Data Privacy, Scaling, and Firm Scope: Evidence from the GDPR

Abstract

The collection and use of personal data has become a key driver of the modern digital economy, and simultaneously data privacy has emerged as a critical strategic and public policy issue. Although prior research has examined many immediate impacts of data privacy protections on firms, the consequences of such protections on the ability of firms to scale has remained unexplored. The impacts of data privacy on scaling is vital for digital firms because it not only affects their core economic model, but may also have follow on strategic implications for important corporate strategies like diversification. To address these questions the current paper exploits the enactment of the GDPR in Europe and examines the effects of privacy protections on the scaling of digital firms, and in turn on firm scope. Differences-in-differences estimates indicate large decreases in scaling associated with stronger privacy protections, and similarly large follow-on impacts on firm scope. These findings have significant implications for the research literatures on data privacy, firm scaling, and firm diversification.

Data Privacy, Scaling, and Firm Scope: Evidence from the GDPR

1. Introduction

The dramatic growth of the digital economy has led to a significant increase in the collection, storage, process, and dissemination of personal data, making data privacy an important and pressing issue for firms, policymakers, and researchers (Fainmesser, Galeotti, & Momot, 2023; Goldfarb & Tucker, 2012). The prior literature on privacy protections has largely documented their negative impacts on organizational outcomes, such as impeding technology adoption and diffusion (Adjerid et al., 2016; Buckman, Adjerid, & Tucker, 2023; Johnson et al., 2023; Miller & Tucker, 2009; Peukert et al., 2022), reducing the efficacy of advertising (Goldfarb & Tucker, 2011), lowering financial performance (Aridor, Che, & Salz, 2020; Goldberg, Johnson, & Shriver, 2019; Ke & Sudhir, 2023; Sun et al., 2023), and hindering startups' fundraising (Jia, Jin, & Wagman, 2021). While the literature has provided empirical evidence of several economic consequences of privacy protections, it has provided only limited insights into their impact on firm scaling and in turn on corporate strategy.

Examining how scaling is affected by privacy protections is important because the ability to scale is a critical determinant of firms' long-term success (Hoffman & Yeh, 2018; Jansen et al., 2023). Conceived as a specific type of growth, scaling is characterized by a firm's creation of additional value without a proportional increase in costs (Giustiziero et al., 2023; Somaya & You, 2024). Prior research highlights that digital firms in particular are more scalable due to the "scale-free" nature of digital assets (Levinthal & Wu, 2010), which allows digital firms to increase their value creation with very limited incremental investments, leading to high gross margins and substantial economies of scale (Giustiziero et al., 2023; Li et al., 2023). However, these are precisely the firms that are likely to be more impacted by privacy protections. Data assets are critical scale-free resources used by digital firms, enabling them to customize their product offerings, derive more accurate forecasts, and operate more efficiently (Helfat et al., 2023). Therefore, understanding how scaling is affected by privacy protections, which inherently limit the collection and use of data assets, is of great theoretical and practical importance.

Drawing on research that highlights firm scaling as having important implications for corporate strategy (Giustiziero et al., 2023), we further extend our inquiry into the impacts of scaling on firm scope. A key rationale for firms to expand their scope is to use the services of their resources more effectively (Penrose, 1959). Prior research suggests that firms expand into financially attractive new markets wherein they can leverage their current resource configurations to generate synergies (Helfat, 1997; Rumelt, 1974; Silverman, 1999). However, in addition to synergies, prior research also highlights the opportunity costs of scarce resources as another key determinant of the scope of the firm (Bennett & Feldman, 2017; Dickler & Folta, 2020; Feldman, 2016; Feldman & Sakhartov, 2022; Kaul, 2012; Wu, 2013). As it relates to scaling, a firm's ability to scale operations within a business increases the opportunity costs of redeploying resources across multiple businesses, pushing firms towards greater specialization and narrower scope (Giustiziero et al., 2023). In line with this logic, when firms experience a decrease in scaling opportunities, they may be more likely to expand their scope due to a reduction in the opportunity cost of spreading resources across multiple businesses. While this theoretical rationale seems plausible, prior research lacks empirical evidence consistent with this predicted relationship between scaling and firm scope.

Accordingly, in this paper, we investigate the following two research questions: How do privacy protections affect firm scaling? And, how do changes in scaling (opportunities) affect firm scope? To address these questions, we employ two related empirical approaches. First, we leverage the implementation of the General Data Protection Regulation (GDPR) as a quasi-natural experiment to examine the impact of privacy protections on scaling. The GDPR became effective in the European Union (E.U.) on May 25, 2018, with an aim to provide E.U. residents with greater control and privacy protections for their personal data. Using a difference-in-differences (DID) approach, we find a 37.7 percentage decrease in scaling by U.S. public firms that were exposed to the GDPR (treated group) relative to firms that were not (control group). Second, to estimate the impact of scaling on firm scope, we use instrumental variable (IV) approaches that leverage the implementation of GDPR as a source of exogenous variation in firm scaling. The IV estimates suggest that a decrease in scaling leads firms to be more likely to expand their scope, as indicated by their increased diversification. Specifically, the decrease in scaling associated with GDPR implementation is correlated with a 36.0% increase in diversification relative to the mean value of diversification pre-GDPR.

The contributions of our study are threefold. First, we extend the emerging stream of research on firm scaling and its determinants (Huang et al., 2017; Li et al., 2023; Varga et al.,

2023) by highlighting how privacy protections and their attendant constraints on the use of data as a firm resource may act as a barrier to firm scaling. Additionally, by providing empirical evidence on the implications of scaling for the scope of the firm, our study responds to a recent call for more attention to the consequences of scaling (Genedy et al., 2024). Second, our study contributes to scholarship on corporate strategy, and in particular on the vital role of opportunity costs in resource allocation decisions and the scope of the firm (Bennett & Feldman, 2017; Dickler & Folta, 2020; Feldman, 2016; Kaul, 2012; Wu, 2013). We add to this prior work by integrating it with the scaling literature and examining empirical evidence for recent theoretical predictions about the relationship between scaling and corporate scope (Giustiziero et al., 2023). Third, our research provides valuable insights for policymakers. While policymakers have increasingly recognized the social value of implementing privacy-protecting regulations, our findings document how these protections can potentially lead to inefficiencies arising from firms' inability to effectively scale. Thus, policymakers may need to weigh the economic costs of privacy regulations against their economic benefits and explore ways to optimize privacy protection in a cost-benefit tradeoff with firm scaling.

2. Related Literatures and Hypotheses

2.1. Data privacy

The increased availability of digital data has caused substantial concerns for the potential misuse of personal data, which, in turn, have led to privacy-protecting actions undertaken by firms and regulators. Prior research has examined the impacts of enhanced privacy protections on several organizational outcomes. For example, there is consensus that enhanced privacy protections add administrative costs to data collection (Adjerid et al., 2016; Godinho de Matos & Adjerid, 2022). Further, prior research highlights that more stringent privacy protections impede the adoption of data-enriched technologies by firms (Adjerid et al., 2016; Buckman et al., 2023; Miller & Tucker, 2009). Along the same line, recent research on GDPR shows that the enforcement of GDPR substantially reduces firms' use of small technology vendors that provide support services to their websites (Johnson et al., 2023; Peukert et al., 2022). In addition, prior work demonstrates that enhanced privacy protections reduce the efficacy of advertising (Goldfarb & Tucker, 2011). Relatedly, emergent research has found a negative impact of privacy protections on web traffic and firm sales (Aridor et al., 2020; Goldberg et al. 2019; Sun et al., 2023). Moreover, it has been found that enhanced privacy protections coincide with lower venture capital received by technology startups (Jia et al., 2021). In response to more stringent privacy regulations, AI startups tend to reallocate limited resources and collaborate with large firms to access data assets (Bessen et al., 2020).

In sum, the empirical literature on privacy protections has largely demonstrated negative outcomes for firms, whereas some work has shown differential effects depending on the specific elements of privacy protections (Miller & Tucker, 2018). Other than these well-documented outcomes, privacy protections also entail a shift in firms' access to data assets that can profoundly affect scaling. However, there has been limited work examining the implications of privacy protections for scaling.

2.2.Firm scaling

Scaling represents a specific type of growth, characterized by a firm's creation of additional value without a proportional increase in costs (Somaya & You, 2024). Prior research emphasizes that one of the key determinants of scaling is the share of scale-free resources in a firm's resource bundles (Giustiziero et al., 2023). Scale-free resources can be used in many different non-competing applications without its value in any one application being substantially impaired (Levinthal & Wu, 2010; Teece, 1980). For example, Uber can leverage its geo-location technology across a wide range of domains without any opportunity cost of its use in one application or another. Compared to industrial firms, digital firms often have a greater ability to scale (Giustiziero et al., 2023). The rationale is that on the supply side, digital resources are often assumed to be of scale-free nature, that is, only small incremental investments are needed to extend their revenue generating potential (Levinthal & Wu, 2010). On the demand side, digital goods are less subject to downward-sloping demand due to the presence of network effects and easy access to large markets through digital distribution channels (Boudreau & Jeppesen, 2015; Wen & Zhu, 2019). As a result, digital firms also experience fewer demand-side constraints to scale (Giustiziero et al., 2023).

Data assets are viewed as critical scale-free resources for digital firms (Jones & Tonetti, 2020), enabling them to customize their product and service offerings, derive more accurate forecasts, and operate more efficiently (Helfat et al., 2023). The scale-free nature of data implies that data assets have high fixed costs of acquisition with low marginal costs of replication, making it an accelerator of scaling for digital firms. Several works have emphasized that digital firms can rapidly scale through leveraging data network effects: better access to data helps firms improve their products, attracting and keeping more users, which, in turn, leads to additional data (Gregory et al., 2022; Ichihashi, 2020). For example, Netflix leverages data network effects as it collects and analyzes data about how its platform is used and then draws on the learning outcomes to continuously improve its content and user interface to increase the perceived value of the streaming services offering through its platform. Further, data-driven learning also enables better informed decision making in that it can help in optimizing supply and demand coordination, streamlining transactions, and reducing costs (Varga et al., 2023). Thus, the leverage of data network effects not only accelerates the value creation for users, but also allows for increasing value creation at relatively lower costs, which is a key driver of scaling.

Since more stringent privacy protections constrain data collection and processing, they limit digital firms' ability to leverage data network effects and data-driven learning to achieve scaling. Even though the digital nature of firms' offerings creates opportunities for them to obtain returns to scale (Giustiziero et al., 2023), those offerings need to be appealing to customers to create value. Because data may lose relevance over time, the flow of new data is a critical driver of digital firms' value creation (Gregory et al., 2022). In the regime of stricter privacy protections, the right granted to customers to withdraw consent at any time increases uncertainty facing firms with respect to the ability to develop new products that are aligned with updated customer preferences. Therefore, we posit that enhanced privacy protections are an inhibitor of firm scaling.

H1. More stringent data privacy protections are associated with a decrease in firm scaling.

2.3. Scaling and the scope of the firm

Based on the notion that the firm is a unique bundle of resources (Penrose, 1959; Wernerfelt, 1984), prior research recognizes that firms are incentivized to expand their scope to exploit excess capacity of resources that have multiple uses but are subject to high transaction costs or market failure (Chang, Eggers, & Keum, 2022; Penrose, 1959; Teece, 1982). A firm's resource base can affect not only the decision of whether to expand but also the choice of product or market domains into which it expands (Chandler, 1962; Mahoney & Pandian, 1992). Specifically, the concept of resource-based synergies plays a key role in predicting the direction of firm expansion. When businesses (a) and (b) share some common factors of production, they achieve synergies or economies of scope because their joint production costs are less than the sum of their stand-alone production costs, namely, Cost (a, b) < Cost (a) + Cost (b) (Panzar & Willig, 1981). Since such synergies are linked to the similarity of resource requirements between businesses, prior studies have shown that firms are more likely to become diversified by entering businesses that are more related to their existing businesses (Helfat, 1997; Rumelt, 1974; Silverman, 1999).

While sharing resources across businesses can create synergies that encourage firms to expand their scope, prior research highlights that the opportunity costs of resource allocation provide an alternative explanation for decisions on firm scope (Bennett & Feldman, 2017; Dickler & Folta, 2020; Feldman, 2016; Feldman & Sakhartov, 2022; Giustiziero et al., 2023). For example, existing research has examined how the opportunity costs of non-scale free capabilities impact diversification decisions (Kaul, 2012). Unlike scale-free resources, non-scale free resources must be allocated among alternative uses and their use generates opportunity costs (Levinthal & Wu, 2010). If a firm relies on non-scale free resources to expand, its businesses have to compete for non-scale free resources, such that some businesses in the portfolio gain success at the expense of others. Under these circumstances, firms should make their expansion decisions according to the opportunity costs of using non-scale free resources in one domain or another. For example, Wu (2013) shows that when facing the decision to participate in multiple markets, firms need to consider the relative demand conditions across alternative markets, which determine the opportunity costs of deploying non-scale free capabilities.

Although digital firms in general possess a large share of non-scale free resources, the co-existence of both scale-free and non-scale free resources implies that these firms also need to allocate their resource bundles to their best use. When scaling is high, any alternative resource-market position is more likely to be inferior to the existing one, since the opportunity costs to use non-scale free resources, such as managerial time and attention, in other markets are very high (Giustiziero et al., 2023). As a result, digital firms with highly scalable resource bundles are likely to specialize in a narrow set of activities because specialization allows them to focus on and maximize returns from their most scalable resources (Giustiziero et al., 2023). In this respect, high scaling often involves highly efficient use of non-scale free resources. That is, there are fewer non-scale free excess resources. Hence, it is reasonable to assume that as scaling goes up, the motivation to search in new markets for a way to use these excess resources goes down. Consistent with this logic, a reduction in scaling increases the opportunity costs of not redeploying non-scale free resources to alternative markets where these resources are fungible. Thuserefore, firms are more likely to diversify when their scaling within existing businesses declines. One may argue that a new market entry can be based on the use of valuable scale-free resources. However, even if a new market entry is based on the exploitation of scale-free resources, it normally requires the use of additional resources that are subject to opportunity costs, such as managerial time and attention, to complement the scale-free resources. In conclusion, these considerations about opportunity costs suggest that the lower the firm's ability to scale, the lower the opportunity costs to redeploy its resources to new markets, thus increasing a firm's propensity to expand its scope.

H2. A reduction in firm scaling is associated with an increase in diversification.

3. Data and Methods

3.1.GDPR background

Designed to safeguard personal data and privacy rights, the GDPR was passed by the E.U. in April 2016, and came into effect on May 25, 2018. While GDPR is a comprehensive regulation that covers multiple aspects of data privacy protections, three aspects are especially relevant for our paper: territorial application, consent and data minimization, and increased compliance risks.

First, the regulation is applicable to both E.U. firms and non-E.U. firms that have business operations in the E.U. (Art. 3(1) & (2) GDPR). Second, under the GDPR a firm needs informed, specific, and unambiguous consent from customers to process their personal data, which requires customers to explicitly opt into data collection (Art. 4(11) GDPR). Further, GDPR requires that customers be able to update or withdraw their consent at any time (Art. 7(3) GDPR). Third, the GDPR has drastically increased compliance risks for privacy violations. For example, to comply with the GDPR, a firm must inform its customers about any transfer of personal data to a third party; otherwise, the firm may be jointly liable for privacy violations (Art. 13(1) & 28(1) GDPR). This implies that GDPR has created an environment where data sharing within firm boundaries is less risky than data sharing across boundaries. Overall, aiming to provide individuals with greater control over their personal data, GDPR represents a significant shift in how data privacy and protection are addressed.

3.2. Identification strategy

Identifying the effect of privacy protections on scaling is challenging as firms' decisions to enhance privacy protections are generally not exogenous; firms that do so are more likely to benefit from it. Specifically, decisions to enhance privacy protections presumptively depend on firms' capability, business model, and technology, which, in turn, can be correlated with their scaling outcomes. Thus, to estimate the effect of privacy protections on scaling, one needs to find an empirical context in which variation in privacy protections arises exogenously. The specific source of exogenous variation we exploit in this paper is the implementation of GDPR. We leverage it as a quasi-natural experiment and use a DID design to control for stable unobserved differences between the comparison groups (Angrist & Pischke, 2009).

To estimate the effects of firm scaling on the scope of the firm, we adopt IV approaches (Wooldridge, 2010). Since scaling is a result of firms' strategic choices, there might exist unobservable attributes that can affect scaling and firm scope simultaneously. Thus, simple regression analyses can be subject to omitted variable biases. To address such endogeneity concerns, we employ a research design that leverages the implementation of GDPR as an IV, which creates an exogenous source of variation for firm scaling. In a nutshell, we use the implementation of GDPR as a quasi-natural experiment to examine the impact of enhanced privacy protections on firm scaling. Additionally, leveraging it as an IV for scaling, we examine the relationship between scaling and firm scope.

3.3.Sample and data

To test our hypotheses, we construct a panel data set of U.S. publicly listed firms using Compustat North America, which contains detailed accounting information for those firms. Our sample is based on U.S. public firms for two main reasons. First, some U.S. firms operate in the E.U. and are thus affected by the shock ("treated" firms), whereas those not serving E.U. customers are unaffected ("control" firms). This geographic distinction serves as a prerequisite for our quasi-experiment design. Second, according to our definition of scaling, which compares increase in value creation relative to increase in resource commitment, a measure of scaling would require data on input and revenue for at least two observation periods. The longitudinal accounting information available to public firms allows us to create consistent and generalizable quantitative assessments of scaling over time.

Due to their substantial dependence on IT and customer data for business operations and strategic decision-making (Cohen, Gurun, & Kominers, 2019; Kim, Gopal, & Hoberg, 2016; Susarla & Mukhopadhyay, 2019), we assemble a sample of firms from four industries identified by two-digit Standard Industrial Classification (SIC) codes: SIC 35 (Computer Equipment), SIC 36 (Electronic & Other Electrical Equipment & Components), SIC 48 (Communications), and SIC 73 (Business Services). Additionally, we include firms in SIC 5961 (E-commerce) in the sample. The industry distribution in our sample is as follows: 3%, 26%, 12%, 56%, and 3%, respectively. While healthcare and finance related industries are also data-intensive, we do not include them primarily because they tend to be subject to industry-specific regulations, which may confound the effect of the GDPR. To identify which firms were affected by the GDPR, we obtain business segment data from Compustat Segments. We further collect information on E.U. sales from Capital IQ to complement missing data and ensure data accuracy. In cases where segment-level data are absent from both databases, we rely on 10-K annual reports (Item 1 and Item 1A) to determine whether a firm operated in the E.U. and was affected by the GDPR. We exclude firms that are headquartered outside the U.S. and drop firm-year observations for which the necessary accounting variables are missing. For the main analysis, our data cover the period between 2015 and 2020. This time frame spans three years before and three years after the GDPR took effect, allowing us to capture the near-term effects of the change in privacy regulation.

These selection criteria yield an initial sample of 725 firms. To address the concern that unobserved differences between treated and control firms may affect our results (see Section 4.2), we use coarsened exact matching (CEM, Iacus, King, & Porro, 2012). Our resulting matched sample consists of 598 firms and 2,882 firm-year observations.

3.4. Measures

3.4.1. Dependent variables

Scaling$_{it}$. We develop an accounting-based measure that derives from our definition of scaling, that is, the increase in value created with the incremental application of resources (Giustiziero et al., 2023; Somaya & You, 2024). *Scaling$_{it}$* is operationalized as $(Revenue_{it} - Revenue_{it-1})/(COGS_{it} - COGS_{it-1})$. The main underlying assumptions of this measure are that the firm's changes in revenue track the changes in total value it creates, and that its changes in cost of revenue track the changes in total resources applied. Under these assumptions, "change in revenues divided by change in COGS" serves as an objective proxy for firm scaling.[1] *Ceteris paribus*, firms that have a higher ratio of change in revenue to change in COGS are more scalable.

Diversification$_{it}$. The diversification measures used in prior research can be categorized into three types: number of segments, multimarket indicator, and concentration of segment

[1] While this measure can take negative values, it is meaningless to discuss scaling when the firm is not growing. To mitigate potential biases caused by negative values, we adjust this measure in three ways: (1) if the firm's revenue decreases, its scaling equals zero regardless of the change in cost of revenue; (2) if the firm's revenue increases and COGS decreases, we drop the observation from the sample; and (3) because the measure is a ratio that can sometimes take extreme values, we drop observations whose scaling is above the 95th percentile.

shares. The number of segments is a count of the business segments with different four-digit SIC codes (Villalonga, 2004). Since the Securities and Exchange Commission's (SEC) requires firms to report business segments that exceed 10% of the firm's sales, assets, or profits in their annual reports, the maximum number of *SIC count*$_{it}$ is 10. The multimarket indicator is a binary variable which captures whether the firm is diversified or specified (Becerra, Markarian, & Santalo, 2020). *Multimarket*$_{it}$ is coded as one if firm *i* is reported as being involved in at least two 4-digit SICs in year *t* and zero otherwise.

The concentration of segment shares within a firm is often measured using Hirschman-Herfindahl index (HHI), as defined by the equation: HHI = $\sum p_i{}^2$, where p_i is the ratio of the segment-level sales (assets) to the firm-level total sales (assets). The HHI-based measure takes a maximum value of 1 when the firm operates in a single segment and decreases with greater dispersion of sales (assets) across business segments. We operationalize *Diversification*$_{it}$ as one minus HHI sales (assets). Other than HHI-based measures, we adopt the entropy index (Palepu, 1985) as an alternative measure of diversification. Specifically, the entropy index is defined as: Entropy = $\sum p_i \cdot \ln{(1/p_i)}$, where p_i is the sales share of segment *i* (in different four-digit SIC codes).

3.4.2. Independent variables

The variable, *Post*$_t$, is an indicator variable that equals 1 for 2018 or thereafter; and 0 otherwise. We use the geographic reach of the GDPR to divide the sample into treated and control groups. Specifically, our treated group consists of firms that reported sales in the E.U.

region in 2017 or prior to it, whereas those firms that did not have market presence in the region during this period serve as our control group. In cases where a firm's geographical segment data are missing, we examine whether the firm discusses the GDPR in its 10-K annual reports. The inclusion of such discussions indicates that the firm falls into the treated group. The variable, $Treated_i$, is an indicator variable that equals 1 if a firm is in the treated group; and 0 otherwise.

3.4.3. Control variables

We include several firm-level variables to account for their potential effects on firm scaling and diversification, including firm size, leverage, return on assets (ROA), organizational slack, and capital intensity. Specifically, *Firm size* is measured as the natural logarithm of one plus assets. *Leverage ratio* is debt divided by assets. *ROA* is net income divided by assets. *Slack* is cash holdings scaled by assets. *Capital intensity* is capital expenditure scaled by assets.

3.5.Summary statistics

Table 1 provides descriptive statistics for the main variables and the corresponding correlation matrix. Our sample consists of 2,882 firm-year observations from 598 unique U.S. public firms over the 2015-2020 period. We have 1,474 treated firm-year observations from 299 unique firms. Our sample firms have mean *Scaling* of 1.46. The standard deviation of scaling is 1.76, which suggests considerable variations. As shown, there is a large positive correlation between various diversification measures. The correlations between *Scaling* and measures of diversification are negative. This is suggestive of Hypothes2, according to which scaling is negatively associated with diversification.

4. DID Analysis: The Effect of Privacy Protections on Scaling

4.1. Econometric models

We run a fixed effects DID regression to empirically test and quantify the effect of an exogenous increase in privacy protections on firm scaling. Specifically, we estimate the following DID regression:

$$y_{it} = \beta_1 \text{Treated}_i * \text{Post}_t + \delta X_{i,t} + \alpha_i + \alpha_t + \varepsilon_{it} \qquad (1)$$

where i indexes firms; t indexes years; α_i and α_t are firm and year fixed effects. The dependent variable of interest is y_{it}, which is firm i's scaling at time t. $Treated_i$ is a dummy variable that equals one if the firm is in the treated group. $Post_t$ is a dummy that equals 1 if the year is in or after 2018. Following existing research on GDPR, we use 2018, the year of GDPR enactment, as the treatment year (Burford et al., 2022; Johnson et al., 2023; Peukert et al., 2022). Additionally, we collect data on search volume related to the topic of GDPR over time using Google Trends. As shown in Figure 1, the implementation of GDPR led to a sharp increase in web searches, in contrast to the negligible change observed after its passage in April 2016. This observation strengthens our rationale for adopting 2018 as the appropriate treatment year. In supplementary analysis, we perform placebo tests and find no significant change in firm scaling after the passage of GDPR in 2016 (see Section 4.4.1). $X_{i,t}$ is the vector of control variables, which are lagged by one year. ε_{it} is the error term. The identification comes from comparing the intertemporal change in scaling in the treated group against the baseline change in scaling in the control group. The regression is estimated by ordinary least squares (OLS). Across all of our specifications we cluster the standard errors at the firm level to account for correlation in errors within the firms. The coefficient of interest is β_1, which represents the average treatment effect of GDPR on scaling across all treated firms.

[Insert Figure 1 about here]

4.2. Validity of the identification strategy

We identify the implementation of GDPR as a suitable event that suddenly shifts firms' privacy protection behavior given that its enforcement is stringent. Failure to comply with GDPR provisions is punishable with fines that can go as high as 20 million euros or 4% of worldwide annual revenue of the prior financial year, whichever is higher (Art. 83(5) & (6) GDPR). While GDPR is legally enforceable, this characteristic alone does not guarantee that they lead to an actual change in firm behavior because the cost of strict GDPR compliance may exceed the maximum fines. We thus search for further suggestive evidence to demonstrate the relevance of this policy change. For example, research has demonstrated that web technology providers that use cookies are requested less following the GDPR, and they tend to adopt more transparent privacy policies following the GDPR (Peukert et al., 2022). Moreover, anecdotal evidence shows that the implementation of GDPR has significant impacts on digital firms' operations,[2] such as Apple updating its privacy terms, Shopify modifying its app permissions for merchants and developers, and Google releasing new consent requirements to developers.

[2] https://techcrunch.com/2018/05/23/apple-introduces-new-privacy-portal-to-comply-with-gdpr/
https://www.shopify.com/partners/blog/gdpr-compliance
https://www.theregister.com/2018/05/24/google_gdpr_mobile_app_devs_choices/

The validity of our identification strategy also depends on the assumption that the timing of the policy change is not systematically endogenous to firm activities. This assumption can be investigated empirically by testing the existence of preexisting trends prior to the treatment. In Figure 2, we show that treated and control firms follow similar trends during the pre-period and there are significant differences between these two groups of firms during the post-period. Further, we examine the dynamics of the treatment effect. If our results are driven by endogenous factors, the policy change should have impacted treated firms already before they had been implemented. However, when we look at the dynamic effect of the treatment, we find no evidence for such preexisting trends (see section 4.3). These findings further alleviate concerns relating to the potential endogeneity of this policy change.

Another potential concern of the identification is that unobserved differences between treated and control firms may affect our results. To address this concern, we construct a sample of matched firms using CEM. We match each treated firm to a control firm on their revenues before the implementation of GDPR. To demonstrate the validity of our matching process, we conduct a balance test to examine whether the treated firms and matched control firms are similar in their firm fundamentals in 2017. Specifically, we test whether variables such as revenue, COGS, assets, net income, growth, gross profit, diversification, and liquidity for treated and control firms are statistically equal at mean. The results presented in Table 2 show that the differences in these variables are generally not statistically significant, ensuring validity of the DID approach.

[Insert Table 2 about here]

4.3.Regression results

The main results of DID estimation are presented in Table 3. The specification in column (1) includes the interaction of treatment and post dummy, firm and year fixed effects. In column (2), we further include control variables. As shown, the coefficient of the interaction term (*Treated* * *Post*) is negative and statistically significant (β = -0.550, p = 0.000), which represents a 37.7% decrease in scaling relative to the mean value of scaling prior to the implementation of GDPR. These results are in line with Hypothesis 1, indicating that enhanced privacy protections are associated with a decrease in firm scaling.

[Insert Table 3 about here]

In column (3), we further assess year-by-year changes in the scaling of treated firms around the implementation of GDPR by adding interaction variables of the *Treated* * *Post* with each of the years in the period 2016-2019, with 2017 being the reference year. As shown, the coefficients of all pretreatment dummies (2015 and 2016) are small and insignificant. This finding provides additional evidence that there is no preexisting trend in the data. The negative and statistically significant coefficients of 2018, 2019, and 2020 suggest that the decrease in scaling starts in 2018, the year of the implementation of the policy, and the policy has a long-lasting effect on firm scaling. Figure 3 exhibits coefficient plots of the yearly treatment tests for firm scaling. Overall, the patterns in the figure mirror the patterns in column (3) of Table 3.

[Insert Figure 3 about here]

4.4.Supplementary analyses

4.4.1. Placebo tests

Consistent with prior research studying the treatment effect of policy change, one of identification assumptions we rely on is that firms did not react to the passage of GDPR in 2016. To examine whether this assumption holds in our study, we run a placebo test in which the placebo treatment is two years before the actual treatment. Specifically, we include "*Treated* * *Placebo*" in Equation (1), where *Placebo* equals one between 2016 and 2018; and zero between 2013 and 2015. The coefficient of this interaction captures the potential treatment effect of the passage of GDPR. The results are reported in column (1) of Table 4. We do not find a significant decrease in firm scaling using this advanced cutoff year, which mitigates the concern that the GDPR effect might have kicked in during pre-treatment period because it was announced in 2016.

Another potential concern with our findings is that GDPR might be confounded by other events in E.U. (e.g., Brexit) and macroeconomic conditions that could have impacted the scaling of firms operating in the E.U. differently than firms operating outside the E.U. To alleviate such concerns, we run another placebo test in which we examine changes in the scaling of firms that are arguably scalable but do not employ substantial digital resources in their businesses. If we observe that the GDPR treatment also produces a difference in estimated scaling of non-digital firms, we might worry that our effects are produced not by changes in privacy protection but rather by other contemporaneous events. Here, we show that this does not appear to be the case. We compile a sample of U.S. public firms in industries whose SIC start with 28 (Chemicals and allied products). The results of this analysis are presented in column (2) of Table 4. As shown, the change in scaling of treated firms is not significantly different from the change in scaling of control firms following the implementation of GDPR. The results help us rule out the possibility of confounding events and further alleviate the concern that we capture a spurious relation between privacy protections and firm scaling.

[Insert Table 4 about here]

4.4.2. Heterogeneous treatment effects

Our findings thus far show that the implementation of the GDPR is overall negatively associated with scaling. Yet there may exist heterogeneity in the effects across different types of firms. For example, although all firms that serve E.U. customers are subject to the GDPR, the exposure may not be uniform across firms; it may vary depending on the extent to which their sales derive from E.U. Specifically, firms' reaction to the GDPR should be larger for those whose businesses rely more heavily on the E.U. market. To explore this heterogeneity, we use the proportion of E.U. sales to total sales in 2017 to create a measure of *EU intensity* and use it as the treatment intensity in our continuous DID estimation, similar to Brynjolfsson, Hui, and Liu (2019). Column (1) of Table 5 shows that scaling decrease is 0.47 percentage points larger for firms with 1% greater EU intensity. This implies a 9.5 percentage points overall scaling decrease given that the average EU intensity is 20.2% in 2017.

Additionally, we explore the heterogeneous effects of the policy change by distinguishing platform firms from non-platform firms. Platform-based firms are more likely to be subject to network effects and thus particularly susceptible to the effects from the GDPR. If the observed scaling changes estimated from Equation (1) are related to the implementation of GDPR, we should expect to see more pronounced effects in firms adopting platform-based business models. We test this hypothesis in column (2) of Table 5 by including a three-way interaction term of *Treated*, *Post*, and *Platform*, where *Platform* is a dummy variable indicating whether the firm adopts a platform-based business model. The negative and significant coefficient suggests that platform-based firms are subject to more decreases in scaling following the GDPR. Further, columns (3) and (4) of Table 5 report the results from applying the baseline DID specification, separately for each of two groups, according to their business models. Column (3) indicates that GDPR is negatively associated with scaling among firms adopting platform-based business models. For the group comprising non-platform firms, column (4) suggests a smaller negative effect on scaling compared with firms in the control group. A seemingly unrelated regression (SUR) test indicates that the coefficients on *Treated * Post* significantly differ across the platform firms and non-platform firms, such that the platform group is associated with larger negative effects following the GDPR's implementation.

We posit that subsample heterogeneity may also result from different industry groups. Columns (5) to (7) of Table 5 report industry-specific effects on scaling for more data-intensive firms and less data-intensive firms. While the implementation of GDPR has significant negative effects across the two industry categories, an application of a SUR test shows that the negative effects are significantly larger for firms that are more data related. Although the GDPR's effects are relatively smaller on firms that are less data related, the still substantial negative effects indicate that the GDPR is widely transformational across the digital technology sector.

[Insert Table 5 about here]

4.4.3. California Consumer Privacy Act (CCPA)

When assessing the impact of a policy intervention, a common concern is the presence of contemporaneous shocks that can affect focal variables of interest. For example, in our context, the implementation of the GDPR coincides with the passage of CCPA. With the aim of enhancing the privacy protection rights of its residents, California was the first state in the U.S. to pass a consumer privacy law comparable to the European GDPR. The CCPA was passed on June 28, 2018, and became effective on January 1, 2020. To examine whether CCPA potentially contributes to the effects we identify, we conduct three additional tests. First, we rely on a sample of treated and control firms that are matched on whether they are headquartered in California, coded as *CA HQ*. The estimates in column (1) of Table 6 show a negative and significant coefficient on *Treated* * *Post*, consistent with our main findings.

Second, if the implementation of CCPA plays a role in our observed patterns of scaling, it should disproportionately affect firms headquartered in California. We test this by introducing a three-way interaction of *Treated*, *Post*, and *CCPA* to the model, where *CCPA* is coded as one if the firm is headquartered in California and the time is in 2020. Column (2) of Table 6 shows that the three-way interaction term is insignificant, which rules out the possibility that our results are driven by CCPA. Third, to further tease out potential confounding effects of CCPA, we drop all observations in 2020 and assess the effect of GDPR on firm scaling within this shorter time horizon. Column (3) of Table 6 suggests that the effects we observe continue to hold if we focus on this subperiod. Although we acknowledge that we cannot completely rule out the potential effect of CCPA, these analyses provide evidence that CCPA is not a primary driver of our results.

[Insert Table 6 about here]

4.4.4. Robustness tests

We conduct several additional analyses to verify the robustness of our primary results. First, to show that our main results are not sensitive to the way scaling measure is adjusted, we limit our sample to firms that have at least one non-zero scaling observation both before and after the shock. Column (1) of Table 7 reports the results and confirms consistency with our primary findings. Second, because there is an increase in scaling in the control group in 2018, there might be concerns that our results are driven by this shift happened in the control group. To rule out this explanation, we adopt a simple before-after design, although we acknowledge that such before-after comparison does not control for potential systematic changes over time. We examine the before-after effect of the GDPR among firms in the treated group to see whether we find a similar pattern there. The results, reported in column (2) of Table 7, are consistent with our main results, implying that our results are unlikely to be driven by changes in the control group after the policy change. Overall, the results in this section give us confidence that our results are robust and are not sensitive to our assumptions and model specifications.

[Insert Table 7 about here]

5. IV Analysis: The Effect of Scaling on Firm Scope

5.1. Econometric models

The instrumental variable, GDPR, is operationalized as a binary indicator that equals one for the treated firms after 2018; and 0 otherwise. With both firm and year fixed effects, the coefficient of this variable represents the DID estimate of the impacts of GDPR on scaling. The two-stage least squares (2SLS) regression models are specified as follows:

$$1^{st} \text{ Stage: } x_{it} = \beta_1 \text{ GDPR}_t + \gamma_1 \text{ Controls}_{i,t} + \alpha_{1i} + \alpha_{1t} + \varepsilon_{1it}$$

$$2^{nd} \text{ Stage: } y_{it} = \beta_2 \hat{x}_{it} + \gamma_2 \text{ Controls}_{i,t} + \alpha_{2i} + \alpha_{2t} + \varepsilon_{2it}$$

where x_{it} refers to scaling of firm i in year t; y_{it} indicates the level of diversification; α_{1i} and α_{2i} capture firm fixed effects; α_{1t} and α_{2t} represent year fixed effects; and ε_{1it} and ε_{2it} are error terms. To test Hypothesis 2, β_2 is the coefficient of interest, which captures the effects of scaling on diversification. To account for the heteroskedasticity and the serial correlations of the errors, the standard errors are clustered at the firm level.

5.2. Validity of the IV

The IV needs to satisfy two conditions: the relevance conditions and the exclusion restriction. We test for the relevance of the instrument by estimating a model with scaling as the dependent variable. This IV satisfies the relevance condition because it is strongly correlated with the endogenous variable scaling. In addition, the IV plausibly satisfies the exclusion restriction condition, which in this case means that the implementation of GDPR is not related to firm scope other than through decreased scaling. While this assumption cannot be directly tested, we conduct auxiliary analyses that rule out some plausible ways in which it could be violated due to other channels (see Section 5.4).

5.3.Regression results

The results of 2SLS regressions are shown in Table 8. Column (1) reports the first-stage regression that predicts the change in scaling following the GDPR. Consistent with Hypothesis 1, the policy change decreases firm scaling (β = -0.533, p = 0.000). The second-stage results from the IV analysis are presented in column (2). Consistent with Hypothesis 2, we find that a one unit increase in scaling is estimated to decrease diversification by 4.7 percentage points (β = -0.047, p = 0.026). Put differently, the decrease in scaling after GDPR implementation is associated with a 36.0% increase in diversification relative to the mean pre-GDPR level, which represents an economically substantial impact of scaling on diversification. In Table 8, we also present the results of 2SLS using alternative measures of diversification as the dependent variable, all of which lead to similar results. Overall, these results indicate that scaling is negatively associated with firm scope.

[Insert Table 8 about here]

5.4.Supplementary analyses

While the implementation of GDPR provides a clean measure of change in scaling, it may affect firm scope in ways other than through reduced scaling, thus violating the exclusion restriction assumption of IV analysis. One possible alternative path through which GDPR affects diversification is increasing operating costs. We test this idea by using logged operating costs and

COGS as dependent variables. The results are shown in columns (1) and (2) of Table 9. If this alternative path is present, we should expect the treatment effects to be positive after GDPR. Interestingly, we find that there is no significant increase in both types of cost, thus alleviating the concern of an invalid IV.

In addition to the mechanism we propose in Hypothesis 2, the observed increase in diversification following the GDPR could be explained by another mechanism: the implementation of GDPR increases transaction costs of exchanging data resources across firm boundaries, which, in turn, may incentivize firms to enter data-related industries. To test whether the transaction cost mechanism explains the observed pattern, we run two analyses that explore whether the results are sensitive to different types of diversification. First, we estimate the effects of GDPR on related versus unrelated diversification, which are created based on entropy measures. The intuition for this analysis is as follows. Assuming that the transaction cost mechanism is at work, we should expect firms that did not operate in data-related industries pre-GDPR to enter those industries to internalize data transfer, thus increasing unrelated diversification. In that case, GDPR should have a similar impact on both related and unrelated diversification. By contrast, if diversification is mainly driven by reduced scaling and thus the need for resource redeployment, GDPR should be associated with related diversification but not with unrelated diversification. The insignificant coefficient of *Treated * Post* in column (4) of Table 9 suggests that the transaction cost mechanism may not be present. Second, we extend this line of inquiry by considering another categorization of diversification. Specifically, we divide all the industries into two segments: data-related segment (i.e., SIC 7374 and SIC 7375) and non-data-related segment. Similar to entropy-based related and unrelated diversification measures, we develop two measures to assess the extent to which firms diversify *within* and *across* these two segments. If the transaction cost mechanism explains diversification, we will find increases in diversification *across* these two segments after the GDPR. That is, firms that were in either of these two segments are now operating in both segments to reduce transaction costs of data transfer. In contrast, as columns (5) and (6) of Table 9 show, we find no significant impact on across diversification but significant increase in within diversification. Together, these results lend further support to the scaling path and the validity of GDPR as an IV for scaling.

[Insert Table 9 about here]

6. Conclusion

This paper examines the impact of privacy protections on firm scaling and how changes in scaling affect the scope of the firm. To address the first question, we exploit a quasi-natural experiment provided by the enactment of GDPR. This policy change constrains firms' ability to collect and process personal data, thus providing a quasi-exogenous variation in the strictness of privacy protections facing U.S. public firms that serve E.U. customers. Using a DID methodology, we find that the implementation of GDPR relates to a 37.7% decrease in scaling relative to the mean value of scaling pre-GDPR. We further posit that a decrease in scaling is associated with increased diversification because it reduces the opportunity cost of spreading resources across multiple businesses. Using an IV approach that leverages the implementation of

GDPR as an IV for scaling, we find that the decrease in scaling associated with GDPR implementation relates to a 36.0% increase in diversification compared to its pre-GDPR mean.

We acknowledge several limitations of our paper, which may provide opportunities for future research. First, we operationalize scaling using change in revenue relative to change in cost of revenue. While this measure is an objective proxy for scaling, it is built upon the assumptions that the value created by a firm is captured by its revenue, and the application of additional resources towards expanding the firm's output is reflected in cost of revenue. Should more detailed data become available, future research could shed light on the validity of this scaling measure. Second, we acknowledge that the implementation of GDPR may not be purely exogeneous. For example, firms may have made compliance efforts after the passage date and before the enforcement deadline. While we cannot completely rule out this possibility, the placebo test helps us mitigate this concern. Third, we measure diversification at the four-digit SIC level whereas overlooking within-industry diversification. Providing evidence on this type of diversification is a challenging task that requires detailed micro data on the firm's operations, but this could be an interesting question that merits further study.

Despite these limitations, our study contributes to several streams of literature. First, we add to the growing body of work on scaling (DeSantola & Gulati, 2017; Giustiziero et al., 2023; Jansen et al., 2023). Although prior studies have begun to examine factors that drive scaling (Huang et al., 2017; Li et al., 2023; Varga et al., 2023), they provide limited insights into how scaling is affected by institutional shifts that change the availability of a critical scale-free input.

Past studies have documented that digital resources are closely associated with scaling potential and success due to their capacity to enable growth with minimal incremental resource commitments (Adner, Puranam, & Zhu, 2019; Giustiziero et al., 2023). While the employment of digital resources may alleviate some of the bottlenecks that would have traditionally constrained firms' ability to achieve rapid scaling, our findings demonstrate that this does not simply remove scaling challenges or may even introduce new challenges. As regulatory frameworks and consumer responses to privacy issues evolve, digital firms may face more constraints on the exploitation of scaling opportunities embedded in digital resources. Additionally, we also contribute to the scaling literature by offering empirical evidence on the implications of scaling for the scope of the firm (Giustiziero et al., 2023). In doing so, our research responds to a recent call for more attention to the consequences of scaling (Genedy et al., 2024). Specifically, our findings indicate that when firms find themselves confronted with a sudden constraint on scaling, they tend to reallocate resources and expand firm scope to overcome such constraints.

Second, our paper complements the large body of literature on corporate strategy. The extant literature has highlighted two distinct mechanisms through which we can understand and explain the scope of the firm. One mechanism has been developed in the transaction cost associated with excess resources that arise from Penrosean growth (Penrose, 1959; Teece, 1982). So far strategic explanations of firm scope choices have been dominated by this mechanism and the associated notion of synergies (Ahuja & Novelli, 2017; Zhou, 2011). A second mechanism, based on opportunity costs, plays an important role when resource allocation is assessed with respect to alternative use within the firm (Bennett & Feldman, 2017; Dickler & Folta, 2020; Feldman, 2016; Feldman & Sakhartov, 2022; Kaul, 2012; Levinthal & Wu, 2010; Wu, 2013). Our focus with respect to firm scope is aligned with this second mechanism. Specifically, we emphasize that there exists a trade-off between the intensive use of resources within a single business and their allocation across multiple business activities. Recognizing that digital firms often possess both scale-free and non-scale free resources that are co-specialized in forming resource bundles to create value, recent research emphasizes the opportunity costs of resource allocation as a critical determinant of firm scope (Giustiziero et al., 2023). However, a robust empirical test of this emerging theory—scaling reduces firm scope—is lacking in part due to some empirical challenges, such as the endogeneity of scaling and the difficulty of operationalizing scaling. We address these challenges by adopting an IV approach and developing an accounting-based measure of scaling. In doing so, our study contributes to this growing literature by offering empirical evidence for the role of scaling in changing corporate scope.

Third, our findings have important policy implications. As data availability has grown exponentially and data analytics has been increasingly adopted by firms, there may be many reasons for legislators to impose restrictions on firms' ability to collect and process personal data. While prior literature sheds light on the potential economic consequences of enhanced privacy protections (Johnson et al., 2023; Ke & Sudhir, 2023; Sun et al., 2023), it largely overlooks their implications for scaling, which is arguably a prioritized objective pursued by most digital firms.

By focusing on the implications for firm scaling, our paper extends this literature and complements the emerging body of work on GDPR. Our findings indicate that privacy is not free. While privacy is valuable to individuals, by limiting the collection and use of personal data, any new privacy regulations will likely make it more challenging for digital firms to scale. Policymakers may thus need to weigh the economic costs of privacy regulations against their economic benefits and explore ways to optimize privacy protection in a cost-benefit tradeoff with firm scaling. For example, privacy protections may have differential effects depending on the specific provisions of privacy laws (e.g., consent requirements, rights over data, etc.).

References

Adjerid, I., Acquisti, A., Telang, R., Padman, R., & Adler-Milstein, J. (2016). The impact of privacy regulation and technology incentives: The case of health information exchanges. *Management Science*, *62*(4), 1042-1063.

Adner, R., Puranam, P., & Zhu, F. (2019). What is different about digital strategy? From quantitative to qualitative change. *Strategy Science*, *4*(4), 253-261.

Ahuja, G., & Novelli, E. (2017). Redirecting research efforts on the diversification-performance linkage: The search for synergy. *Academy of Management Annals*, *11*(1), 342-390.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press: Princeton.

Aridor, G., Che, Y. K., & Salz, T. (2020). The economic consequences of data privacy regulation: Empirical evidence from GDPR. *NBER working paper 26900*.

Becerra, M., Markarian, G., & Santalo, J. (2020). The effect of import competition on product diversification revisited. *Strategic Management Journal*, *41*(11), 2126-2152.

Bennett, V. M., & Feldman, E. R. (2017). Make room! Make room! A note on sequential spinoffs and acquisitions. *Strategy Science*, *2*(2), 100-110.

Bessen, J. E., Impink, S. M., Reichensperger, L., & Seamans, R. (2020). GDPR and the importance of data to AI startups. *NYU Stern School of Business*.

Boudreau, K. J., & Jeppesen, L. B. (2015). Unpaid crowd complementors: The platform network effect mirage. *Strategic Management Journal*, *36*(12), 1761-1777.

Brynjolfsson, E., Hui, X., & Liu, M. (2019). Does machine translation affect international trade? Evidence from a large digital platform. *Management Science*, *65*(12), 5449-5460.

Buckman, J. R., Adjerid, I., & Tucker, C. (2023). Privacy regulation and barriers to public health. *Management Science*, *69*(1), 342-350.

Burford, N., Shipilov, A. V., & Furr, N. R. (2022). How ecosystem structure affects firm performance in response to a negative shock to interdependencies. *Strategic Management Journal*, *43*(1), 30-57.

Chandler A. 1962. *Strategy and structure: Chapters in the history of the American industrial enterprise*. MIT Press: Cambridge, MA.

Chang, S., Eggers, J. P., & Keum, D. D. (2022). Bottleneck resources, market relatedness, and the dynamics of organizational growth. *Organization Science*, *33*(3), 1049-1067.

Cohen, L., Gurun, U. G., & Kominers, S. D. (2019). Patent trolls: Evidence from targeted firms. *Management Science*, *65*(12), 5461-5486.

DeSantola, A., & Gulati, R. (2017). Scaling: Organizing and growth in entrepreneurial ventures. *Academy of Management Annals*, *11*(2), 640-668.

Dickler, T. A., & Folta, T. B. (2020). Identifying internal markets for resource redeployment. *Strategic Management Journal*, *41*(13), 2341-2371.

Fainmesser, I. P., Galeotti, A., & Momot, R. (2023). Digital privacy. *Management Science*, *69*(6), 3157-3173.

Feldman, E. R. (2016). Corporate spin-offs and capital allocation decisions. *Strategy Science*, *1*(4), 256-271.

Feldman, E. R., & Sakhartov, A. V. (2022). Resource redeployment and divestiture as strategic alternatives. *Organization Science*, *33*(3), 926-945.

Genedy, M., Hellerstedt, K., Naldi, L., & Wiklund, J. (2024). Growing pains in scale-ups: How scaling affects new venture employee burnout and job satisfaction. *Journal of Business Venturing*, *39*(2), 106367.
Giustiziero, G., Kretschmer, T., Somaya, D., & Wu, B. (2023). Hyperspecialization and hyperscaling: A resource-based theory of the digital firm. *Strategic Management Journal*, *44*(6), 1391-1424.
Godinho de Matos, M., & Adjerid, I. (2022). Consumer consent and firm targeting after GDPR: The case of a large telecom provider. *Management Science*, *68*(5), 3330-3378.
Goldberg, S., Johnson, G., & Shriver, S. (2019). Regulating privacy online: The early impact of the GDPR on European web traffic & e-commerce outcomes. *Available at SSRN*, *3421731*.
Goldfarb, A., & Tucker, C. E. (2011). Privacy regulation and online advertising. *Management Science*, *57*(1), 57-71.
Goldfarb, A., & Tucker, C. (2012). Shifts in privacy concerns. *American Economic Review*, *102*(3), 349-353.
Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2022). Data network effects: Key conditions, shared data, and the data value duality. *Academy of Management Review*, *47*(1), 189-192.
Helfat, C. E. (1997). Know-how and asset complementarity and dynamic capability accumulation: the case of R&D. *Strategic Management Journal*, *18*(5), 339-360.
Hoffman, R., & Yeh, C. (2018). *Blitzscaling: The lightning-fast path to building massively valuable companies*. Currency.
Huang, J., Henfridsson, O., Liu, M. J., & Newell, S. (2017). Growing on steroids: Rapidly scaling the user base of digital ventures through digital innovation. *MIS Quarterly, 41*(1): 301-314.
Iacus, S., King, G., & Porro, G. (2012) Causal inference without balance checking: Coarsened exact matching. *Political Analysis. 20(1):*1-24.
Ichihashi, S. (2020). Online privacy and information disclosure by consumers. *American Economic Review*, *110*(2), 569-595.
Jansen, J. J., Heavey, C., Mom, T. J., Simsek, Z., & Zahra, S. A. (2023). Scaling-up: Building, leading and sustaining rapid growth over time. *Journal of Management Studies*, *60*(3), 581-604.
Jia, J., Jin, G. Z., & Wagman, L. (2021). The short-run effects of the general data protection regulation on technology venture investment. *Marketing Science*, *40*(4), 661-684.
Johnson, G. A., Shriver, S. K., & Goldberg, S. G. (2023). Privacy and market concentration: intended and unintended consequences of the GDPR. *Management Science*.
Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, *110*(9), 2819-2858.
Kaul, A. (2012). Technology and corporate scope: Firm and rival innovation as antecedents of corporate transactions. *Strategic Management Journal*, *33*(4), 347-367.
Ke, T. T., & Sudhir, K. (2023). Privacy rights and data security: GDPR and personal data markets. *Management Science*, *69*(8), 4389-4412.
Kim, K., Gopal, A., & Hoberg, G. (2016). Does product market competition drive CVC investment? Evidence from the US IT industry. *Information Systems Research*, *27*(2), 259-281.

Levinthal, D. A., & Wu, B. (2010). Opportunity costs and non-scale free capabilities: Profit maximization, corporate scope, and profit margins. *Strategic Management Journal*, *31*(7), 780-801.
Li, J., Pisano, G., Xu, Y., & Zhu, F. (2023). Marketplace scalability and strategic use of platform investment. *Management Science*, *69*(7), 3958-3975.
Mahoney, J. T., & Pandian, J. R. (1992). The resource-based view within the conversation of strategic management. *Strategic Management Journal*, *13*(5), 363-380.
Miller, A. R., & Tucker, C. (2009). Privacy protection and technology diffusion: The case of electronic medical records. *Management Science*, *55*(7), 1077-1093.
Miller, A. R., & Tucker, C. (2018). Privacy protection, personalized medicine, and genetic testing. *Management Science*, *64*(10), 4648-4668.
Palepu, K. (1985). Diversification strategy, profit performance and the entropy measure. *Strategic Management Journal*, *6*(3), 239-255.
Panzar, J. C., & Willig, R. D. (1981). Economies of scope. *The American Economic Review*, *71*(2), 268-272.
Penrose, E. T. 1959. *The theory of the growth of the firm*. Wiley: New York.
Peukert, C., Bechtold, S., Batikas, M., & Kretschmer, T. (2022). Regulatory spillovers and data governance: Evidence from the GDPR. *Marketing Science*, *41*(4), 746-768.
Rumelt RP. 1974. *Strategy, structure, and economic performance*. Harvard University Press: Boston, MA.
Silverman, B. S. (1999). Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science*, *45*(8), 1109-1124.
Somaya, D., & You, J. (2024). Scalability, venture capital availability, and unicorns: Evidence from the valuation and timing of IPOs. *Journal of Business Venturing, 39(1), 106345*.
Sun, T., Yuan, Z., Li, C., Zhang, K., & Xu, J. (2023). The value of personal data in internet commerce: A high-stakes field experiment on data regulation policy. *Management Science*.
Susarla, A., & Mukhopadhyay, T. (2019). Can outsourcing of information technology foster innovations in client organizations? An empirical analysis. *MIS Quarterly*, *43*(3), 929-949.
Teece, D. J. (1980). Economies of scope and the scope of the enterprise. *Journal of Economic Behavior & Organization*, *1*(3), 223-247.
Teece, D. J. (1982). Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior & Organization*, *3*(1), 39-63.
Varga, S., Cholakova, M., Jansen, J. J., Mom, T. J., & Kok, G. J. (2023). From platform growth to platform scaling: The role of decision rules and network effects over time. *Journal of Business Venturing*, *38*(6), 106346.
Villalonga, B. (2004). Diversification discount or premium? New evidence from the business information tracking series. *The Journal of Finance*, *59*(2), 479-506.
Wen, W., & Zhu, F. (2019). Threat of platform-owner entry and complementor responses: Evidence from the mobile app market. *Strategic Management Journal*, *40*(9), 1336-1367.
Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*(2), 171-180.
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press: Cambridge, MA.

Wu, B. (2013). Opportunity costs, industry dynamics, and corporate diversification: Evidence from the cardiovascular medical device industry, 1976–2004. *Strategic Management Journal*, *34*(11), 1265-1287.

Zhou, Y. M. (2011). Synergy, coordination costs, and diversification choices. *Strategic Management Journal*, *32*(6), 624-639.

Table 1. Descriptive Statistics and Correlation Matrix

	Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Scaling	1.46	1.76	0.00	8.23												
(2) Treated	0.51	0.50	0.00	1.00	0.11											
(3) Post	0.50	0.50	0.00	1.00	0.01	0.01										
(4) Assets(ln)	0.68	0.87	0.00	5.30	0.15	0.10	0.05									
(5) Leverage ratio	0.32	1.65	0.00	53.50	-0.05	-0.04	0.01	0.01								
(6) ROA	-0.00	0.05	-2.30	0.31	0.04	0.05	0.02	0.05	-0.58							
(7) Slack	0.20	0.20	0.00	1.00	-0.01	0.04	-0.01	-0.31	0.04	-0.16						
(8) Capital intensity	0.04	0.25	0.00	13.37	-0.01	0.01	0.01	-0.01	-0.01	-0.01	-0.02					
(9) Diversification	0.07	0.16	0.00	0.88	-0.05	0.06	0.01	0.14	-0.01	0.03	-0.16	-0.01				
(10) Entropy	0.12	0.26	0.00	2.21	-0.05	0.07	0.00	0.15	-0.01	0.03	-0.17	-0.01	0.98			
(11) SIC count	1.34	0.75	1.00	10.00	-0.06	0.06	-0.01	0.19	-0.00	0.03	-0.16	-0.01	0.79	0.86		
(12) 1-HHI_Asset	0.05	0.14	0.00	0.87	-0.07	0.03	-0.01	0.12	-0.00	0.02	-0.18	-0.01	0.87	0.88	0.81	
(13) Multimarket	0.25	0.43	0.00	1.00	-0.08	0.04	-0.00	0.16	-0.02	0.03	-0.18	-0.01	0.79	0.79	0.80	0.81

Table 2. Comparison of control and treatment groups in year before treatment

	mean_control	mean_treated	mean_differences	t-statistic	p-value
Before matching					
Revenue	1217.70	5038.25	-3820.55	-3.24	0.00
COGS	629.31	2569.90	-1940.59	-3.08	0.00
Assets	2681.73	9570.77	-6889.04	-2.98	0.00
Net income	0.10	0.61	-0.51	-2.39	0.02
Growth	0.10	0.19	-0.09	-1.72	0.09
Gross profit	588.39	2468.34	-1879.96	-3.23	0.00
Diversification	0.06	0.08	-0.02	-1.32	0.19
Liquidity	0.20	0.20	-0.01	-0.38	0.71
After matching					
Revenue	1279.92	1336.40	-56.48	-0.19	0.85
COGS	661.24	680.77	-19.53	-0.12	0.91
Assets	2821.53	2453.35	368.18	0.48	0.63
Net income	0.11	0.09	0.02	0.32	0.75
Growth	0.11	0.20	-0.09	-1.36	0.18
Gross profit	618.68	655.63	-36.95	-0.25	0.80
Diversification	0.06	0.07	-0.01	-0.81	0.42
Liquidity	0.20	0.21	-0.01	-0.60	0.55

Table 3. Regression results of DID (Dependent variable: Scaling)

VARIABLES	(1)	(2)	(3)
Treated * Post	-0.529***	-0.550***	
	(0.109)	(0.110)	
2015			-0.003
			(0.172)
2016			0.019
			(0.164)
2018			-0.604***
			(0.158)
2019			-0.562***
			(0.166)
2020			-0.659***
			(0.179)
Observations	2,882	2,882	2,882
R-squared	0.018	0.029	0.032
Number of firms	598	598	598
Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. In column (3), 2017 (the year before the shock) is the reference year.

Table 4. Placebo tests

VARIABLES	(1) Placebo year	(2) SIC 28
Treated * Placebo year	-0.105	
	(0.108)	
Treated * Post		0.147
		(0.211)
Observations	2,824	898
R-squared	0.011	0.031
Number of firms	593	206
Controls	Yes	Yes
Firm FE	Yes	Yes
Year FE	Yes	Yes

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. In column (1), the sample period is 2013-2018, and 2016 is the placebo treatment year. *Placebo year* equals 1 if it is in or after 2016. In column (2), the sample consists of matched U.S. public companies whose SIC starts with 28 (Chemicals and allied products) during 2015-2020.

Table 5. Heterogenous treatment effects

VARIABLES	(1) Continuous treatment	Platform BM Tripple DID	Platform=1	Platform=0	Data intensive SICs Tripple DID	Data SIC=1	Data SIC=0
EU intensity * Post	-0.469***						
	(0.140)						
Treated * Post		-0.435***	-1.255***	-0.437***	-0.378***	-1.074***	-0.380***
		(0.119)	(0.253)	(0.119)	(0.115)	(0.269)	(0.115)
Platform * Post		-0.169					
		(0.190)					
Treated * Post * Platform		-0.848***					
		(0.271)					
Data SIC * Post					0.268		
					(0.197)		
Treated * Post * Data SIC					-0.683**		
					(0.286)		
SUR test (p-value)			0.003***			0.017**	
Observations	2,196	2,882	473	2,409	2,882	776	2,106
R-squared	0.028	0.040	0.150	0.025	0.033	0.069	0.023
Number of firms	458	598	104	494	598	180	418
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. In column (1), *EU intensity* is calculated as the proportion of E.U. sales to total sales. In column (2), *Platform* indicates whether the firm adopts a platform-based business model, which is coded using business descriptions provided by Pitchbook. In columns (3) and (4), We split the sample into two groups based on the adoption of platform business models and test the treatment effect in these two subsamples. The results of DID analysis are shown in columns (3) and (4). The seemingly unrelated regression (SUR) test for these two groups is given between columns (3) and (4). *Data SIC* equals 1 if the firm belongs to data intensive industries: SIC 7370 (Computer programming and data processing, such as Adobe, Twitter, Yelp), SIC 7374 (Data processing and preparation, such as PayPal and Square), and SIC 4899 (Communication services, such as Altigen and Telenav). The subsample analyses are shown in columns (6) and (7). The SUR test for the data intensive industries and other industries is between columns (6) and (7).

Table 6. Test potential confounding effects of California Consumer Privacy Act (CCPA)

VARIABLES	(1) CA matched	(2) CA post-2020	(3) 2015-2019
Treated * Post	-0.504***	-0.469***	-0.410***
	(0.108)	(0.109)	(0.115)
Treated * Post * CCPA		-0.664	
		(0.451)	
CCPA		0.456	
		(0.371)	
Observations	2,879	2,879	2,430
R-squared	0.029	0.031	0.023
Number of firms	598	598	595
Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. In column (1), we use an indicator of California headquarters as an additional matching variable, leading to balanced treated and control groups in terms of *CA HQ*. In column (2), *CCPA* equals 1 if the firm is headquartered in California and the time is in 2020. In column (3), the sample period is between 2015 and 2019 given that CCPA became effective on January 1, 2020.

Table 7. Supplementary analysis: DID estimation

VARIABLES	(1) Restricted sample	(2) 1st difference
Treated * Post	-0.584***	
	(0.129)	
Post		-0.504***
		(0.157)
Observations	1,942	1,474
R-squared	0.041	0.029
Number of firms	377	299
Controls	Yes	Yes
Firm FE	Yes	Yes
Year FE	Yes	Yes

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. In column (1), restricted sample consists of firms that have at least one non-zero scaling observations both before and after the GDPR. In column (2), the sample only consists of treated firms.

Table 8. Regression results of 2SLS

	(1)	2nd stage				
VARIABLES	1st stage	Diversification	Entropy	ln(1+SIC count)	1-HHI_Asset	Multimarket
Treated * Post	-0.533***					
	(0.116)					
Scaling		-0.047**	-0.070**	-0.048**	-0.033**	-0.089*
		(0.021)	(0.031)	(0.024)	(0.014)	(0.049)
Observations	2,678	2,678	2,678	2,678	2,434	2,678
Number of firms	579	579	579	579	550	579
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-test/ Chi-square	6.410***	40.62***	44.79***	1685***	50.10***	49.10***

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 9. Validity of the instrumental variable

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Operating cost	COGS	DR	DU	Within	Across
Treated * Post	0.001	0.031	0.024**	0.013	0.026**	0.009
	(0.035)	(0.040)	(0.011)	(0.010)	(0.012)	(0.008)
Observations	3,318	3,228	2,678	2,678	2,462	2,462
R-squared	0.249	0.188	0.013	0.005	0.011	0.027
Number of firms	592	582	579	579	534	534
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. In column (1), the dependent variable is log-transformed operating cost, which is the sum of COGS and SG&A. The values of operating cost are winsorized (at the 1% level) to mitigate the potential impact of outliers. In column (2), the dependent variable is log-transformed COGS. The values of COGS are winsorized (at the 1% level) to mitigate the potential impact of outliers. In column (3), related diversification $DR_j = \sum_{i\varepsilon j} P_i^j \ln(1/P_i^j)$, where P_i^j is the share of the industry i of segment j in the total sales of the segment. $DR = \sum_{j=1}^{M} DR_j P^j$, where P^j is the share of the j segment sales in the total sales of the firm. In column (4), unrelated diversification $DU = \sum_{j=1}^{M} P^j \ln(1/P^j)$, which is the weighted average of all the segment shares. In columns (5) and (6), we divide all the industries into two segments: data-related segment (i.e., SIC 7374 and SIC 7375) and non-data-related segment. Within-segment diversification $Within_j = \sum_{i\varepsilon j} P_i^j \ln(1/P_i^j)$, where P_i^j is the share of the industry i of segment j in the total sales of the segment. $Within = \sum_{j=1}^{M} Within_j P^j$, where P^j is the share of the j segment sales in the total sales of the firm. Across-segment diversification $Across = P^d \ln(1/P^d) + P^n \ln(1/P^n)$, where P^d is the share of the data-related segment sales in the total sales of the firm and P^n is the share of the non-data-related segment sales in the total sales of the firm. Columns (5) and (6) drop firms that operate in both segments pre-GDPR.

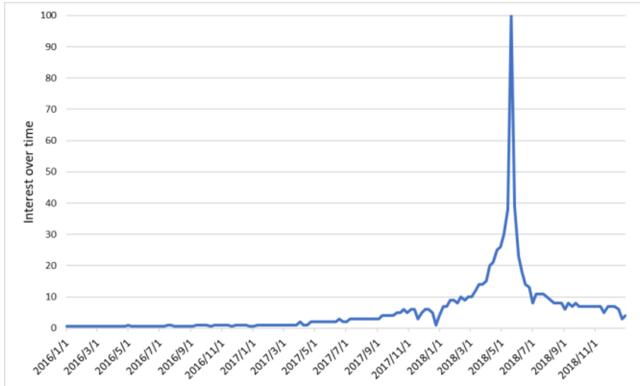

Figure 1. Google trends (web search in the U.S.)

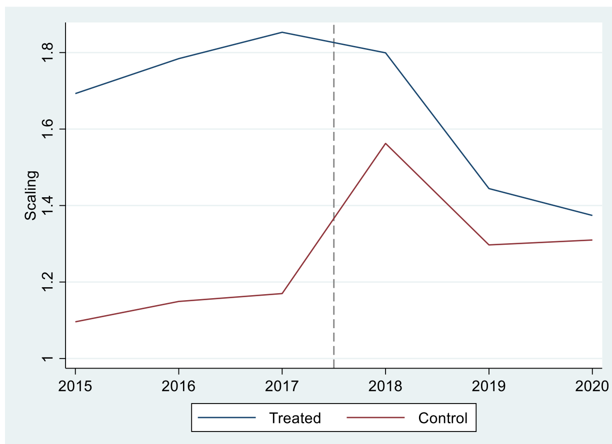

Figure 2. Plot of firm scaling in a matched sample

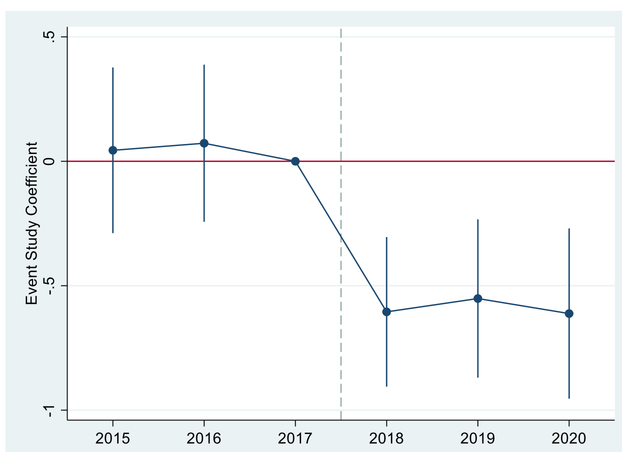

Figure 3. Estimated impacts of GDPR on scaling. 2017 is the reference year. *Vertical bands represent ±1.96 times the standard error of each point estimate. *Clustered standard errors, as in main OLS regressions.