, 4%, and 3% of revenue for firms with less than 50 million, 50-200 million, and over 200 million CNY annual sales, respectively. At least 60% of total revenue must come from qualified high-tech products or services. The company must also hold core proprietary intellectual property, pass an innovation evaluation, and have had no major safety, quality, or environmental violations in the prior year.
Indep Dir	The proportion of independent directors, calculated as the number of independent directors divided by the total number of directors.
Leverage	Total liabilities divided by total assets.
Ln(Assets)	The natural logarithm of firms' total assets, defined as the market value of equity plus the book value of debt.
Ln(Avg Pay Ee)	The natural logarithm of the average employee pay, defined as labor expense minus management compensation, divided by the number of employees (excluding management).
Ln(Avg Pay Mgmt)	The natural logarithm of the average manager pay, defined as total management compensation divided by the number of executive managers (excluding in independent directors).
Ln(CapEx)	The natural logarithm of capital expenditure on PP&E, intangible assets, and other fixed assets.
Ln(Covid Cases)	The natural logarithm of total Covid-19 cases in the firm's headquarter province in 2020.

Table A1 *(continued)*

Variable	Definition
Ln(Covid Deaths)	The natural logarithm of total Covid-19 deaths in the firm's headquarter province in 2020.
Ln(Ee Number)	The natural logarithm of the number of employees the firm has, excluding executives.
Ln(ESOP Ee Num)	The natural logarithm of one plus the number of employees participating in ESOPs.
Ln(Firm Age)	The natural logarithm of the firm's age.
Ln(Mgmt Number)	The natural logarithm of the number of executives the firm has.
Ln(Rev)	The natural logarithm of revenue.
LS	Labor share, defined as cash labor expense (including salaries, wages, and benefits) divided by revenue.
LS Ee	Employee labor share, defined as the cash labor expense minus management compensation, divided by revenue.
LS Mgmt	Management labor share, defined as management compensation divided by revenue.
Mgmt Hold	Management share holding, defined as common shares held by managers divided by the firm's total outstanding common shares.
Min 5/10 Pct	The number of advertised positions requiring at least five/ten years of experience, divided by the total number of advertised job positions.
PPE/Assets	Property, Plant, and Equipment divided by total assets.
Postgrad Pct	The number of advertised positions requiring postgraduate degrees, divided by the total number of advertised job positions.
R&D Pct	The number of advertised R&D-related positions (including IT research, development, engineering, testing, automation, and system design), divided by the total number of advertised job positions.
R&D/Rev	R&D Investment divided by revenue.
Rev Growth	The percentage change in revenue from year $t-1$ to year t.
Rev/R&D Ee	Revenue (in millions CNY) per R&D employee.
ROA	Return on asset, defined as net profits divided by total assets.

Table A2 Cybersecurity Law's Effect on Labor Share of Digital-Core Industries

The dependent variable is *LS*, defined as cash labor expense (including salaries, wages, and benefits) divided by revenue. *Core* is an indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China. *Post CSL* is an indicator for years including and after 2017 (the Cybersecurity Law was enacted in November 2016). Sample period is between 2014 and 2019. Standard errors are clustered by firm; *t*-statistics are in parentheses.

Dep Variable:	LS	
	(1)	(2)
Core × Post CSL	0.010***	0.008***
	(3.12)	(2.73)
Ln(Assets)		-0.000
		(-0.24)
ROA		-0.074***
		(-4.40)
Leverage		-0.012
		(-1.36)
Mgmt Hold		-0.007
		(-0.63)
Indep Dir		-0.015
		(-0.96)
Rev Growth		-0.016***
		(-8.27)
Ln(Firm Age)		-0.021
		(-1.04)
PPE/Assets		0.015
		(1.28)
N	15294	15294
Adj R-sq	0.01	0.05
Firm FE	Y	Y
Year FE	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A3 Data Protection Regulations' Effect on Capital Investment of Digital-Core Industries
The dependent variables are *Ln(CapEx)*, defined as the natural logarithm of capital expenditure on PP&E, intangible assets, and other fixed assets (columns 1 and 2), capital expenditure divided by revenue (columns 3 and 4), and capital expenditure divided by total assets (columns 5 and 6). *Core* is an indicator for industries core to the digital economy as classified by the National Bureau of Statistics of China. *Post* is an indicator for years including and after 2021 (i.e., enactment of the PIPL). Standard errors are clustered by firm; *t*-statistics are in parentheses.

Dep Variable:	Ln(CapEx)		CapEx/Rev		CapEx/Assets	
	(1)	(2)	(3)	(4)	(5)	(6)
Core × Post	-0.026	0.013	-0.003	-0.002	-0.000	0.000
	(-0.57)	(0.33)	(-0.80)	(-0.53)	(-0.22)	(0.23)
Ln(Assets)		0.615***		0.009***		0.001
		(21.56)		(2.89)		(0.68)
ROA		1.148***		0.021		0.007
		(7.18)		(1.42)		(1.57)
Leverage		-0.156		-0.105***		-0.047***
		(-1.22)		(-7.13)		(-10.33)
Mgmt Hold		0.602***		-0.013		-0.000
		(3.63)		(-0.84)		(-0.07)
Indep Dir		0.065		0.005		-0.002
		(0.23)		(0.18)		(-0.23)
Rev Growth		0.146***		-0.010***		0.004***
		(5.54)		(-3.24)		(4.25)
Ln(Firm Age)		-0.104		-0.046		-0.019
		(-0.26)		(-1.14)		(-1.32)
PPE/Assets		-2.505***		-0.306***		-0.123***
		(-12.95)		(-11.92)		(-14.29)
N	16127	16127	16127	16127	16127	16127
Adj R-sq	0.01	0.14	0.00	0.05	0.01	0.08
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$