Regulating Data: Evidence from Corporate America*

Fabio Motoki
Norwich Business School
University of East Anglia
f.motoki@uea.ac.uk

Jedson Pinto†
School of Management
University of Texas at Dallas
jedson.pinto@utdallas.edu

May 2024

* Acknowledgments: We are grateful to Heng Wang (Emily) for her help understanding and providing part of the data breach dataset. We also thank Wanli Zhao (Editor), an anonymous reviewer, Gil Sadka, Gaizka Ormazabal, Gabriel Voelcker, Tsafrir Livne, Matteo Binfare, Victor Almeida, Cameron Holstead, Victor Rodrigues, Cristiano Costa, Roberto Decourt, and workshop participants at Unisinos for their helpful suggestions and feedback. All remaining errors are our own. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We declare no conflict of interest and the data that support the findings of this study is available through public sources listed in the paper.
† Corresponding author. University of Texas at Dallas, 800 West Campbell Rd, SM 41, Richardson, TX 75080, phone: 972-883-5065, email: jedson.pinto@utdallas.edu.

Regulating Data: Evidence from Corporate America

Abstract

Using the enforcement of the General Data Protection Regulation (GDPR) as our empirical setting, we examine how stricter data privacy and data protection requirements affect shareholder wealth, firms' investment decisions, and data breaches. Consistent with consumer privacy negatively affecting firms, we find that U.S. firms exposed to the GDPR lose 0.7-1.1% in market value relative to unexposed firms in the week in which the regulation became enforceable. We find that the decrease in market value is partially attributable to a decrease in sales growth. GDPR-exposed firms increase their investment above that of control firms and become less likely to report a data breach post-regulation. The decrease in data breach likelihood is statistically and economically significant, resulting in up to 34 million records not being leaked, which costs between $205 million and $561 million to firms in breach mitigation expenses per year. The results of this study should be of interest to academics and regulators worldwide by examining the costs and benefits of regulating data.

JEL: D80, G14, G18, G11, M40

Keywords: Privacy, Data, Real Effects, Securities Regulation, GDPR.

"Privacy is one of the biggest problems in this new electronic age. At the heart of the Internet culture is a force that wants to find out everything about you. And once it has found out everything about you and two hundred million others, that's a very valuable asset, and people will be tempted to trade and do commerce with that asset. This wasn't the information that people were thinking of when they called this the information age."

− Andrew Grove, Co-founder and former CEO of Intel Corporation

1 Introduction

Data privacy and security are two of the main concerns of the modern era. To address these concerns, the European Union (EU) approved the General Data Protection Regulation (GDPR), which became enforceable on May 25th, 2018. The GDPR requires greater transparency in how firms collect consumer data by establishing clear opt-in consent for data collection, imposing stricter data management and control, and assigning substantial penalties and liability risks for data processing or data flow violations. The new regulation is expected to affect how firms gather, process, and use consumer data, and is likely to impose high compliance costs on firms.[1] However, little is known about the broad consequences of regulatory data. Although there is a growing body of knowledge addressing the costs imposed by the GDPR (Aridor et al., 2023; Chen et al., 2022; Ferracuti et al., 2024; Goldberg et al., forthcoming; Jia et al., 2021; Johnson et al., 2023), there is a paucity of evidence regarding the intended goals of increasing data security (Goldfarb & Que, 2023). Nevertheless, understanding the costs and benefits is key to improving public policies (Arrow et al., 1996; Cochrane, 2014; Posner & Weyl, 2013). Therefore, to fill this void, we study how the GDPR affected not only the performance of publicly traded firms but also the probability of corporate data breaches.

[1] For example, in 2018 Microsoft disclosed having more than 1,600 engineers working with GDPR compliance projects.

The GDPR requires any enterprise that controls or processes EU residents' data to abide by new rules, *regardless* of its location. Therefore, firms across the world, including those in the United States, might be subject to the GDPR to the extent that they process EU residents' personal data.[2] Given its extraterritorial reach, we focus on the effects of the GDPR on U.S. firms for several reasons.

First, the U.S. setting allows us to compare firms exposed to the regulation (treated group) with similar firms that are not exposed to the regulation (control group). Hence, we can use firms in the control group as a counterfactual to what would have happened if the regulation had not been created. Second, the *within*-U.S. analysis allows us to hold constant important variables, such as legal and financial systems, which are likely to affect equity markets and firms' investment decisions. Third, the enforcement of the GDPR was bundled with other major events in Europe (i.e., Markets in Financial Instruments Directive II (MIFID II) in 2018 and Brexit). These concurrent regulatory changes have led to significant changes in firms' investment and market structures (Bloom et al., 2019; Guo & Mota, 2021; Hassan et al., 2023; Lang et al., 2023), which could make it difficult for researchers and regulators to isolate the effects of the GDPR on European firms. However, given its clear extraterritorial reach, we can study the effects of GDPR on publicly traded firms in the United States. Fourth, the data available for U.S. companies also enables us to study how GDPR affects data breaches, which is one of the goals of the regulation. Finally, many states in the U.S. and countries across the world are also debating or enacting laws,

[2] An EU individual accessing a firm's website is not a sufficient condition to become liable under the GDPR. "The organization would have to target a data subject in an EU country. Generic marketing doesn't count. For example, a Dutch user who Googles and finds an English-language webpage written for U.S. consumers or B2B customers would not be covered under the GDPR. However, if the marketing is in the language of that country and there are references to EU users and customers, then the webpage would be considered targeted marketing and the GDPR will apply." (Faitelson, 2017).

such as the GDPR (e.g., Brazil, China, and Canada). Therefore, it is important to assess the consequences of GDPR outside the EU.

In our first analysis, we examine how GDPR affects firms' stock prices, a key metric that summarizes investors' expectations of firms' future performance. Specifically, we conduct an event study analysis around the week of the initial enforcement of the GDPR, where we compare the stock returns of U.S. firms exposed to the GDPR (*treated* group) relative to the stock returns of a group of firms not exposed to the GDPR (*control* group). We used firms' annual report discussions of the GDPR in 2018 to identify firms exposed to the GDPR. U.S. publicly traded companies are required to report and discuss significant risk factors in their annual reports (10-Ks) (Campbell et al., 2014; SEC, 2005). Therefore, firms that are directly or indirectly exposed to GDPR must disclose such risks to investors. These disclosures allow us to precisely identify firms that are exposed to the regulation, as we can identify even firms without direct sales linked to the EU. We assume that the firm is not exposed to the GDPR if it does not mention the regulation in its 2018 10-K.[3]

Using raw and size-adjusted measures of stock returns, we find that treated firms exhibit significantly negative returns around the enforcement of the GDPR relative to control firms. Our control group includes firms that, besides not mentioning GDPR-related risks, are from Healthcare, Banking, or Insurance industries that already had strong data protection regulations pre-GDPR (HIPAA and GLBA).[4] Our treated firms are companies that mention GDPR-related risks but are not from any of these three industries. The difference in stock returns is robust when controlling

[3] One possibility is that firms might still be subject to the GDPR but they do not discuss the regulation on their annual filing. This would introduce measurement error in our analyses that could bias our results *against* finding any significant effects of the regulation. Nevertheless, we also run robustness tests using Compustat Segments data (untabulated), and results are qualitatively similar.

[4] In the United States, the Health Insurance Portability and Accountability Act (HIPAA) of 1996 and the Gramm-Leach-Bliley Act (GLBA) of 1999 are two industry-specific cyber privacy laws that precede the GDPR.

for differences in location or adjusting for firm size. Furthermore, placebo regressions using the same week in the previous (2017) or following (2019) years suggest that the results are unlikely to be spurious. In terms of total market capitalization loss, the drop in firm value for treated firms adds up from $42 to $76 billion around the enforcement of the GDPR. These results are consistent with investors anticipating a substantial negative effect of stricter data privacy laws on firms' future cash flow.

To understand the mechanism behind the negative market stock reaction, we examine how the EU GDPR affects firms' accounting performance, specifically firms' sales growth. Using a difference-in-differences design, we find that U.S. firms exposed to the EU GDPR exhibit statistically slower sales growth than U.S. firms not exposed to GDPR. The decrease in sales growth is robust to year-quarter fixed time effects, industry trends, state trends, and firm fixed effects, which could explain the differences in firms' performance. In terms of economic significance, U.S. firms affected by the EU GDPR have their sales grow 5.8-6.6 percentage points (p.p.) slower than control firms after the law went into effect. This result is consistent with consumer privacy limiting firms' ability to generate new sales and echoes recent studies documenting significant costs associated with GDPR.[5]

To understand how firms react to stricter data protection regulations and whether the GDPR achieved one of the intended goals of increasing consumer data protection, we find that U.S. firms exposed to the EU GDPR invest more and become less likely to disclose a data breach in a given quarter post-GDPR. The decrease in data breach likelihood is mostly driven by a decrease in data

[5] This result is also consistent with the event study analysis that suggests that investors anticipate a negative effect of the EU GDPR on firms' future expected cash flows.

breaches associated with hacking and malware, suggesting that firms are better at preventing cyberattacks.[6]

In terms of economic magnitude, the decrease in data breach likelihood represents 10 fewer data breaches in a year, which could prevent up to 34 million records from being breached yearly. In 2023, the cost per record for a medium-sized breach (up to 101,200 records) was estimated to be approximately $165 per record, with large breaches having lower per-record estimates but elevated total economic costs (IBM Security, 2023).[7] This would imply that GDPR helped save between $205 and $561 million dollars in data breach costs every year.[8] These results suggest that the GDPR may have achieved one of its intended goals of enhancing consumer data protection and privacy. To the best of our knowledge, this is the first study to document the potential benefits associated with recent efforts to regulate consumer data protection and data privacy.

We also examine whether the EU GDPR affected how investors reacted to the announcement of a data breach using data from 62 data breaches in the sample period. Using a five-day return window around the data breach announcement, we find evidence that post-GDPR investors may react more negatively to a data breach for firms with stricter data protection requirements. The effect is economically significant, with a data breach being associated with up to a 5.3% drop in stock prices in the five days around the announcement of the data breach for EU GDPR firms

[6] *Conditional* on a firm having a data breach, we find no significant evidence that the GDPR affected the size of a cyber-attack.

[7] Note that these estimates only include costs incurred by firms to mitigate effects of the data breach and ignore personal losses from individuals affected by the breach. Although we could not find an estimate for the average loss for the an individual involved in a breach, the FBI reports $10.3 billion in losses from 800,944 complaints (FBI, 2022). As an exercise, if 100 breached records give rise to one complaint, then the GDPR potentially prevented 340,000 complaints, avoiding about $4.4 billion in losses for individuals.

[8] One single 34 million records breach results in about $205 million in costs. However, 34 separate 100k records breaches would cost $16.5 million each, totaling $561 million

relative to control firms.[9] These results are consistent with investors anticipating significant litigation costs associated with the EU GDPR fines in the case of a data breach.

This study contributes to the literature in several ways. First, it is one of the first studies to examine potential extraterritorial capital market consequences of the EU GDPR on publicly traded companies. Specifically, we find that the EU GDPR negatively affects U.S. firms' market value. The adverse effect seems to be related to stricter data privacy and security laws slowing firms' sales growth. Our study adds to the growing body of literature documenting the costs of GDPR, such as a decrease in EU venture capital investment (Jia et al., 2021) especially when ventures and lead investors are not in the same state or union (Jia et al., 2020); decreased use and increased market concentration for web technology vendors (Johnson et al., 2023; Peukert et al., 2022); decreased online firms' revenues (Goldberg et al., forthcoming); and reduced the amount and quality of information for firms to make investing and operating decisions (Ferracuti et al., 2024). These results are consistent with Goldfarb and Tucker (2011), who found that privacy laws can decrease marketing efficiency.

The current study also provides the first attempt to examine whether the GDPR has achieved its intended consequences of increasing data protection. Specifically, we find that U.S. firms subject to the GDPR are less likely to report a data breach after enforcement of the regulation. The decrease in data breach likelihood appears to be driven by a decrease in data breaches associated with hacking and malware, which might be attributed to greater investment in cybersecurity. This greater investment might be due to the increased attention of a more specialized board to oversee cybersecurity risks in U.S. firms due to the GDPR (Klein et al., 2022). We extend Klein et al.

[9] These results are also consistent with market reactions to data breaches that are followed by class-action lawsuits and other litigations (Kamiya et al., 2021).

(2022) by showing that this increased attention to cybersecurity risks had practical effects, potentially preventing more than 34 million records from being breached per year. Collectively, these results suggest that the GDPR might have achieved one of the intended consequences of greater data protection. To the best of our knowledge, this study is the first to provide evidence of the potential benefits of GDPR.

Finally, this study contributes to the growing literature on cybersecurity. Prior studies have examined the potential characteristics associated with the likelihood of a firm disclosing a data breach, as well as the potential consequences of a data breach (Kamiya et al., 2021). We add to the literature by examining the effects of regulatory scrutiny on the likelihood of data breach. Using the enforcement of the GDPR as a quasi-natural experiment, we find that stricter privacy laws substantially reduce firms' likelihood of reporting data breaches. Additionally, our results suggest that the market may react negatively to these breaches in the presence of GDPR, extending Richardson et al. (2019), who found that the pre-GDPR market reaction to data breaches is unimportant. Our results indicate that GDPR may have changed how the market perceives these breaches, potentially changing executives' incentives to protect customer data. These results are consistent with regulations being an alternative way to address the recent concerns of data privacy and security. These results should be of interest to regulators worldwide that have enacted or enacted laws similar to the EU GDPR.

2 Institutional Background

Data privacy and security are two of the main concerns of the modern era. As computers become capable of gathering and processing large amounts of data, companies are increasingly trying to collect and benefit from consumer data (Goldfarb & Que, 2023; Goldfarb & Tucker, 2011). If digital strategies are successfully implemented, firms can use big data to improve their

marketing strategies, increase their revenues, and grow faster (Ferracuti et al., 2024). In addition, because of the non-rivalry characteristic of the data, firms can also collect consumer data to sell to third parties (Jones & Tonetti, 2020). The benefits of (big) data create incentives for firms to collect any information from individuals that could be relevant to a business (e.g., individuals' gender, age, location, consumption habits, address, retweets, Facebook likes, etc.). Consumers might not be aware of what information a company collects and uses (Rainie, 2018).

Given its inherent economic value, consumer data have also become targets of cyberattacks.[10] For example, in 2018, the Federal Bureau of Investigation (FBI) Internet Crime Complaint Center reported 351,937 consumer complaints related to cyberattacks. The cost of losses from consumer complaints was estimated at $2.7 billion. These figures have risen to 800,944 complaints totaling losses of $10.3 billion in 2022, highlight the increasing significance of protecting data (FBI, 2022). Importantly, companies are also targets of costly cyberattacks. Kamiya et al. (2021) estimated that U.S. firms lost millions of dollars in cyber-attacks. Echoing the importance of data privacy, U.S. CEOs consistently rank cyber threats as one of the top five risks in doing business worldwide (PwC, 2019, 2020, 2021, 2022, 2023).

Motivated by concerns about how firms collect, store, use, and protect consumer data, the EU created the GDPR, which became enforceable on May 25th, 2018. The GDPR requires higher transparency in how firms collect consumer data by establishing clear opt-in consent for data collection, imposing stricter data management and control, and assigning substantial penalties and liability risks for data processing or data flow violations. The GDPR replaced the EU's Data Protection that went into effect in 1995 and changed the definition of personal data to reflect

[10] A cyberattack is any attempt to expose, alter, disable, destroy, steal, or gain unauthorized access to or make unauthorized use of an asset. Cyberattacks pose a significant risk for the modern economy.

changes in technology and the ways in which organizations collect data about people. Under the 1995 Directive, personal data are defined as names, addresses, and personal identification numbers. Under the GDPR, personal data are defined as any information that could be used, on its own or together with other data, to identify an individual.

The new regulation established clear individual rights related to privacy. First, the GDPR forces firms to ask individuals to explicitly opt-in for the processing of any personal data (*Opt-in and Consent).* Before, firms could have a default opt-in policy but would give the option for individuals to opt-out. Second, to ensure that EU residents have more transparency and power over their data, the GDPR gives data subjects the Right to Access their personal data (*Right to Access).* This allows consumers to better understand and monitor the information companies collect and how they handle consumer data. Third, residents of the EU can ask companies to either transfer their data to another service provider (*Right to data portability*) or to erase all their data (*Right to be Forgotten)*.

It is important to note that the GDPR affects any entity that controls or processes information regarding EU residents. This includes firms outside the European Union, such as U.S. firms. In addition, under the GDPR, both data processors and data controllers are jointly responsible for complying with the new rules and are liable for violations. Therefore, many U.S. firms could be subject to GDPR scrutiny to the extent that they process data from EU residents. However, an EU resident simply accessing a website does not necessarily create liability under the GDPR. For instance, if a website does not target EU individuals and does not collect or process their data, GDPR does not apply (EU, 2018a; Faitelson, 2017).

The GDPR also enhances transparency by requiring organizations to report data breaches to individuals whose data were compromised and to the supervisory authority within 72 hours of

becoming aware of the breach. In breach notification, firms are required to notify the nature of the breach, the approximate number of data subjects affected, describe the likely consequences of the personal data breach, and describe the measures taken to address the breaches (EU, 2018b, 2018c). Thus, the GDPR created uniformity, because before it, firms would have to research and ensure compliance with each member state.[11]

The new EU regulation imposes hefty fines on firms that violate new requirements. Specifically, firms could be charged up to either 4% of their global turnover or €20 million, whichever is higher (EU 2018d). Violations of the GDPR include a lack of consent to process data, violating consumers' privacy by design, or failing to notify authorities and individuals of a data breach.

In summary, the GDPR is one of the biggest privacy and security laws in modern areas. The new law is expected to affect how firms operate, and is likely to impose significant compliance costs on firms. Nevertheless, little is known about the consequences of the regulation for publicly traded companies or whether the GDPR has achieved one of its intended goals of increasing data privacy and security. This study aimed to fill this gap.

We focus on the effects of GDPR on U.S. firms for several reasons. First, the U.S. setting allows us to compare firms exposed to stricter data privacy and security laws with similar firms that are not exposed to the regulation. Second, the within-U.S. analysis allows us to hold constant important variables, such as legal and financial systems, which are likely to affect equity markets and firms' investment decisions. This allows us to identify firms that are exposed to the GDPR

[11] In the United States, state laws require firms to disclose the existence of a data breach to the individuals affected (Obaydin et al., 2024). However, firms are not required to make a widespread public announcement to capital markets and have significant time (30 to 45 days) to disclose the breach. The EU GDPR requires a timelier and more detailed disclosure regarding the data breach.

either directly or indirectly. Third, the enforcement of the GDPR was bundled with other major events in Europe, such as the enforcement of Markets in Financial Instruments Directive II (MIFID II) in 2018 and Brexit. Both events have led to significant changes in firms' investment and market structure (Bloom et al., 2019; Campello et al., 2022; Guo & Mota, 2021; Hassan et al., 2023; Lang et al., 2023), which could make it difficult for researchers and regulators to isolate the effects of the GDPR on European firms. However, given its clear extraterritorial reach, we can study the effects of GDPR on publicly traded firms in the United States. Fourth, the data available for American firms enable a cleaner measure of exposition to the GDPR and access detailed firm-level data breaches. For instance, all publicly traded companies in the U.S. are required to file and discuss significant risk factors in their annual reports (10-Ks). Finally, many states in the U.S. and countries worldwide are debating or enacting laws similar to the GDPR.[12] Thus, it is important to assess the consequences of GDPR outside the EU.

3 Sample and Research Design

We collect accounting data from Compustat Quarterly files, market data from CRSP, and 10-K information from the WRDS SEC Analytics Suite. Following Kamiya et al. (2021) and Iyer et al. (2020), we collect data breach data from the Privacy Rights Clearinghouse (Privacy Rights Clearinghouse, 2020). The initial sample starts with all firms with 10-K information in 2018 that exists on the GVKEY-CIK Link table of the WRDS SEC Analytics Suite. These criteria result in an initial sample of 5,283 unique firms and 116,281 observations covering 2014 to 2019.

[12] In the United States, several states enacted data breach disclosure legislation to improve consumer data protection (Obaydin et al., 2024). The EU GDPR differs from these laws by imposing higher degrees of transparency in every step of how the firm gathers and uses the consumer data. The EU GDPR also imposes clear and significant penalties for lack of compliance and has significant extraterritorial reach. Similar to the data breach disclosure state legislations, the EU GDPR might also limit the amount of discretion that firms had in handling and disclosing information to capital markets.

We exclude foreign firms directly listed in the United States that still file 10-K and are headquartered outside the United States. We focus primarily on U.S. firms headquartered in the country to control for potential differences in regulations, enforcement, and market development. This reduces the sample to 4,671 unique firms and 86,540 firm-quarter observations. Next, we remove the sample firms without sufficient market data in 2018 to compute the market tests associated with the enforcement of the GDPR. This further reduces the sample to 3,771 unique firms, or 71,550 firm-quarter observations. To avoid the potential effects of the global pandemic and to keep the sample period homogenous, we drop observations before 2016Q2 and after 2019Q4, following the literature that indicates that the pandemic started affecting U.S. firms in February 2020 (Fahlenbrach et al., 2021).[13] This results in a sample of 3,771 unique firms and 49,003 firm-quarter observations. Next, we delete firms in states without other observations in the same industry-state pair. This filter helps to increase the statistical power of the estimations and to examine the effects of the GDPR within a state and within an industry.[14] Because we want to study how a data protection regulation might affect US companies, we choose only companies from the Healthcare, Banking, and Insurance industries as the control group, as these firms already operate under the stricter data protection laws of HIPAA and GBLA (United States, 1996, 1999).[15] If any treatment firm is from one of these industries, we drop it from the sample. The final sample has 1,013 unique firms and 12,909 firm-quarter observations. Table 1 Panel A summarizes the sample selection process used in this study.

[13] The main inferences of the paper remain unchanged if we include 2020Q1.
[14] The inferences of the paper remain unchanged if we don't include this filter. Furthermore, the inferences remain similar if we use annual data versus quarterly data.
[15] Industries 11, 44, and 45 from the Fama-French 48-industry classification.

3.1 Identifying exposure to the GDPR

We use the discussion of the GDPR in firms' annual reports in 2018 to identify firms subject to the GDPR (*treated* group). All publicly traded companies in the U.S. are required to file and discuss significant risk factors in their annual reports (Campbell et al., 2014; SEC, 2005). Importantly, cybersecurity risks are especially prominent, with specific guidelines from the SEC detailing which kind of cyber risks must be disclosed and how they relate to other mandatory disclosure items such as the Management Discussion and Analysis (MD&A) and Description of Business sections of the 10-K filing (Klein et al., 2022; SEC, 2011). Consequently, firms that are directly or indirectly exposed to the GDPR must disclose such risk to investors, allowing us to precisely identify firms that are exposed to the regulation, because we can identify even firms without direct sales linked to the EU. In addition to mandated disclosure, there is evidence that managers may have an incentive to disclose specific risk factors because there appears to be a positive market reaction to more specific risk factor disclosures (Hope et al., 2016). Our approach follows prior studies that have used similar disclosures to identify U.S. firms exposed to Brexit (Campello et al., 2022; Hassan et al., 2023). Conversely, the control group is U.S. firms that are not exposed to GDPR-associated risks and operate in the Healthcare, Banking or Insurance industries.

One concern could be that even though firms are required to disclose existing risks on their 10-Ks, managers might still want to hide their exposure to increased data protection and data privacy regulations. If that is the case, then our way of identifying GDPR exposure firms would *underestimate* how many firms are exposed, resulting in "exposed" firms being included in the control group. Ultimately, this approach would bias us against finding results, as a subsample of the control group could be economically affected by the GDPR.

Using text-based measures to identify firms' exposure to GDPR is significantly better than using segmented data. First, firms have substantial discretion in their *segment reporting* on how they disclose their exposure to European countries, which creates a potential measurement error in identifying exposed firms (Botosan et al., 2021),[16] while the disclosure of risks is mandatory in the U.S. (SEC, 2005), with specific guidance regarding cybersecurity risks (SEC, 2011). Second, firms are still subject to the GDPR, as long as they collect or process personal data from people in the EU. Therefore, firms do not necessarily need to have significant revenues linked to European countries to be subject to GDPR.[17] Nevertheless, firms must disclose potential business risks on their 10-Ks. Hence, text-based measures should identify firms exposed to GDPR more accurately. Appendix 1 contains excerpts with anecdotal evidence of how firms disclose their GDPR risks.

3.2 Market returns

We used the 10-Ks filed in 2018 to identify treated firms because the GDPR became enforceable in 2018. Therefore, all firms exposed to the new regulation should make investors aware of their potential effects on companies' business. One could argue that, as the GDPR was enacted in 2016, investors and firms have potential anticipatory effects. However, if parties anticipate all the effects before enforcement, no effect should be detectable in 2018. Therefore, we could potentially bias our analyses to identify any potential changes in stock prices or accounting performance. Our choice follows the literature using enforcement as the event (Autor et al., 2006;

[16] For example, some firms might report a geographic segment as "Asian and European Countries", others might report a segment as "Europe", while others might report only certain European countries. It is unclear whether the firms in this example are exposed the same way to the GDPR. Furthermore, SFAS 131 (ASC 280) gives firms discretion on how to report their segments (business, geographic, matrix – business X geographic), potentially introducing even more noise as firms that have EU-generated revenue may not report geographic segments.

[17] For example, a firm with no significant European revenues that collects data of European citizens to sell to other U.S. firms has to comply with the GDPR.

Chen et al., 2022; Demirer et al., 2024; Johnson et al., 2023; Peukert et al., 2022; Watzinger et al., 2020).

Specifically, we compare the stock price returns of U.S. firms exposed to the GDPR relative to the stock price returns of a group of firms not exposed to the GDPR from May 20th to May 28th, the week the GDPR became enforceable. There is robust evidence that investors' attention towards data protection and data privacy spikes in that week, as evidenced by Internet search results. Figure 1 shows a clear spike in Google searches of the term "General Data Protection Regulation" in the week of GDPR enforcement. Interestingly, the spike occurs in both the United States (Panel A) and worldwide (Panel B). Nevertheless, focusing on enforcement date biases, we find no stock market reaction if investors had already anticipated and fully incorporated the effects of GDPR on firms' prices.

<Figure 1 about here>

Table 1 summarizes the distribution of firms exposed to the GDPR by industry (Panel B) and state (Panel C). As noted in Panel B, GDPR affected firms in multiple industries, from Business Services to Utilities and Shipping containers. The industries most affected by GDPR seem to be Business Services (174 unique firms), Pharmaceutical Products (61 unique firms), and trading (31 firms). Panel C shows that the GDPR affected firms headquartered in several U.S. states. Perhaps unsurprisingly, the state with the greatest number of firms affected is California (149 unique firms), followed by New York (63 unique firms), and Massachusetts (50 unique firms).

<Table 1 about here>

Using this dataset, we estimate the following linear regression:

$$Return_i = \beta_0 + \beta_1 I(Treated) + \gamma_{state} + \epsilon_{it} \quad (1)$$

Where $Return_{it}$ is one of two measures of change in shareholder value: the cumulative raw returns (RawRet) for firm *i* for a given week and the cumulative size-adjusted return (*SizeAdj Ret*) for firm *i* for that given week. The variable of interest is *I(Treated),* which is an indicator variable that equals one for firms that discussed the GDPR in their 2018 10-Ks and zero otherwise. *I(Treated)* capture the average difference in returns for treated firms relative to control firms. We also include state-fixed effects to control for potential state-specific time-invariant determinants of market returns.[18]

GDPR might limit firms' ability to gather, process, and use consumer data, which could result in lower future cash flows for the firm. Therefore, investors might price a negative effect of GDPR on firm value. This would be captured by a negative β_1. Alternatively, stricter data protection and data privacy laws might bolster consumers' confidence in the economy, which could have a positive effect on the firms' future sales, and thus future prices (Ke & Sudhir, 2023). In this scenario, β_1 would be non-negative. Which effect dominates is an empirical question.

3.3 Accounting performance

To understand the mechanism behind the negative market stock reaction, we examine how GDPR affect firms' accounting performance, as measured by sales growth. We focus on sales growth, given that consumer data play a critical role in attracting future revenue, and recent

[18] The state of California enacted the California Consumer Privacy Act (CCPA) in Jan/2018 and the state of New York enacted the Stop Hacks and Improve Electronic Data Security Act (SHIELD) on July 25, 2019. However, both became enforceable only in 2020 (Field, 2020; Tierling & Lanois, 2020). Therefore, both enactment and enforceability of SHIELD are outside of our event study window, while only enforceability of CCPA falls out of the window. Nevertheless, as here we are measuring the cumulative return of a single week, the state-level fixed effect absorbs specific characteristics of each state, including the enacted CCPA.

evidence suggests that EU firms respond to the GDPR by storing less data and reducing data processing (Demirer et al., 2024). Specifically, we estimate the following difference-in-differences model.

$$SalesGrowth_{it} = \beta_0 + \beta_1 I(Treated) * I(Post) + \beta_2 I(Treated) + FEs + \epsilon_{it} \quad (2)$$

Where $SalesGrowth_{it}$ is the logarithm of current sales (Compustat item *saleq*) divided by sales from the same quarter in the previous year ($saleq_{t-4}$). The variable of interest is *I(Treated)*I(Post),* which captures the average treatment effect of GDPR on firms' revenue. Specifically, *I(Treated)* is an indicator variable that equals one if the firm discusses GDPR on its 2018 10-K and zero otherwise. *I(Post)* is an indicator variable that equals one for observations after 2018Q1, and zero otherwise. It only appears in the interaction due to the year-quarter fixed effects. *FEs* indicates a set of fixed effects. We run increasingly conservative specifications, including fixed effects for (a) year-quarter and industry trends, (b) adding state trends, and (c) adding firm-level FEs. These additional controls allow us to account for potential state-specific time trends (such as new regulations) but also to account for firm-specific characteristics that might be correlated with firms' sales growth (such as culture).

If the GDPR limits the ability of firms to attract new customers, we would expect a decrease in firms' sales growth. This would be evidenced by a negative β_1. Alternatively, the increase in data protection and privacy might bolster customers' confidence in the economy, which could result in greater sales growth (β_1 would be positive). For example, after the EU GDPR, customers might feel more comfortable buying products from certain websites since they would expect their data to remain private (Janakiraman et al., 2018).

3.4 Investment effects of GDPR

This subsection examines how GDPR affects firms' investment decisions. Anecdotal evidence suggests that the EU GDPR resulted in firms redirecting capital from capital expenditures to IT expenses.[19] Specifically, we estimate a difference-in-differences model of how the EU GDPR affects three different measures of investment: total investment *(Total_Inv*), capital expenditures (*CAPEX*), and research and development (*R&D*). Specifically, we estimate the following model:

$$Inv_{it} = \beta_0 + \beta_1 I(Treated) * I(Post) + \beta_2 I(Treated) + FEs + \epsilon_{it} \quad (3)$$

Inv_{it} is one of three different investment measures. The first is *Total_Inv*, measured as the sum of capital expenditure and R&D in the last four quarters divided by total assets. The second is *CAPEX*, measured as total capital expenditure in the last four quarters divided by total assets. However, WRDS Compustat does not provide quarterly level CAPEX directly, only year-to-date CAPEX (*capxy*). Therefore, we calculate the variable *capexq* from *capxy* as the variation in *capxy* in two consecutive quarters in the same fiscal year. Finally, for the first quarter of the fiscal year, we make *capxq = capxy*. The third is *R&D*, measured as total expenditure related to intangible capital formation in the last four quarters divided by total assets. Following prior studies, we assume that reported R&D (*xrdq*) equals zero if it is missing. Additionally, we add a proportion of SG&A expenses (*xsgaq*) to *xrdq* according to the firm's Fama–French five-industry classification, following Ewens et al. (2023), to obtain intangible capital formation expenditures (*rnd*).

The variable of interest is *I(Treated)*I(Post),* which captures the average treatment effect of GDPR on firms' investment. Specifically, *I(Treated)* is an indicator variable equal to one if the firm discusses GDPR on its 2018 10-K, and zero otherwise. *I(Post)* is an indicator variable that

[19] For example, FBD Holdings discussed in its 2018 Q1 conference call that GDPR required the firm to largely increase investment in IT. Similarly, the company SABRE discussed in its 2018 Q2 conference call that GDPR forced the firm to rotate total investment from CapEx to OpEx (See Appendix 1).

equals one for observations after 2018Q1 and zero otherwise. Note that it only appears in the interaction due to the year-quarter fixed effects. *FEs* indicates a set of fixed effects. Similar to sales growth, we run increasingly conservative specifications, including fixed effects for (a) year-quarter and industry trends, (b) adding state trends, and (c) adding firm-level FEs.

Stricter data privacy and protection regulations could lead to three possible outcomes for firms' investment. The EU GDPR could impose significant costs on firms (Obaydin et al., 2024), which could result in them deciding to leave the European market. This would be evidenced by a decrease in firms' investment ($\beta_1 < 0$). However, the EU GDPR could also lead to significant increases in investments on data protection and IT to ensure compliance with the regulation ($\beta_1 > 0$). Finally, the regulation could have no effect on firms' investments if firms already meet the requirements of the new rules.

3.5 Data breach analysis

Did the EU GDPR achieve one of the intended consequences of the increased data protection? To address this question, we examine whether the EU GDPR affected the probability of a firm disclosing a data breach (*extensive margin analysis)*. Specifically, we estimate a difference-in-differences model to examine whether treated firms are less (or more) likely to disclose a data breach post-GDPR. We estimate the following linear probability model:

$$Data\_Breach_{it} = \beta_0 + +\beta_1 I(Treated) * I(Post) + \beta_2 I(Treated) + \beta_3 I(Post) + FEs + \epsilon_{it} \quad (4)$$

Data_breach is an indicator variable that equals one if the firm discloses a data breach in a given quarter and zero otherwise. Similar to prior analyses, the variable of interest is *I(Treated)*I(Post),* which captures the effect of GDPR on the probability of a firm reporting a data breach after 2018Q1. *FEs* indicate the fixed effects. Following Kamiya et al. (2021), we include

industry and year fixed effects to control for differences in the likelihood of data breaches across industries and over time. In the most conservative specifications, we also include state fixed effects to control for potential geographic determinants that could affect the likelihood of a data breach and drop industry and state fixed effects in favor of more flexible firm fixed effects.

If the GDPR reached one of its intended objectives of increasing data protection and privacy, we would expect to observe a significant drop in the likelihood of a firm disclosing a data breach after the enforcement of the GDPR ($\beta_1 < 0$). However, the GDPR increases transparency by requiring firms to notify individuals and authorities of a data breach in a timely manner (EU, 2018b, 2018c). Moreover, prior studies suggest that firms might withhold bad news (Jung & Kwon, 1988; Kothari et al., 2009), including data breaches (Amir et al., 2018). If, before the GDPR, firms withheld data breaches, then we might find that the GDPR led to an increase in the number of disclosed data breaches because firms now face harsher consequences if they fail to disclose these breaches. Thus, β_1 is an estimate of the joint effect of improvements in data security and improvements in disclosure (Johnson, 2022). Consequently, it is a lower-bound estimate of the effect of data security brought about by the GDPR.

The PRC data breach dataset also contains information on the type of data breach. Specifically, the PRC identifies whether the data breach is related to Payment Card Fraud (*CARD*), Unintended Disclosure (*DISC*), Hacking or Malware (*HACK*), insiders (*INSD*), physical loss (*PHYS*), portable devices (*PORT*), stationary devices (*STAT*), and unknown (*UNKN*). However, the sample size prevents meaningful estimations from examining how GDPR affects *CARD*, *INSD, and STAT*. Thus, we focus primarily on the effect of the EU GDPR on the following types of data breach: *HACK, DISC, PHYS, PORT and UNKN*. Specifically, we estimate difference-in-

differences models to examine the effects of EU GDPR on each type of data breach mentioned above.

3.5.1 Data breach properties and market reaction to data breaches

We also examine whether the EU GDPR affected the properties of a data breach *conditional* on a firm having a data breach. Specifically, we examine whether the EU GDPR affected the size of the data breach (*intensive margin analysis)*. We estimate the following regression:

$$Size\_Breach_{it} = \beta_0 + +\beta_1 I(Treated) * I(Post) + \beta_2 I(Treated) + \beta_3 I(Post) + \epsilon_{it} \quad (5)$$

where *Size_Breach* is the natural logarithm of the number of data records estimated to be affected by a data breach, as presented in the PRC dataset. One limitation of this analysis is that firms might not know how many individuals are affected by a data breach. Therefore, the analysis is limited to observations in which the PRC can identify the number of breached records.

On the one hand, if data protection regulation helps firms prevent and minimize threats to consumer data, then firms exposed to the EU GDPR might report smaller data breaches post-GDPR. This would be evidenced by a negative average treatment effect. On the other hand, we might find that firms exposed to the EU GDPR report larger data breaches post-GDPR. This might happen if 1) firms were previously underreporting or omitting records breached or 2) if firms managed to only successfully prevent smaller data breaches. Therefore, even though the *likelihood* of the data breach might decrease, the *size* of the data breach could increase.

As our final analysis, we examine whether the EU GDPR affected how investors react to the announcement of a data breach. Specifically, we estimate a difference-in-differences model examining the 3-days market reaction to a data breach for the treated and control firms before and

after the enforcement of the GDPR. Pre-GDPR, existing evidence indicates that data breaches do not have an important market effect (Richardson et al., 2019). However, the EU GDPR imposes fines of €20 million, or 4% of firms' global revenues, on firms that fail to protect European consumers' data. Significant fines and compliance costs could negatively affect future expected cash flows. Therefore, after enforcement of the GDPR, investors might react more negatively to the announcement of a data breach for firms exposed to the EU GDPR.

3.6 Descriptive Statistics

Table 2 presents summary statistics for the market and accounting variables of the paper. All variables are defined in Appendix 2. To eliminate the potential effects of outliers in the analysis, we winsorize all the accounting variables by year-quarter at the 1% level.

<Table 2 about here>

Table 2, Panel A, shows that the average stock return for the control group during the enforcement week of the GDPR was slightly positive. In contrast, the average stock return for firms exposed to the GDPR was slightly negative. Table 2, Panel B, suggests that both treated and control firms present positive sales growth, although control firms sales grew faster. Treated firms seem to have higher levels of total investment, CAPEX, and R&D.

4 Results

In this section, we present the market consequences and real effects of the enforcement of the EU GDPR on US-headquartered firms. In Section 4.1, we conduct an event study analysis of the enforcement of the EU GDPR. In Section 4.2, we explore the more direct consequences of increased data protection and privacy requirements by examining how the EU GDPR affected

firms' accounting revenue. In Section 4.3, we examine whether the EU GDPR affects firms' investment decisions. In section 4.4, we examine whether the EU GDPR achieved some of its intended consequences by affecting the likelihood of a firm disclosing a data breach and the type of data breach (extensive margin analysis). Section 4.5 shows the results of the EU GDPR on the size of a data breach (intensive margin analysis). Finally, Section 4.6 shows the results of the EU GDP on the announcement of a data breach.

4.1 Event Study Analysis

The descriptive statistics in Table 2 suggest that U.S. firms exposed to the EU GDPR exhibited a negative average return, while the control firms presented slightly positive returns. While univariate analysis offers a first and important look at the data, a more sophisticated analysis is required to evaluate the difference in stock returns across types of firms.

Table 3 presents the results of the linear regressions that formally compare the stock returns of the treated and control firms. The coefficient of interest is that of *I(Treated),* which captures the extent to which stock returns for firms exposed to the EU GDPR differ from those of firms not exposed to the EU GDPR. Columns 1 and 3 of Panel A present the baseline model with constant. In Columns 2 and 4, we estimate a more conservative model that includes state-fixed effects, which controls for location-specific determinants of stock performance. As this is a one-week window event study, there is no need to control for time effects.[20] Following O'Donovan et al. (2019), we clustered the standard errors by state.[21] Across all the columns, the coefficient of *I(Treated)* is

[20] Ideally, we would like to include industry-related fixed effects. However, the way the sample is composed, with only three industries in the control group, and several industries with only a handful of firms in the treatment, lead to collinearity problems involving I(Treated) when including industry-state or state fixed effects. However, in untabulated results with the full sample containing all industries in the control group and industry-state fixed effects, results are qualitatively similar.

[21] Results are stronger if we use robust standard errors. We find similar results if we cluster the standard errors by industry.

negative and significant. Interestingly, the coefficient is larger in magnitude for the most conservative models (-1.0% for column 2 and -1.1% for column 4). In terms of total market capitalization loss, the drop in firm value for treated firms adds up to $42 to $76 billion around the enforcement of the GDPR.[22] These results are consistent with investors anticipating a negative effect of stricter data protection and privacy requirements on firms' future cash flow. These results are also consistent with recent evidence from Obaydin et al. (2024), who document significant costs associated with data breach disclosure requirements for U.S. firms.

<Table 3 about here>

One concern is that the difference in stock returns observed in Panel A could be due to a spurious correlation. In Figure 2, we show two placebo regressions with the same firms in the same week but in the year *before* (2017) or the year *after* (2019) the enforcement of the EU GPDR. If the results are potentially driven by seasonality, similar negative returns should be observed in the placebo tests. However, as evidenced by the plot, GDPR-exposed firms exhibited a negative return only in 2018, when the GDPR became enforceable.[23]

<Figure 2 about here>

We also estimate several additional model specifications to rule out the concern that our results are driven by specific research design choices. Specifically, we re-estimate the market

[22] These values are calculated by summing all the estimated incremental drop in firm value for treated firms. The drop in firm value for each firm is calculated as the coefficient of treated in Table 3 multiplied by the treated firm stock prices at the end of the prior year.

[23] On May 18th, 2017, the Federal Communications Commission adopted a Notice of Proposed Rulemaking (FCC 17-60, WC Docket No. 17-108, on Restoring Internet Freedom) ending public-utility regulation of the Internet and reverting net neutrality rules from the Obama administration. The Notice was published on May 23rd. Since it potentially affects revenues from Internet Service Providers (ISPs) and companies that use the Internet to generate revenues, the positive results for 2017 should be interpreted with care.

reaction to the GDPR by either using robust standard errors or clustering them by industry or using the Fama-French three-factor model plus momentum (FF4) factor-adjusted return measure. Panel B of Table 3 presents the results. In all the model specifications, except the FF4 models with fixed effects, we find that the coefficient of *I(Treated)* is negative and statistically significant at conventional levels. However, the data requirements for calculating FF4 are more demanding, leading to a decrease in the sample size and thus increasing the estimator variance. This increased variance is at least partially responsible for the significant loss of the coefficients. Nevertheless, the *t* stats remain close to the cutoff for the usual 10% level. Altogether, these results suggest that U.S. firms exposed to GDPR lose significant market value around the enforcement of regulations. The analyses in Figure 2 and Panel B of Table 3 bolster the confidence that the results in Panel A are not attributed to seasonality, are unlikely to be spurious, and are robust to different model specifications.

Stricter data protection requirements might affect firms that are more likely to depend on intangible data capital, such as social media companies or trading companies. Specifically, the potential negative effects of the EU GDPR may be more pronounced for firms with high asset intangibility levels (Binfarè, 2019; Kamiya et al., 2021). To test this conjecture, we split the sample based on levels of asset intangibility in the year prior to the enforcement of the EU GDPR and compare the stock returns of treated and control firms. Panel C of Table 3 presents the results. Across both measures of shareholder change in wealth, we find that the coefficient of *I(Treated)* is negative and highly significant for firms with high asset intangibility (above the median). There are no significant stock return differences between the treated and control firms in the low asset intangibility group. These results are consistent with the asset structure playing an important role in investors' reactions to the EU GDPR.

4.2 Firm Revenue

The results in Table 3 suggest that investors anticipate that EU GDPR will negatively affect firms' cash flows. To understand the potential mechanism behind this result, we investigate how the EU GDPR affected firms' sales growth in a difference-in-differences model.

The results are presented in Table 4. The coefficient of interest is that of *I(Treated)*I(Post),* which captures the average effect of stricter data protection and privacy requirements on firms' sales growth. Column 1 includes the year-quarter and industry-year fixed effects to control for potential common time shocks and industry trends during the sample period. In Column 2, we add state-year fixed effects to control for potential changes in state regulations or geographic shocks that could explain changes in firms' sales growth. Finally, in Column 3, we add firm fixed effects to control for firm-specific determinants of sales growth. Across all columns, the coefficient of I(Treated) × I(Post) is negative and significant. Specifically, firms affected by the EU GDPR have sales grow between 5.8 and 6.6 percentage points (p.p.) slower than control firms after the law went into effect. These results are consistent with consumer privacy, which limits firms' ability to generate new sales and achieve higher return on investment. These results are also consistent with the market price event analysis, which suggests that investors anticipate a negative effect of the EU GDPR on firms' future expected cash flows.

<Table 4 about here>

4.3 Investment Effects of GDPR

Prior analyses document how investors react to the enforcement of the European Union General Data Protection Regulation and how it affects firms' sales growth. In this subsection, we examine whether the EU GDPR affects firms' investment decisions. Following prior studies

(Campello et al., 2022; Hassan et al., 2023), we focus on how U.S. firms change their capital expenditure and research and development investment following European regulations. The results are presented in Table 5. We focus on the most conservative specification with the same full set of fixed effects, as shown in Table 4. We omit the main effects for brevity.

The coefficient of interest is that of *I(Treated)*I(Post).* Column 1 suggests an increase in investment after 2018Q1 for treatment firms, which seems to be driven mostly by increases in R&D (Column 3), with some contribution from the CAPEX (Column 2). This result is in line with the survey and anecdotal evidence that the new regulation imposes significant compliance costs on the firms. For example, Microsoft said in a press release in May 2018 that it would allocate more than 1,600 engineers to work in compliance with GDPR. Moreover, a PwC survey of 200 companies with more than 500 employees found that 68% planned to spend $1 to $10 million to meet the regulations' requirements.

<Table 5 about here>

Taken together, these results suggest that firms react to EU GDPR mostly by increasing investment in intangible capital. These results complement the findings of Klein et al. (2022), who document that, after the enactment of the GDPR, firms add more directors with cyber/IT expertise. Specifically, our results suggest that firms ramp up their investments in R&D, potentially by redesigning their products to comply with the new regulation. Our finding is also consistent with Chen et al. (2022), who find an indirect effect on patents and, while we find a direct effect on investment.

4.4 Data Breach Analysis

One of the intended consequences of the EU GDPR is for firms to be better able to protect consumers' data. If that is the case, then firms subject to the EU GDPR should be less likely to report a data breach after the enforcement of the EU GDPR (*extensive margin analysis)*. We test this conjecture using a difference-in-differences linear probability model.[24] The results are presented in Table 6. The variable of interest is *I(Treated)*I(Post).* In all specifications, the coefficient of *I(Treated)*I(Post)* is negative and significant at the usual levels. This result suggests that firms subject to stricter data protection requirements are less likely to report a data breach in any quarter post-GDPR. The decrease in the likelihood of data breach is robust to the inclusion of industry, state, and firm fixed effects. This suggests that the effect of GDPR persists, even when comparing firms within the same industry or within the same state. The decrease in data breach likelihood represents approximately 10 fewer data breaches in a year, which could save up to 34 million records from being breached on average.[25] However, remember from the discussion in Section 3 that this is most likely a lower bound estimate, since the coefficient is the joint effect of (a) more reporting of breaches due to increased penalties for not reporting and (b) less data breaches due to increased data protection practices.

[24] We estimated Linear Probability Models (LPMs) instead of a logit because LPMs yield good approximations of the average marginal effects, are consistent, and inference can be simply adjusted by using robust standard errors (Wooldridge, 2010). Another reason for the use of LPMs is that it better accommodates the usage of fixed effects, especially when there might be groups without a lot of variation in the dependent variable (Breuer and deHaan, 2023). Specifically, prior studies have shown that the non-linear estimators, such as the maximum-likelihood estimator (MLE) is typically biased and inconsistent when using firm fixed effects (Neyman and Scott, 1948; Lancaster, 2000 – "incidental parameters problem"). As Breuer and deHaan (2023) argue, the common strategy adopted by the finance and accounting literature to deal with this issue is to use OLS

[25] There are 516 treated firms in the sample. The EU GDPR is associated with a 0.5% drop in the likelihood of a data breach for a given firm-quarter, which represents about 2.5 fewer data breaches in a quarter or 10 fewer data breaches in a year. The average number of records breached in a cyberattack in the treated sample is 3,446,504 (Table 8, Panel A). Hence, 10*3,446,504 is approximately 34 million records.

<Table 6 about here>

Figure 3 shows the effect decomposed by year. This supports the findings presented in Table 6. There is no clear parallel trend violation of pre-GDPR, and there is evidence that post-GDPR (2019) breaches become less common in GDPR-exposed companies.

<Figure 3 about here>

4.4.1 Data breach types

We also examine whether the effects of EU GDPR on different types of data breaches. In the sample, we have sufficient variation to estimate the effects of the EU GDPR on five types of data breach: *Hack* (hacking or malware), *InvDisc* (involuntary disclosure), *Port* (portable device), *Phys* (loss of physical object), and *Unkn* (unknown). All types of data breaches are explained in Section 3 and on the PRC website.[26] Following Table 6, we estimate a difference-in-differences linear probability model where the dependent variables are indicator variables of whether the firm disclosed a certain type of data breach in a given quarter and zero otherwise. The results are presented in Table 7. The main coefficient of interest is that of *I(Treated)*I(Post).* The average treatment effect is negative and significant only for *Hack*. This result suggests that the decrease in data breach likelihood is driven by decreases in data breaches associated with hacking or malware. This result is consistent with the prevention of cyberattacks by improving data security after the enforcement of the EU GDPR.

<Table 7 about here>

[26] https://privacyrights.org/data-breaches

4.5 Data Breach size

Another important question is whether the EU GDPR affected the properties of a data breach, *conditional* on the occurrence of a data breach (*intensive margin analysis).* Stricter data protection requirements might not only affect the likelihood of a firm having a data breach but also the size of a cyber-attack. To examine this question, we estimate a difference-in-differences model, where the dependent variable is the natural logarithm of the number of records breached by a given cyber-attack, as disclosed by the PRC data breach dataset. One limitation of this analysis is that a significant part of the data breaches has missing (unknown) value for the size of the data breach. Another limitation is that the entire sample is already small, and when limiting to controls from Healthcare, Insurance, and Banking, the sample becomes very small. Therefore, inferences may be limited. Table 8 presents the results of the study. In Panel A of Table 8, we present descriptive statistics of the average size of a data breach in the sample and the average 5-days market reaction to a cyber-attack. Panel B presents the difference-in-differences estimates.

<Table 8 about here>

The descriptive statistics in Panel A of Table 8 indicate that the average data breach in the sample affected 1,914,924 records. However, the data are highly skewed, a phenomenon also observed in industry reports (IBM Security, 2023): the median size of a data breach in the sample is 818 records, and the skewness of the total number of records breached is 7.55 (untabulated). Therefore, we log-transformed the data to limit the potential effects of outliers on the multivariate analyses. Panel A of Table 8 also suggests that, on average, a data breach is associated with negative market reactions for treated firms. In Panel B of Table 8, we formally test whether the EU GDPR affects the size of the data breach. The coefficient of interest is that of

*I(Treated)*I(Post).* Across all columns, the average treatment effect is positive, but insignificant. These results suggest that firms exposed to the EU GDPR exhibit data breaches of similar size as control firms, *conditional* on having a data breach.[27] Note that the results in Tables 6 and 7 suggest that firms become less likely to report a data breach post-GDPR. Taken together, these results are consistent with stricter data protection requirements, decreasing the *expected* value of data records affected by a data breach.

4.6 Market Reaction to a data breach

Prior analyses suggest that firms subject to stricter data protection regulations invest more in IT and cybersecurity training and are less likely to disclose a data breach post-GDPR. In this subsection, we estimate a difference-in-differences linear model to examine whether the EU GDPR affected how investors reacted to the announcement of a data breach using data from 62 data breaches in the sample period (from 2016 to 2019).[28] Table 9 presents the results of the study. Columns (1)–(3) use raw returns as the independent variable, with specifications becoming more conservative by adding year fixed effects (Column 2) and breach-type fixed effects (Column 3). Columns (4)–(6) are similar, but for size-adjusted returns. The coefficient of interest is that of *I(Treated)*I(Post).* Although the coefficients of interest are negative in five out of the six specifications, only one (Column 1) is significant at the usual level. However, lack of power seems to play an important role, as in the untabulated results, the same specifications using the full sample of 133 breaches yield negative coefficients in all six specifications, with only the less conservative raw returns specification (Column 1), which is not statistically significant.

[27] Results using the full sample of 133 breaches lead to similar inferences.
[28] The sample size is similar to Amir et al. (2018) who examine data breaches from 2010 to 2015 and document 46 breaches per year for listed firms from all industries.

These results are consistent with stock prices reacting more negatively to data breaches for firms with stricter data protection requirements post-GDPR. This effect is also economically significant. Specifically, a data breach can cause an incremental drop of up to 5.3% in stock prices in the five days around the announcement of the data breach for EU GDPR firms, relative to control firms. These results are consistent with investors anticipating the significant costs imposed by regulators when consumer data is breached. We also extend Richardson et al. (2019), who found that pre-GDPR, the market reaction to data breaches, was unimportant. Our results suggest that the GDPR may have changed how the market perceives these breaches, potentially changing managers' incentives regarding customer data protection.

<Table 9 about here>

5 Conclusion

In this paper, we investigate the broad consequences of the General Data Protection Regulation (GDPR) on capital markets and firm investment decisions. The GDPR imposes greater data protection and privacy requirements for firms that gather, process, or use European consumer data. To identify the effects of GDPR on investors' wealth and firm investment decisions, we compare firms exposed to GDPR (as evidenced by their 10-K disclosures) with firms not exposed to GDPR. We use several different model specifications and high-order fixed effects to control for potential confounding explanations.

We find that U.S. firms exposed to the EU GDPR exhibit a significant drop in firm value at the time that the requirements went into effect. This result is not driven by a specific industry or state, and is unlikely to be spurious. Difference-in-differences analysis suggests that the drop in firm value might be attributed to a potential slowdown in sales growth. Furthermore, firms exposed

to GDPR invest more than control firms, which is consistent with anecdotal evidence that firms spend sizable sums in information technology (IT) to become compliant and the number of patents evidence from Chen et al. (2022).

Consistent with the GDPR achieving one of the intended consequences, we find that firms subject to higher data protection requirements are less likely to report data breaches. To the best of our knowledge, this is a novel result and calls for more research on the impact of the GDPR on data breaches (Johnson, 2022). The decrease in the likelihood of a data breach is driven by a decrease in hacking and malware breaches. These results are consistent with firms redesigning their operations and infrastructure to meet the GDPR requirements of enhanced data security, which might also be accomplished by investing more in employee training that helps prevent cyber-attacks. Point estimates suggest that the EU GDPR may have prevented approximately 34 million data records from being breached per year. As we measure the joint effect of both more data breaches reporting and fewer data breaches, this estimate is most likely a lower bound. We found no evidence that GDPR affects the size of data breaches.

Collectively, our results provide one of the first large-scale evidence of the costs and benefits of the GDPR on public firms. These results should be of interest to regulators worldwide that have adopted (United States, Canada, Brazil, and China) or are thinking about adopting rules similar to the European Union General Data Protection Regulation.

6 References

Amir, E., Levi, S., & Livne, T. (2018). Do firms underreport information on cyber-attacks? Evidence from capital markets. *Review of Accounting Studies*, *23*(3), 1177–1206. https://doi.org/10.1007/s11142-018-9452-4

Aridor, G., Che, Y., & Salz, T. (2023). The effect of privacy regulation on the data industry: Empirical evidence from GDPR. *The RAND Journal of Economics*, *54*(4), 695–730. https://doi.org/10.1111/1756-2171.12455

Arrow, K., Cropper, M., Eads, G., Hahn, R., Lave, L., Noll, R., Portney, P., Russel, M., Schmalensee, R., Smith, V. K., & Stavins, R. (1996). *Benefit-Cost Analysis in Environmental, Health, and Safety Regulation: A Statement of Principles*. American Enterprise Institute for Public Policy Research. https://aei.org/wp-content/uploads/2014/04/-benefitcost-analysis-in-environmental-health-and-safety-regulation_161535983778.pdf

Autor, D. H., Donohue, J. J., & Schwab, S. J. (2006). The Costs of Wrongful-Discharge Laws. *The Review of Economics and Statistics*, *88*(2), 211–231. https://doi.org/10.1162/rest.88.2.211

Binfarè, M. (2019). *The Real Effects of Operational Risk: Evidence from Data Breaches* (SSRN Scholarly Paper 3411553). https://doi.org/10.2139/ssrn.3411553

Bloom, N., Bunn, P., Chen, S., Mizen, P., Smietanka, P., & Thwaites, G. (2019). *The Impact of Brexit on UK Firms* (Working Paper 26218). National Bureau of Economic Research. https://doi.org/10.3386/w26218

Botosan, C. A., Huffman, A., & Stanford, M. H. (2021). The State of Segment Reporting by U.S. Public Entities: 1976–2017. *Accounting Horizons*, *35*(1), 1–27. https://doi.org/10.2308/HORIZONS-19-104

Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H., & Steele, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, *19*(1), 396–455. https://doi.org/10.1007/s11142-013-9258-3

Campello, M., Cortes, G. S., d'Almeida, F., & Kankanhalli, G. (2022). Exporting Uncertainty: The Impact of Brexit on Corporate America. *Journal of Financial and Quantitative Analysis*, *57*(8), 3178–3222. https://doi.org/10.1017/S0022109022000308

Chen, C., Frey, C. B., & Presidente, G. (2022). *Privacy Regulation and Firm Performance: Estimating the GDPR Effect Globally*. https://www.oxfordmartin.ox.ac.uk/publications/privacy-regulation-and-firm-performance-estimating-the-gdpr-effect-globally/

Cochrane, J. H. (2014). Challenges for Cost-Benefit Analysis of Financial Regulation. *The Journal of Legal Studies*, *43*(S2), S63–S105. https://doi.org/10.1086/678351

Demirer, M., Jiménez Hernández, D. J., Li, D., & Peng, S. (2024). *Data, Privacy Laws and Firm Production: Evidence from the GDPR* (Working Paper 32146). National Bureau of Economic Research. https://doi.org/10.3386/w32146

EU. (2018a). *Art. 3 GDPR - Territorial scope*. https://gdpr.eu/article-3-requirements-of-handling-personal-data-of-subjects-in-the-union/

EU. (2018b). *Art. 33 GDPR - Notification of a personal data breach to the supervisory authority*. https://gdpr.eu/article-33-notification-of-a-personal-data-breach/

EU. (2018c). *Art. 34 GDPR - Communication of a personal data breach to the data subject*. https://gdpr.eu/article-34-communication-of-a-personal-data-breach/

EU. (2018d). *Art. 83 GDPR - General conditions for imposing administrative fines*. https://gdpr.eu/article-83-conditions-for-imposing-administrative-fines/

Ewens, M., Peters, R. H., & Wang, S. (2023). *Measuring Intangible Capital with Market Prices* (SSRN Scholarly Paper 3287437). https://doi.org/10.2139/ssrn.3287437

Fahlenbrach, R., Rageth, K., & Stulz, R. M. (2021). How Valuable Is Financial Flexibility when Revenue Stops? Evidence from the COVID-19 Crisis. *The Review of Financial Studies*, *34*(11), 5474–5521. https://doi.org/10.1093/rfs/hhaa134

Faitelson, Y. (2017, December 4). Council Post: Yes, The GDPR Will Affect Your U.S.-Based Business. *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2017/12/04/yes-the-gdpr-will-affect-your-u-s-based-business/

FBI. (2022). *Internet Crime Report 2022*. Federal Bureau of Investigation. https://www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf

Ferracuti, E., Koo, M., Lee, M., & Stubben, S. (2024). *Acquisition of Customer Information and Corporate Decision Making* (SSRN Scholarly Paper 4682350). https://doi.org/10.2139/ssrn.4682350

Field, E. (2020). United States Data Privacy Law: The Domino Effect After the GDPR. *North Carolina Banking Institute*, *24*(1), 481.

Goldberg, S. G., Johnson, G. A., & Shriver, S. K. (forthcoming). Regulating Privacy Online: An Economic Evaluation of the GDPR. *American Economic Journal: Economic Policy*. https://doi.org/10.1257/pol.20210309

Goldfarb, A., & Que, V. F. (2023). The Economics of Digital Privacy. *Annual Review of Economics*, *15*(Volume 15, 2023), 267–286. https://doi.org/10.1146/annurev-economics-082322-014346

Goldfarb, A., & Tucker, C. E. (2011). Privacy Regulation and Online Advertising. *Management Science*, *57*(1), 57–71. https://doi.org/10.1287/mnsc.1100.1246

Guo, Y., & Mota, L. (2021). Should information be sold separately? Evidence from MiFID II. *Journal of Financial Economics*, *142*(1), 97–126. https://doi.org/10.1016/j.jfineco.2021.05.037

Hassan, T. A., Hollander, S., Lent, L. V., & Tahoun, A. (2023). The Global Impact of Brexit Uncertainty. *The Journal of Finance*. https://doi.org/10.1111/jofi.13293

Hope, O.-K., Hu, D., & Lu, H. (2016). The benefits of specific risk-factor disclosures. *Review of Accounting Studies*, *21*(4), 1005–1045. https://doi.org/10.1007/s11142-016-9371-1

IBM Security. (2023). *Cost of a Data Breach Report*. https://www.ibm.com/reports/data-breach

Iyer, S. R., Simkins, B. J., & Wang, H. (2020). Cyberattacks and impact on bond valuation. *Finance Research Letters*, *33*, 101215. https://doi.org/10.1016/j.frl.2019.06.013

Janakiraman, R., Lim, J. H., & Rishika, R. (2018). The Effect of a Data Breach Announcement on Customer Behavior: Evidence from a Multichannel Retailer. *Journal of Marketing*, *82*(2), 85–105. https://doi.org/10.1509/jm.16.0124

Jia, J., Jin, G. Z., & Wagman, L. (2020). *GDPR and the Localness of Venture Investment* (SSRN Scholarly Paper 3436535). https://doi.org/10.2139/ssrn.3436535

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. https://doi.org/10.1287/mksc.2020.1271

Johnson, G. A. (2022). Economic Research on Privacy Regulation: Lessons from the GDPR and Beyond. In *The Economics of Privacy*. University of Chicago Press.

https://www.nber.org/books-and-chapters/economics-privacy/economic-research-privacy-regulation-lessons-gdpr-and-beyond

Johnson, G. A., Shriver, S. K., & Goldberg, S. G. (2023). Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR. *Management Science*, *69*(10), 5695–5721. https://doi.org/10.1287/mnsc.2023.4709

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, *110*(9), 2819–2858. https://doi.org/10.1257/aer.20191330

Jung, W.-O., & Kwon, Y. K. (1988). Disclosure When the Market Is Unsure of Information Endowment of Managers. *Journal of Accounting Research*, *26*(1), 146–153. JSTOR. https://doi.org/10.2307/2491118

Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., & Stulz, R. M. (2021). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics*, *139*(3), 719–749. https://doi.org/10.1016/j.jfineco.2019.05.019

Ke, T. T., & Sudhir, K. (2023). Privacy Rights and Data Security: GDPR and Personal Data Markets. *Management Science*, *69*(8), 4389–4412. https://doi.org/10.1287/mnsc.2022.4614

Kezdi, G. (2004). Robust Standard Error Estimation in Fixed-Effects Panel Models. *Hungarian Statistical Review*, *Special number 9*.

Klein, A., Manini, R., & Shi, Y. (Crystal). (2022). Across the Pond: How US Firms' Boards of Directors Adapted to the Passage of the General Data Protection Regulation†. *Contemporary Accounting Research*, *39*(1), 199–233. https://doi.org/10.1111/1911-3846.12735

Kothari, S. P., Shu, S., & Wysocki, P. D. (2009). Do Managers Withhold Bad News? *Journal of Accounting Research*, *47*(1), 241–276. https://doi.org/10.1111/j.1475-679X.2008.00318.x

Lang, M., Pinto, J., & Sul, E. (2023). MiFID II unbundling and sell-side analyst research. *Journal of Accounting and Economics*, 101617. https://doi.org/10.1016/j.jacceco.2023.101617

Obaydin, I., Xu, L., & Zurbruegg, R. (2024). The unintended cost of data breach notification laws: Evidence from managerial bad news hoarding. *Journal of Business Finance & Accounting*. https://doi.org/10.1111/jbfa.12794

O'Donovan, J., Wagner, H. F., & Zeume, S. (2019). The Value of Offshore Secrets: Evidence from the Panama Papers. *The Review of Financial Studies*, *32*(11), 4117–4155. https://doi.org/10.1093/rfs/hhz017

Peukert, C., Bechtold, S., Batikas, M., & Kretschmer, T. (2022). Regulatory Spillovers and Data Governance: Evidence from the GDPR. *Marketing Science*, *41*(4), 746–768. https://doi.org/10.1287/mksc.2021.1339

Posner, E., & Weyl, E. G. (2013). Benefit-Cost Analysis for Financial Regulation. *American Economic Review*, *103*(3), 393–397. https://doi.org/10.1257/aer.103.3.393

Privacy Rights Clearinghouse. (2020). *Data Breach Chronology Database* (January 13th 2020) [dataset].

PwC. (2019). *22nd Annual Global CEO Survey*. https://www.pwc.com/gx/en/ceo-survey/2019/report/pwc-22nd-annual-global-ceo-survey.pdf

PwC. (2020). *23rd Annual Global CEO Survey*. https://www.pwc.com/gx/en/ceo-survey/2020/reports/pwc-23rd-global-ceo-survey.pdf

PwC. (2021). *24th Annual Global CEO Survey*. https://www.pwc.com/gx/en/ceo-survey/2021/reports/pwc-24th-global-ceo-survey.pdf

PwC. (2022). *25th Annual Global CEO Survey*. https://www.pwc.com/gx/en/ceo-survey/2022/main/content/downloads/25th_CEO_Survey.pdf

PwC. (2023). *26th Annual Global CEO Survey*. https://www.pwc.com/gx/en/issues/c-suite-insights/ceo-survey-2023.html

Rainie, L. (2018, March 27). Americans' complicated feelings about social media in an era of privacy concerns. *Pew Research Center*. https://www.pewresearch.org/short-reads/2018/03/27/americans-complicated-feelings-about-social-media-in-an-era-of-privacy-concerns/

Richardson, V. J., Smith, R. E., & Watson, M. W. (2019). Much Ado about Nothing: The (Lack of) Economic Impact of Data Privacy Breaches. *Journal of Information Systems*, *33*(3), 227–265. https://doi.org/10.2308/isys-52379

SEC. (2005). *Securities and exchange commission final rule, release no. 33–8591 (FR-75)*. https://www.sec.gov/files/rules/final/33-8591.pdf

SEC. (2011). *CF Disclosure Guidance: Topic No. 2—Cybersecurity*. https://www.sec.gov/divisions/corpfin/guidance/cfguidance-topic2.htm

Tierling, E., & Lanois, P. (2020, April 13). *NY's SHIELD Act has taken effect—What does this mean for your business?* [IAPP]. https://iapp.org/news/a/new-yorks-shield-act-has-taken-effect-what-does-this-mean-for-your-business/

United States. (1996). *An act to amend the Internal Revenue Code of 1986 to improve portability and continuity of health insurance coverage in the group and individual markets, to combat waste, fraud, and abuse in health insurance and health care delivery, to promote the use of medical savings accounts, to improve access to long-term care services and coverage, to simplify the administration of health insurance, and for other purposes.* (104–191). https://www.govinfo.gov/app/details/PLAW-104publ191

United States. (1999). *An act to enhance competition in the financial services industry by providing a prudential framework for the affiliation of banks, securities firms, insurance companies, and other financial service providers, and for other purposes.* (106–102). https://www.govinfo.gov/app/details/PLAW-106publ102

Watzinger, M., Fackler, T. A., Nagler, M., & Schnitzer, M. (2020). How Antitrust Enforcement Can Spur Innovation: Bell Labs and the 1956 Consent Decree. *American Economic Journal: Economic Policy*, *12*(4), 328–359. https://doi.org/10.1257/pol.20190086

Appendix 1 – *Anecdotal evidence of 10-Ks mentioning the General Data Protection Regulation*

VERISK ANALYTICS Inc. (10-K, 2017)

Regulatory developments could negatively impact our business.

The Company has implemented various measures to comply with the Data Protection Directive and the forthcoming General Data Protection Regulation, however, there can be no assurances that such methods will not be invalidated as well. If the Company is unable to comply with the transfer mechanisms adopted pursuant to the Data Protection Directive and the forthcoming General Data Protection Regulation, it will impede the ability to conduct business between the U.S. and the E.U. which could have a material adverse effect on our business, financial position, results of operations or cash flows.

https://www.sec.gov/Archives/edgar/data/0001442145/000144214518000008/vrsk10k12312017.htm

SNAP Inc. (10-K, 2017)

Because we store, process, and use data, some of which contains personal information, we are subject to complex and evolving federal, state, and foreign laws and regulations regarding privacy, data protection, content, and other matters. (…)

Several proposals have recently been adopted or are currently pending before federal, state, and foreign legislative and regulatory bodies that could significantly affect our business. The General Data Protection Regulation in the European Union, which will go into effect on May 25, 2018, may require us to change our policies and procedures and, if we are not compliant, may seriously harm our business.

https://www.sec.gov/Archives/edgar/data/0001564408/000156459018002721/snap-10k_20171231.htm

ACHILLION PHARMACEUTICALS, INC. (10-K, 2018)

Given the breadth and depth of changes in data protection obligations, preparing for and complying with the GDPR's requirements has required and will continue to require significant time, resources and a review of our technologies, systems and practices, as well as those of any third-party collaborators, service providers, contractors or consultants that process or transfer personal data collected in the European Union. The GDPR and other changes in laws or regulations associated with the enhanced protection of certain types of sensitive data, such as healthcare data or other personal information from our clinical trials, could require us to change our business practices

https://www.sec.gov/Archives/edgar/data/0001070336/000119312519067172/d678011d10k.htm

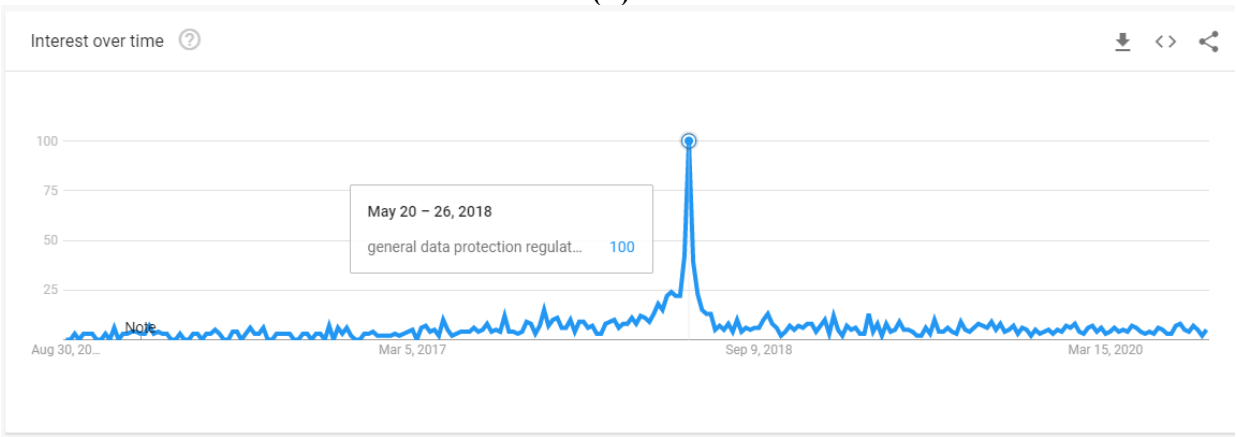
Appendix 2 – Variable definitions

Variable name	Variable Definition	Source
RawRet	Cumulative raw return for a given firm around the enforceability of the General Data Protection Regulation (GDPR). Returns are from May 21st to May 28th.	CRSP
SizeAdj Ret	Size-adjusted cumulative return for a given firm around the enforceability of the GDPR. Returns are from May 21st to May 28th.	CRSP
*I(Treated)*I(Post)*	Indicator variable capturing the average treatment effect. I(Treated)*I(Post) is the product of I(Treated) and I(Post) defined below.	Computed in the paper
I(Treated)	Indicator function equals one if firm discusses the General Data Protection Regulation (GDPR) on its 2018 10-K, zero otherwise.	WRDS SEC Analytics
I(Post)	Indicator variable equals one after 2018Q1, zero otherwise. In the case of investments, it equals one after 2016Q1, zero otherwise.	Computed in the paper
SalesGrowth	Sales growth. SalesGrowth is measured at the natural logarithm of current quarterly sales divided by lagged sales from the same quarter from the previous year.	Compustat Quarterly
Total_Inv	Total investment for a given firm in a given quarter. Total investment is measured as the sum of capital expenditures plus research and development expenses divided by total assets.	Compustat Quarterly
CAPEX	Capital expenditures for a given firm in a given quarter. CAPEX is measured as sum of capital expenditures in the current quarter and the previous three quarters divided by total assets.	Compustat Quarterly
R&D	Research and Development expenses. R&D is measured as the sum of research and development expenditures, plus a fraction of SG&A expenses according to the firm's industry (Ewens et al., 2023), in the current plus last three quarters, divided by total assets.	Compustat Quarterly
DataBreach	Indicator variable equals one if the firm discloses a data breach in the quarter, zero otherwise.	PRC Data Breach
SizeBreach	The natural logarithm of total number of records affected by a data breach as disclosed by the PRC Data Breach dataset	PRC Data Breach

Figures

Figure 1 Google Trends for searches related to "general data protection regulation" between 2015 and 2020, U.S. (Panel A) and Worldwide (Panel B).

(A)

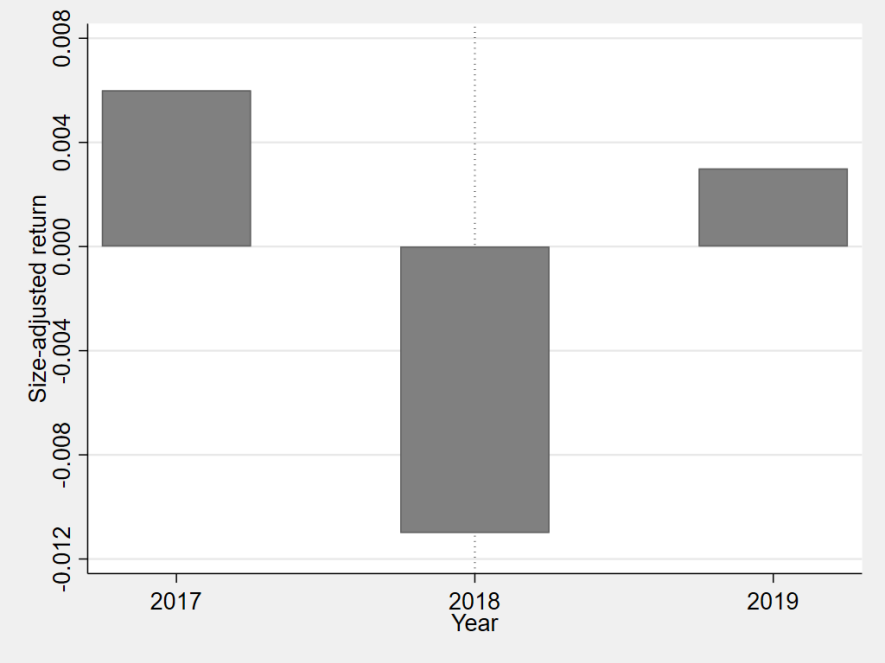


Notes: United States. Source: https://trends.google.com/trends/explore?date=today%205-y&geo=US&q=general%20data%20protection%20regulation

(B)

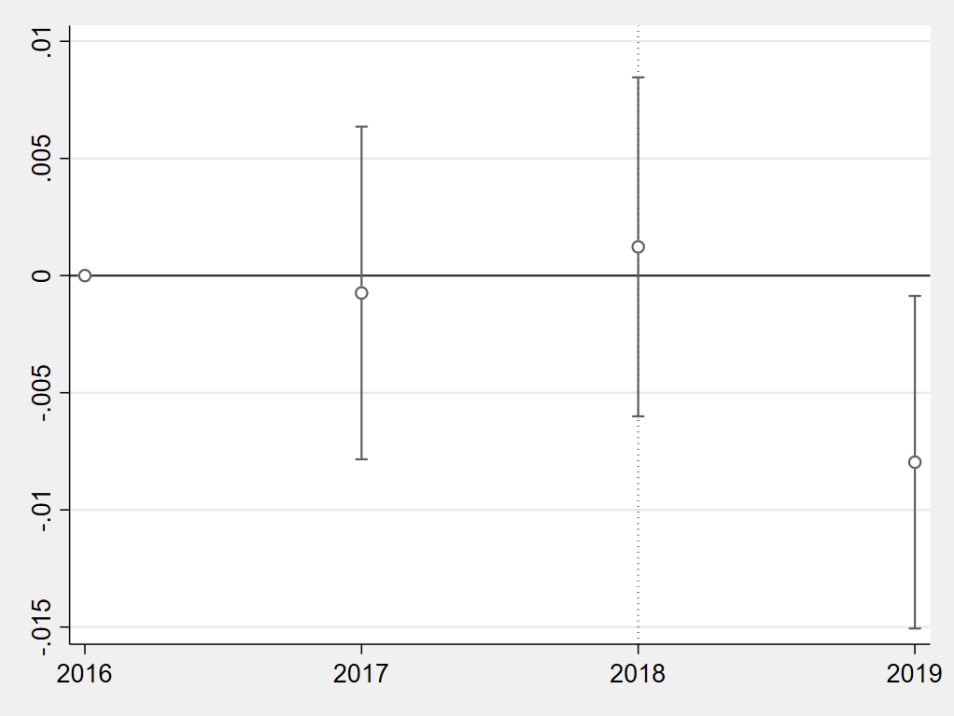
Notes: Worldwide. Source: https://trends.google.com/trends/explore?date=today%205-y&q=general%20data%20protection%20regulation

Figure 2: Effect of the GDPR on firms' stock returns



Notes: This figure shows the results of comparing the size-adjusted stock returns of GDPR-exposed firms in relation to not-exposed firms on the week GDPR became enforceable (May 21st to May 28th of 2018). The bar graphs represent the coefficients of the variable I(Treated). We also control for state fixed effects. We use the same week of the year for the 2017 and 2019 placebo tests to filter out seasonality effects.

Figure 3: Effect of the GDPR in the likelihood of data breaches



Notes: This figure shows the results of estimating the effect of the EU GDPR on the data breach likelihood around the enactment of the EU GDPR. The base year is 2016. Standard errors clustered by state. We include the main effects and firm and year Fes. Bars indicate the 90% confidence interval.

Tables

TABLE 1
Sample description

Panel A: Sample Selection

Sample and Filters	*Firm-quarter*	*UniqueFirms*	*Treated*
Compustat and CRSP Firms in 2018. Sample period initially ranges from 2014 to 2019.	116,281	5,284	688
After dropping firms headquartered outside the USA	86,540	4,671	624
After dropping firms with missing market data	71,550	3,771	603
After dropping data before 2016Q1 and after 2019Q4	49,003	3,771	603
After dropping firms without another observation in the state-industry.	43,499	3,323	551
After keeping only Healthcare, Banking, and Insurance as controls, and dropping treatment firms from these industries	12,909	1,013	516

Panel B: Sample Distribution by Fama-French 48 Industries

Industry	*Unique Firms*	*%*	*Industry*	*Unique Firms*	*%*
Business Services	174	33.72	Printing and Publishing	4	0.78
Pharmaceutical Products	61	11.82	Steel Works	4	0.78
Trading	31	6.01	Business Supplies	4	0.78
Electronic Equipment	29	5.62	Construction Materials	3	0.58
Medical Equipment	27	5.23	Electrical Equipment	3	0.58
Computers	26	5.04	Automobiles and Trucks	3	0.58
Other	19	3.68	Rubber and Plastic Products	2	0.39
Wholesale	16	3.10	Aircraft	2	0.39
Retail	15	2.91	Real Estate	2	0.39
Communication	14	2.71	Almost Nothing	2	0.39
Machinery	10	1.94	Construction	1	0.19
Measuring and Control Equipment	10	1.94	Utilities	1	0.19
Consumer Goods	8	1.55	Shipping Containers	1	0.19
Apparel	7	1.36			
Chemicals	7	1.36			
Transportation	7	1.36			
Recreation	5	0.97			
Personal Services	5	0.97			
Restaurants, Hotels, Motels	5	0.97			
Food Products	4	0.78			
Entertainment	4	0.78			
Total	516	100.0			

Panel C: Treated Sample distribution by State

State	*Unique Firms*	%	*State*	*Unique Firms*	%
CA	149	28.88	IN	4	0.78
NY	63	12.21	MN	4	0.78
MA	50	9.69	NV	4	0.78
TX	27	5.23	UT	4	0.78
IL	22	4.26	DC	3	0.58
NJ	22	4.26	WI	3	0.58
CO	16	3.10	KY	2	0.39
PA	16	3.10	NH	2	0.39
VA	16	3.10	SC	2	0.39
GA	14	2.71	TN	2	0.39
OH	14	2.71	AR	1	0.19
WA	14	2.71	KS	1	0.19
FL	13	2.52	LA	1	0.19
CT	12	2.33	ME	1	0.19
NC	11	2.13	MO	1	0.19
AZ	8	1.55	OR	1	0.19
MD	7	1.36			
MI	6	1.16			
Total	516	100			

TABLE 2
Descriptive Statistics

Panel A: Market returns tests

Variable	*Control*			*Treated*			*All*		
	Mean	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*
Raw return	0.006	0.03	497	-0.000	0.05	516	0.003	0.05	1,013
Size-adj return	0.007	0.03	497	-0.000	0.05	516	0.003	0.05	1,013

Panel B: Accounting-based performance tests

Variable	*Control*			*Treated*			*All*		
	Mean	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*
SaleGrowth	0.124	0.19	6,373	0.104	0.32	6,418	0.114	0.26	12,791
TotalInv	0.020	0.07	5,047	0.221	0.20	6,141	0.130	0.19	11,188
CAPEX	0.005	0.01	5,047	0.029	0.03	6,141	0.018	0.02	11,188
R&D	0.015	0.06	5,047	0.192	0.20	6,141	0.112	0.18	11,188

Notes: This table presents the descriptive statistics for the key variables of the study. *Treated* firms are not from Healthcare, Banking and Insurance that did mention GDPR in their 2018 10-Ks. *Control* were firms from Healthcare, Banking, and Insurance that did not mention GDPR in their 2018 10-Ks. *All* firms were both *Control* and *Treated* firms. All the variables are defined in Appendix 2.

TABLE 3
Event study – Market reaction to the start date of GDPR.

Panel A: Enforcement week

	RetRaw		*SizeAdj Ret*	
	(1)	*(2)*	*(3)*	*(4)*
I(Treated)	-0.006**	-0.010***	-0.008***	-0.011***
	(-2.59)	(-2.81)	(-2.90)	(-3.06)
Constant	0.006***	0.008***	0.007***	0.009***
	(4.69)	(4.54)	(4.93)	(5.00)
N	1013	1012	1013	1012
R^2	0.01	0.02	0.01	0.03
State FE	No	Yes	No	Yes

Panel B: Alternative model specifications for the market reaction to the enforceability of GDPR

Dependent Variable	Model Specification	$\beta_{\text{I(Treated)}}$	T-stat	*N*	R^2
RawRet	Robust SEs, State FE	-0.011***	(-3.02)	1,012	0.027
SizeAdj Ret	Robust SEs, State FE	-0.010***	(-2.65)	1,012	0.023
RawRet	Cluster by Ind, State FE	-0.011**	(-2.37)	1,012	0.027
SizeAdj Ret	Cluster by Ind, State FE	-0.010*	(-2.03)	1,012	0.023
FF4 Ret	Cluster by State, No FE model	-0.009**	(-2.42)	837	0.009
FF4 Ret	Cluster by State, State FE	-0.007	(-1.55)	835	0.044
FF4 Ret	Robust SEs, State FE	-0.007	(-1.54)	835	0.044
FF4 Ret	Cluster by Ind, State FE	-0.007	(-1.48)	835	0.044

Panel C: Cross-sectional test on market reaction to the enforceability of GDPR

	RawRet		SizeAdj Ret	
	Low Intan	*High Intan*	*Low Intan*	*High Intan*
	(1)	(2)	(3)	(4)
I(Treated)	0.002	-0.007***	0.000	-0.008***
	(0.21)	(-3.26)	(0.02)	(-3.58)
Constant	-0.002	0.007***	-0.001	0.008***
	(-0.25)	(5.51)	(-0.07)	(5.54)
Diff High – Low (*t*)	-0.009 (-0.93)		-0.008 (-0.83)	
N	251	762	251	762
R^2	0.000	0.008	0.000	0.011
State FE	No	No	No	No

Notes: This table presents the OLS results of regressing two different measures of change in shareholder wealth on firm indicators for firms exposed to GDPR during the enforcement week of the regulation (May 21st to May 28th, 2018). The two measures of change in shareholder wealth are the cumulative raw return (*RawRet*) and size-adjusted cumulative return (*SizeAdj Ret*). *I(Treated*) is an indicator variable that equals one if the firm discusses GDPR on its 10-K scale and zero otherwise. All the variables are defined in Appendix 2. Standard errors were clustered by state. ***, **, * indicates significance at the 0.01, 0.05, 0.10 level, respectively.

TABLE 4
Revenue effects of GDPR

	SalesGrowth		
	(1)	*(2)*	*(3)*
I(Treated)*I(Post)	-0.066***	-0.058***	-0.060***
	(-3.86)	(-2.99)	(-3.09)
N	12,791	12,727	12,719
R^2	0.07	0.11	0.37
Main Effects	Yes	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes
State-Year FE	No	Yes	Yes
Firm FE	No	No	Yes

Notes: This table presents the OLS results of regressing firm sales growth on firm indicators for firms exposed to the GDPR and post-GDPR enforceability quarters. The main coefficient of interest is that of *I(Treated)*I(Post)*. *I(Treated)* is an indicator variable that equals one if the firm discusses GDPR on its 10-K scale and zero otherwise. *I(Post)* is an indicator variable that equals one after May 2018 and zero otherwise. The main effects are included in the regression, but are not reported due to collinearity with fixed effects. All the variables are defined in Appendix 2. Standard errors were clustered by state. ***, **, * indicates significance at the 0.01, 0.05, 0.10 level, respectively.

TABLE 5
Investment decisions

	(1)	*(2)*	*(3)*
	Total inv.	*CAPEX*	*R&D*
I(Treated)*I(Post)	0.011***	0.002***	0.009***
	(3.17)	(4.95)	(2.75)
N	10,805	10,805	10,805
R^2	0.92	0.88	0.92
Main Effects	Yes	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes

Notes: This table presents the OLS results of regressing various measures of firm investment on firm indicators for firms exposed to the GDPR and post-GDPR enforceability quarters. The three measures of investment are the sum of capital expenditures and R&D by the firm (Total_Inv), total capital expenditures (CapExp), and total R&D expenses (R&D). All dependent variables were scaled by lagged total assets. The main coefficient of interest is that of *I(Treated)*I(Post)*. *I(Treated)* is an indicator variable that equals one if the firm discusses GDPR on its 10-K scale and zero otherwise. *I(Post)* is an indicator variable that equals one after 2018Q1 and zero otherwise. The main effects are included in the regression, but are not reported due to collinearity with fixed effects**.** All the variables are defined in Appendix 2. Standard errors were clustered by state. ***, **, * indicates significance at the 0.01, 0.05, 0.10 level, respectively.

TABLE 6
Data Breach Likelihood

	DataBreach *(1)*	*DataBreach* *(2)*	*DataBreach* *(3)*
I(Treated)*I(Post)	-0.005*	-0.005*	-0.005**
	(-1.92)	(-1.79)	(-2.24)
N	13,258	13,253	13,251
R^2	0.03	0.04	0.36
Main effects	Yes	Yes	Yes
Industry FE	Yes	Yes	No
State FE	No	Yes	No
Firm FE	No	No	Yes
Year FE	Yes	Yes	Yes

Notes: This table presents the linear probability model results of regressing the indicator variable of whether the firm reports data breaches in a quarter on firm indicators for firms exposed to the GDPR and post-GDPR enforceability quarters. The main coefficient of interest is that of *I(Treated)*I(Post)*. *I(Treated)* is an indicator variable that equals one if the firm discusses GDPR on its 10-K scale and zero otherwise. *I(Post)* is an indicator variable that equals one after May 2018 and zero otherwise. All the variables are defined in Appendix 2. Standard errors were clustered by state. ***, **, * indicates significance at the 0.01, 0.05, 0.10 level, respectively. main effects are included in the regression, but are not reported due to collinearity with fixed effects.

TABLE 7
Type of Data Breach

	Hack	*InvDisc*	*Port*	*Phys*	*Unkn*
I(Treated)*I(Post)	-0.005***	-0.000	0.001	-0.001	0.000
	(-2.77)	(-0.25)	(1.33)	(-0.92)	(0.04)
N	13,251	13,251	13,251	13,251	13,251
R^2	0.23	0.38	0.20	0.20	0.25
Main effects	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: This table presents the linear probability model results of regressing various types of data breaches on firm indicators for U.S. firms exposed to GDPR and post-GDPR enforceability quarters. The main types of data breach are *Hack, InvDisc, Port, Phys, Unkn. Hack* is an indicator variable equal to one if the data breach refers to hacking or malware and zero otherwise. *InvDisc* is an indicator variable that equals one if the data breach refers to an involuntary disclosure and zero otherwise. *Port* is an indicator variable equal to one if the data breach refers to a data breach related to a portable device and zero otherwise. *Phys* is an indicator variable that equals one if the data breach refers to a data breach related to the loss of a physical object and zero otherwise. Finally, *Unkn* is an indicator variable equal to one if the data breach refers to an unknown type of breach and zero otherwise. The main coefficient of interest is that of *I(Treated)*I(Post)*. *I(Treated)* is an indicator variable that equals one if the firm discusses GDPR on its 10-K scale and zero otherwise. *I(Post)* is an indicator variable that equals one after May 2018 and zero otherwise. All the variables are defined in Appendix 2. Standard errors were clustered by state. ***, **, * indicates significance at the 0.01, 0.05, 0.10 level, respectively.

TABLE 8
Data Breach Size

Panel A: Descriptive Statistics

	Control		*Treated*		*Total*	
	Mean	*Median*	*Mean*	*Median*	*Mean*	*Median*
Size_Breach	42,994	1,214	3,446,504	500	1,914,924	818
ln(Size_Breach)	7.130	7.102	5.080	6.217	6.002	6.708
RawRet	0.008	0.004	-0.006	-0.000	0.000	0.003
SizeAdj Ret	0.010	0.002	-0.004	-0.003	0.002	0.000
N			62			

Panel B: Regression Analysis

	ln(Size_Breach) *(1)*	*ln(Size_Breach)* *(2)*	*ln(Size_Breach)* *(3)*
I(Treated)*I(Post)	3.530	4.291	5.112
	(1.22)	(1.34)	(1.25)
I(Treated)	†	†	†
I(Post)	0.865		
	(0.61)		
N	50	50	48
R^2	0.55	0.59	0.63
Industry FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Year FE	No	Yes	Yes
BreachType FE	No	No	Yes

Notes: This table presents the descriptive statistics of the number of records and market reaction associated with data breaches in the sample. Panel A presents the descriptive statistics for *Size_Breach (unlogged and logged)*, *RawRet* and *SizeAdj Ret*. Panel B presents the OLS results of examining the effect of GDPR on the size of disclosed data breaches. *Size_Breach* is defined as the total number of breached records (logarithm of). The main coefficient of interest in Panel B is that of *I(Treated)*I(Post)*. *I(Treated)* is an indicator variable that equals one if the firm discusses GDPR on its 10-K scale and zero otherwise. *I(Post)* is an indicator variable that equals one after May 2018 and zero otherwise. All the variables are defined in Appendix 2. Robust standard errors exist because there are too few states to cluster (Kezdi, 2004). ***, **, * indicates significance at the 0.01, 0.05, 0.10 level, respectively. †*I(Treated)* drops due to collinearity with Industry FEs.

TABLE 9
Market Reaction to Data Breach

	RawRet	*RawRet*	*RawRet*	*SizeAdj Ret*	*SizeAdj Ret*	*SizeAdj Ret*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
I(Treated)*I(Post)	-0.053*	-0.019	-0.013	-0.039	-0.005	0.000
	(-1.81)	(-0.67)	(-0.40)	(-1.57)	(-0.21)	(0.02)
I(Treated)	†	†	†	†	†	†
I(Post)	0.024			0.016		
	(1.52)			(1.01)		
N	52	52	50	52	52	50
R^2	0.35	0.43	0.55	0.25	0.37	0.49
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	No	Yes	Yes
BreachType FE	No	No	Yes	No	No	Yes

Notes: This table presents the OLS results of regressing two different measures of change in shareholder wealth on firm indicators for firms exposed to GDPR and post-GDPR enforceability quarters. The two measures of change in shareholder wealth are the cumulative raw return (*RawRet*) and size-adjusted cumulative return (*SizeAdj Ret*) around a given data breach. The main coefficient of interest is that of *I(Treated)*I(Post)*. *I(Treated)* is an indicator variable that equals one if the firm discusses GDPR on its 10-K scale and zero otherwise. *I(Post)* is an indicator variable that equals one after May 2018 and zero otherwise. All the variables are defined in Appendix 2. Robust standard errors exist because there are too few states to cluster (Kezdi, 2004). ***, **, * indicates significance at the 0.01, 0.05, 0.10 level, respectively. †*I(Treated)* drops due to collinearity with Industry FEs.