# Tracing Out International Data Flow: The Value of Data and Privacy

**Junjun Quan** [*]

December 5, 2023

**Abstract**

I measure firms' value of data and consumers' privacy preferences by analyzing the supply and demand-side reactions to the EU's General Data Protection Regulation (GDPR). While previous research has focused on consumer reactions to privacy regulations, my study also incorporates firm responses. After GDPR limits firms' access to data, the EU sales share of US data-intensive firms declines by 8%. EU consumers, who can choose to share less data, suffer a 6% deterioration in user experience as measured by app ratings. I develop a general equilibrium model to map these empirical findings and estimate the value of data and privacy. Privacy-conscious consumers gain from privacy protection. However, the quantitative model reveals that the digital welfare of other consumers declines because firms also use data to enhance productivity and customize digital products. In aggregate, EU digital welfare declines by 4.05%.

**Keywords:** Value of Data and Privacy, Multinational Firms, Consumer Privacy Protection, Data Economy

**JEL classifications:** G30, D12, D22, O34, D62, D18, F20

[*] Columbia Business School. Email: jq2291@columbia.edu. Kravis Hall, 665 W 130th St, New York.

# 1 Introduction

Consumer data has become an important form of intangible capital in the digital era. The fast advancement in computing power and artificial intelligence has led to a massive leap in data processing capacity. As companies increasingly mine vast troves of consumer data, privacy concerns have been rising. In response, legislators worldwide have enacted privacy regulations, such as the EU's General Data Protection Regulation (GDPR), China's Personal Information Protection Law (PIPL), and California Consumer Privacy Act (CCPA). Despite differences in privacy protection practices, the universal challenge lies in striking the right balance: harnessing data to innovate while being mindful of the privacy implications.

To inform future policy, we need to assess the importance of data for companies and consumer privacy preferences at the aggregate level. The EU's General Data Protection Regulation provides a useful policy experiment to study the effects of privacy laws, and I examine its impact on US multinational firms and their customers. While earlier studies have considered the value of data and consumer responses to privacy laws in isolation, my paper integrates both the supply-side and demand-side responses and estimates the value of data and privacy from an equilibrium model. I first show novel reduced-form evidence of the regulation's effects on firms and consumers. GDPR prompts US data-intensive companies to reallocate their business operations across geographical segments. Their sales share from the European market declines, where data access is more restrictive, but their sales from other regions of the world increase. EU consumers grapple with a conundrum: less data sharing for better privacy protection means sacrificing personalized digital experiences, highlighting the key trade-off on the consumer side.

The two main findings on firm business shifting and user experience deterioration are equilibrium outcomes that reflect the adjustment from both the firm and consumer side, and they cannot be directly mapped to the underlying value of data and privacy. This calls for a model that incorporates data as a key input in the production function and consumer utility to disentangle the supply and demand forces that drive what we observe in the data. I then use the calibrated model to quantify the welfare implications of GDPR.

One main contribution of this paper lies in evaluating the impact of a privacy regulation from both the firms' and consumers' perspectives while accounting for equilibrium interactions between these two parties. Firms use data to improve productivity and customize consumer products, and they price in the data sharing behaviors of consumers. Privacy-conscious consumers trade off the personalized digital experiences for better privacy protection. Consumer privacy preferences affect firms' data value through a feedback loop: more data collected today increases productivity tomorrow, thereby catalyzing further data generation tomorrow. A consumer base with heightened privacy awareness results in less data accumulation each period.

Data also exhibits characteristics of a public good, as firms can utilize the data harvested from one consumer to glean insights about other consumers with similar characteristics. Moreover, data boosts firm-level productivity and lowers the production cost, which benefits all consumers. However, firms' inability to perfectly price discriminate at the individual level, along with individuals' failure to internalize the positive externalities their data sharing bestows upon others, leads to an under-provision of data. Upon the enactment of GDPR, privacy-conscious consumers free-ride on their non-privacy-conscious counterparts, leading to a distributional welfare impact. They get better privacy protection while enjoying good digital service quality from the data shared by others. Additionally, the GDPR exerts spillover effects on US consumers through the product quality of US multinational firms.

I use the GDPR as a policy experiment to empirically test the importance of data for firms and the privacy preferences of consumers. The GDPR, enacted in the European Union in May 2018, is the most stringent and comprehensive privacy regulation to date,[1] serving as a blueprint for the privacy regulations around the world. Post-GDPR, firms intending to collect and process data from EU residents are mandated to secure explicit consent, clearly informing consumers about the purpose of data usage. While primarily aimed at safeguarding EU residents' personal data,[2] GDPR's scope extends beyond the EU as any multinational firm that operates in the European market needs to comply. The fact that I examine the

[1] For severe violations, as listed in Art. 83(5) GDPR, a company can be fined up to 20 million euros or 4% of their total global turnover of the preceding fiscal year, whichever is greater. For less severe violations, as defined in Art. 83(4) GDPR, a company will still face fines of up to 10 million euros or 2% of its entire global turnover of the preceding fiscal year, whichever is greater.

[2] GDPR also applies to Iceland, Norway, and Liechtenstein, which belong to the European Economic Area (EEA), not EU. As of 2021, the United Kingdom retains the law in identical form despite no longer being an EU member state.

impact of GDPR on US multinational firms alleviates the policy endogeneity concerns as most of US firms are not directly involved in the legislation process of GDPR. However, GDPR has significant impact on US multinational firms. In recent years, there has been an increasing number of US public firms disclosing privacy-related risk factors in their 10-K filings, as shown in Figure A1. Since 2016, the disclosure of such risks has become more specific by directly mentioning privacy regulations like GDPR and CCPA.

Data-intensive firms are more exposed to GDPR, so to quantify the exposure to the privacy regulation, I construct a novel measure of data intensiveness, which proxies for firms' dependence on data in their production process. The measure consists of two dimensions: skill and technology. For the skill dimension, I leverage the Lightcast job posting data to measure firms' demand for AI and data management related skills. For the technology dimension, I exploit the USPTO patent data to estimate the market and scientific value of the data processing patents these firms own. I average over these two dimensions and come up with a comprehensive measure of data intensiveness.

I find a compositional shift in the share of sales that US multinational firms derive from each part of the world, with data-intensive firms shifting away from the EU market. Employing a difference-in-differences (DID) design, I observe an 8% drop in the share of sales generated from the European market among US data-intensive firms. Firms collect data from the consumers that they sell products to.[3] However, since consumer demand is not fully elastic, consumers from outside of the EU cannot fully absorb the unmet "data demand" by US multinational firms. There is a negative productivity shock to the US firms because of the restricted access to data in the EU market, which leads to a lower demand for data scientists, data engineers, software engineers, and machine learning engineers by US data-intensive firms with positive EU exposure.

To properly account for the demand response from consumers, I also look into the digital experiences of EU and US consumers. I web-scraped the entire history of user reviews, both numerical and textual, for 4,883 popular apps on the Google Play Store, spanning all 32 app categories. I collected the reviews separately for US and EU users by visiting the Google Play Store by country and language. I classify apps by the purpose of data collection, e.g.,

[3] Here I mainly discuss the first-party data that firms directly collect from consumers.

personalization and advertisement. I employ a difference-in-differences strategy and show that, for apps that collect data for personalization purposes, there is a 6% decline in EU user ratings post-GDPR, while their US counterparts experience a much smaller impact. The difference in the changes of user experiences between the US and EU is both statistically and economically significant. Privacy protection is generally not meant to eliminate data sharing and reaching a state of secrecy. It is generally framed as giving consumers the choice to share more or less data as they desire because data sharing also improves the quality of digital products.

As firms' response in resource reallocation induces changes in the consumer composition of US multinational firms, the estimated effects on user ratings also reflect the supply-side adjustment. To disentangle the supply and demand side effects from the empirical findings, I build a two-economy equilibrium model, in which US multinational firms offer goods and services to both domestic and foreign consumers. One significant challenge in bridging the theoretical and empirical work on the data economy is the measurement of data stock. Drawing inspiration from the theoretical literature, I link the production and consumption processes with the data generation process, constructing a dynamic model to capture the accumulation of data over time. Data is modeled as a byproduct of economic activities. First, data facilitates the development of new technology and enhances productivity; in addition, it can be harnessed to predict consumer preferences and deliver more tailored products. Due to the second feature, consumers weigh the advantages of sharing data for personalized experiences against the costs of privacy infringement. When GDPR grants EU consumers greater control over their own data, they assess the personal benefits of data sharing against privacy violation costs, without fully internalizing the positive externality[4] their data could offer to others. US multinational data-intensive firms, in response to the data-sharing regulation change, divert from the European market and reallocate resources to other geographical segments. EU local digital firms are constrained to the European market, bearing the brunt of GDPR's impact. Depending on demand elasticity, these costs

[4] When an individual shares their data, the benefits extend beyond personal gains. It also helps refine predictive models that discern the preferences of others with similar characteristics, thereby enhancing the overall efficiency of the matching processes. Additionally, as firms accumulate more data to train their algorithms, they achieve greater productivity overall.

on firms will be partially passed on to EU consumers, and their US counterparts might also be negatively impacted due to the externality of data sharing.

GDPR, a regulation intended to improve consumer welfare by giving consumers more autonomy over data sharing, inadvertently ends up harming both US and EU consumers, with EU privacy-conscious consumers being the only group benefiting from this regulation. Although EU consumers gain better privacy protection, the positive impact on welfare is tempered by market forces—data sharing is priced in by digital firms. This regulation not only stifles the growth of digital firms but also imposes a negative welfare impact on consumers in the aggregate. Furthermore, when data-sharing choices are given back to individuals, the phenomenon of the "tragedy of commons" emerges. Individuals undersupply data because they do not consider the positive externality on others. The EU digital welfare declines by 4.05%, and the US digital welfare declines by 1.37%. Given the immense growth potential of the digital economy, it is imperative that privacy regulations balance the efficiency gains from data sharing against consumer privacy concerns. This paper is the first to examine the various forces that could affect the welfare implications of privacy regulations from a multinational perspective and by accounting for interactions between firms and consumers in equilibrium.

**Related Literature:** My paper contributes to the rapidly growing field of the data economy. The literature in this area highlights the concept that data, a by-product of economic activities, can be traded as an asset and plays a vital role as an input in the production process (Admati and Pfleiderer 1990; Veldkamp 2005; Van Nieuwerburgh and Veldkamp 2006; Ordonez 2013; Fajgelbaum et al. 2017; Begenau et al. 2018; Farboodi et al. 2018; Choi et al. 2019; Farboodi et al. 2019; Veldkamp and Chung 2019; Cong et al. 2020; Farboodi and Veldkamp 2020; Jones and Tonetti 2020; Farboodi and Veldkamp 2021; Cong et al. 2022; Eeckhout and Veldkamp 2022; Chang et al. 2023; Farboodi and Veldkamp 2023; Veldkamp 2023). My paper is closely related to the research exploring the integration of data technology/AI with human labor and its consequent impact on firms (Abis and Veldkamp 2020; Cao et al. 2020, 2021; Acemoglu et al. 2022a; Babina et al. 2022b) and the research measuring the value of data for asset management companies and other market participants (Farboodi et al. 2022;

Bai et al. 2023). My paper is also connected to the theoretical literature that delves into the role of data intermediaries, consumers' privacy preferences, and their data sharing decisions (Bordalo et al. 2016; Bergemann et al. 2019; Braghieri 2019; Kirpalani and Philippon 2020; de Montjoye et al. 2021; Acemoglu et al. 2022b; Chen 2022; Argenziano and Bonatti 2023; Liu et al. 2023). My paper focuses on the interactions between firms and consumers in the context of privacy regulation, and the interactions come from the dual roles of data: data can be used by firms to increase productivity and customize digital products. I develop a general equilibrium model to disentangle the supply and demand-side responses following a privacy regulation. I estimate the value of data for firms and show that consumers' privacy preferences play an important role in influencing the marginal value of data.

My paper also contributes to the broader discussion on the impact of privacy regulations. Previous work has shed light on the impact of privacy regulations/policies on digital marketing, VC funding, web tracking, web technologies, financial security, firm performance, and market competition (Evans 2009; Goldfarb and Tucker 2011; Johnson 2013; Lenard and Rubin 2013; Benkler et al. 2018; Jia et al. 2018; Choi et al. 2019; Martin et al. 2019; Aridor et al. 2020; Bleier et al. 2020; Jia et al. 2020; Johnson et al. 2020; Zhuo et al. 2021; Canayaz et al. 2022; Janssen et al. 2022; Johnson 2022; Peukert et al. 2022; Godinho de Matos and Adjerid 2022; Bian et al. 2023; Johnson et al. 2023). Other papers look at firm privacy disclosure, data access, and consumer data sharing behaviors (Ramadorai et al. 2020; Chen et al. 2021; Babina et al. 2022a). Johnson (2022) provides a comprehensive literature review of the empirical evidence on the impact of GDPR. Aridor et al. (2020) leverages the data from a travel intermediary and shows that GDPR results in a 12.5% drop in the intermediary observed EU consumers. Goldfarb and Tucker (2011) study the effects of the European E-Privacy Directive, which limited firms' ability to track users' online behavior and show that online display advertisements in the EU became less effective than other areas after the directive was enacted. Canayaz et al. (2022) study the negative impact of CCPA on the profitability of conversational AI firms. I use GDPR as a policy experiment to measure of the value of data for firms and the privacy preferences of consumers. I provide further evidence on the impact of a regional privacy regulation (GDPR) from a global perspective and focus on both firms and consumers. Furthermore, I quantify the welfare impact of GDPR on different types of consumers. I show that GDPR leads to a distributional welfare impact, with EU privacy-conscious consumers gaining at the cost of other consumers.

Moreover, my paper is connected to the literature that seeks to quantify the value of privacy. Tang (2019) runs a lending experiment on a Chinese fintech platform. The paper links loan application completion rate with borrowers' privacy preferences, and measures the value of loans that borrowers are willing to give up to avoid disclosing sensitive information (social network ID or employer). Bian et al. (2021) studies how Apples' app privacy disclosures affect app users' willingness to download an app, and its negative impact on revenue. My paper shows indicative evidence that privacy-conscious consumers weigh the benefits of data sharing for personalized digital experiences against privacy concerns. While previous studies mainly focus on consumer reactions to privacy regulations, my research also incorporates firm responses, which affect digital product quality and effective prices. Specifically, I estimate the value of privacy for consumers from an equilibrium model.

The rest of the paper is structured as follows. In Section 2, I describe the data and measurement methods used in the paper. In Section 3, I analyze the demand for data by firms. In Section 4, I study the demand for privacy by consumers. In Section 5, I set up a theoretical framework to map the empirical findings. In Section 6, I calibrate the model and perform a welfare analysis. Section 7 concludes.

# 2 Data and Measurement

## 2.1 Data Sources

### 2.1.1 Lightcast US Online Job Posting Data

Lightcast US online job postings data covers more than 200 million electronic job postings in the US from Jan 1, 2010 to May 31, 2020. Burning Glass web-scraped job posting information from around 40,000 company websites and online job boards, and they apply a de-duplication algorithm to avoid counting the same job posting multiple times. They parse the raw textual data and extract detailed information on the Employer, location, occupation, industry, wages, and skills required. Carnevale et al. (2014) estimate that the job posting data covers around 60% - 70% of all vacancies in the United States. The detailed skill requirements in the job posting data will enable me to measure US firms' demand for different types of talent. Following Abis and Veldkamp (2020), Acemoglu et al. (2020), and Babina et al. (2020), I classify jobs into AI-related postings and data-management-related postings.[5] Firms' demand for data managers, data scientists, and machine learning engineers can help me measure how a firm's business model depends on consumers' data. I can also study how the workforce composition of US firms changes in response to privacy regulations.

### 2.1.2 Accounting, Financial, and Geographical Segment Data

I obtain accounting and financial data of US public firms from Compustat North America Fundamentals Quarterly and CRSP, including total assets, total debt, total sales, gross profits, net profits, market capitalization, daily stock prices, etc.

Furthermore, Compustat Geographical Segment data supplements the firm-level accounting data with revenue, costs, investment compositions by geographical regions. FASB[6] 131, effective December 15, 1997, requires public business enterprises to report financial information and descriptive information about their Operating segments.[7] This Statement requires that a public business enterprise report a measure of segment profit or loss, certain specific revenue and expense items, and segment assets. It requires reconciliations of total segment revenues, total segment profit or loss, total segment assets, and other amounts disclosed for segments to corresponding amounts in the enterprise's general-purpose financial statements. It requires that all public business enterprises report information about the revenues derived from the enterprise's products or services (or groups of similar products and services), about the countries in which the enterprise earns revenues and holds assets, and about major customers regardless of whether that information is used in making operating decisions. However, this Statement does not require an enterprise to report information that is not

[5] The keyword list used for classification can be found in Appendix D.

[6] Financial Accounting Standards Board.

[7] This Statement supersedes FASB Statement No.14, Financial Reporting for Segments of a Business Enterprise, but retains the requirement to report formation about major customers. It amends FASB Statement No.94, Consolidation of All Majority-Owned Subsidiaries, to remove the special disclosure requirements for previously unconsolidated subsidiaries. This Statement does not apply to nonpublic business enterprises or to not-for-profit organizations. See https://www.fasb.org/page/PageContent?pageId=/reference-library/superseded-standards/summary-of-statement-no-131.html for more details.

prepared for internal use if reporting it would be impracticable.

The S&P Global Market Intelligence parses the 10-K filing textual data and tabulates the segment disclosure in a structured format. The Compustat Business Information files were designed to allow for restated data in conjunction with changes in disclosure requirements. The Segment Item Value File provides the historical data and up to 2 data source years of restated data back to 1998. The number of records for each data year depend on whether the company restates the period with a subsequent source. [8] For each year, I keep the data when it was first reported (historical data). During the sample period 2010-2021, around 72% of US public firms disclose their geographical revenue compositions each year, and 60% of US public firms generate revenue from international sources.

The segment data enables me to measure the fraction of revenue coming from and the strategic importance of each geographical region for US public firms. I am particularly interested in how US multinational firms reallocate their businesses across geographical segments.

### 2.1.3 Innovation

Patent data are from the United States Patent and Trademark Office. Kogan et al. (2017) have introduced a new measure of the economic value of patents. They use the stock market response to patent granting to estimate the economic value of patents. They have made the data available online thorough a GitHub repository.[9] They have also matched the patent data to the the CRSP firm/security level identifier.

### 2.1.4 Google Play Store Data

I collect app review data from Google Play Store to measure user experiences. The review data contains both numerical ratings and textual comments. The numerical rating is on a scale of 1 (low) to 5 (high). In the textual comments, consumers share details about their

[8] For example, if XYZ Corp reported their 1998 business segment data on the 1998 10K, there would be one record for that year. In 1999, XYZ Corp restates their 1998 data with the 1999 10K, there would be one record for 1999 and two records for 1998: one with the Source Year of 1998 and the other with 1999. In 2000, they restate both 1999 and 1998 data. There would be one record for 2000, two records for 1999 (one historical [Source Year = 1999] and one restated [Source Year = 2000]), and three records for 1998 (one historical [Source Year = 1998] and two restated [Source Year = 1999, 2000].

[9] https://github.com/KPSS2017/Technological-Innovation-Resource-Allocation-and-Growth-Extended-Data

experiences while using the apps. When we rank the reviews by relevance, the ones at the top are usually very informative about apps' main products or services. By switching the region of the Google Play Store, I collect the data separately for US users and EU users. Since companies often offer different versions of products in different markets, app user experiences can differ across countries. Moreover, the quality of digital services will be affected by the amount of data users share with app developers.

My analysis focuses on 4,883 popular apps on Google Play Store. To compile this list of apps, I start with the 250 most popular apps recommended by Google in each app category, including Art and Design, Auto and Vehicles, Beauty, Books and Reference, Business, Comics, Communication, Dating, Education, Entertainment, Events, Finance, Food and Drink, Health and Fitness, House and Home, Libraries and Demo, Lifestyle, Maps and Navigation, Medical, Music and Audio, News and Magazines, Parenting, Personalization, Photography, Productivity, Shopping, Social, Sports, Tools, Travel and Local, Video Players and Editors, and Weather. Then I extend from this initial list and search for relevant apps associated with each app, and this process brings me to around 20,401 apps.

In 2021, Google announced that all developers on the Google Play platform are required to disclose their apps' privacy and security practices in a Data Safety section of their apps' store listing page. The measure is aimed at helping Google Play users understand how the apps collect and share their data before they download.[10] This information helps users make more informed choices when deciding which apps to install. Section B.3 provides several screenshots from Instagram's Data Safety section on what information it collects from users and for what purposes. By July 20, 2022, all developers must declare how they collect and handle user data for the apps they publish on Google Play and provide details about how they protect this data through security practices like encryption. This includes data collected and handled through any third-party libraries or SDKs used in their apps.

To be included in my sample, an app needs to have a valid Data Safety disclosure and have at least ten reviews before and after GDPR came into effect. These two criteria bring the sample from 20,402 apps to 4,883 in the main analysis.

[10] Apple App Store also has a similar change in 2021, named privacy nutrition labels. These labels fall into three categories: "Data Used to Track You", "Data Linked to You", and "Data Not Linked to You".

### 2.1.5 Coresignal LinkedIn Profile Data

In addition to job posting data, I utilize online professional profile data provided by Coresignal. Coresignal collects detailed profile information of LinkedIn members, including attributes like names, job titles, locations, work experiences, educational backgrounds, etc.

I leverage the data on job titles and associated companies from each member's experience section to identify their respective employers and the functions they serve within those companies. These LinkedIn profiles also specify the start and end dates of their work experiences. By aggregating this information across all profiles affiliated with a specific company, I can deduce both the size and composition of its workforce over time. Specifically, I examine the total workforce size, the "data staff", who are mainly responsible for data collection, management, and analysis (such as machine learning engineers, data engineers, software engineers, and data analysts), and the personnel devoted to customer service and support functions.

### 2.1.6 Risk Disclosures in Annual 10-K Filing

Under Regulation S-K Item 105, US public firms are required to provide, under the caption "Risk Factors" in their 10-K filings to the SEC, a discussion of the material factors that make an investment in the registrant or offering speculative or risky. They need to concisely explain how each risk affects the registrant or the securities being offered. Campbell et al. (2014) find that managers faithfully disclose the risk they face, and firms facing greater risk disclose more risk factors. I use textual analysis tools to extract corporate risk disclosures from their annual 10-K filings.

## 2.2 Measurement

### 2.2.1 Data Intensiveness

The data intensiveness measure assesses the degree to which a firm's business operations depend on consumer data collection and the extent to which this data can be used to improve its products, technology, and marketing strategies. Notable examples include information technology firms such as Google, Meta, and Netflix. These companies gather vast amounts of data to refine their algorithms, enhance their products, and function as digital platforms that facilitate advertising campaigns for smaller businesses.

However, the digital economy extends far beyond these well-known tech giants. Rapid advancements in computing power and artificial intelligence have enabled a growing number of firms to collect, process, and exploit large volumes of consumer data, sparking a digital transformation across various industries. Retail giants like Walmart and Target, while not traditionally seen as technology firms, have started hiring data scientists and machine learning engineers in response to the increasing need for consumer data analysis. Likewise, the automotive industry is experiencing a digital revolution, with Alphabet's Waymo and GM's Cruise heavily investing in AI talent for their research and development teams working on autonomous vehicles.

It is clear that relying solely on industry classifications is inadequate for understanding the digital economy. Investment in digital assets has shifted from physical infrastructure to talent acquisition in data management and analysis, as well as research and development of data processing technology. I propose a measure of data intensiveness based on the talent employed by firms and the market and scientific value of their data processing technologies. Data processing technology refers to patents with the Cooperative Patent Classification (CPC) code G06F, G06N, G06Q, G06T, G06V, or G16, which pertain to data processing, computing, image processing, video recognition, etc. A detailed description of each category can be found on the USPTO website[11] and Section C.1. Among the patents[12] matched to public firms, 38.5% are classified as data-intensive. I assess their scientific value by the number of forward citations these patents receive (adjusted for patent "age") and measure market value using the method proposed by Kogan et al. (2017). In each year-quarter, I compute the scientific value of data processing technology using the following formula:

$$\text{Scientific Value}_{i,t} = \frac{\sum_p \text{Forward Citations (Newly Granted Data-Intensive Patents)}_{i,p,t}}{\sum_{p'} \text{Total Forward Citations of All Newly Granted Patents}_{i,p',t}} \tag{1}$$

[11] https://www.uspto.gov/web/patents/classification/cpc/html/cpc-G.html

[12] I only consider patents with CPC code starting with G (Physics) and H (Electricity).

and compute the market value of data processing technology as

$$\text{Market Value}_{i,t} = \sum_{p} \frac{\text{Market Value of Patent}_{i,p,t}}{\text{Market Capitalization}_{i,t}} \tag{2}$$

In each year-quarter, the market value of patent p is scaled by the market capitalization of firm i. These two variables capture the first dimension of data intensiveness: data processing technology.

For the second dimension of data-intensiveness, I use the keywords identified by Abis and Veldkamp (2020), Acemoglu et al. (2020), and Babina et al. (2020) and classify jobs that require AI and data management skills. The list of AI skills includes machine learning, computer vision, deep learning, virtual agents, image recognition, natural language processing, speech recognition, and neural networks, among others. Data management skills encompass Apache Hive, information retrieval, data warehousing, SQL Server, data visualization, database management, data governance, and database administration, among others. The complete list of keywords for AI skills and data management skills can be found in Appendix D. For each year-quarter, I compute the percentage of job postings that require AI related skills and data management related skills.

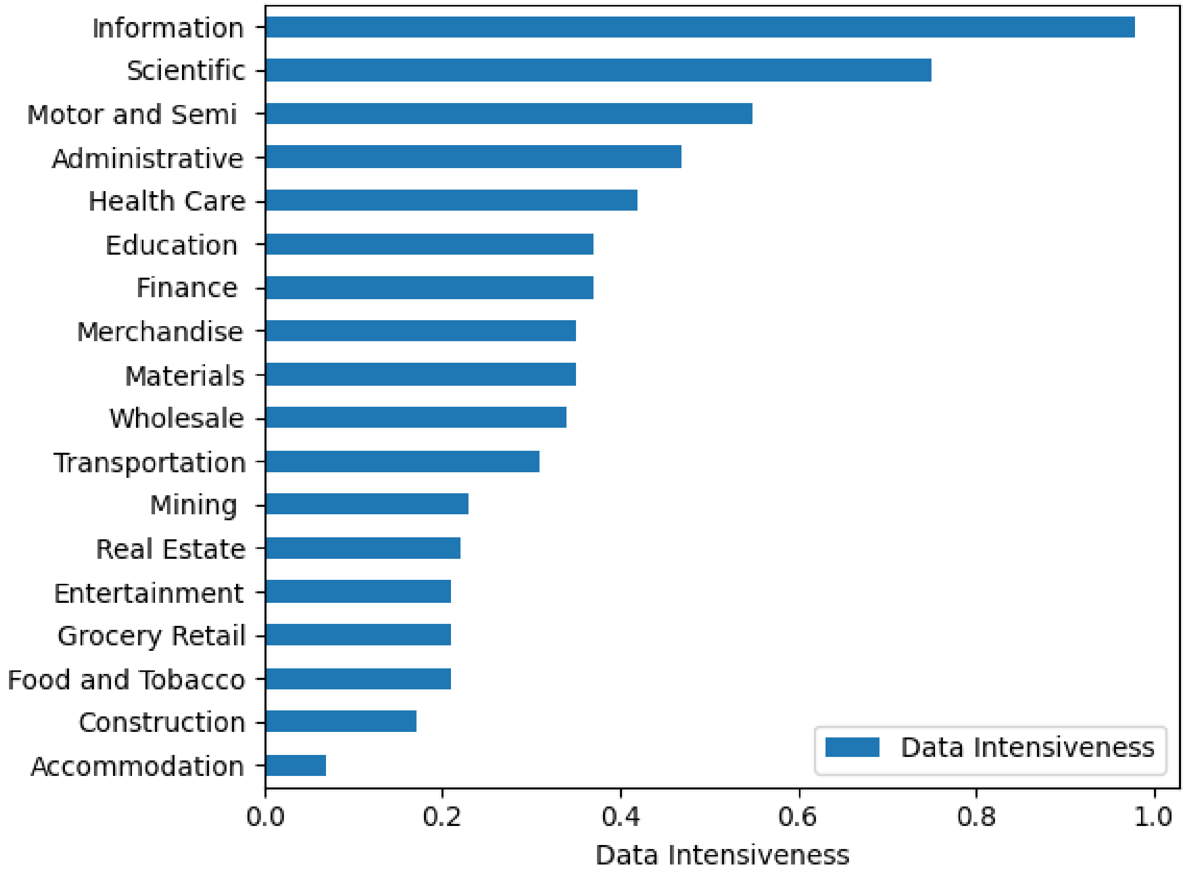
$$\text{AI Talent Demand}_{i,t} = \frac{\text{Job Postings Requiring AI Related Skill}_{i,t}}{\text{Total Job Posting}_{i,t}} \tag{3}$$

$$\text{Data Management Demand}_{i,t} = \frac{\text{Data Management Related Posting}_{i,t}}{\text{Total Job Posting}_{i,t}} \tag{4}$$

I integrate information from these multiple dimensions of data intensiveness, scale them, extract the first principal component from the scaled vectors, and generate a comprehensive measure for data intensiveness.

I compute the pre-2018 average (prior to GDPR implementation) of this data-intensive measure. Firms are ranked based on this comprehensive measure, with the median serving as the cutoff. Firms above the median are classified as data-intensive, while those below the median are categorized as non-data-intensive. Figure 1 displays the industry average of this

Figure 1: Data Intensiveness By Industry



*Source*: Lightcast US Job Posting, USPTO Patent.
*Notes*: The figure shows the unweighted average of the data-intensiveness measure constructed in section 2.2.1. Two factors are taken into consideration while constructing the data-intensiveness measure, skill and technology. For the skill dimension, I measure the hiring demand for AI-related and data-management-related skills from the job posting data by Lightcast. For the technology dimension, I measure the market value and scientific value of the data processing patents from the USPTO patent data. As shown in the figure, the sorting of industries roughly align with our intuition. On the top of the list, there are information and scientific industries, while at the bottom of the list, there are construction and accommodation industries.

data intensiveness measure.

## 2.3 Institutional Background on GDPR

The General Data Protection Regulation (GDPR), adopted by the EU in 2016 as a successor to the 1995 Data Protection Directive, reflects a significant leap in data protection.[13] It came into effect in May 2018, establishing a unified data protection framework across the EU, ensuring that companies, irrespective of their location, adhere to a singular set of rules when operating within the EU.[14] Its structure comprises eleven chapters, addressing various facets including general provisions, rights of the data subject, duties of data controllers or processors, data transfers to third countries, supervisory authorities, and penalties for breaches.

GDPR was forged through extensive deliberations, with a notable milestone on December 17, 2015, when the European Parliament's Committee for Civil Liberties, Justice and Home Affairs committee formally adopted the negotiated text of the GDPR. It's a manifestation of the EU's commitment to safeguarding individuals' fundamental rights in the digital age, while delineating the obligations of data processors, ensuring compliance, and stipulating sanctions for non-compliance.[15]

A core tenet of the GDPR is the broadened scope of personal data, encompassing not just identifiable information but extending to pseudonymous and online identifiers. It has notably enhanced individual rights, such as the "right to be forgotten," and introduced firm obligations like timely data breach notifications. GDPR sets a clear legal framework for data processing, anchored on six legal bases, with consent being one among them. It also outlines provisions for data transfers outside the EU, ensuring such transfers align with the EU's data protection standards.

The enforcement of GDPR is entrusted to the Data Protection Authorities (DPAs) of each EU member state, empowered to impose hefty fines on non-compliant entities. The fines can reach up to the greater of €20 million or 4% of a firm's global annual revenue for severe

[13] Source: edps.europa.eu
[14] Source: commission.europa.eu
[15] Source: consilium.europa.eu

infractions. The regulation also provides a “one-stop-shop” mechanism for multinational firms, simplifying their interactions with EU regulators by allowing them to choose a lead regulator based on their headquarters’ location.

Moreover, GDPR underscores a high standard for consent, with individuals having the right to withdraw consent as easily as they gave it. This regulation significantly raises the legal and financial stakes for firms engaged in data processing, marking a paradigm shift towards a more privacy-centric business environment. The GDPR’s intricate framework has spawned extensive discussions among stakeholders, reflecting its profound implications on the legal landscape and operational modalities of firms within and beyond the EU.

In a nutshell, the GDPR signifies a monumental stride in fortifying data protection, augmenting individual rights, and imposing stringent obligations on data processing entities, thereby fostering a culture of accountability and transparency in the digital realm.

# 3 The Demand for Data by Firms

In Section 2.2.1, we observe that the demand for data scientists and machine learning engineers varies among firms. If data is combined with talent to create knowledge and enhance production technology (Abis and Veldkamp 2020), a negative shock to the data available to firms is likely to impact their production processes.

This section examines how US multinational corporations react to the General Data Protection Regulation (GDPR), a regional privacy regulation. US multinational firms have access to both EU and non-EU markets. GDPR stands as the most comprehensive and stringent privacy regulation worldwide. In the US, there is no federal-level comprehensive privacy law, aside from industry-specific privacy standards such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA).[16] GDPR grants EU consumers greater control over their data and enhances their role as the supplier of data. Prior research (Aridor et al. 2020; Goldberg et al. 2019) demonstrates that following GDPR’s implementation, European households shared less data with firms and made it more difficult for firms to

[16] Several US states have passed state-level privacy laws, including California (effective January 1, 2020), Virginia (effective January 1, 2023), Colorado (effective July 1, 2023), and Utah (effective December 31, 2023).

track them online. As a result, this regulation has limited US firms' access to European data. In this sense, GDPR acts as a data supply shock, enabling us to examine the data demand of US multinational firms.[17]

## 3.1 Cross-Market Business Adjustment

For US multinational firms, the European Union represents a significant foreign market, accounting for a substantial portion of their internationally originated sales. Specifically, when considering US firms with an EU segment, the region contributes to around 15.3% of their total sales. Historically and culturally, consumer preferences in the European market closely align with those in the US domestic market. Consequently, acquiring insights into EU consumers' preferences also enables US technology firms to better understand their domestic customers. Thus, the EU market serves as a crucial data source for US firms.

Table 1: Summary Statistics for Data-Intensive and Other Firms

| | Data-Intensive Firms | | Non-Data-Intensive Firms | | t-test | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Diff | t-stat |
| EU Sales Share (%) | 15.5 | 14.2 | 15.1 | 15.4 | 0.4 | (1.4) |
| EU Sales ($m) | 1,234 | 6,063 | 541 | 1,485 | 693 | (7.2) |
| US Sales ($m) | 3,997 | 15,186 | 1,956 | 5,525 | 2,040 | (8.2) |
| Other Sales (incl. US, $m) | 6,971 | 24,189 | 3,017 | 7,193 | 3,955 | (10.2) |
| Total Assets ($m) | 12,772 | 45,394 | 4,454 | 10,155 | 8,318 | (11.9) |
| Book to Market | 0.50 | 0.30 | 0.63 | 0.31 | -0.12 | (-18.7) |
| Observations | 4,448 | | 4,583 | | 9,031 | |

Table 1 shows the summary statistics of the key variables for the data-intensive and non-data-intensive firms. As we can see, the two groups are very similar in terms of the share of sales coming from the European market, even though data-intensive firms are generally larger and have lower book to market ratio. Therefore, I will control for lagged firm size (total assets) and book to market ratio in the empirical analysis. I will also interact the controls with the time binary variable, which indicates the GDPR's enactment, to control for the differential impact of the regulation that is sorted on size and market valuation.

[17] The regulation was drafted by EU legislators and passed by the European Parliament, making it less likely to be influenced by lobbying efforts from US corporations.

Following the implementation of the GDPR, as documented by the literature, US firms' access to EU consumer data has been limited. In response to this, US multinational firms may strategically shift portions of their businesses away from the European market and towards other regions, particularly the US domestic market, to capitalize on the more lenient regulatory environment. This section tests this hypothesis, examining the potential impact of GDPR on the geographical sales compositions of US multinational firms in the context of data access.

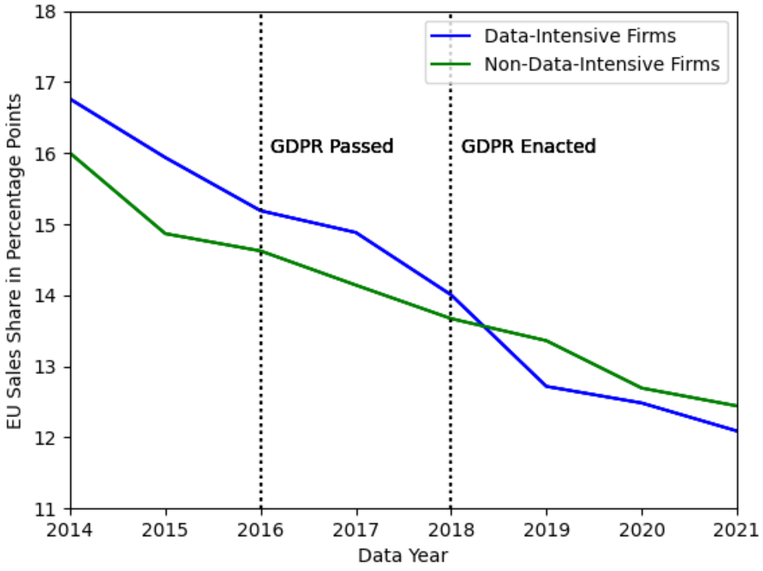


Figure 2: Differential Change in EU Sales Share for Data-Intensive Firms Post-GDPR

Figure 2 shows the fraction of sales (revenue) generated by US firms from the European market. The sample is divided into two groups based on their data intensiveness: the data-intensive ones (above the median measure of data-intensiveness) and the non-data-intensive group (below the median measure of data-intensiveness). The measure of data intensiveness, as defined in Section 2.2.1, serves as a crucial factor in assessing the potential exposure to the data privacy regulation.

The EU sales share for both groups declines, but there is a clear break in 2018, when the data-intensive firms experience a more pronounced decrease. Their EU sales share crosses below the non-data-intensive firms. To empirically estimate this effect after 2018, I employ a difference-in-differences design and estimate the following equation.

$$\frac{\text{EU Sales}_{i,t}}{\text{Total Sales}_{i,t}} \times 100\% = \alpha_{d,t} + \phi_i + \beta_{\text{data}} \cdot \text{GDPR-Effective}_t \times \text{Data-Intensive}_i + \boldsymbol{\gamma} \boldsymbol{X}_{i,t} + \varepsilon_{i,t} \quad (5)$$

The dependent variable is the share of EU sales by US multinational firm $i$ in year $t$; $\alpha_{d,t}$ and $\phi_i$ are industry-by-year and firm fixed effects; $\boldsymbol{X}_{i,t}$ is a vector of time-varying firm-level characteristics, including book to market ratio and log(total assets) at $t-1$. GDPR-Effective$_t$ is a binary variable that equals one if time $t$ is after GDPR's enactment date in 2018. Data-Intensive$_i$ is a binary variable that equals one if firm $i$ is in the data-intensive category as defined in Section 2.2.1.

The results are shown in Table 2, where our primary interest lies in the coefficient $\beta_{\text{data}}$ before the interaction term in equation 5. Panel A reveals that the EU sales share of data-intensive firms decreases by 1.154 to 1.335 percentage points following GDPR's implementation. The mid-point of this range is 1.245 Considering the unconditional mean of EU sales share of data-intensive firms before 2018 stands at 16.25 percentage points, this coefficient corresponds to an 8 (1.245/16.25) percent decline in EU business size. In column (1), (2), and (3), I use a binary measure of data-intensiveness. As we can see, the results are robust to different specifications after controlling for firm fixed effects, year fixed effects, industry-by-year fixed effects, and firm-level time varying characteristics like size (logarithm of one-period lagged assets) and one-period lagged book to market ratio. In column (4), I show the results from an alternative specification where I use the original continuous measure of data-intensiveness. The results help us understand the effects on the intensive margin. As we can see, more data-intensive a firm is, the bigger the decrease in the European business segment.

To understand what drives the decrease in the EU sales share, I look into the sales change by geographical segments. I estimate equation 5 but replace the dependent variable with log(1 + EU Sales) and log(Other Sales). The results are shown in Panel B. EU sales experience a significant decline for US data-intensive firms, and the sales in other geographical segments increase. Cohn et al. (2022) points out that using $\log(1+x)$ in regressions does not have apparent economic meaning and can potentially lead to biased estimates. Therefore, I

Table 2: Cross-Market Business Adjustment

*Notes:* I employ a difference-in-differences design to examine the impact of GDPR on the geographical revenue distribution of US multinational firms. Since most US firms report their geographical revenue compositions at an annual frequency in their 10-K filings, the observations are at the firm-year level. I estimate the following equation

$$Y_{i,t} = \alpha_{d,t} + \phi_i + \beta_{\text{data}} \cdot \text{GDPR-Effective}_t \times \text{Data-Intensive}_i + \boldsymbol{\gamma} \boldsymbol{X}_{i,t} + \varepsilon_{i,t}$$

The dependent variable in Panel A is $Y_{i,t} = \frac{\text{EU Sales}_{i,t}}{\text{Total Sales}_{i,t}} \times 100\%$, which measures the EU sales share for firm $i$ at time $t$ in percentage points; $\phi_i$ and $\alpha_{d,t}$ are firm and industry-by-year fixed effects; $\boldsymbol{X}_{i,t}$ is a vector of time-varying firm-level characteristics, including book to market ratio and log(total assets) at $t-1$. GDPR-Effective$_t$ is a binary variable that equals one if time $t$ is after GDPR's enactment date. Data-Intensive$_i$ is a binary variable that equals one if firm $i$ is in the data-intensive category. I also examine one specification where I include the continuous measure of data-intensiveness, which is defined in section 2.2.1. In Panel B, the dependent variable is log(1 + EU Sales) in column (1) and log(Other Sales) in column (2). The standard errors are clustered at the industry level, and t-statistics are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: EU Sales Share Declines After GDPR

| Dependent Variable: | EU Sales Share (%) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| GDPR Effective × Data-Intensive (binary) | -1.179** | -1.154** | -1.335** | |
| | (-2.598) | (-2.632) | (-2.128) | |
| GDPR Effective × Data-Intensive (value) | | | | -0.785* |
| | | | | (-1.867) |
| Log(Assets) (lagged) | | | 0.480 | 0.494 |
| | | | (0.550) | (0.560) |
| Book to Market (lagged) | | | 0.677 | 0.719 |
| | | | (1.001) | (1.087) |
| Controls × GDPR Effective | No | No | Yes | Yes |
| Year FE | Yes | No | No | No |
| Firm FE | Yes | Yes | Yes | Yes |
| Industry by Year FE | No | Yes | Yes | Yes |
| $R^2$ | 0.767 | 0.776 | 0.779 | 0.779 |
| Observations | 8,540 | 8,521 | 8,275 | 8,275 |

Panel B: EU Sales Decrease While Sales From Other Regions Increase

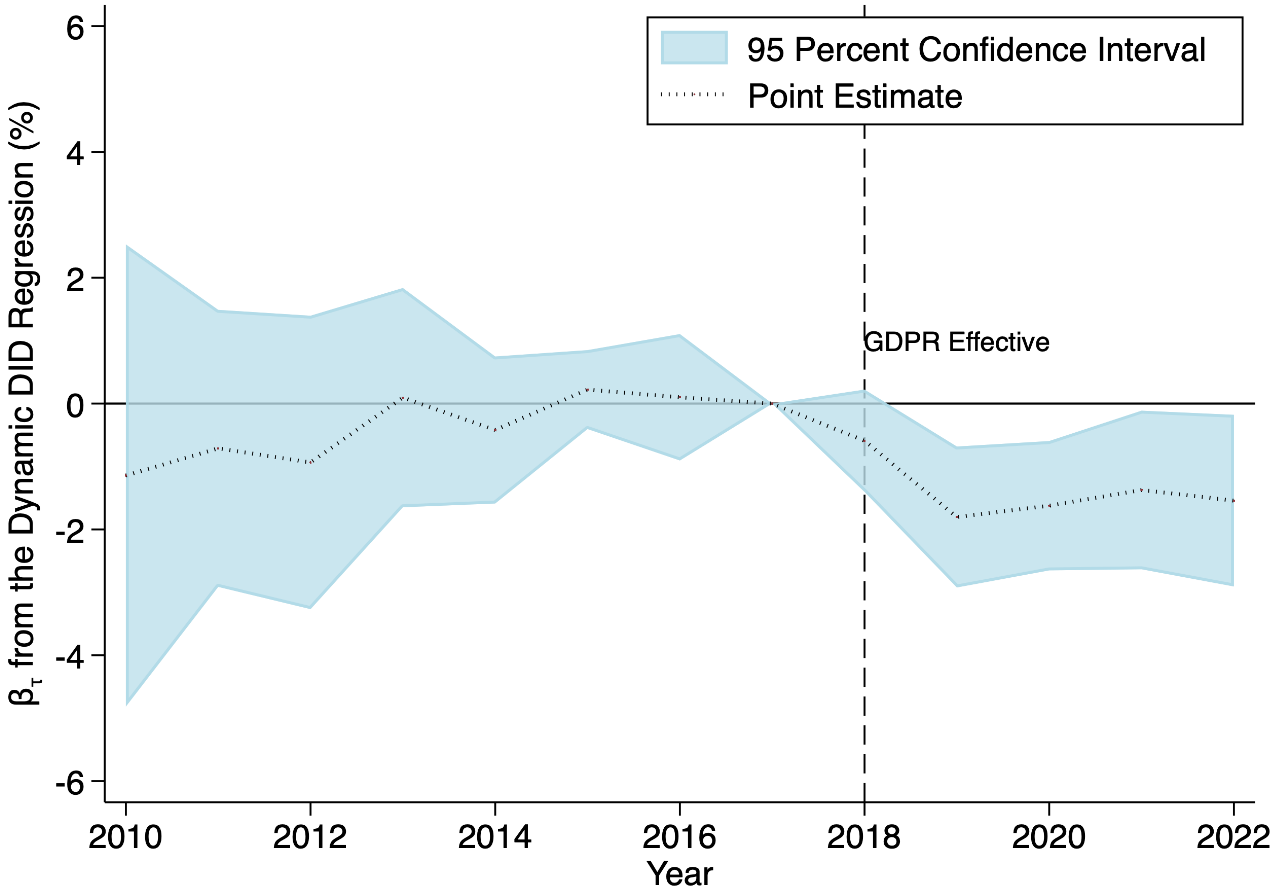
| Dependent Variable: | log(1 + EU Sales) | log(Other Sales) |
|---|---|---|
| | (1) | (2) |
| GDPR Effective × Data-Intensive (binary) | -0.230** | 0.073** |
| | (-2.371) | (2.234) |
| Controls | Yes | Yes |
| Controls × GDPR Effective | Yes | Yes |
| Firm FE | Yes | Yes |
| Industry by Year FE | Yes | Yes |
| $R^2$ | 0.746 | 0.981 |
| Observations | 8,275 | 8,267 |

Figure 3: The Dynamics of Cross-Market Adjustment

*Notes:* I extend the regression in equation 5 to a dynamic difference-in-difference setting so that I can check for the pre-trend and examine when the effect of GDPR kicks in. I run the following regression.

$$Y_{i,t} = \alpha_{d,t} + \phi_i + \sum_{\tau \neq 2018} \beta_\tau \cdot \boldsymbol{I}(t = \tau) \times \text{Data-Intensive}_i + \boldsymbol{\gamma X}_{i,t} + \varepsilon_{i,t}$$

$\alpha_{d,t}$ is the industry-by-year fixed-effect, $\phi_i$ is the firm fixed-effect, and $\boldsymbol{X}_{i,t}$ is a vector of time-varying firm-level characteristics, including book to market ratio and firm size. $\boldsymbol{I}(t = \tau)$ is a binary variable that equals one if year $t = \tau$. Data-Intensive$_i$ is a binary variable that equals one if firm $i$ is in the data-intensive category. I define data intensiveness in section 2.2.1. The standard errors are clustered at the industry level.



estimate a fixed-effect Poisson model in Table A2 and show that the results are qualitatively similar.

The drop in EU sales may be attributed to either a reduction in business size in real terms or a decline in the profitability of the EU segment. To further investigate this question, I examine the profitability of the EU segment and at the firm level for both data-intensive and non-data-intensive firms. The results are displayed in Table A3, where I consider two measures of profitability: gross profit margin (GPM) and operating profit margin (OPM). As evident from the table, the coefficient preceding the interaction term is nearly zero and lacks statistical significance. Consequently, no discernible change in profitability exists between data-intensive firms and non-data-intensive firms, either for the EU segment or at the firm level.

There might be further concerns that what I document here is simply capturing a common trend in the tech sector. Since I have already controlled for industry-by-year fixed effects in Table 2, this should be less of a concern. I perform an additional robustness check in Table A4. I introduce an additional interaction term, involving the time indicator GDPR-Effective$_t$ and a binary variable Tech$_i$, which equals one when a firm $i$'s North American Industry Classification System Code (NAICS) begins with 51. As we can see, the results are robust to this additional control. The impact of GDPR remains statistically and economically similar. This inclusion helps alleviate the concern that the findings in Table 2 arise from a common trend within the tech sector, as opposed to firms' reliance on data.

I further examine the dynamic impact of GDPR on EU sales and extend the regression in equation 5 to a dynamic difference-in-differences framework. This approach allows me to check for pre-trends and investigate when the effect of GDPR begins and how persistent it is. I run the following regression:

$$Y_{i,t} = \alpha_{d,t} + \phi_i + \sum_{\tau \neq 2018} \beta_{\text{data},\tau} \cdot \boldsymbol{I}(t = \tau) \times \text{Data-Intensive}_i + \boldsymbol{\gamma} \boldsymbol{X}_{i,t} + \varepsilon_{i,t} \tag{6}$$

The notations in the above equation are similar to those in equation 5, with the exception that we now include by-period interaction terms and analyze the coefficients $\beta_{\text{data},\tau}$. Figure 3 plots the coefficients from the regression in equation 6, along with a 95 percent confidence band. The figure clearly shows no pre-trend, and the negative impact of GDPR only emerges after 2018, gradually deepening over time.

## 3.2 Complementarity Between Data and Labor

If data is combined with talent to create knowledge and improve production technology, a negative shock to the amount of data available to firms will likely change their production process. In this section, I look into the complementarity and substitutability between data and human capital.

I look into the LinkedIn profile data and examine how GDPR affects the hiring of data analysts, data engineers, software engineers, and machine learning engineers for data-intensive versus non-data-intensive firms after GDPR. Similar to the previous section, I adopt a difference-in-differences design and estimate the following equation.

$$\log(Y_{i,t}) = \alpha_{k,t} + \phi_i + \beta_1 \cdot \text{GDPR-Effective}_t \times \text{Data-Intensive}_i + \boldsymbol{\gamma} \boldsymbol{X}_{i,t} + \varepsilon_{i,t} \tag{7}$$

where $Y_{i,t}$ is the number of employees that are either data analysts, data engineers, software engineers, or machine learning engineers. $\alpha_{k,t}$ is the industry-by-year-quarter fixed-effect, $\phi_i$ is firm fixed-effect, and $\boldsymbol{X}_{i,t}$ captures time-varying firm-level characteristics, including book to market ratio and firm assets. GDPR-Effective$_t$ equals one if time $t$ is after 2018. Data-Intensive$_i$ is a binary variable that equals one if firm $i$ is classified into the data-intensive category.

As depicted in the first two columns of Table 3 Panel A, US data-intensive firms without EU exposure exhibit an uptrend in hiring data-related personnel over time, while this pattern is missing for data-intensive firms with EU exposure. In column (3), I run a triple difference regression and examine the differential impact on firms with substantial EU businesses as opposed to those with none. As seen in the first row, US data-intensive firms with EU exposure experience a negative impact on their demand for data-related talents. The results imply that there exists a complementarity between data and labor. The GDPR, by restricting US firms' access to European data, precipitates lower productivity which in turn leads to a diminished demand for data-related talents.

Table 3: Complementarity Between Data and Labor

*Notes:* I adopt a difference-in-differences design and estimate the following equation.

$$\log(Y_{i,t}) = \alpha_{k,t} + \phi_i + \beta_1 \cdot \text{GDPR-Effective}_t \times \text{Data-Intensive}_i + \boldsymbol{\gamma}\boldsymbol{X}_{i,t} + \varepsilon_{i,t} \tag{8}$$

where $Y_{i,t}$ is the number of employees that are either data analysts, data engineers, software engineers, or machine learning engineers. $\alpha_{k,t}$ is the industry-by-year-quarter fixed-effect, $\phi_i$ is firm fixed-effect, and $\boldsymbol{X}_{i,t}$ captures time-varying firm-level characteristics, including book to market ratio and firm assets. GDPR-Effective$_t$ equals one if time $t$ is after May 2018. Data-Intensive$_i$ is a binary variable that equals one if firm $i$ is classified into the data-intensive category. t-statistics are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Data Processing Staff Grows Slower for EU Exposed Firms

| Dependent Variable: | Log(Data Staff) | | |
|---|---|---|---|
| | EU Exposed | Non-EU-Exposed | Total |
| | (1) | (2) | (3) |
| GDPR Effective × Data-Intensive (binary) × EU Exposed | | | -0.121** |
| | | | (-2.175) |
| GDPR Effective × Data-Intensive (binary) | -0.012 | 0.074** | 0.074* |
| | (-0.317) | (2.424) | (2.028) |
| GDPR Effective × EU Exposed | | | 0.013 |
| | | | (0.349) |
| Controls | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes |
| Firm FE | Yes | Yes | Yes |
| Industry by Year-Quarter FE | Yes | Yes | Yes |
| R$^2$ | 0.928 | 0.919 | 0.923 |
| Observations | 32,963 | 33,908 | 55,976 |

Panel B: Customer Support Staff Grows Slower for EU Exposed Firms

| Dependent Variable: | Log(Customer Support) | | |
|---|---|---|---|
| | EU Exposed | Non-EU-Exposed | Total |
| | (1) | (2) | (3) |
| GDPR Effective × Data-Intensive (binary) × EU Exposed | | | -0.063** |
| | | | (-2.527) |
| GDPR Effective × Data-Intensive (binary) | 0.000 | 0.046*** | 0.054*** |
| | (0.012) | (2.845) | (3.412) |
| GDPR Effective × EU Exposed | | | 0.038* |
| | | | (1.867) |
| Controls | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes |
| Firm FE | Yes | Yes | Yes |
| Industry by Year-Quarter FE | Yes | Yes | Yes |
| R$^2$ | 0.913 | 0.917 | 0.918 |
| Observations | 30,225 | 33,660 | 53,423 |

Table 3 Panel B shows that data-intensive firms with significant EU exposure also witness a slower growth in customer support staff, albeit less pronounced compared to data staff. This is understandable as a policy regulation impacting data supply is more likely to affect employees directly engaged in data collection, cleaning, and analysis. For customer support staff, the effect is indirect. As firm growth is negatively impacted either in the US or EU market, there's a likelihood of firms scaling down their hiring for other support staff as well.

# 4 The Demand for Privacy by Consumers

Privacy protection is not solely about limiting data sharing, but about granting consumers the autonomy to decide the extent of data sharing. Indeed, sharing data often comes with rewards, either pecuniary or otherwise. When it comes to monetary rewards, many are happy to provide phone numbers and email addresses in exchange for discounts. For instance, we might share our contact information to get 10 percent off on an online shopping site. On the non-monetary side, we permit platforms like social media and streaming services to access our personal data and online behavior. This, in turn, allows them to refine and personalize our user experiences. The allure is clear: imagine a TikTok stream impeccably tailored to one's taste or a Netflix dashboard highlighting favorite shows. Yet, the balance between data sharing and privacy is delicate. When companies push boundaries or misuse personal data, consumer welfare might be impaired.

In this section, I explore how consumers weigh the benefits of data sharing against the need for privacy protection. As with section 3, the introduction of GDPR acts as a natural experiment, altering the "supply of privacy." This regulatory change empowers EU consumers with greater data autonomy, letting them decide how much data they share with companies. By analyzing review data from the Google Play Store, I aim to understand how the user experiences of EU and non-EU consumers change post-GDPR.

## 4.1 Google Play Store Review Data

Consumers evaluate Apps along three primary dimensions. Firstly, they look at an app's functionality, placing emphasis on how well it performs its intended tasks and the intuitiveness of its interface. Secondly, they consider the advertisements present within the app, with a keen eye on their relevance and intrusiveness to the user experience. Lastly, any additional offerings such as in-app purchases or subscription options are also taken into account.

In Google Play Store, app users can leave both numerical ratings (on a scale of 1-5) and textual comments. People comment on all three aspects of user experiences as mentioned above. We can visit different versions of the Google Play Store by changing the country and language options. This provides us with a way to differentiate between the comments left by EU and non-EU users. For example, when you use the url, "https://play.google.com/store/apps/details?id=com.instagram.android&hl=en_US&gl=US," you can visit the US version of the Instagram page. The Ratings and Reviews section will show the reviews left by US users. When you change the language and country option, from "&hl=en_US&gl=US" to "&hl=fr&gl=FR", you can visit the French version of the Instagram page. The comments section will only show the reviews from French users. Since GDPR applies to all EEA countries, I gather reviews from the EEA countries in one subsample, while US reviews are compiled separately.

Apps vary in their reliance on consumer data. Drawing from the data safety disclosures discussed in Section 2.1.4 from the Google Play Store, I have classified apps into two categories: those that are heavily data-driven for personalization and those that operate with minimal user information. Users interacting with the former are likely to notice a significant change in their experience if they opt to share less data, while the impact is much more limited for users of the latter group.

Before delving into an in-depth analysis of this review data, I will first present some summary statistics to set the context. To be included in my sample, an app needs to have a valid Data Safety disclosure and have at least ten reviews before and after GDPR came into effect. There are 4,883 apps in the main analysis.

## 4.2 App Ratings

For EU users, choosing to share less data with mobile app providers can have implications on their user experience. This is especially the case for apps that rely heavily on data for personalization. Such apps often seek a diverse range of information to tailor user experi-

Table 4: Reviews Summary Statistics

*Notes:* I show the summary statistics of the review data below, including average daily rating, annual total reviews with advertisement complaints, annual totals reviews that mention in-app purchases or subscriptions.

| | Daily Average Score | | Total Annual Ads Complaints | | Total Annual Purchase Comments | |
|---|---|---|---|---|---|---|
| | EU | US | EU | US | EU | US |
| Obs | 6,457,975 | 7,373,199 | 36,635 | 41,081 | 36,635 | 41,081 |
| Mean | 4.04 | 3.91 | 39.05 | 20.34 | 46.20 | 34.74 |
| SD | 1.08 | 1.19 | 389.93 | 181.72 | 260.74 | 177.32 |
| Min | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 25% | 3.67 | 3.33 | 0.00 | 0.00 | 0.00 | 0.00 |
| 50% | 4.40 | 4.30 | 1.00 | 1.00 | 1.00 | 2.00 |
| 75% | 5.00 | 5.00 | 7.00 | 5.00 | 10.00 | 14.00 |
| Max | 5.00 | 5.00 | 34,723 | 25,527 | 11,774 | 16,571 |

ences. This can encompass basic details like names and email addresses, but may extend to more sensitive data such as political or religious beliefs, sexual orientations, and health metrics. Additional data, like browsing histories and in-app activities, also contribute to this personalization process.

To test for this hypothesis, I employ a difference-in-differences design and study how limited access to data in the European market affects the quality of service provided by mobile apps, measured by the daily average user numeric ratings. I run the following regression.

$$Y_{i,m,t} = \alpha_m + \phi_i + \beta_{\text{service}} \cdot \text{GDPR}_m \times \text{Personalization Collected}_i + \boldsymbol{\gamma}\boldsymbol{X}_{i,m,t} + \varepsilon_{i,m,t} \qquad (9)$$

where $Y_{i,m,t}$ is the daily average rating for app $i$ on day $t$. $\alpha_m$ is the year-month fixed-effect, $\phi_i$ is the app fixed-effect. $\text{GDPR}_m$ is a binary variable that equals one if time $t$ is after GDPR's enactment month, May 2018. $\boldsymbol{X}_{i,m,t}$ is a vector of time-varying app characteristics, including the total number of daily review (winsorized at the one percent level on both ends) and app data sharing practice. Personalization $\text{Collected}_i$ is a binary variable that equals one if app $i$ collects user data for personalization purposes. I analyze the reviews by the EU and US users separately.

The results, presented in Table 5, reveal the impact of GDPR on the quality of digital apps. Columns (1) and (2) indicate that apps collecting user information for personalization

Table 5: Daily Average Rating

*Notes:* I employ a difference-in-differences design and study how limited access to data in the European market affects the quality of service provided by mobile apps, measured by the daily average user numeric ratings. The observations of the sample used in this table are at the app-day level. In columns (1) and (2), I run the following regression.

$$Y_{i,m,t} = \alpha_m + \phi_i + \beta_{\text{service}} \cdot \text{GDPR}_m \times \text{Personalization Collected}_i + \boldsymbol{\gamma X}_{i,m,t} + \varepsilon_{i,m,t}$$

where $Y_{i,m,t}$ is the daily average rating for app $i$ on day $t$. $\alpha_m$ is the year-month fixed-effect, $\phi_i$ is the app fixed-effect. $\text{GDPR}_m$ is a binary variable that equals one if time $t$ is after GDPR's enactment month, May 2018. Personalization $\text{Collected}_i$ is a binary variable that equals one if app $i$ collects user data for personalization purposes. In column (3), I run a triple difference regression.

$$\begin{aligned} Y_{i,m,k,t} = &\alpha_m + \phi_i + \psi_k + \beta_1^* \cdot \text{GDPR}_m \times \text{Personalization Collected}_i \times \text{EU}_k \\ &+ \beta_2 \cdot \text{GDPR}_m \times \text{Personalization Collected}_i + \beta_3 \cdot \text{GDPR}_m \times \text{EU}_k \\ &+ \beta_4 \cdot \text{Personalization Collected}_i \times \text{EU}_k + \boldsymbol{\gamma X}_{i,m,k,t} + \varepsilon_{i,m,k,t} \end{aligned}$$

where $Y_{i,m,k,t}$ is the average daily rating by users from region $k$ for app $i$ on day $t$. $\psi_k$ is the region (US or EU) fixed-effect. $\text{EU}_k$ is an indicator variable that equals one if the reviews come from the EU users. The coefficient $\beta_1^*$ before the triple interaction term captures differential change in app quality between the EU and US users after GDPR for apps collecting personalization information. t-statistics are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

| Dependent Variable: | EU Users | US Users | All |
|---|---|---|---|
| Daily Average Rating | (1) | (2) | (3) |
| GDPR Effective × Personalization Collected | -0.082*** | -0.017*** | -0.014*** |
| | (-43.505) | (-8.600) | (-7.443) |
| GDPR Effective × Personalization Collected × EU | | | -0.071*** |
| | | | (-29.978) |
| GDPR Effective × EU | | | 0.044*** |
| | | | (26.039) |
| Personalization Collected × EU | | | 0.073*** |
| | | | (38.290) |
| Controls | Yes | Yes | Yes |
| Year-Month FE | Yes | Yes | Yes |
| App FE | Yes | Yes | Yes |
| Region FE | No | No | Yes |
| $\text{R}^2$ | 0.246 | 0.242 | 0.235 |
| Observations | 5,757,960 | 6,548,836 | 12,306,796 |

experience a decrease in user ratings in both the US and EU regions. However, this decline is significantly more pronounced for EU users compared to their US counterparts. These findings suggest that post-GDPR, while EU users may opt to share less data, this choice correlates with a noticeable drop in the quality of digital services, as reflected by app ratings. The data in column (2) also hint at potential spillover effects impacting the US user base.[18] I will discuss the possible explanations for the spillover effects in the theory section. User ratings are typically concentrated around 4.0, with the inter-quartile range for EU users being 3.67-5.0. Thus, a 0.08 decrease in user ratings signifies a 6 percent reduction in the inter-quartile range. In contrast, for US users, a 0.017 drop in ratings equates to a 1 percent decrease in the inter-quartile range.

To further examine the differential impact on the two user groups, in column (3), I run a triple difference regression.

$$\begin{aligned} Y_{i,m,k,t} = &\alpha_m + \phi_i + \psi_k + \beta_1^* \cdot \text{GDPR}_m \times \text{Personalization Collected}_i \times \text{EU}_k \\ &+ \beta_2 \cdot \text{GDPR}_m \times \text{Personalization Collected}_i + \beta_3 \cdot \text{GDPR}_m \times \text{EU}_k \\ &+ \beta_4 \cdot \text{Personalization Collected}_i \times \text{EU}_k + \boldsymbol{\gamma} \boldsymbol{X}_{i,m,k,t} + \varepsilon_{i,m,k,t} \end{aligned}$$

where $Y_{i,m,k,t}$ is the average daily rating by users from region $k$ for app $i$ on day $t$. $\psi_k$ is the region (US or EU) fixed effect. $\text{EU}_k$ is an indicator variable that equals one if the reviews come from the EU users. $\boldsymbol{X}_{i,m,k,t}$ is a vector of time-varying app characteristics, including the total number of daily review. The coefficient $\beta_1^*$ before the triple interaction term captures the differential change in user quality between the EU and US users after GDPR for apps collecting personalization information. As we can see from the results, the negative impact of GDPR on app service quality is much larger for EU users, and it is statistically significant.

## 4.3 Advertisement Complaints

As of March 2023, 97% of the apps on Google Play Store, and 94.5% of the apps on Apple App Store are free to download.[19] Moreover, for most of the apps, we can enjoy basic

[18] Since I control for app fixed effects, the comparison essentially lies between US and EU users using the same app.

[19] https://www.statista.com/statistics/263797/number-of-applications-for-mobile-phones/

functions without paying a penny. Then how do app developers make money?

Of course, app owners are not running charities. There are multiple ways for them to monetize their users, including advertisements and in-app purchases and subscriptions. When we use an app, we devote our attention to the content displayed in the user interface. Like the television industry, user or viewer attention can be exploited for advertising. App owners can incorporate and auction off advertisement slots in their apps. Common types of mobile advertisements include banners, pop-up windows, native ads, and rewarded videos. As advertisement publishers, app owners work with advertisement networks and delegate the advertisement auctions to them.

However, most of people do not like advertisements and find them extremely annoying. The average numeric rating is 3.0 when mobile app users mention advertisement related keywords in their reviews, compared to 4.0 for all types of reviews.

In-app advertisement is one important income source for most free apps. Oftentimes, we are also asked to register for an account with our email addresses or phone numbers. And for a lot of social media apps, we willingly provide personal information like names, genders, birthdays, home addresses, places of births, etc. The list goes on, and sometimes we would be surprised at how much we have shared with the internet. When we use these apps, we also reveal our own preferences through app activities. These are all valuable data that can be used by these apps to build a digital profile of us. Moreover, the data we shared with different apps can also be linked together using either our device unique identifiers or other individual identifiers like email addresses or phone numbers. All these valuable information can further be used for targeted advertising, which shows different advertisements to people with different preferences.

When users choose to share less information with apps and they choose to block third-party advertisement tracking, we might see a decrease in advertisement effectiveness. I define apps that engage in target advertising as the ones that collect personal information for advertising and marketing purposes and apps that collect device specific IDs for cross-app tracking. I do not have micro-level data on the number of advertisements each app puts into their user interface, but I do observe the comments that are related to complaints about advertisements.

## Table 6: Annual Advertisement Complaints

*Notes:* I employ a difference-in-differences design and study how limited access to data in the European market affects the number of advertisements complaints by mobile app users. The observations in this analysis are at the app-year level. In columns (1) and (2), I run the following regression:

$$\ln(1 + Y_{i,t}) = \alpha_t + \phi_i + \beta_1 \cdot \text{GDPR}_t \times \text{Target Advertising}_i + \varepsilon_{i,t}$$

where $Y_{i,t}$ is the total number of advertisement related complaints for app $i$ in year $t$. I take the logarithm of the annual number of complaints to the natural base. $\alpha_t$ is the year fixed effect. $\phi_i$ is the app fixed effect. $\text{GDPR}_t$ is a binary variable that equals one if time $t$ is after GDPR's enactment year, 2018. $\text{Target Advertising}_i$ is a binary variable that equals one if app $i$ collects user data for targeted advertising purposes. I analyze reviews left by the EU and US users separately. In column (3), I run a triple difference regression:

$$\begin{aligned} \ln(1 + Y_{i,k,t}) = &\alpha_t + \phi_i + \psi_k + \beta_1^* \cdot \text{GDPR}_t \times \text{Target Advertising}_i \times \text{EU}_k \\ &+ \beta_2 \cdot \text{GDPR}_t \times \text{Target Advertising}_i + \beta_3 \cdot \text{GDPR}_t \times \text{EU}_k \\ &+ \beta_4 \cdot \text{Target Advertising}_i \times \text{EU}_k + \varepsilon_{i,k,t} \end{aligned}$$

where $Y_{i,k,t}$ is the total number of advertisement related complaints for app $i$ in year $t$. $\psi_k$ is the region (US or EU) fixed effect. $\text{EU}_k$ is an indicator variable that equals one if the reviews come from the EU users. The coefficient $\beta_1^*$ before the triple interaction term captures the differential change in advertisement intensity between the EU and US mobile app markets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

| Dependent Variable: | EU Users | US Users | All |
|---|---|---|---|
| ln(1+Annual # of Advertising Complaints) | (1) | (2) | (3) |
| GDPR Effective × Target Advertising × EU | | | 0.098*** |
| | | | (3.308) |
| GDPR Effective × Target Advertising | 0.259*** | 0.189*** | 0.179*** |
| | (7.942) | (7.302) | (6.719) |
| GDPR Effective × EU | | | -0.016 |
| | | | (-0.923) |
| Target Advertising × EU | | | -0.206*** |
| | | | (-4.950) |
| Year FE | Yes | Yes | Yes |
| App FE | Yes | Yes | Yes |
| Region FE | No | No | Yes |
| $R^2$ | 0.804 | 0.800 | 0.703 |
| Observations | 33,328 | 37,247 | 70,575 |

If app owners try to make up for the loss in advertisement effectiveness, they might put in more advertisements. If the number of user complaints about advertisements are proportional to the number of ads being put into these apps, we will very likely see an increase in the advertisement related complaints. Table 6 confirms this hypothesis, we see a significant increase in advertising related complaints for both EU and US consumers. However, the increase is much larger for EU users. The results translate into a 10 percent increase in advertisement intensity for EU mobile apps after GDPR came into effect.

Another way apps can generate revenue is through in-app purchases and subscriptions. I do not have data on in-app purchases and subscriptions, but users often write reviews about these type of expenses. I identify these type of comments through textual review data. The analysis results are shown in Table A5. We can see that there is a much larger increase in purchase related comments among EU users than their US counterparts. This implies that mobile apps are switching to other sources of revenue after data privacy regulations render advertisement less effective.

In Table A6, we can see that the results in Table 6 and Table A5 are not driven by larger increase in active user base in the European market. The growth in total number of reviews are very similar across the two markets.

# 5 Model

This study endeavors to bridge the gap between the empirical evidence and theoretical work on data and privacy. I provide a framework to quantitatively measure the value of data, privacy preferences, and the welfare implications of privacy regulations. To accomplish this, I develop a two-economy equilibrium model to better understand the strategic choices made by US multinational firms when facing regional privacy regulations like the GDPR in the European Union.

The model's structure is illustrated in Figure A2. US firms provide goods and services to both European customers and US customers (or, more accurately, customers from the rest of the world). Data is a byproduct of economic activities, and it serves two roles. First, data is used in the production process to increase productivity. We can think of firms using data in the R&D process to create new technology. Second, firms collect and analyze consumers' data to learn about their preferences. The more data firms have, the better they can tailor their products to consumers' preferences.

While consumers enjoy the advantages of personalized recommendations and enhanced service quality, they have concerns about sharing personal data with firms. Privacy concerns may stem from the psychological costs or social stigma of disclosing excessive personal information, as well as from predatory advertising or pricing tactics employed by firms.

## 5.1 Firms

**Digital Firms** There is a continuum of digital firms, denoted by $j \in [0,1]$. Digital firm $j$ combines technology $A_{jt}$ and labor $L_{jt}$ to produce products in each period $t$. Data accumulated at the end of period $t-1$ determines the level of productivity $A_{jt}$ in period $t$. We can think of firms using data to develop new technology, create new products, train algorithms, or streamline production and sales processes. At time $t$, firm $j$'s production function is

$$Y_{jt} = D_{j,t-1}^{\eta} L_{jt}^{1-\eta} = A_{jt} L_{jt}^{1-\eta}, \quad \eta \in (0,1) \tag{10}$$

Data is a byproduct of economic activities. One unit of consumption generates one unit of data. When $Y_{jt}$ units of goods are consumed, equal amount of data is generated. $x_{jt}Y_{jt}$ of the generated data is collected by firm $j$, which will be used to increase productivity in the next period. $x_{jt} \in [0,1]$ is a control variable that may be determined by either firm $j$ or its customers, depending on data sharing or privacy regulations. Data depreciates in each period and accumulates from period to period.

$$D_{jt} = (1-\kappa)\underbrace{D_{j,t-1}}_{\text{old data}} + \underbrace{x_{jt}Y_{jt}}_{\text{new data}} \tag{11}$$

where $\kappa$ is the data depreciation rate. Data accumulated at the end of period $t$ will then be used in the next period $t+1$.

**Non-digital Firms** Since my main focus is on the digital sector, I assume there is one large non-digital product producer that produces non-digital products locally, and non-digital products serve as the numeraire in the economy. The consumption of non-digital products does not generate data, which is the main difference from digital products.

## 5.2 Households

At time t, a continuum of households, denoted by $i \in [0, 1]$, exists within each economy $k$, where $k \in \{\text{US}, \text{EU}\}$. Household $i$ chooses between two consumption types: digital goods and non-digital goods. Examples of digital goods include social media platforms, streaming services, online shopping, and any other digital services that will potentially document your digital footprints. In contrast, non-digital goods represent other outside options, such as purchasing groceries at a local market, attending concerts, or engaging in offline entertainment services. Households select from a wide array of digital products, with $j \in [0, 1]$. The main difference between digital products and non-digital products is that digital firms can collect consumer data to personalize their experiences.

**Data and Product Preferences** We denote household $i$'s preferences for company $j$'s product as $\boldsymbol{\theta_{ijt}}$, and it consists of two components: a common component $\theta_{jt}$ and an idiosyncratic component $\theta_{ijt}$.

$$\boldsymbol{\theta_{ijt}} = (\theta_{jt}, \theta_{ijt})^T \tag{12}$$

We assume the preferences for companies' products are transitory and independently drawn from normal distributions.

$$\theta_{jt} \sim \mathcal{N}(\theta_j, \sigma_j^2), \quad \theta_{ijt} \sim \mathcal{N}(\theta_{ij}, \sigma_{ij}^2) \tag{13}$$

These preferences are not directly observable by firms, but firms can learn about these preferences from the data shared by consumers. In the spirit of Farboodi and Veldkamp (2021), data are information or signals about consumers' preferences. Suppose consumer $i$ shares $x_{ijt}$ units of data, $\{s_{ijt,l}\}$, with firm $j$. Each unit of data, $s_{ijt,l}$, lends firm $j$ insights about consumer $i$'s idiosyncratic preferences,

$$s_{ijt,l} = \theta_{ijt} + \varepsilon_{ijt,l}, \quad \varepsilon_{ijt,l} \sim \mathcal{N}(0, s_{ij}^2) \tag{14}$$

Firm $j$ also uses the information collected from all consumers to learn about their general taste, $\theta_{jt}$. There are $\bar{x}_{jt} = \int_i x_{ijt} di$ units of data in total, and each data point is a signal $s_{jt,l}$ about the common taste $\theta_{jt}$.

$$s_{jt,l} = \theta_{jt} + \varepsilon_{jt,l}, \quad \varepsilon_{jt,l} \sim \mathcal{N}(0, s_j^2) \tag{15}$$

Given the information, firms $j$ chooses product features $\boldsymbol{a_{ijt}}$. The squared distance between the chosen features and consumer preferences is given by

$$(a_{jt} - \theta_{jt})^2 \text{ and } (a_{ijt} - \theta_{ijt})^2 \tag{16}$$

Firm $j$ chooses the best action to minimize the quadratic loss. The best action is the conditional mean of the common component and the idiosyncratic component.

$$\mathbb{E}\left[\theta_{jt} | \{s_{jt,l}\}\right] \text{ and } \mathbb{E}\left[\theta_{ijt} | \{s_{ijt,l}\}\right] \tag{17}$$

Suppose the quality of the product is inversely related to the conditional expectation of the squared distance, and I choose the following formulation.

$$q_{ijt}(\boldsymbol{a_{ijt}}) = \gamma \mathbb{E}\left[\left(\mathbb{E}\left[\theta_{jt} | \{s_{jt,l}\}\right] - \theta_{jt}\right)^2 | \{s_{jt,l}\}\right]^{-1} + (1-\gamma) \mathbb{E}\left[\left(\mathbb{E}\left[\theta_{ijt} | \{s_{ijt,l}\}\right] - \theta_{ijt}\right)^2 | \{s_{ijt,l}\}\right]^{-1} \tag{18}$$

where $\gamma$ determines the contribution of common preferences relative to idiosyncratic preferences. Assuming the prior contains is uninformative, then we can rewrite the quality expression as

$$q_{ijt} = \gamma \bar{x}_{jt} + (1-\gamma) x_{ijt} \tag{19}$$

where $\bar{x}_{jt} = \int_i x_{ijt} di$ is the aggregate level of data sharing among firm $j$'s customer base, and $x_{ijt}$ is the amount of data shared by consumer $i$ with firm $j$.[20]

**Privacy Preferences** Firms can potentially track consumers across platforms and learn about every aspect of their preferences beyond the reasonable use of data. Excessive data collection by firms can lead to predatory advertising and pricing practices. There is also a social cost associated with the revelation of sensitive personal information, e.g. medical records, marital status, sexual orientation, religious beliefs, and immigration status, especially for people from disadvantaged socioeconomic backgrounds. Sometimes, firms do not even need to directly obtain such information because machine learning algorithms can make inferences from other observable personal traits, including but not limited to searching, browsing, and shopping histories. The probability of a privacy breach event is positively related to the amount data being collected from consumers. We can think of the arrival of a privacy breach event as a Poisson shock, and the probability of arrival is $x_{ijt}^2$. The number of privacy breach events follow a Poisson distribution

$$\Pr(N_{ijt} = k) = \frac{\left(x_{ijt}^2\right)^k e^{-x_{ijt}^2}}{k!} \tag{20}$$

When such events occur, it causes a damage of $\delta_i$ to household $i$'s digital experience. The expected damage from privacy breach while sharing $x_{ijt}$ units of data is

$$\mathbb{E}\left[\delta_i N_t\right] = \delta_i x_{ijt}^2 \tag{21}$$

Within the context of this paper, I do not distinguish among the various mechanisms that underlie consumers' privacy preferences.

**Consumer Heterogeneity in Privacy Preferences** I assume that a fraction $\lambda$ of consumers are privacy-conscious, and they incur a cost of $\delta_i x_{ijt}^2$ when they share $x_{ijt}$ amount of data. Therefore, they value the option to share less data. A fraction $1 - \lambda$ of consumers are

[20] Here $\bar{x}_{jt}$ is shown equal-weighted for demonstration purposes. Later, when we discuss the firm-level product quality, it is weighted by consumption units.

non-privacy-conscious, and their digital experiences exclusively depend on the data sharing to improve product customization. For household $i$, her digital experience, depending on its type $z_i$, is given by

$$\gamma \bar{x}_{jt} + (1-\gamma) x_{ijt} - \mathbb{1}_{\{z_i=1\}} \delta_i x_{ijt}^2 \tag{22}$$

where $z_i = 1$ when household $i$ is privacy conscious.

**Digital and Non-Digital Consumption** Household $i$'s digital consumption is modified by the personalized product quality and privacy concerns. We assume

$$u_{ijt,\text{digital}} = (\overbrace{q_{ijt}}^{\text{quality}} - \overbrace{\mathbb{1}_{\{z_i=1\}} \delta_i x_{ijt}^2}^{\text{privacy concern}}) \ln c_{ijt} = \left(\gamma \bar{x}_{jt} + (1-\gamma) x_{ijt} - \mathbb{1}_{\{z_i=1\}} \delta_i x_{ijt}^2\right) \ln c_{ijt} \tag{23}$$

In contrast, non-digital consumption neither generates data nor leads to privacy concerns.

$$u_{it,\text{non-digital}} = \ln c_{it}^{\text{nd}} \tag{24}$$

## 5.3 Multinational Setting

Now we extend the baseline setting to a two-economy equilibrium model. US firms provide goods and services to both European customers and US customers. EU customers also have access to local EU digital products.

**US Multinational Digital Firms** For US multinational digital firms, they decide the total volume of production, the resources (human capital) allocated to each geographical segment ($L_{jt,\text{us}}^{\text{us}}$ and $L_{jt,\text{eu}}^{\text{us}}$), and the data collection practice in each market ($x_{jt,\text{us}}^{\text{us}}$ and $x_{jt,\text{eu}}^{\text{us}}$).[21] Their optimization problems are given by

$$\max_{\{L_{jt,\text{us}}^{\text{us}}\},\{L_{jt,\text{eu}}^{\text{us}}\},\{x_{jt,\text{us}}^{\text{us}}\},\{x_{jt,\text{eu}}^{\text{us}}\}} V_j^{\text{us}}(D_{j0}^{\text{us}}) = \sum_{t=1}^{\infty} \left(p_{jt,\text{us}}^{\text{us}} Y_{jt,\text{us}}^{\text{us}} + p_{jt,\text{eu}}^{\text{us}} Y_{jt,\text{eu}}^{\text{us}} - w_{t,\text{us}} L_{jt,\text{us}}^{\text{us}} - w_{t,\text{eu}} L_{jt,\text{eu}}^{\text{us}}\right) \tag{25}$$

[21] For clarification, the superscript "us" denotes the location/headquarter of the company, while the subscript "us" denotes the location of the consumption. I first lay out the framework for the pre-GDPR regime first. That is why I assign the data sharing choices $x_{jt,\text{us}}^{\text{us}}$ and $x_{jt,\text{eu}}^{\text{us}}$ to firms.

subject to

$$\begin{aligned}
Y^{\mathrm{us}}_{jt,\mathrm{us}} &= \left(D^{\mathrm{us}}_{j,t-1}\right)^{\eta}\left(L^{\mathrm{us}}_{jt,\mathrm{us}}\right)^{1-\eta}\\
Y^{\mathrm{us}}_{jt,\mathrm{eu}} &= \left(D^{\mathrm{us}}_{j,t-1}\right)^{\eta}\left(L^{\mathrm{us}}_{jt,\mathrm{eu}}\right)^{1-\eta}\\
D^{\mathrm{us}}_{jt} &=(1-\kappa_{\mathrm{us}})D^{\mathrm{us}}_{j,t-1} + x^{\mathrm{us}}_{jt,,\mathrm{us}}Y^{\mathrm{us}}_{jt,\mathrm{us}} + x^{\mathrm{us}}_{jt,\mathrm{eu}}Y^{\mathrm{us}}_{jt,\mathrm{eu}}\\
x^{\mathrm{us}}_{jt,\mathrm{us}} &\in[0,1], \quad x^{\mathrm{us}}_{jt,,\mathrm{eu}} \in [0,1]
\end{aligned}$$

**EU Local Digital Firms** For EU local digital firms, they only serve the EU local market, and their optimization problems are given by

$$\max_{\{L^{\mathrm{eu}}_{jt,\mathrm{eu}}\},\{x^{\mathrm{eu}}_{jt,\mathrm{eu}}\}} V^{\mathrm{eu}}_{j}(D^{\mathrm{eu}}_{j0}) = \sum_{t=1}^{\infty}\left(p^{\mathrm{eu}}_{jt,\mathrm{eu}}Y^{\mathrm{eu}}_{jt,\mathrm{eu}} - w_{t,\mathrm{eu}}L^{\mathrm{eu}}_{jt,\mathrm{eu}}\right) \tag{26}$$

subject to

$$\begin{aligned}
Y^{\mathrm{eu}}_{jt,\mathrm{eu}} &= \left(D^{\mathrm{eu}}_{j,t-1}\right)^{\eta}\left(L^{\mathrm{eu}}_{jt,\mathrm{eu}}\right)^{1-\eta}\\
D^{\mathrm{eu}}_{jt} &=(1-\kappa_{\mathrm{eu}})D^{\mathrm{eu}}_{j,t-1} + x^{\mathrm{eu}}_{jt,\mathrm{eu}}Y^{\mathrm{eu}}_{jt,\mathrm{eu}}\\
x^{\mathrm{eu}}_{jt,\mathrm{eu}} &\in[0,1]
\end{aligned}$$

**US Households** For the US households, their optimization problem is given by

$$\begin{aligned}
\max_{\{c^{\mathrm{us}}_{ijt,\mathrm{us}}\},\{c^{\mathrm{nd}}_{it,\mathrm{us}}\}} u_{it,\mathrm{us}} =&K\underbrace{\int_0^1\left(\overbrace{\gamma\bar{x}^{\mathrm{us}}_{jt} + (1-\gamma)x^{\mathrm{us}}_{ijt,\mathrm{us}}}^{\text{benefits of data sharing}} - \overbrace{\mathbb{1}_{\{z_i=1\}}\delta_i\left(x^{\mathrm{us}}_{ijt,\mathrm{us}}\right)^2}^{\text{privacy concerns}}\right)\ln c^{\mathrm{us}}_{ijt,\mathrm{us}}\,dj}_{\text{consumption of US digital products}}\\
&+(1-K)\underbrace{\ln c^{\mathrm{nd}}_{it,\mathrm{us}}}_{\text{conusmption of non-digital products}}
\end{aligned} \tag{27}$$

subject to the budget constraint

$$\underbrace{\int_0^1 p^{\mathrm{us}}_{jt,\mathrm{us}}c^{\mathrm{us}}_{ijt,\mathrm{us}}dj}_{\text{payment for US digital products}} + \underbrace{c^{\mathrm{nd}}_{it,\mathrm{us}}}_{\text{payment for non-digital products}} \leq \int_0^1 \Pi^{\mathrm{us}}_{ijt,\mathrm{us}}dj + \Pi^{\mathrm{nd}}_{it,\mathrm{us}} + w_{t,\mathrm{us}}l_{it,\mathrm{us}} \tag{28}$$

They choose their consumption basket of digital products and non-digital products. Parameter $K$ determines the relative importance of digital consumption to non-digital consumption. $\bar{x}^{\text{us}}_{jt}$ is the average level of data sharing (weighted by consumption) among all consumers of US digital firm $j$, including both US and EU users. $z_i = 1$ if household $i$ is privacy conscious. The total budget for household $i$ in the US is given by

$$\Pi_{it,\text{us}} = \underbrace{\int_0^1 \Pi^{\text{us}}_{ijt,\text{us}} dj}_{\text{profit distribution from US digital firms}} + \underbrace{\Pi^{\text{nd}}_{it,\text{us}}}_{\text{profit distribution from US non-digital firms}} + \underbrace{w_{t,\text{us}} l_{it,\text{us}}}_{\text{labor income}} \tag{29}$$

which consists of the profits distributed from the digital and non-digital firms and labor income.

**EU Households** For the European households, their optimization problem is given by

$$\begin{aligned}\max_{\{c^{\text{us}}_{ijt,\text{eu}}\},\{c^{\text{eu}}_{ij't,\text{eu}}\},\{c^{\text{nd}}_{it,\text{eu}}\}} u_{it,\text{eu}} =& K\beta \underbrace{\int_0^1 \left( \overbrace{\gamma \bar{x}^{\text{us}}_{jt} + (1-\gamma) x^{\text{us}}_{ijt,\text{eu}}}^{\text{benefits of data sharing}} - \overbrace{\mathbb{1}_{\{z_i=1\}} \delta_i \left(x^{\text{us}}_{ijt,\text{eu}}\right)^2}^{\text{privacy concerns}} \right) \ln c^{\text{us}}_{ijt,\text{eu}} \, dj}_{\text{consumption of US digital products}} \\ &+ K(1-\beta) \underbrace{\int_0^1 \left( \gamma \bar{x}^{\text{eu}}_{j't} + (1-\gamma) x^{\text{eu}}_{ij't,\text{eu}} - \mathbb{1}_{\{z_i=1\}} \delta_i \left(x^{\text{eu}}_{ij't,\text{eu}}\right)^2 \right) \ln c^{\text{eu}}_{ij't,\text{eu}} \, dj'}_{\text{consumption of EU local digital products}} \\ &+ (1-K) \underbrace{\ln c^{\text{nd}}_{it,\text{eu}}}_{\text{conusmption of non-digital products}} \end{aligned} \tag{30}$$

subject to the budget constraint

$$\begin{aligned} &\underbrace{\int_0^1 p^{\text{us}}_{jt,\text{eu}} c^{\text{us}}_{ijt,\text{eu}} dj}_{\text{payment for US digital products}} + \underbrace{\int_0^1 p^{\text{eu}}_{j't,\text{eu}} c^{\text{eu}}_{ij't,\text{eu}} dj'}_{\text{payment for EU local digital products}} + \underbrace{c^{\text{nd}}_{it,\text{eu}}}_{\text{payment for non-digital products}} \\ &\leq \int_0^1 \Pi^{\text{us}}_{ijt,\text{eu}} dj + \int_0^1 \Pi^{\text{eu}}_{ij't,\text{eu}} dj' + \Pi^{\text{nd}}_{it,\text{eu}} + w_{t,\text{eu}} l_{it,\text{eu}} \end{aligned} \tag{31}$$

$\bar{x}^{\text{us}}_{jt}$ is the average level of data sharing (weighted by consumption) among all consumers of multinational US digital firm $j$. Because of this setting, the data sharing level of EU consumers can affect the digital utility of US consumers, and vice versa. $\bar{x}^{\text{eu}}_{j't}$ is the average level of data sharing (weighted by consumption) among all consumers of EU firm $j'$. EU households' consumption basket consists of US digital products, EU digital products, and non-digital products. Parameter $\beta$ determines the preferences of EU households for US digital products relative to EU digital products. Households are hand-to-mouth, and they neither save nor make inter-temporal consumption decisions. This is a reasonable assumption given my main focus is on consumers' digital consumption and data sharing choices. The total budget for household $i$ in the EU is given by

$$\Pi_{it,\text{eu}} = \underbrace{\int_0^1 \Pi^{\text{us}}_{ijt,\text{eu}} dj}_{\text{profit distribution from the EU subsidiary of US digital firms}} + \underbrace{\int_0^1 \Pi^{\text{eu}}_{ij't,\text{eu}} dj'}_{\text{profit distribution from EU digital firms}}$$
$$+ \underbrace{\Pi^{\text{nd}}_{it,\text{eu}}}_{\text{profit distribution from EU non-digital firms}} + \underbrace{w_{t,\text{eu}} l_{it,\text{eu}}}_{\text{labor income}} \tag{32}$$

## 5.4 Equilibrium Definition

An equilibrium consists of quantities and prices such that

1. In the pre-GDPR regime, US Multinational digital firms choose a sequence of production decisions $\{L^{\text{us}}_{jt,\text{us}}, L^{\text{us}}_{jt,\text{eu}}\}$ and data collection decisions $\{x^{\text{us}}_{jt,\text{us}}, x^{\text{us}}_{jt,\text{eu}}\}$ to maximize the discounted value of all future profits.

$$\sum_{t=1}^{\infty} \left( p^{\text{us}}_{jt,\text{us}} Y^{\text{us}}_{jt,\text{us}} + p^{\text{us}}_{jt,\text{eu}} Y^{\text{us}}_{jt,\text{eu}} - w_{t,\text{us}} L^{\text{us}}_{jt,\text{us}} - w_{t,\text{eu}} L^{\text{us}}_{jt,\text{eu}} \right) \tag{33}$$

2. In the pre-GDPR regime, EU local digital firms choose a sequence of production decisions $\{L^{\text{eu}}_{jt,\text{eu}}\}$ and data collection decisions $\{x^{\text{eu}}_{jt,\text{eu}}\}$ to maximize the discounted value of all future profits.

$$\sum_{t=1}^{\infty} \left( p^{\text{eu}}_{jt,\text{eu}} Y^{\text{eu}}_{jt,\text{eu}} - w_{t,\text{eu}} L^{\text{eu}}_{jt,\text{eu}} \right) \tag{34}$$

3. US households choose a sequence of consumption decisions $\{c^{\text{us}}_{ijt,\text{us}}, c^{\text{nd}}_{it,\text{us}}\}$ to maximize their utility each period.

4. EU households choose a sequence of consumption decisions $\{c^{\text{us}}_{ijt,\text{eu}}, c^{\text{eu}}_{ij't,\text{eu}}, c^{\text{nd}}_{it,\text{eu}}\}$ to maximize their utility each period.

5. In the post-GDPR regime, EU consumers regain control over their own data, and make data sharing decisions $\{x^{\text{us}}_{ijt,\text{eu}}, x^{\text{eu}}_{ijt,\text{eu}}\}$ on top of consumption decisions in each period. US digital firms retain control of US data $\{x^{\text{us}}_{jt,\text{us}}\}$ but lose control of EU data $\{x^{\text{us}}_{jt,\text{eu}}\}$. Similarly, EU local digital firms lose control of EU data $\{x^{\text{eu}}_{jt,\text{eu}}\}$.

6. $\{p^{\text{us}}_{jt,\text{us}}, p^{\text{us}}_{jt,\text{eu}}, p^{\text{eu}}_{jt,\text{eu}}\}$ clear the goods market.

7. $\{w_{t,\text{us}}, w_{t,\text{eu}}\}$ clear the labor market.

## 5.5 Model Solution

### 5.5.1 Pre-GDPR

In the baseline scenario, referred to as the pre-GDPR regime, both US and EU digital firms exercise control over the extent of consumer data collection. The impact of firms' data collection decisions on households is illustrated in the utility functions. Equations 27 and 30 show how data choices by firms shape households' digital experiences and, consequently, their digital consumption choices. It's important to note that firms will not necessarily opt for maximal data collection (setting $x^{\text{us}}_{jt,\text{us}} = 1$, $x^{\text{us}}_{jt,\text{eu}} = 1$, and $x^{\text{eu}}_{jt,\text{eu}} = 1$). Instead, their strategies are nuanced, taking into account the prevalence of privacy-conscious consumers and the value these consumers place on their privacy.

**US Multinational Digital Firms** The challenge in determining firms' data collection strategies lies in their dependency on consumer preferences. To effectively navigate this challenge, my approach involves a two-step process. I first solve for firms' production problem by assuming fixed values for $x^{\text{us}}_{jt,\text{us}}$ and $x^{\text{us}}_{jt,\text{eu}}$. Then I solve for consumers' optimal digital consumption choices. Subsequently, I reintegrate the data collection decisions into the analysis. This is achieved by applying market clearing conditions, which ensures that the model accounts for the interplay between consumer preferences and firms' data strategies.

With the first-step simplification, for US multinational firms, they make a sequence of production decisions to maximize firm value. Their HJB equation can be written as

$$V_j^{\text{us}}(D_{j,t-1}^{\text{us}}) = \max_{\{L_{jt,\text{us}}^{\text{us}}\},\{L_{jt,\text{eu}}^{\text{us}}\}} \left(p_{jt,\text{us}}^{\text{us}} Y_{jt,\text{us}}^{\text{us}} + p_{jt,\text{eu}}^{\text{us}} Y_{jt,\text{eu}}^{\text{us}} - w_{t,\text{us}} L_{jt,\text{us}}^{\text{us}} - w_{t,\text{eu}} L_{jt,\text{eu}}^{\text{us}}\right) + \frac{1}{1+r} V_j^{\text{us}}(D_{jt}^{\text{us}}) \tag{35}$$

Since I have already determined the data choices for the firms, firm $j$ only needs to keep track of the state variable, data stock $D_{j,t-1}^{\text{us}}$, and the production decisions $\{L_{jt,\text{eu}}^{\text{us}}, L_{jt,\text{eu}}^{\text{us}}\}$ in each period. First, I can derive the first order conditions w.r.t. the labor choices $L_{jt,\text{us}}^{\text{us}}$ and $L_{jt,\text{eu}}^{\text{us}}$ and solve for the optimal production decisions. [22]

$$(L_{jt,\text{us}}^{\text{us}})^{\eta} = \frac{(1-\eta)(D_{j,t-1}^{\text{us}})^{\eta}((1+r)p_{jt,\text{us}}^{\text{us}} + \overbrace{V_j^{\text{us}'}(D_{jt}^{\text{us}})x_{jt,\text{us}}^{\text{us}})}^{\text{future value of data}}}{(1+r)w_{t,us}} \tag{36}$$

$$(L_{jt,\text{eu}}^{\text{us}})^{\eta} = \frac{(1-\eta)(D_{j,t-1}^{\text{us}})^{\eta}\left((1+r)p_{jt,\text{eu}}^{\text{us}} + V_j^{\text{us}'}(D_{jt}^{\text{us}})x_{jt,\text{eu}}^{\text{us}}\right)}{(1+r)w_{t,eu}} \tag{37}$$

The complementarity between data and labor is empirically demonstrated in Section 3.2. As a firm acquires more data, labor productivity increases, leading to more labor being deployed and higher production. It's also important to note how the future value of data influences a firm's production decisions. When firm $j$ produces more in the current period, it accumulates more data, which in turn enhances its productivity in the subsequent period. Consequently, firm $j$ will produce more than what would have been determined by the marginal revenue $\{p_{jt,\text{us}}^{\text{us}}, p_{jt,\text{eu}}^{\text{us}}\}$ and marginal cost $\{w_{t,us}, w_{t,eu}\}$.

Data resembles capital in a conventional growth model (Solow 1956; Swan 1956). However, there is no adjustment cost for data in my model.[23] I solve the model on the balanced growth path and suppose the stock of data grows at the constant rate $b_j^{\text{us}}$.

$$D_{jt}^{\text{us}} = (1 + b_j^{\text{us}})D_{j,t-1}^{\text{us}} \tag{38}$$

The equilibrium output grows at the same rate as the state variable data stock $D_{jt}^{\text{us}}$. I guess

[22] Details of the derivation can be found in the Appendix A.

[23] Extra data storage space either locally or on the cloud is not the major cost of data analysis for most companies.

and verify the equilibrium value function as

$$V_j^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}}) = B_j^{\mathrm{us}} \cdot D_{j,t-1}^{\mathrm{us}} \tag{39}$$

Then I can derive an expression for $B_j^{\mathrm{us}}$

$$B_j^{\mathrm{us}} = \frac{(1+r)\eta\left(p_{jt,\mathrm{us}}^{\mathrm{us}} C_{j,\mathrm{us}}^{\mathrm{us}} + p_{jt,\mathrm{eu}}^{\mathrm{us}} C_{j,\mathrm{eu}}^{\mathrm{us}}\right)}{r + \kappa_{\mathrm{us}} - \eta(b_j^{us} + \kappa_{\mathrm{us}})} \tag{40}$$

where $C_{j,\mathrm{us}}^{\mathrm{us}}$ and $C_{j,\mathrm{eu}}^{\mathrm{us}}$ are constants,[24] and

$$C_{j,\mathrm{us}}^{\mathrm{us}} = \frac{\left(L_{jt,\mathrm{us}}^{\mathrm{us}}\right)^{1-\eta}}{\left(D_{j,t-1}^{\mathrm{us}}\right)^{1-\eta}}, \quad C_{j,\mathrm{eu}}^{\mathrm{us}} = \frac{\left(L_{jt,\mathrm{eu}}^{\mathrm{us}}\right)^{1-\eta}}{\left(D_{j,t-1}^{\mathrm{us}}\right)^{1-\eta}} \tag{41}$$

Then

$$V_j^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}}) = \frac{(1+r)\eta\left(p_{jt,\mathrm{us}}^{\mathrm{us}} C_{jt,\mathrm{us}}^{\mathrm{us}} + p_{jt,\mathrm{eu}}^{\mathrm{us}} C_{jt,\mathrm{eu}}^{\mathrm{us}}\right)}{r + \kappa_{\mathrm{us}} - \eta(b_j^{us} + \kappa_{\mathrm{us}})} \cdot D_{j,t-1}^{\mathrm{us}} \tag{42}$$

Equation 42 resembles the Gordon growth model except that the data feedback loop enters the expression through $\eta$. Higher $\eta$ means higher productivity from data. More production today means more data being accumulated for tomorrow's production, and $\frac{\partial V_j^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}})}{\partial \eta} > 0$. Also very intuitively, higher depreciation rate leads to lower firm value, and $\frac{\partial V_j^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}})}{\partial \kappa_{\mathrm{us}}} < 0$. I will solve for the equilibrium numerically in Section 6.1.

**EU Local Digital Firms** Similar to the US digital firms, I can set up the HJB equation for EU local digital firms.

$$V_j^{\mathrm{eu}}(D_{j,t-1}^{\mathrm{eu}}) = \max_{\{L_{jt,\mathrm{eu}}^{\mathrm{eu}}\}} \left(p_{jt,\mathrm{eu}}^{\mathrm{eu}} Y_{jt,\mathrm{eu}}^{\mathrm{eu}} - w_{t,\mathrm{eu}} L_{jt,\mathrm{eu}}^{\mathrm{eu}}\right) + \frac{1}{1+r} V_{j,\mathrm{eu}}(D_{jt}^{\mathrm{eu}}) \tag{43}$$

Then I can solve for the optimal production decisions.

$$(L_{jt,\mathrm{eu}}^{\mathrm{eu}})^{\eta} = \frac{(1-\eta)(D_{j,t-1}^{\mathrm{eu}})^{\eta}\left((1+r)p_{jt,\mathrm{eu}}^{\mathrm{eu}} + V_j^{\mathrm{eu}\prime}(D_{jt}^{\mathrm{eu}}) x_{jt,\mathrm{eu}}^{\mathrm{eu}}\right)}{(1+r)w_{t,eu}} \tag{44}$$

[24] Equation 40 illustrates the intuition for the value function, but we have to acknowledge that $p_{jt,\mathrm{us}}^{\mathrm{us}}$ and $p_{jt,\mathrm{us}}^{\mathrm{us}}$ are equilibrium outcomes. The final expression will also depend on the output elasticity of labor through constants $C_{j,\mathrm{us}}^{\mathrm{us}}$ and $C_{j,\mathrm{eu}}^{\mathrm{us}}$. Here, I can perform a partial equilibrium analysis to understand the comparative statics.

Again, I solve the model on the balanced growth path and suppose the stock of data grows at the constant rate $b_j^{\text{eu}}$. I also guess and verify that the value function

$$V_j^{\text{eu}}(D_{j,t-1}^{\text{eu}}) = B_j^{\text{eu}} \cdot D_{j,t-1}^{\text{eu}} \tag{45}$$

Then I can derive an expression for $B_j^{\text{eu}}$

$$B_j^{\text{eu}} = \frac{(1+r)\eta p_{jt,\text{eu}}^{\text{eu}} C_{j,\text{eu}}^{\text{eu}}}{r + \kappa_{\text{eu}} - \eta(b_j^{eu} + \kappa_{\text{eu}})} \tag{46}$$

and

$$V_j^{\text{eu}}(D_{j,t-1}^{\text{eu}}) = \frac{(1+r)\eta p_{jt,\text{eu}}^{\text{eu}} C_{jt,\text{eu}}^{\text{eu}}}{r + \kappa_{\text{eu}} - \eta(b_j^{eu} + \kappa_{\text{eu}})} \cdot D_{j,t-1}^{\text{eu}} \tag{47}$$

where

$$C_{j,\text{eu}}^{\text{eu}} = \frac{\left(L_{jt,\text{eu}}^{\text{eu}}\right)^{1-\eta}}{\left(D_{j,t-1}^{\text{eu}}\right)^{1-\eta}} \tag{48}$$

**US Household** Here I solve for a symmetric case and suppose all privacy-conscious consumers have the same privacy preferences $\delta$. I can set up the Lagrangian of the US households' optimization problem.

$$\begin{aligned}\mathcal{L}_{it,\text{us}} =& K \int_0^1 \left(\gamma \bar{x}_{jt}^{\text{us}} + (1-\gamma) x_{ijt,\text{us}}^{\text{us}} - \mathbb{1}_{\{z_i=1\}} \delta \left(x_{ijt,\text{us}}^{\text{us}}\right)^2\right) \ln c_{ijt,\text{us}}^{\text{us}} dj + (1-K) \ln c_{it,\text{us}}^{\text{nd}} \\ &+ \mu_{it,\text{us}} \left(\int_0^1 \Pi_{ijt,\text{us}}^{\text{us}} dj + \Pi_{it,\text{us}}^{\text{nd}} + w_{t,\text{us}} l_{it,\text{us}} - \int_0^1 p_{jt,\text{us}}^{\text{us}} c_{ijt,\text{us}}^{\text{us}} dj - c_{it,\text{us}}^{\text{nd}}\right) \end{aligned} \tag{49}$$

Here I aim to solve for an interior solution. I can derive the first order conditions and solve for the optimal non-digital consumption as

$$\begin{aligned} c_{it,\text{us}}^{\text{nd}} =& \frac{(1-K)\Pi_{it,\text{us}}}{K \int_0^1 \left(\gamma \bar{x}_{jt}^{\text{us}} + (1-\gamma) x_{ijt,\text{us}}^{\text{us}} - \mathbb{1}_{\{z_i=1\}} \cdot \delta \left(x_{ijt,\text{us}}^{\text{us}}\right)^2\right) dj + (1-K)} \\ =& \frac{(1-K)\Pi_{it,\text{us}}}{K X_{it,\text{us}} + (1-K)} \end{aligned} \tag{50}$$

where

$$X_{it,\mathrm{us}} = \int_0^1 \left(\gamma \bar{x}_{jt}^{\mathrm{us}} + (1-\gamma) x_{ijt,\mathrm{us}}^{\mathrm{us}} - \mathbb{1}_{\{z_i=1\}} \delta \left(x_{ijt,\mathrm{us}}^{\mathrm{us}}\right)^2\right) dj \tag{51}$$

The optimal digital consumption is given by

$$c_{ijt,\mathrm{us}}^{\mathrm{us}} = \frac{K\left(\gamma \bar{x}_{jt}^{\mathrm{us}} + (1-\gamma) x_{ijt,\mathrm{us}}^{\mathrm{us}} - \mathbb{1}_{\{z_i=1\}} \delta \left(x_{ijt,\mathrm{us}}^{\mathrm{us}}\right)^2\right) \Pi_{it,\mathrm{us}}}{p_{jt,\mathrm{us}} \left(K X_{it,\mathrm{us}} + (1-K)\right)} \tag{52}$$

**EU Household** Similarly, I can solve for the optimal consumption choices of EU households. For European households, their optimal consumption of US digital products will be

$$c_{ijt,\mathrm{eu}}^{\mathrm{us}} = \frac{K\beta\left(\gamma \bar{x}_{jt}^{\mathrm{us}} + (1-\gamma) x_{ijt,\mathrm{eu}}^{\mathrm{us}} - \mathbb{1}_{\{z_i=1\}} \delta \left(x_{ijt,\mathrm{eu}}^{\mathrm{us}}\right)^2\right) \Pi_{it,\mathrm{eu}}}{p_{jt,\mathrm{eu}}^{\mathrm{us}} \left(K\left(\beta X_{it,\mathrm{eu}}^{\mathrm{us}} + (1-\beta) X_{it,\mathrm{eu}}^{\mathrm{eu}}\right) + (1-K)\right)} \tag{53}$$

where

$$\begin{aligned} X_{it,\mathrm{eu}}^{\mathrm{us}} &= \int_0^1 \left(\gamma \bar{x}_{jt}^{\mathrm{us}} + (1-\gamma) x_{ijt,\mathrm{eu}}^{\mathrm{us}} - \mathbb{1}_{\{z_i=1\}} \delta \left(x_{ijt,\mathrm{eu}}^{\mathrm{us}}\right)^2\right) dj \\ X_{it,\mathrm{eu}}^{\mathrm{eu}} &= \int_0^1 \left(\gamma \bar{x}_{j't}^{\mathrm{eu}} + (1-\gamma) x_{ij't,\mathrm{eu}}^{\mathrm{eu}} - \mathbb{1}_{\{z_i=1\}} \delta \left(x_{ij't,\mathrm{eu}}^{\mathrm{eu}}\right)^2\right) dj' \end{aligned} \tag{54}$$

and their consumption of EU local digital products will be

$$c_{ijt,\mathrm{eu}}^{\mathrm{eu}} = \frac{K(1-\beta)\left(\gamma \bar{x}_{j't}^{\mathrm{eu}} + (1-\gamma) x_{ij't,\mathrm{eu}}^{\mathrm{eu}} - \mathbb{1}_{\{z_i=1\}} \delta \left(x_{ij't,\mathrm{eu}}^{\mathrm{eu}}\right)^2\right) \Pi_{it,\mathrm{eu}}}{p_{j't,\mathrm{eu}}^{\mathrm{eu}} \left(K\left(\beta X_{it,\mathrm{eu}}^{\mathrm{us}} + (1-\beta) X_{it,\mathrm{eu}}^{\mathrm{eu}}\right) + (1-K)\right)} \tag{55}$$

EU households' consumption of non-digital products is given by

$$c_{it,\mathrm{eu}}^{\mathrm{nd}} = \frac{(1-K)\Pi_{it,\mathrm{eu}}}{K\left(\beta X_{it,\mathrm{eu}}^{\mathrm{us}} + (1-\beta) X_{it,\mathrm{eu}}^{\mathrm{eu}}\right) + (1-K)} \tag{56}$$

**Market Clearing** The market clearing conditions for US digital goods is

$$Y_{jt,\mathrm{us}}^{\mathrm{us}} = \int_0^1 c_{ijt,\mathrm{us}}^{\mathrm{us}} di, \quad Y_{jt,\mathrm{eu}}^{\mathrm{us}} = \int_0^1 c_{ijt,\mathrm{eu}}^{\mathrm{us}} di \tag{57}$$

The market clearing condition for EU local digital product is

$$Y^{\text{eu}}_{j't,\text{eu}} = \int_0^1 c^{\text{eu}}_{ij't,\text{eu}} di \tag{58}$$

The market clearing condition for the labor markets are

$$L_{t,\text{us}} = \int_0^1 L^{\text{us}}_{jt,\text{us}} dj, \quad L_{t,\text{eu}} = \int_0^1 L^{\text{us}}_{jt,\text{eu}} dj + \int_0^1 L^{\text{eu}}_{j't,\text{eu}} dj' \tag{59}$$

Then with the market clearing conditions, I can go back and solve for the optimal data collection choices by firms. It will be solved numerically in the calibration section.

### 5.5.2 Post-GDPR

In the post-GDPR regime, EU households regain control of their own data and choose $x^{\text{us}}_{ijt,\text{eu}}$. As a result, digital firms only set the data sharing decisions for households from the US. While choosing their desired level of data sharing, EU households do not incorporate the positive externality they have on other households. Since each individual is atomistic, $\frac{\partial \bar{x}_{jt}}{\partial x_{ijt}} = 0$. Given other people's data sharing choice, EU privacy-conscious households' optimal level of data sharing is

$$x^{*}_{ijt,\text{eu}} = \frac{1-\gamma}{2\delta} \tag{60}$$

When $\delta > \frac{1-\gamma}{2}$, $x^{*}_{ijt,\text{eu}} < 1$. This is reflected in two extra sets of control variables, $\{x^{\text{us}}_{ijt,\text{eu}}\}$ and $\{x^{\text{us}}_{ijt,\text{eu}}\}$, for EU households. For US households, their optimization problems remain the same.

For US multinational digital firms and EU local digital firms, they lose the control over EU consumers data. Alternatively, I can view that EU consumers' decisions put an upper bound on the amount of data digital firms can collect.

$$x^{\text{us}}_{ijt,\text{eu}} \in \left[0, \frac{1-\gamma}{2\delta}\right], \quad x^{\text{eu}}_{ijt,\text{eu}} \in \left[0, \frac{1-\gamma}{2\delta}\right] \tag{61}$$

Following the same procedure as in the previous section, I can solve for equilibrium outcomes under the alternative post-GDPR regime. The setup of the model is in part inspired by the empirical section. There are two key findings in the empirical section. First, after GDPR came into effect, US multinational firms in the data-intensive category reduce their exposure to the European market. Second, EU consumers see a decline in their user ratings on digital platforms. These empirical findings match the predictions of the model. I can use the quantitative findings to estimate the two key parameters in the model, the output elasticity of data $\eta$ and privacy preferences of consumers $\delta$. I will explain in the following section how I plan to calibrate these two parameters.

# 6 Calibration and Welfare Analysis

## 6.1 Calibration

In this subsection, I will calibrate the model to match the empirical moments documented in Sections 3 and 4. Section 3 highlights an 8% decline in EU businesses for US data-intensive firms, while observing a rise in sales from other global regions. Moreover, Section 4 reveals a less satisfactory consumer experience for EU consumers compared to their US counterparts post-GDPR, implying a trade-off between privacy protection and data-dependent user experiences. The findings in these sections represent equilibrium outcomes reflecting GDPR-induced responses from both firms and consumers. Utilizing the model, as described in Section 5, enables disentangling supply and demand effects. I will solve the model numerically on the balanced growth path pre- and post-GDPR, compute the changes in equilibrium outcomes across these regimes, and match with empirical evidence. The targeted moments are summarized in Table 7. I also abstract away from the labor market and choose a wage rate matching the labor income share of firm revenue.

The targeted moments correspond to the main empirical findings from Section 3 and Section 4. These include the shifting in revenue from the EU to US after GDPR, the pre-GDPR EU sales share for data-intensive firms, and the decline in service quality for EU users and US users after GDPR. Since, in the model, the data growth rate is directly linked to the firm growth rate, I can calibrate the data depreciation rate in the US and EU to the firm growth rate in the US and EU. These six moments will be used to jointly pin down six parameters $(\eta, \delta, \beta, \gamma, \kappa_{\text{us}}, \kappa_{\text{eu}})$ that are internally calibrated. $\eta$ is the output elasticity of data, which determines how data contributes to firm productivity. $\delta$ captures consumers' privacy preferences. The more privacy-conscious consumers are, the less they are willing to share data. $\beta$ captures EU consumers preferences for US digital products. The parameter $\gamma$ captures how common preferences component contributes to digital experiences, which determines the size of the spillover effects from the EU to the US market.

For US multinational digital firms, we define the share of EU sales as

$$\psi^{\text{us}}_{jt,\text{eu}} = \frac{Y^{\text{us}}_{jt,\text{eu}}}{Y^{\text{us}}_{jt,\text{us}} + Y^{\text{us}}_{jt,\text{eu}}} \tag{62}$$

The pre-GDPR EU sales share of US digital firms is $\psi^{\text{us}}_{\text{eu,pre-GDPR}} = 0.1625$, while the post-GDPR EU sales share of US digital firms is $\psi^{\text{us}}_{\text{eu,post-GDPR}} = 0.1501$. The quality of digital products, which is dependent on the aggregate level and individual level of data sharing, is given by

$$s_{ijt} = \gamma \bar{x}_{jt} + (1 - \gamma) x_{ijt} \tag{63}$$

That is, one individual's data sharing behaviors not only depends on her own data sharing behavior but also on the data sharing behaviors of other people. From Section 4, we find that the overall digital experiences of EU users decline by 6% while the overall user experiences of US users decline by 1%. We can use these two moments to help us understand the privacy preferences of consumers, $\delta$, and the contribution of common preferences in digital products, $\gamma$. I also calibrate the data depreciation rate to the growth rates of the US and EU tech sectors respectively. In my model, the data growth rate is directly linked to the firm growth rate, and the data depreciation rate affects the data growth rate. The data depreciation rates calibrated are 31.1% per year for US firms and 34.5% per year for EU firms, aligning closely with the numbers used in the literature.

In my model, consumers exhibit heterogeneous privacy preferences. To determine the share of privacy-conscious consumers, I delve into existing literature. Studies by Aridor et al. (2020), Goldberg et al. (2019), and Zhao et al. (2021) indicate that post-GDPR, the "observability" of EU consumers diminishes by approximately 12%-16%. I match this to a

Table 7: Calibration Parameters

*Notes*: In this Table, I describe the targeted moments and the corresponding calibrated parameters.

| Parameter | Explanation | Value | Source/Target |
|---|---|---|---|
| **Internally Calibrated Parameters:** | | | |
| $\eta$ | Output elasticity of data | 0.348 | EU sales declines by 8% post-GDPR (Section 3) |
| $\delta$ | Privacy preferences | 0.482 | EU user rating declines by 6 percent post-GDPR (Section 4) |
| $\beta$ | EU preferences for US digital products | 0.042 | pre-GDPR EU sales share equals 16.25% (Section 3) |
| $\gamma$ | The positive externality of data | 0.571 | US user ratings declines by 1 percent post-GDPR (Section 4) |
| $\kappa_{\text{us}}$ | US data depreciation rate per year | 0.311 | US tech industry growth rate $b^{\text{us}}_{j,\text{pre}} = 0.082$ (Yahoo Finance) |
| $\kappa_{\text{eu}}$ | EU data depreciation rate per year | 0.345 | EU tech industry growth rate $b^{\text{eu}}_{j,\text{pre}} = 0.085$ (Deloitte) |
| **Externally Chosen Parameters:** | | | |
| $\lambda$ | Share of privacy sensitive consumers | 0.28 | Consumer Observability declines by 12% - 16% post-GDPR Aridor et al. (2020); Goldberg et al. (2019) Zhao et al. (2021) |
| $K$ | Relative preference for digital consumption | 0.09 | Digital economy accounts for 9% of total GDP in 2018 (BEA) |
| $r_{\text{us}}$ | Discount rate | 0.108 | Cost of capital for US Tech firms (link). |
| $r_{\text{eu}}$ | Discount rate | 0.102 | Cost of capital for EU Tech firms (link). |
| **Normalization Parameters** | | | |
| $\Pi_{\text{eu}}/\Pi_{\text{us}}$ | EU to US economy size ratio | 0.78 | EU to US GDP ratio in 2018 (0.78) |
| $w_{\text{us}}$ | US wage rate | 0.4 | Labor share of income in the US 68% |
| $w_{\text{eu}}$ | EU wage rate | 0.3 | Labor share of income in the EU 65% |
| $D_{\text{eu}}/D_{\text{us}}$ | EU to US equilibrium data stock ratio | 0.7 | EU to US tech sector size ratio in 2018 (62.6% Statista) |

decline in the average level of data sharing among EU consumers, which, in turn, aids in determining the fraction of privacy-sensitive consumers. Regarding the digital economy's size, I reference a report by the Bureau of Economic Analysis, which presents a broad definition of the digital economy, encompassing ecommerce and other partially digital activities. I set the share of digital consumption $K = 0.09$. For discount rates, I utilize the cost of capital calculations from Damodaran's website for US and EU tech firms respectively. Furthermore, I align the size ratio of the US to the EU economy based on their GDP ratio in 2018, and adjust the wage rate in each country to match the labor share of income. Lastly, I match the equilibrium level of data stock in both countries to the EU to US tech sector size ratio as cited in a Statista report.

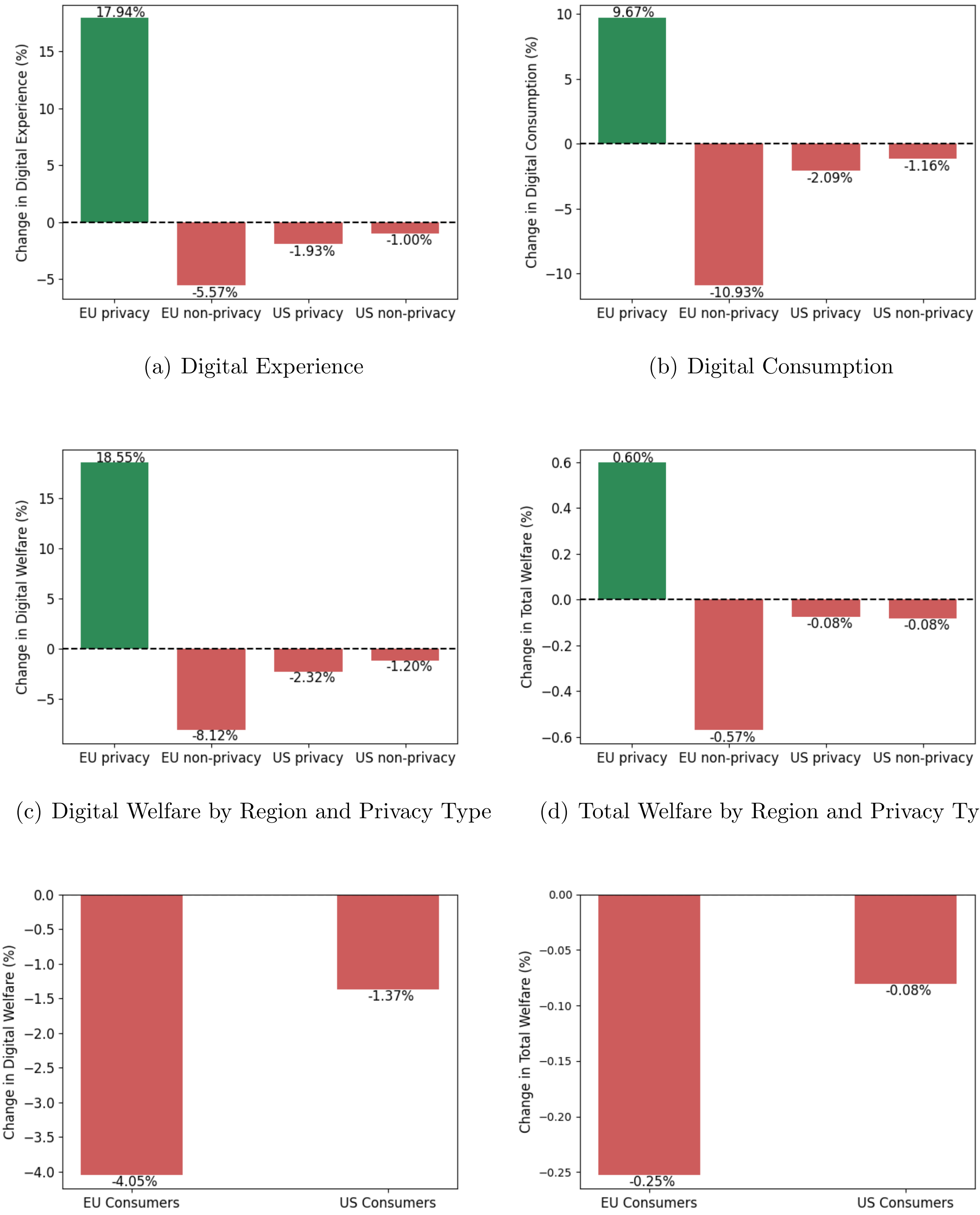
## 6.2 Welfare Analysis

### 6.2.1 Digital and Total Welfare

In this section, I explore the welfare implications of GDPR. Since firms redistribute profits to consumers in each economy, digital welfare is defined as the consumer surplus from digital consumption, and total welfare is defined as the consumer surplus from both digital consumption and non-digital consumption.

The calibration reveals that GDPR has an uneven welfare impact on different groups of consumers. EU privacy-conscious consumers enjoy a substantial increase in their digital experiences after GDPR came into effect. Their EU non-privacy-conscious counterparts incur a large welfare loss in digital experiences as they use not only the same US multinational digital products, but also the local EU digital products. Both US privacy-conscious and non-privacy-conscious consumers also face negative impact from GDPR through spillover effects. As shown in Figure 4, EU privacy conscious consumers significantly improve their digital service quality and consume more digital products. They feel "safer" consuming more digital products when their privacy is properly protected, aligning with the initial motivation of the EU regulations.

As depicted in panel (a) of Figure 4, as EU privacy conscious consumers choose to share less data, their digital experiences improve due to better privacy protection. How-



(a) Digital Experience

(b) Digital Consumption

(c) Digital Welfare by Region and Privacy Type

(d) Total Welfare by Region and Privacy Type

(e) Digital Welfare by Region

(f) Total Welfare by Region

Figure 4: Digital Experiences, Consumption, and Welfare Change

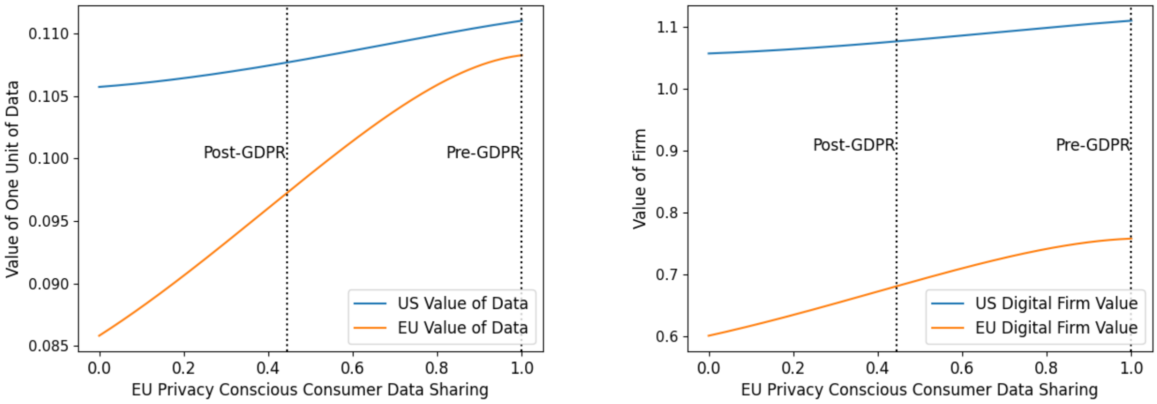
ever, individuals do not properly internalize the positive externality of data sharing on other consumers, leading to firms receiving less precise signals about the common component of preference across consumers. As a result, the digital experiences of all other three categories of consumers deteriorate, with EU non-privacy-conscious consumers bearing the brunt of the negative impact as their digital consumption bundle overlaps the most with EU privacy-conscious consumers. The negative impact on US consumers stems from the business operations of US multinational digital firms, as EU consumers are also part of their customer base.

As illustrated in panel (b) of Figure 4, the change in digital experiences does not map one-to-one to the change in digital consumption, even though they have homothetic preferences. This is because firms price in the data sharing behaviors of consumers. Unable to perfectly price discriminate against individuals, all consumers bear the cost of less data sharing. As today's data collection contributes to tomorrow's productivity improvement, less data sharing leads to a lower level of production and increases the cost of production. Consequently, firms raise prices on all consumers. In my model, firms can price discriminate at the regional level. Therefore, as EU privacy-conscious consumers choose to share less data, US multinational digital firms increase the price on EU consumers, as do EU local digital firms. As a result, EU privacy-conscious consumers increase their digital consumption less than what would have been predicted by the improvement in digital experiences. Due to the change in firms' pricing behaviors, the negative impact on the other three groups of consumers gets amplified. On net, the overall digital consumption decreases by 5.5%.

In panel (c) of Figure 4, I compute the change in consumer surplus for digital consumption, reflecting the composite effects from panels (a) and (b). As expected, EU privacy-conscious consumers receive a substantial boost from the privacy protection, improving their overall digital welfare by almost 18.6%. However, this improvement comes at the cost of other consumer groups, with their EU non-privacy-conscious peers bearing most of the negative impact. US consumers also feel the impact from this regulation due to the operations of US multinational firms. As I calibrate to the state of the world where digital consumption only accounts for around 9 percent of total consumption, the impact of the regulation on total welfare (including both digital and non-digital consumption) across all consumers is less dramatic but still significant. In panel (d), similar to panel (c), we see that EU privacy-conscious consumers emerge as the sole winners from this privacy regulation. When aggregating across privacy types to the regional level, a negative welfare impact on all consumers is apparent. From panels (e) and (f), EU consumers suffer the most from this regulation, with negative spillover effects reaching the US market through the business operations of US multinational firms.

### 6.2.2 The Value of Data and Digital Firms

As shown in Section 5, due to the data feedback loop, the value of one unit of data depends on consumers' data sharing choices. In my framework, data is a byproduct of economic activities, and a less privacy-conscious consumer base leads to higher data accumulation each period. A digital firm's data stock is directly linked to its productivity. The essence of the data feedback loop (as in Farboodi and Veldkamp (2021); Jones and Tonetti (2020)) is that more data leads to more production and more production generates more data. There is a multiplier effect of data as shown in the value functions of digital firms, equation 42 and 47.



(a) Value of Data (b) Value of Firm

Figure 5: The Value of Data and Digital Firms

Consumers play an important role in the data collection process. When a privacy regulation like GDPR gives EU consumers more control over their privacy, they can choose to share less data with firms, which reduces the multiplier effect. In Figure 5 panel (a), I plot the value of one unit of data for US and EU digital firms against the data sharing choices of EU privacy-conscious consumers. As we can see, the value of one unit of data increases monotonically with the data sharing level of EU privacy-conscious consumers. Since EU privacy-conscious consumers account for 28% of the EU digital firms' customer base but only 4% of the US digital firms' customer base, the increase in the value of data is most pronounced for EU digital firms. At full data sharing, the gap between US and EU data value comes from the difference in data depreciation rate and discount rate as calibrated in the previous section.

In panel (b) of Figure 5, I plot the value of EU and US digital firm against the data sharing level of EU privacy-conscious consumers. It largely resembles the pattern in panel (a), more data sharing leads to higher digital firm value. The gap between the US and EU digital firms comes from the different levels of equilibrium data stock as calibrated in section 6.1.

## 6.3 Discussion on Social Optimum

### 6.3.1 EU Consumer Welfare

When data sharing decisions are given back to EU individuals, EU privacy conscious consumers fail to internalize the positive externality they have on others. In contrast, EU non-privacy-conscious consumers' personal incentives are aligned with social incentives.

Apparently, the data sharing choices made by EU individuals are not socially optimal. The question is whether it is at least optimal for EU privacy-conscious consumers, the intended group that the EU regulators aim to protect. In this section, I first perform a counterfactual analysis, where I vary the level of data sharing choices by EU privacy-conscious consumers and examine the welfare consequences on different consumer groups.

As it turns out, the individual data sharing choices by EU privacy-conscious consumers are not even optimal for themselves. There are two competing forces that determine the welfare levels of EU privacy-conscious consumers, sharing less data for improved personal privacy protection versus losing on the positive externality of data. Depending on which force dominates, the welfare level of EU-privacy-conscious consumers is a non-monotone function in their data sharing decisions.

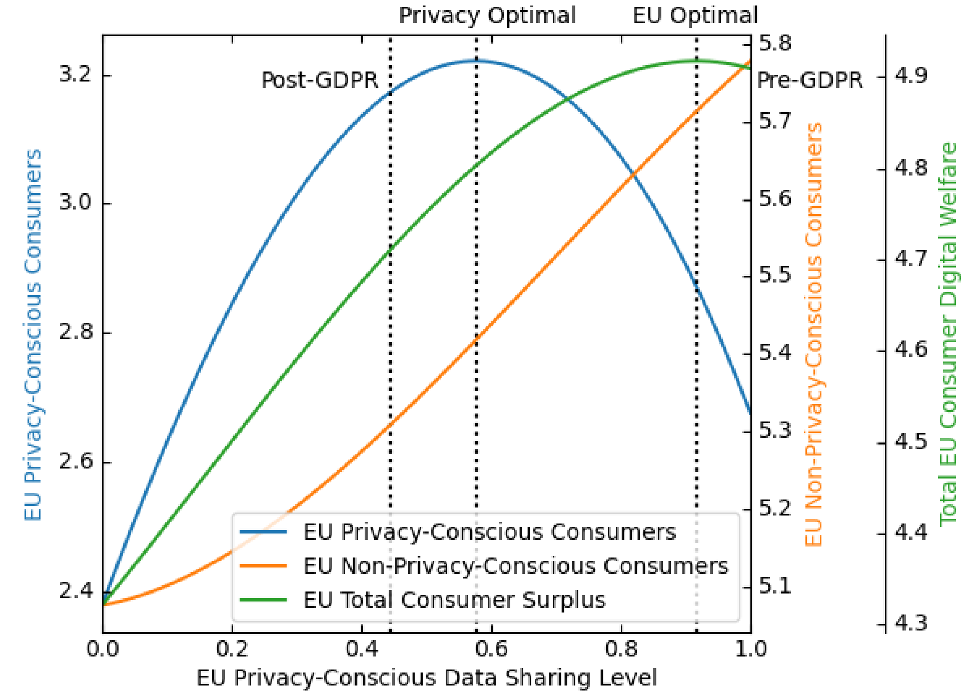
In Figure 6 panel (a), I plot the digital welfare of EU consumers on data sharing levels of EU privacy conscious consumers. The pre-GDPR regime is the full-data sharing regime, which is strictly preferred by the non-privacy-conscious consumers. The post-GDPR regime, which is characterized by too much data withholding, is neither optimal for EU privacy conscious consumers nor socially optimal. This is because the post-GDPR regime loses too much on the positive externality of data, where the marginal benefit of data sharing dominates the marginal cost of privacy loss for EU privacy-conscious consumers. The socially optimal level of digital consumption is achieved when we balance the privacy-conscious consumers and non-privacy-conscious consumers, which is between the optimal point for EU privacy-conscious consumers and pre-GDPR full data sharing regime.

In Figure 6 panel (b), I explore the total welfare, which includes both the digital consumption and non-digital consumption. To maximize total welfare, EU privacy-conscious consumers should share more data than they would have if they only maximize digital welfare. This is because individual data sharing exerts positive externality not only through improving the matching efficiency of common preferences but also through increasing the productivity of firms. When firms can produce at a lower cost, consumers will also have a higher budget to consume non-digital consumption. This extra benefit of data sharing leads to the optimal level of data sharing for total welfare being higher than the digital welfare. This again highlights the importance of analyzing the privacy regulations from an equilibrium perspective so that we can account for all the equilibrium forces that affect consumer welfare.

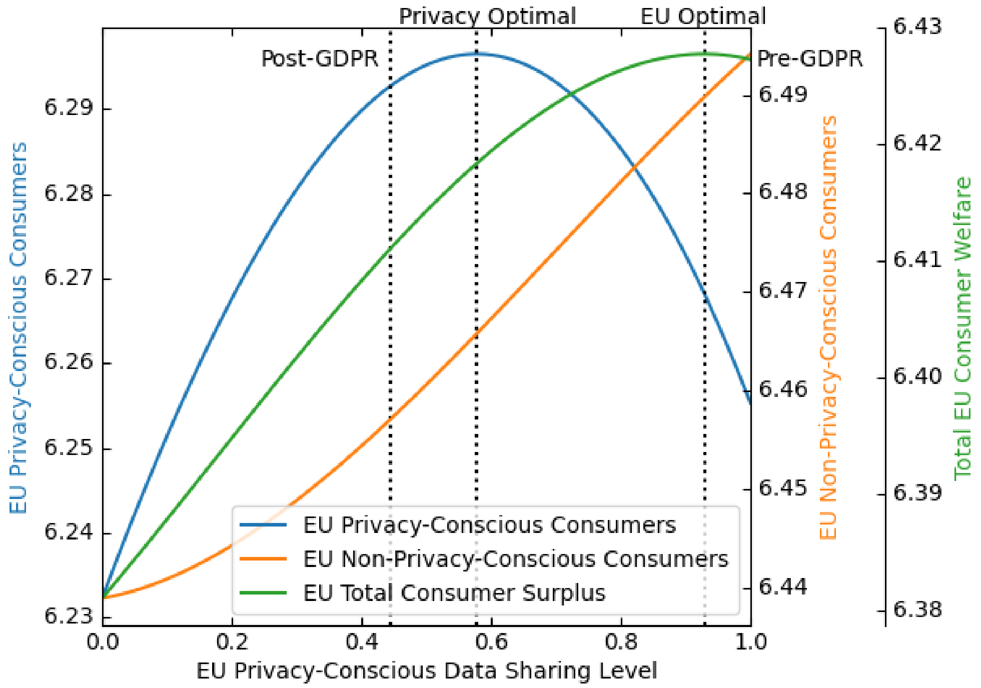
### 6.3.2 GDPR Type Regulations in Both Regimes

How about the US consumers? EU regulators only have their constituents in mind. Suppose we have a benevolent regulator that deeply cares about the welfare of both states. What would be the optimal level of data sharing for EU and US privacy-conscious consumers?

In this section, we first explore a counterfactual scenario where both US and EU enact GDPR. As it turns out, it is even worse than the post-GDPR regime. The "tragedy of

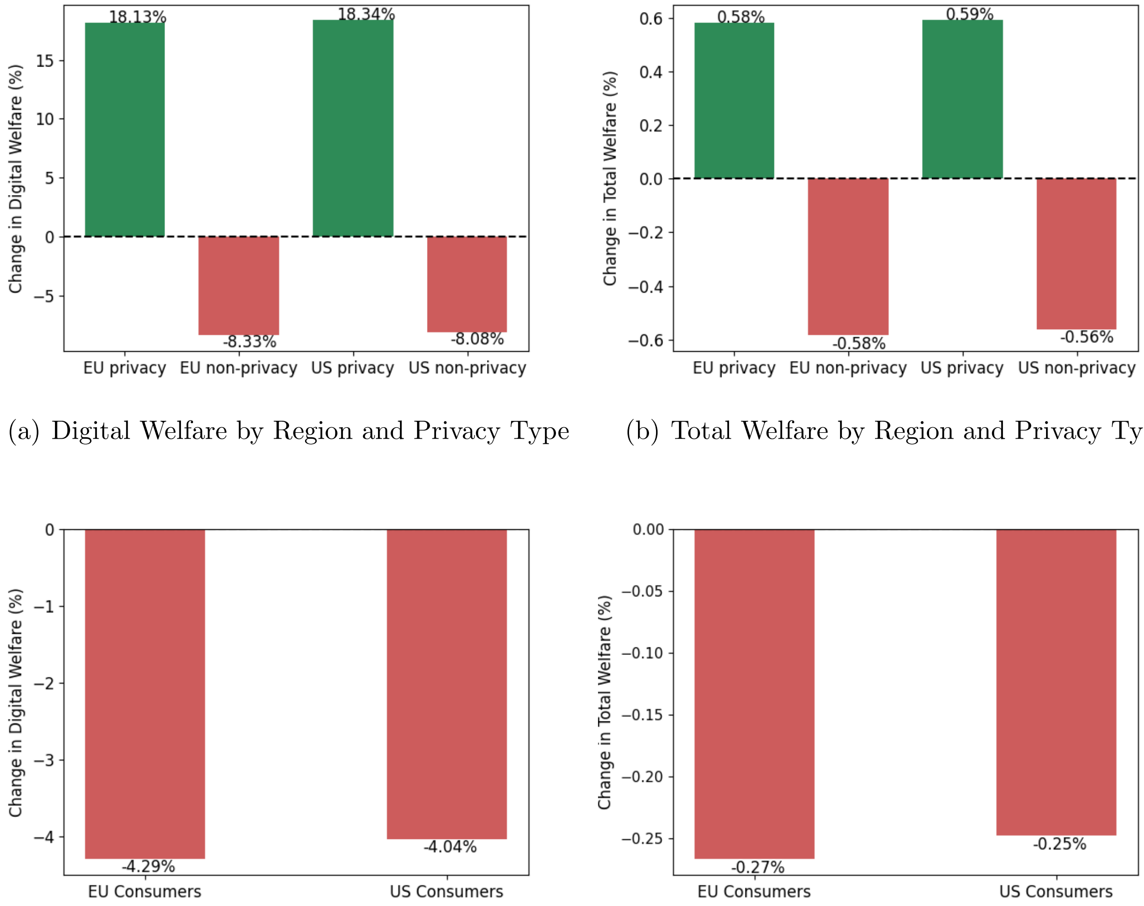


(a) Digital Welfare



(b) Total Welfare

Figure 6: Welfare Dependence on Data Sharing



(a) Digital Welfare by Region and Privacy Type

(b) Total Welfare by Region and Privacy Type

(c) Digital Welfare by Region

(d) Total Welfare by Region

Figure 7: Counterfactual Digital Experiences, Consumption, and Welfare Change (GDPR in Both Regions)

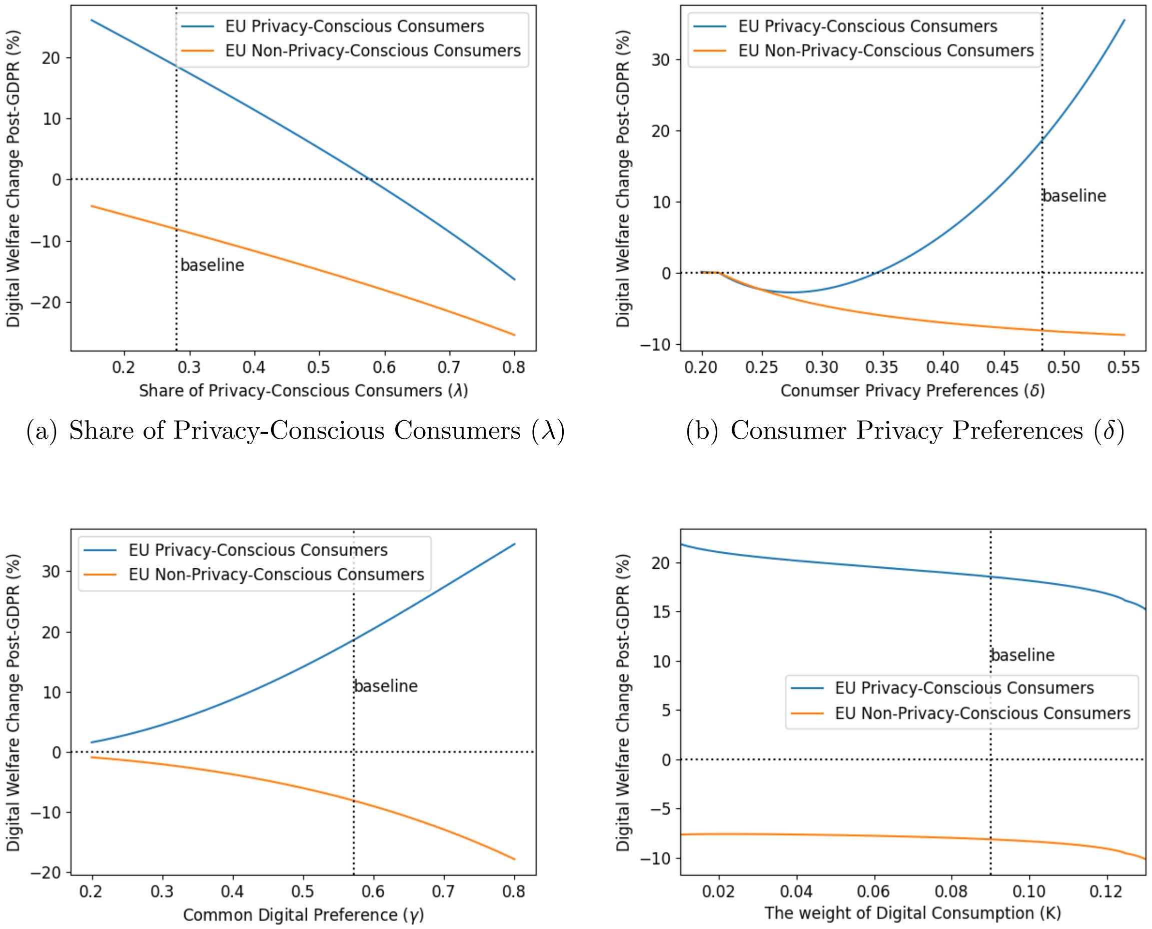
commons" gets further amplified as US privacy-conscious consumers also get to free-ride on other consumers. We obtain a quasi-symmetric impact on US and EU consumers. Similar to the welfare analysis in Section 6.2.1, there is a large digital welfare gain for privacy-conscious consumers in both EU and US, while the welfare gains come at a cost to non-privacy-conscious consumers. The magnitude of the welfare improvement is similar for both regions. The impact on total welfare is of much smaller scale because digital consumption only accounts for around 9 percent of total consumption. However, the aggregate negative impact on US and EU digital and total welfare get amplified.

## 6.4 Sensitivity of Welfare Outcome to Main Parameters

This section examines the impact of key parameter selections on welfare analysis outcomes. First, I investigate the effect of varying the proportion of privacy-conscious consumers ($\lambda$) on welfare. Figure 8 panel (a) illustrates that an increased proportion of privacy-conscious consumers results in diminished digital welfare for both EU privacy-conscious and non-privacy-conscious individuals. This decline occurs because consumers do not internalize the positive externality of their data-sharing decisions, a problem that intensifies with a more privacy-conscious consumer base. Notably, when their proportion exceeds 60%, both consumer groups experience welfare losses following GDPR implementation.

I then keep other key parameters constant to assess how changes in consumer privacy preferences influence the welfare effects of GDPR. As depicted in Figure 8 panel (b), at privacy preference levels below 0.35, GDPR does not improve welfare for privacy-conscious consumers. In this lower preference range, the negative impact from a lower aggregate level of data sharing overshadows the benefits of increased privacy protection. However, as privacy preferences rise, the benefits of enhanced privacy protection begin to outweigh the negative effects, leading to a non-monotone pattern in panel (b).

It is natural to think that as the importance of the common digital preference ($\gamma$) increases, both EU privacy-conscious and non-privacy-conscious consumers would fare worse from GDPR as the data free-riding problem intensifies. Figure 8 panel (c) tells us otherwise. A higher $\gamma$ also means that the idiosyncratic digital preference component becomes less important, and the digital experiences of privacy conscious consumers are largely determined



(a) Share of Privacy-Conscious Consumers ($\lambda$)

(b) Consumer Privacy Preferences ($\delta$)

(c) Common Digital Preferences ($\gamma$)

(d) The Weight of Digital Consumption ($K$)

Figure 8: Sensitivity of Welfare Outcome to Main Parameters

by privacy protection. Consequently, as $\gamma$ increases, EU privacy-conscious consumers experience increased welfare gains from GDPR. However, for non-privacy-conscious consumers, an increase in $\gamma$ exacerbates the free-riding issue by privacy conscious consumers, resulting in greater welfare losses.

In Figure 8 panel (d), I explore how the relative importance of digital consumption ($K$) affects digital welfare. The changes in digital welfare post-GDPR is not very sensitive to $K$. The digital welfare of EU consumers slightly decreases as $K$ increase. This could be driven by the fact that, as the importance of digital consumption grows, the pricing power of digital firms also grow.

# 7 Conclusion

This paper delves into the impact of the General Data Protection Regulation (GDPR) from both firms' and consumers' perspectives, unveiling complex equilibrium effects. I find that US multinational firms, recognizing the adverse impact of GDPR, reallocate their businesses across geographical segments. I also observe that EU consumers endure a less satisfactory consumer experience compared to their US counterparts, suggesting a trade-off between privacy protection and data-dependent user experiences.

The paper introduces a tractable estimation framework that uncovers the value of data and privacy within an equilibrium model, shedding light on the welfare impact of GDPR. Although GDPR is a EU regulation, all firms with business operations in the EU or handling EU consumers' data must comply. Through US multinational firms, GDPR also affects the welfare of US consumers. GDPR, a "pro-choice" regulation envisioned to enhance consumer welfare, inadvertently compromises the welfare of both US and EU consumers, with EU privacy-conscious consumers being the sole beneficiaries. While EU consumers benefit from enhanced privacy protection, the positive welfare impact is dampened by market forces —– data sharing is priced in by digital firms. When data-sharing choices are given back to individuals, they under-supply data due to the failure of internalizing the positive externality on others. Given the burgeoning potential of the digital economy, it's paramount that privacy regulations strike a balance between the efficiency gains from data sharing and consumer privacy protection.

The project augments the expansive discourse on the data economy and privacy regulations from an international perspective. Several US states, following the EU's lead, have enacted similar privacy regulation frameworks, including California, Colorado, Connecticut, Utah, and Virginia. More US states are pondering such regulations, with a new iteration of federal-level privacy regulation, the American Data Privacy and Protection Act, on the horizon. It is crucial to understand the potential impact of data privacy regulations on the delicate interplay between firms and consumers.

# References

Abis, S. and L. Veldkamp (2020): "The Changing Economics of Knowledge Production," *Available at SSRN 3570130.*

Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo (2020): "AI and jobs: Evidence from online vacancies," Tech. rep., National Bureau of Economic Research.

——— (2022a): "Artificial intelligence and jobs: evidence from online vacancies," *Journal of Labor Economics*, 40, S293–S340.

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2022b): "Too much data: Prices and inefficiencies in data markets," *American Economic Journal: Microeconomics*, 14, 218–256.

Admati, A. R. and P. Pfleiderer (1990): "Direct and indirect sale of information," *Econometrica: Journal of the Econometric Society*, 901–928.

Argenziano, R. and A. Bonatti (2023): "Data Markets with Privacy-Conscious Consumers," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 113, 191–196.

Aridor, G., Y.-K. Che, and T. Salz (2020): "The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR," Tech. rep., National Bureau of Economic Research.

Babina, T., G. Buchak, and W. Gornall (2022a): "Customer data access and fintech entry: Early evidence from open banking," .

Babina, T., A. Fedyk, A. X. He, and J. Hodson (2020): "Artificial intelligence, firm growth, and industry concentration," *Firm Growth, and Industry Concentration (November*, 22, 2020.

——— (2022b): "Firm investments in artificial intelligence technologies and changes in workforce composition," *Available at SSRN 4060233.*

Bai, J., J. Li, and A. Manela (2023): “The value of data to fixed income investors,” *Available at SSRN 4343095.*

Begenau, J., M. Farboodi, and L. Veldkamp (2018): “Big data in finance and the growth of large firms,” *Journal of Monetary Economics*, 97, 71–87.

Benkler, Y., R. Faris, and H. Roberts (2018): *Network propaganda: Manipulation, disinformation, and radicalization in American politics*, Oxford University Press.

Bergemann, D., A. Bonatti, and T. Gan (2019): “The economics of social data,” .

Bian, B., X. Ma, and H. Tang (2021): “The supply and demand for data privacy: Evidence from mobile apps,” *Available at SSRN 3987541.*

Bian, B., M. Pagel, and H. Tang (2023): “Consumer surveillance and financial fraud,” Tech. rep., National Bureau of Economic Research.

Bleier, A., A. Goldfarb, and C. Tuckerc (2020): “Consumer privacy and the future of data-based innovation and marketing,” *International Journal of Research in Marketing.*

Bordalo, P., N. Gennaioli, and A. Shleifer (2016): “Competition for attention,” *The Review of Economic Studies*, 83, 481–513.

Braghieri, L. (2019): “Targeted advertising and price discrimination in intermediated online markets,” *Available at SSRN 3072692.*

Campbell, J. L., H. Chen, D. S. Dhaliwal, H.-m. Lu, and L. B. Steele (2014): “The information content of mandatory risk factor disclosures in corporate filings,” *Review of Accounting Studies*, 19, 396–455.

Canayaz, M., I. Kantorovitch, and R. Mihet (2022): “Consumer Privacy and Value of Consumer Data,” *Swiss Finance Institute Research Paper.*

Cao, S., W. Jiang, J. L. Wang, and B. Yang (2021): “From Man vs. Machine to Man+ Machine: The Art and AI of Stock Analyses,” Tech. rep., National Bureau of Economic Research.

Cao, S., W. Jiang, B. Yang, and A. L. Zhang (2020): “How to Talk When a Machine is Listening: Corporate Disclosure in the Age of AI,” Tech. rep., National Bureau of Economic Research.

Carnevale, A. P., T. Jayasundera, and D. Repnikov (2014): “Understanding online job ads data,” *Georgetown University, Center on Education and the Workforce, Technical Report (April).*

Chang, Q., L. W. Cong, L. Wang, and L. Zhang (2023): “Production, Trade, and Cross-Border Data Flows,” Tech. rep., National Bureau of Economic Research.

Chen, D. (2022): “The market for attention,” *Available at SSRN 4024597.*

Chen, L., Y. Huang, S. Ouyang, and W. Xiong (2021): "The data privacy paradox and digital demand," Tech. rep., National Bureau of Economic Research.

Choi, J. P., D.-S. Jeon, and B.-C. Kim (2019): "Privacy and personal data collection with information externalities," *Journal of Public Economics*, 173, 113–124.

Cohn, J. B., Z. Liu, and M. I. Wardlaw (2022): "Count (and count-like) data in finance," *Journal of Financial Economics*, 146, 529–551.

Cong, L. W., W. Wei, D. Xie, and L. Zhang (2022): "Endogenous growth under multiple uses of data," *Journal of Economic Dynamics and Control*, 141, 104395.

Cong, L. W., D. Xie, and L. Zhang (2020): "Knowledge Accumulation, Privacy, and Growth in a Data Economy," *Privacy, and Growth in a Data Economy (October 8, 2020)*.

de Montjoye, Y.-A., T. Ramadorai, T. Valletti, and A. Walther (2021): "Privacy, adoption, and truthful reporting: a simple theory of contact tracing applications," *Economics Letters*, 198, 109676.

Eeckhout, J. and L. Veldkamp (2022): "Data and market power," Tech. rep., National Bureau of Economic Research.

Evans, D. S. (2009): "The online advertising industry: Economics, evolution, and privacy," *Journal of economic perspectives*, 23, 37–60.

Fajgelbaum, P. D., E. Schaal, and M. Taschereau-Dumouchel (2017): "Uncertainty traps," *The Quarterly Journal of Economics*, 132, 1641–1692.

Farboodi, M., A. Matray, and L. Veldkamp (2018): "Where has all the big data gone?" *Available at SSRN 3164360*.

Farboodi, M., R. Mihet, T. Philippon, and L. Veldkamp (2019): "Big data and firm dynamics," in *AEA papers and proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 109, 38–42.

Farboodi, M., D. Singal, L. Veldkamp, and V. Venkateswaran (2022): "Valuing financial data," Tech. rep., National Bureau of Economic Research.

Farboodi, M. and L. Veldkamp (2020): "Long-run growth of financial data technology," *American Economic Review*, 110, 2485–2523.

——— (2021): "A model of the data economy," Tech. rep., National Bureau of Economic Research.

——— (2023): "Data and markets," *Annual Review of Economics*, 15, 23–40.

Godinho de Matos, M. and I. Adjerid (2022): "Consumer consent and firm targeting after GDPR: The case of a large telecom provider," *Management Science*, 68, 3330–3378.

Goldberg, S., G. Johnson, and S. Shriver (2019): “Regulating privacy online: An economic evaluation of the GDPR,” *Available at SSRN 3421731.*

Goldfarb, A. and C. Tucker (2011): “Online display advertising: Targeting and obtrusiveness,” *Marketing Science*, 30, 389–404.

Janssen, R., R. Kesler, M. E. Kummer, and J. Waldfogel (2022): “GDPR and the lost generation of innovative apps,” Tech. rep., National Bureau of Economic Research.

Jia, J., G. Z. Jin, and L. Wagman (2018): “The short-run effects of gdpr on technology venture investment,” Tech. rep., National Bureau of Economic Research.

——— (2020): “GDPR and the Localness of Venture Investment,” *Available at SSRN 3436535.*

Johnson, G. (2013): “The impact of privacy policy on the auction market for online display advertising,” .

——— (2022): “Economic research on privacy regulation: Lessons from the GDPR and beyond,” .

Johnson, G. A., S. K. Shriver, and S. Du (2020): “Consumer privacy choice in online advertising: Who opts out and at what cost to industry?” *Marketing Science*, 39, 33–51.

Johnson, G. A., S. K. Shriver, and S. G. Goldberg (2023): “Privacy and market concentration: intended and unintended consequences of the GDPR,” *Management Science.*

Jones, C. I. and C. Tonetti (2020): “Nonrivalry and the Economics of Data,” *American Economic Review*, 110, 2819–58.

Kirpalani, R. and T. Philippon (2020): “Data sharing and market power with two-sided platforms,” Tech. rep., National Bureau of Economic Research.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017): “Technological innovation, resource allocation, and growth,” *The Quarterly Journal of Economics*, 132, 665–712.

Lenard, T. M. and P. H. Rubin (2013): “The big data revolution: Privacy considerations,” *Technology Policy Institute*, 1–2.

Liu, Z., M. Sockin, and W. Xiong (2023): “Data Privacy and Algorithmic Inequality,” Tech. rep., National Bureau of Economic Research.

Martin, N., C. Matt, C. Niebel, and K. Blind (2019): “How data protection regulation affects startup innovation,” *Information systems frontiers*, 21, 1307–1324.

Ordonez, G. (2013): “The asymmetric effects of financial frictions,” *Journal of Political Economy*, 121, 844–895.

Peukert, C., S. Bechtold, M. Batikas, and T. Kretschmer (2022): "Regulatory spillovers and data governance: Evidence from the GDPR," *Marketing Science*, 41, 746–768.

Ramadorai, T., A. Uettwiller, and A. Walther (2020): "The market for data privacy," *Available at SSRN 3352175*.

Solow, R. M. (1956): "A contribution to the theory of economic growth," *The quarterly journal of economics*, 70, 65–94.

Swan, T. W. (1956): "Economic growth and capital accumulation," *Economic record*, 32, 334–361.

Tang, H. (2019): "The value of privacy: Evidence from online borrowers," *Available at SSRN*.

Van Nieuwerburgh, S. and L. Veldkamp (2006): "Learning asymmetries in real business cycles," *Journal of monetary Economics*, 53, 753–772.

Veldkamp, L. (2023): "Valuing data as an asset," *Review of Finance*, rfac073.

Veldkamp, L. and C. Chung (2019): "Data and the aggregate economy," in *Annual Meeting Plenary*, Society for Economic Dynamics, 2019-1.

Veldkamp, L. L. (2005): "Slow boom, sudden crash," *Journal of Economic theory*, 124, 230–257.

Zhao, Y., P. Yildirim, and P. K. Chintagunta (2021): "Privacy regulations and online search friction: Evidence from GDPR," *Available at SSRN 3903599*.

Zhuo, R., B. Huffaker, S. Greenstein, et al. (2021): "The impact of the general data protection regulation on internet interconnection," *Telecommunications Policy*, 45, 102083.

# Online Appendix

for *Tracing Out International Data Flow: The Value of Data and Privacy*

Junjun Quan

Columbia Business School

[Please Click Here for the Latest Version]

## Contents

# A Proofs

## A.1 Pre-GDPR Solution

In this appendix, I expand the solution displayed in Section 5.5.

**US Multinational Digital Firms** For US multinational firms, their HJB equation can be written as

$$V_j^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}}) = \max_{\{L_{jt,\mathrm{us}}^{\mathrm{us}}\},\{L_{jt,\mathrm{eu}}^{\mathrm{us}}\}} \left(p_{jt,\mathrm{us}}^{\mathrm{us}} Y_{jt,\mathrm{us}}^{\mathrm{us}} + p_{jt,\mathrm{eu}}^{\mathrm{us}} Y_{jt,\mathrm{eu}}^{\mathrm{us}} - w_{t,\mathrm{us}} L_{jt,\mathrm{us}}^{\mathrm{us}} - w_{t,\mathrm{eu}} L_{jt,\mathrm{eu}}^{\mathrm{us}}\right) + \frac{1}{1+r} V_j^{\mathrm{us}}(D_{jt}^{\mathrm{us}}) \tag{64}$$

The first order condition w.r.t. $L_{jt,\mathrm{us}}^{\mathrm{us}}$

$$\underbrace{p_{jt,\mathrm{us}}^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}})^\eta (1-\eta)(L_{jt,\mathrm{us}}^{\mathrm{us}})^\eta}_{\text{current marginal product of labor}} + \underbrace{\frac{1}{1+r} V_j^{\mathrm{us}\prime}(D_{jt}^{\mathrm{us}}) \frac{\partial D_{jt}^{\mathrm{us}}}{\partial L_{jt,\mathrm{us}}^{\mathrm{us}}}}_{\text{future value of data}} = w_{t,\mathrm{us}} \tag{65}$$

where

$$\frac{\partial D_{jt}^{\mathrm{us}}}{\partial L_{jt,\mathrm{us}}^{\mathrm{us}}} = x_{jt,\mathrm{us}}^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}})^\eta (1-\eta)(L_{jt,\mathrm{us}}^{\mathrm{us}})^{-\eta} \tag{66}$$

Substitute the above expression into equation 65 and we can get

$$p_{jt,\mathrm{us}}^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}})^\eta (1-\eta)(L_{jt,\mathrm{us}}^{\mathrm{us}})^{-\eta} + \frac{1}{1+r} V_j^{\mathrm{us}\prime}(D_{jt}^{\mathrm{us}}) x_{jt,\mathrm{us}}^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}})^\eta (1-\eta)(L_{jt,\mathrm{us}}^{\mathrm{us}})^{-\eta} = w_{t,\mathrm{us}} \tag{67}$$

and

$$(L_{jt,\mathrm{us}}^{\mathrm{us}})^\eta = \frac{(1-\eta)(D_{j,t-1}^{\mathrm{us}})^\eta \left((1+r)p_{jt,\mathrm{us}}^{\mathrm{us}} + V_j^{\mathrm{us}\prime}(D_{jt}^{\mathrm{us}}) x_{jt,\mathrm{us}}^{\mathrm{us}}\right)}{(1+r)w_{t,us}} \tag{68}$$

Similarly, we can get the first order condition w.r.t $L_{jt,\mathrm{eu}}^{\mathrm{us}}$.

$$p_{jt,\mathrm{eu}}^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}})^\eta (1-\eta)(L_{jt,\mathrm{eu}}^{\mathrm{us}})^{-\eta} + \frac{1}{1+r} V_j^{\mathrm{us}\prime}(D_{jt}^{\mathrm{us}}) x_{jt,\mathrm{eu}}^{\mathrm{us}}(D_{j,t-1}^{\mathrm{us}})^\eta (1-\eta)(L_{jt,\mathrm{eu}}^{\mathrm{us}})^{-\eta} = w_{t,\mathrm{eu}} \tag{69}$$

$$(L_{jt,\mathrm{eu}}^{\mathrm{us}})^\eta = \frac{(1-\eta)(D_{j,t-1}^{\mathrm{us}})^\eta \left((1+r)p_{jt,\mathrm{eu}}^{\mathrm{us}} + V_j^{\mathrm{us}\prime}(D_{jt}^{\mathrm{us}}) x_{jt,\mathrm{us}}^{\mathrm{us}}\right)}{(1+r)w_{t,eu}} \tag{70}$$

Take the first order derivative of the value function w.r.t. $D^{\text{us}}_{j,t-1}$. By the Envelope Theorem

$$V_j^{\text{us}'}(D^{\text{us}}_{j,t-1}) = p^{\text{us}}_{jt,\text{us}}\eta\left(D^{\text{us}}_{j,t-1}\right)^{\eta-1}\left(L^{\text{us}}_{jt,\text{us}}\right)^{1-\eta} + p^{\text{us}}_{jt,\text{eu}}\eta\left(D^{\text{us}}_{j,t-1}\right)^{\eta-1}\left(L^{\text{us}}_{jt,\text{eu}}\right)^{1-\eta} + \frac{1}{1+r}V_j^{\text{us}'}(D^{\text{us}}_{jt})\frac{\partial D^{\text{us}}_{jt}}{\partial D^{\text{us}}_{j,t-1}} \tag{71}$$

where

$$\frac{\partial D^{\text{us}}_{jt}}{\partial D^{\text{us}}_{j,t-1}} = 1-\kappa_{\text{us}} + x^{\text{us}}_{jt,\text{us}}\eta\left(D^{\text{us}}_{j,t-1}\right)^{\eta-1}\left(L^{\text{us}}_{jt,\text{us}}\right)^{1-\eta} + x^{\text{us}}_{jt,\text{eu}}\eta\left(D^{\text{us}}_{j,t-1}\right)^{\eta-1}\left(L^{\text{us}}_{jt,\text{eu}}\right)^{1-\eta} \tag{72}$$

Therefore,

$$\frac{1}{1+r}V_j^{\text{us}'}(D^{\text{us}}_{jt}) = \frac{V_j^{\text{us}'}(D^{\text{us}}_{j,t-1}) - p^{\text{us}}_{jt,\text{us}}\eta\left(D^{\text{us}}_{j,t-1}\right)^{\eta-1}\left(L^{\text{us}}_{jt,\text{us}}\right)^{1-\eta} - p^{\text{us}}_{jt,\text{eu}}\eta\left(D^{\text{us}}_{j,t-1}\right)^{\eta-1}\left(L^{\text{us}}_{jt,\text{eu}}\right)^{1-\eta}}{1-\kappa_{\text{us}} + x^{\text{us}}_{jt,\text{us}}\eta\left(D^{\text{us}}_{j,t-1}\right)^{\eta-1}\left(L^{\text{us}}_{jt,\text{us}}\right)^{1-\eta} + x^{\text{us}}_{jt,\text{eu}}\eta\left(D^{\text{us}}_{j,t-1}\right)^{\eta-1}\left(L^{\text{us}}_{jt,\text{eu}}\right)^{1-\eta}} \tag{73}$$

We solve the model on the balanced growth path, suppose the stock of data grows at the constant rate $b_{j,\text{us}}$.

$$D^{\text{us}}_{jt} = (1+b^{\text{us}}_j)D^{\text{us}}_{j,t-1} \tag{74}$$

$$\frac{\Delta D^{\text{us}}_{jt}}{D^{\text{us}}_{j,t-1}} = x^{\text{us}}_{jt,\text{us}}\left(D^{\text{us}}_{j,t-1}\right)^{\eta-1}\left(L^{\text{us}}_{jt,\text{us}}\right)^{1-\eta} + x^{\text{us}}_{jt,\text{eu}}\left(D^{\text{us}}_{j,t-1}\right)^{\eta-1}\left(L^{\text{us}}_{jt,\text{eu}}\right)^{1-\eta} - \kappa_{\text{us}} \tag{75}$$

As with the Solow-Swan model (Solow 1956; Swan 1956), we need equilibrium labor to grow proportionally with the data stock.

$$C^{\text{us}}_{j,\text{us}} = \frac{\left(L^{\text{us}}_{jt,\text{us}}\right)^{1-\eta}}{\left(D^{\text{us}}_{j,t-1}\right)^{1-\eta}}, \quad C^{\text{us}}_{j,\text{eu}} = \frac{\left(L^{\text{us}}_{jt,\text{eu}}\right)^{1-\eta}}{\left(D^{\text{us}}_{j,t-1}\right)^{1-\eta}} \tag{76}$$

where $C^{\text{us}}_{j,\text{us}}$ and $C^{\text{us}}_{j,\text{eu}}$ are constants. Then

$$(1+b^{\text{us}}_j)^{1-\eta} = (1+g^{\text{us}}_{j,\text{us}})^{1-\eta} = (1+g^{\text{us}}_{j,\text{eu}})^{1-\eta} \tag{77}$$

where $g^{\text{us}}_{j,\text{us}}$ and $g^{\text{us}}_{j,\text{eu}}$ are the growth rates of equilibrium labor choices $L^{\text{us}}_{jt,\text{us}}$ and $L^{\text{us}}_{jt,\text{eu}}$. Then the growth rates of output $Y^{\text{us}}_{jt,\text{us}}$ and $Y^{\text{us}}_{jt,\text{eu}}$ are

$$g^{\text{us}}_{j,\text{us},y} = g^{\text{us}}_{j,\text{eu},y} = (1+b^{\text{us}}_j)^{\eta}(1+g^{\text{us}}_{j,\text{us}})^{1-\eta} - 1 = b^{\text{us}}_j \tag{78}$$

That is, the equilibrium output grows at the same rate as the state variable data stock. We guess and verify the equilibrium value function as

$$V_j^{\text{us}}(D_{j,t-1}^{\text{us}}) = B_j^{\text{us}} \cdot D_{j,t-1}^{\text{us}} \tag{79}$$

Then we can derive an expression for $B_j^{\text{us}}$

$$B_j^{\text{us}} = \frac{(1+r)\eta\left(p_{jt,\text{us}}^{\text{us}} C_{j,\text{us}}^{\text{us}} + p_{jt,\text{eu}}^{\text{us}} C_{j,\text{eu}}^{\text{us}}\right)}{r + \kappa_{\text{us}} - \eta(b_j^{us} + \kappa_{\text{us}})} \tag{80}$$

Then we have

$$V_j^{\text{us}}(D_{j,t-1}^{\text{us}}) = \frac{(1+r)\eta\left(p_{jt,\text{us}}^{\text{us}} C_{jt,\text{us}}^{\text{us}} + p_{jt,\text{eu}}^{\text{us}} C_{jt,\text{eu}}^{\text{us}}\right)}{r + \kappa_{\text{us}} - \eta(b_j^{us} + \kappa_{\text{us}})} \cdot D_{j,t-1}^{\text{us}} \tag{81}$$

**EU Local Digital Firms** Similar to the US digital firms, we can set up the HJB equation for EU local digital firms.

$$V_j^{\text{eu}}(D_{j,t-1}^{\text{eu}}) = \max_{\{L_{jt,\text{eu}}^{\text{eu}}\}} \left(p_{jt,\text{eu}}^{\text{eu}} Y_{jt,\text{eu}}^{\text{eu}} - w_{t,\text{eu}} L_{jt,\text{eu}}^{\text{eu}}\right) + \frac{1}{1+r} V_j^{\text{eu}}(D_{jt}^{\text{eu}}) \tag{82}$$

Following the same procedures, we can get the first order condition w.r.t. $L_{jt,\text{eu}}^{\text{eu}}$

$$p_{jt,\text{eu}}^{\text{eu}}(D_{j,t-1}^{\text{eu}})^{\eta}(1-\eta)(L_{jt,\text{eu}}^{\text{eu}})^{-\eta} + \frac{1}{1+r} V_{j,\text{eu}}'(D_{jt,\text{eu}}) x_{jt,\text{eu}}^{\text{eu}}(D_{j,t-1}^{\text{eu}})^{\eta}(1-\eta)(L_{jt,\text{eu}}^{\text{eu}})^{-\eta} = w_{t,\text{eu}} \tag{83}$$

Then we can solve for the optimal production decisions.

$$(L_{jt,\text{eu}}^{\text{eu}})^{\eta} = \frac{(1-\eta)(D_{j,t-1}^{\text{eu}})^{\eta}\left((1+r)p_{jt,\text{eu}}^{\text{eu}} + V_j^{\text{eu}\prime}(D_{jt}^{\text{eu}}) x_{jt,\text{eu}}^{\text{eu}}\right)}{(1+r)w_{t,eu}} \tag{84}$$

Take the first order derivative of the value function w.r.t. $D_{j,t-1}^{\text{eu}}$.

$$\frac{1}{1+r} V_j^{\text{eu}\prime}(D_{jt,\text{eu}}) = \frac{V_j^{\text{eu}\prime}(D_{j,t-1}^{\text{eu}}) - p_{jt,\text{eu}}^{\text{eu}}\eta\left(D_{j,t-1}^{\text{eu}}\right)^{\eta-1}\left(L_{jt,\text{eu}}^{\text{eu}}\right)^{1-\eta}}{1 - \kappa_{\text{eu}} + x_{jt,\text{eu}}^{\text{eu}}\eta\left(D_{j,t-1}^{\text{eu}}\right)^{\eta-1}\left(L_{jt,\text{eu}}^{\text{eu}}\right)^{1-\eta}} \tag{85}$$

Again, we solve the model on the balanced growth path and suppose the stock of data grows at the constant rate $b_j^{\text{eu}}$.

$$D_{jt}^{\text{eu}} = (1 + b_j^{\text{eu}}) D_{j,t-1}^{\text{eu}} \tag{86}$$

Let

$$C_{j,\text{eu}}^{\text{eu}} = \frac{\left(L_{jt,\text{eu}}^{\text{eu}}\right)^{1-\eta}}{\left(D_{j,t-1}^{\text{eu}}\right)^{1-\eta}} \tag{87}$$

We also guess and verify that the value function takes the form

$$V_j^{\text{eu}}(D_{j,t-1}^{\text{eu}}) = B_j^{\text{eu}} \cdot D_{j,t-1}^{\text{eu}} \tag{88}$$

Then we can derive an expression for $B_j^{\text{eu}}$

$$B_j^{\text{eu}} = \frac{(1+r)\eta p_{jt,\text{eu}}^{\text{eu}} C_{j,\text{eu}}^{\text{eu}}}{r + \kappa_{\text{eu}} - \eta(b_j^{eu} + \kappa_{\text{eu}})} \tag{89}$$

and

$$V_j^{\text{eu}}(D_{j,t-1}^{\text{eu}}) = \frac{(1+r)\eta p_{jt,\text{eu}}^{\text{eu}} C_{jt,\text{eu}}^{\text{eu}}}{r + \kappa_{\text{eu}} - \eta(b_j^{eu} + \kappa_{\text{eu}})} \cdot D_{j,t-1}^{\text{eu}} \tag{90}$$

**US Household** We can set up the Lagrangian of the US households' optimization problem.

$$\begin{aligned}\mathcal{L}_{it,\text{us}} =& K \int_0^1 \left(\gamma \bar{x}_{jt}^{\text{us}} + (1-\gamma) x_{ijt,\text{us}}^{\text{us}} - \mathbb{1}_{\{z_i=1\}} \delta \left(x_{ijt,\text{us}}^{\text{us}}\right)^2\right) \ln c_{ijt,\text{us}}^{\text{us}} dj + (1-K) \ln c_{it,\text{us}}^{\text{nd}} \\ & \mu_{it,\text{us}} \left(\int_0^1 \Pi_{ijt,\text{us}}^{\text{us}} dj + \Pi_{it,\text{us}}^{\text{nd}} + w_{t,\text{us}} l_{it,\text{us}} - \int_0^1 p_{jt,\text{us}}^{\text{us}} c_{ijt,\text{us}}^{\text{us}} dj - c_{it,\text{us}}^{\text{nd}}\right)\end{aligned} \tag{91}$$

Here we aim to solve for an interior solution. We can derive the first order conditions

$$\text{US digital goods: } \frac{\partial \mathcal{L}_{it,\text{us}}}{\partial c_{ijt,\text{us}}^{\text{us}}} = K \left(\gamma \bar{x}_{jt}^{\text{us}} + (1-\gamma) x_{ijt,\text{us}}^{\text{us}} - \mathbb{1}_{\{z_i=1\}} \delta \left(x_{ijt,\text{us}}^{\text{us}}\right)^2\right) \left(c_{ijt,\text{us}}^{\text{us}}\right)^{-1} - \mu_{it,\text{us}} p_{jt,\text{us}}^{\text{us}} = 0 \tag{92}$$

$$\text{non-digital goods: } \frac{\partial \mathcal{L}_{it,\text{us}}}{\partial c_{it,\text{us}}^{\text{nd}}} = (1-K) \left(c_{it,\text{us}}^{\text{nd}}\right)^{-1} - \mu_{it,\text{us}} = 0 \tag{93}$$

$$\text{budget constraint: } \frac{\partial \mathcal{L}_{it,\text{us}}}{\partial \mu_{it,\text{us}}} = \int_0^1 \Pi_{ijt,\text{us}}^{\text{us}} dj + \Pi_{it,\text{us}}^{\text{nd}} + w_{t,\text{us}} l_{it,\text{us}} - \int_0^1 p_{jt,\text{us}}^{\text{us}} c_{ijt,\text{us}}^{\text{us}} dj - c_{it,\text{us}}^{\text{nd}} \tag{94}$$

From equation 92 and 93, the optimal digital consumption for US households follows

$$c^{\text{us}}_{ijt,\text{us}} = \frac{K\left(\gamma \bar{x}^{\text{us}}_{jt} + (1-\gamma)x^{\text{us}}_{ijt,\text{us}} - \mathbb{1}_{\{z_i=1\}}\delta\left(x^{\text{us}}_{ijt,\text{us}}\right)^2\right)c^{\text{nd}}_{it,\text{us}}}{(1-K)p^{\text{us}}_{jt,\text{us}}} \tag{95}$$

Along with the budget constraint, we can get the optimal non-digital consumption

$$\begin{aligned} c^{\text{nd}}_{it,\text{us}} =& \frac{(1-K)\Pi_{it,\text{us}}}{K\int_0^1\left(\gamma \bar{x}^{\text{us}}_{jt} + (1-\gamma)x^{\text{us}}_{ijt,\text{us}} - \mathbb{1}_{\{z_i=1\}}\delta\left(x^{\text{us}}_{ijt,\text{us}}\right)^2\right)dj + (1-K)} \\ =& \frac{(1-K)\Pi_{it,\text{us}}}{KX_{it,\text{us}} + (1-K)} \end{aligned} \tag{96}$$

where

$$X_{it,\text{us}} = \int_0^1 \left(\gamma \bar{x}^{\text{us}}_{jt} + (1-\gamma)x^{\text{us}}_{ijt,\text{us}} - \mathbb{1}_{\{z_i=1\}}\delta\left(x^{\text{us}}_{ijt,\text{us}}\right)^2\right)dj \tag{97}$$

Then it follows from equation 95 that the optimal digital consumption is given by

$$c^{\text{us}}_{ijt,\text{us}} = \frac{K\left(\gamma \bar{x}^{\text{us}}_{jt} + (1-\gamma)x^{\text{us}}_{ijt,\text{us}} - \mathbb{1}_{\{z_i=1\}}\delta\left(x^{\text{us}}_{ijt,\text{us}}\right)^2\right)\Pi_{it,\text{us}}}{p_{jt,\text{us}}\left(KX_{it,\text{us}} + (1-K)\right)} \tag{98}$$

**EU Household** Similarly, we can solve for the optimal consumption choices of EU households. For European households, their optimal consumption of US digital products will be

$$c^{\text{us}}_{ijt,\text{eu}} = \frac{K\beta\left(\gamma \bar{x}^{\text{us}}_{jt} + (1-\gamma)x^{\text{us}}_{ijt,\text{eu}} - \mathbb{1}_{\{z_i=1\}}\delta\left(x^{\text{us}}_{ijt,\text{eu}}\right)^2\right)\Pi_{it,\text{eu}}}{p^{\text{us}}_{jt,\text{eu}}\left(K\left(\beta X^{\text{us}}_{it,\text{eu}} + (1-\beta)X^{\text{eu}}_{it,\text{eu}}\right) + (1-K)\right)} \tag{99}$$

where

$$\begin{aligned} X^{\text{us}}_{it,\text{eu}} &= \int_0^1 \left(\gamma \bar{x}^{\text{us}}_{jt} + (1-\gamma)x^{\text{us}}_{ijt,\text{eu}} - \mathbb{1}_{\{z_i=1\}}\delta\left(x^{\text{us}}_{ijt,\text{eu}}\right)^2\right)dj \\ X^{\text{eu}}_{it,\text{eu}} &= \int_0^1 \left(\gamma \bar{x}^{\text{eu}}_{jt} + (1-\gamma)x^{\text{eu}}_{ijt,\text{eu}} - \mathbb{1}_{\{z_i=1\}}\delta\left(x^{\text{eu}}_{ijt,\text{eu}}\right)^2\right)dj \end{aligned} \tag{100}$$

and their consumption of EU local digital products will be

$$c_{ijt,\mathrm{eu}}^{\mathrm{eu}} = \frac{K(1-\beta)\left(\gamma \bar{x}_{jt}^{\mathrm{eu}} + (1-\gamma) x_{ijt,\mathrm{eu}}^{\mathrm{eu}} - \mathbb{1}_{\{z_i=1\}} \delta \left(x_{ijt,\mathrm{eu}}^{\mathrm{eu}}\right)^2\right) \Pi_{it,\mathrm{eu}}}{p_{jt,\mathrm{eu}}^{\mathrm{eu}} \left(K\left(\beta X_{it,\mathrm{eu}}^{\mathrm{us}} + (1-\beta) X_{it,\mathrm{eu}}^{\mathrm{eu}}\right) + (1-K)\right)} \tag{101}$$

EU households' consumption of non-digital products is given by

$$c_{it,\mathrm{eu}}^{\mathrm{nd}} = \frac{(1-K)\Pi_{it,\mathrm{eu}}}{K\left(\beta X_{it,\mathrm{eu}}^{\mathrm{us}} + (1-\beta) X_{it,\mathrm{eu}}^{\mathrm{eu}}\right) + (1-K)} \tag{102}$$

# B Additional Figures

## B.1 Privacy-Related Risk Factor Disclosure

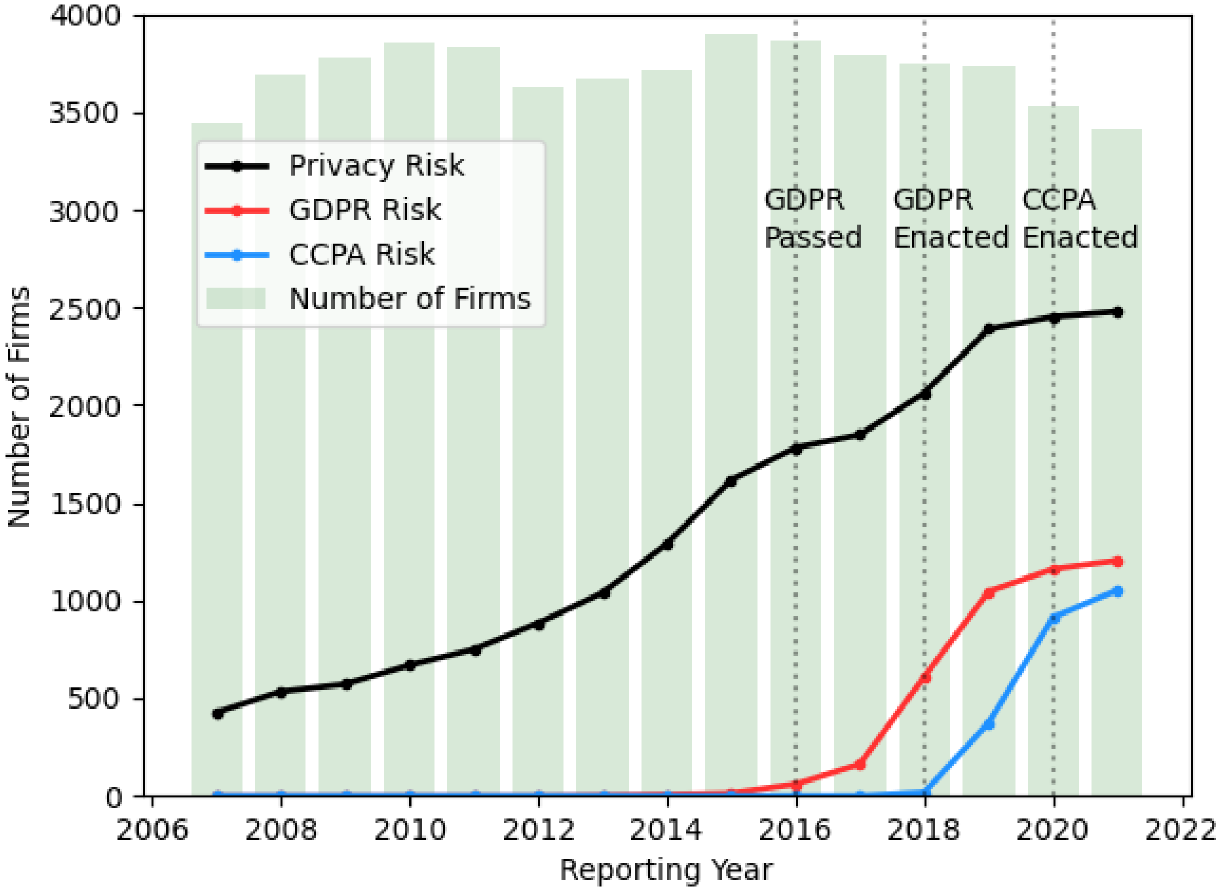


Figure A1: Increasing Number of US Public Firms Disclosing Privacy-Related Risk Factors

*Notes*: The light green bar shows the number of US public firms with valid risk factor disclosures (Item 1A) in their annual 10-K filings. The number of US public firms is around 3500-4000 from 2006 to 2021, so the sample covers most of the US public firms. The black line shows the number of US public firms that disclose any privacy related risk; the red line shows the number of US public firms that disclose GDPR related risk; and the blue line shows the number of US public firms that disclose CCPA related risk.

## B.2 The Two Roles of Data

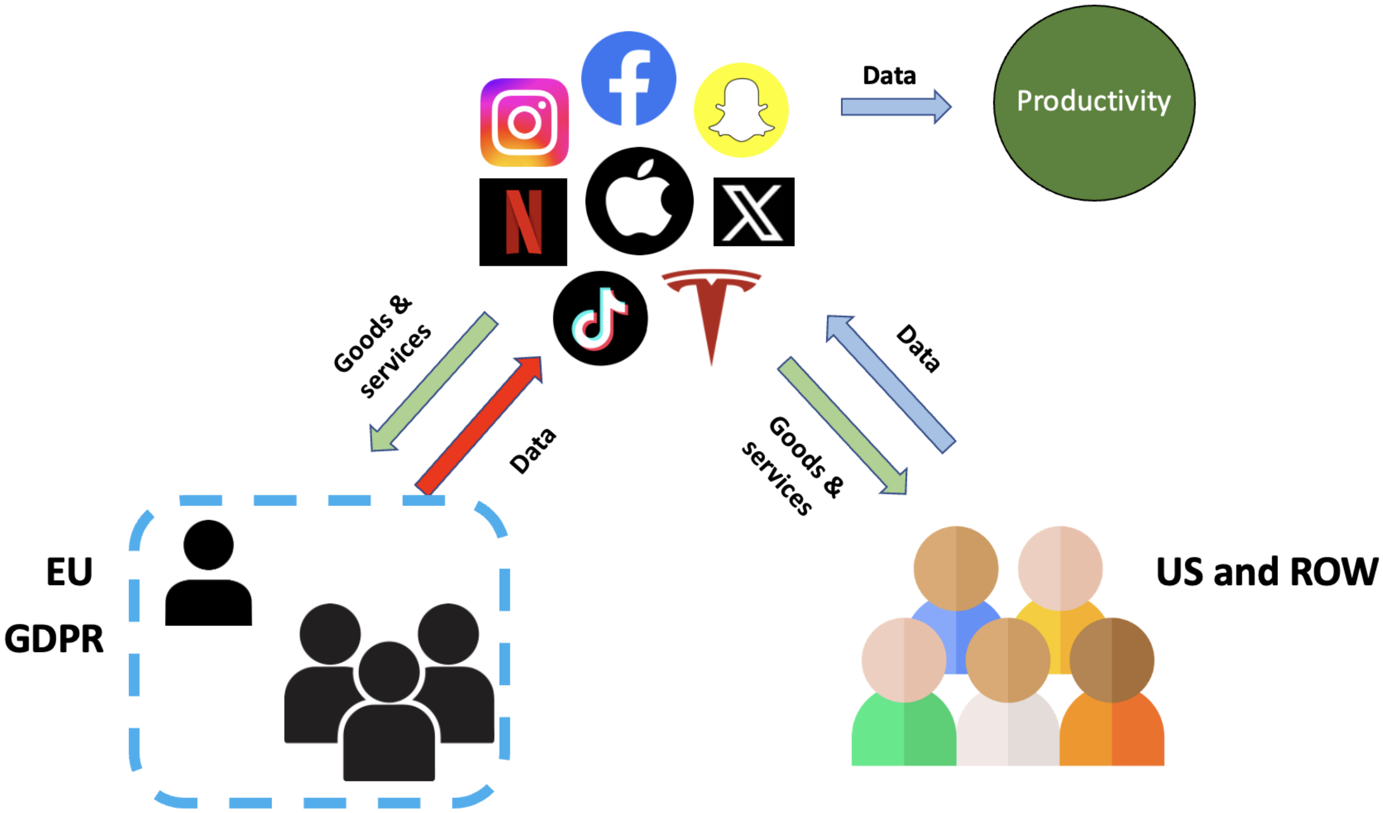
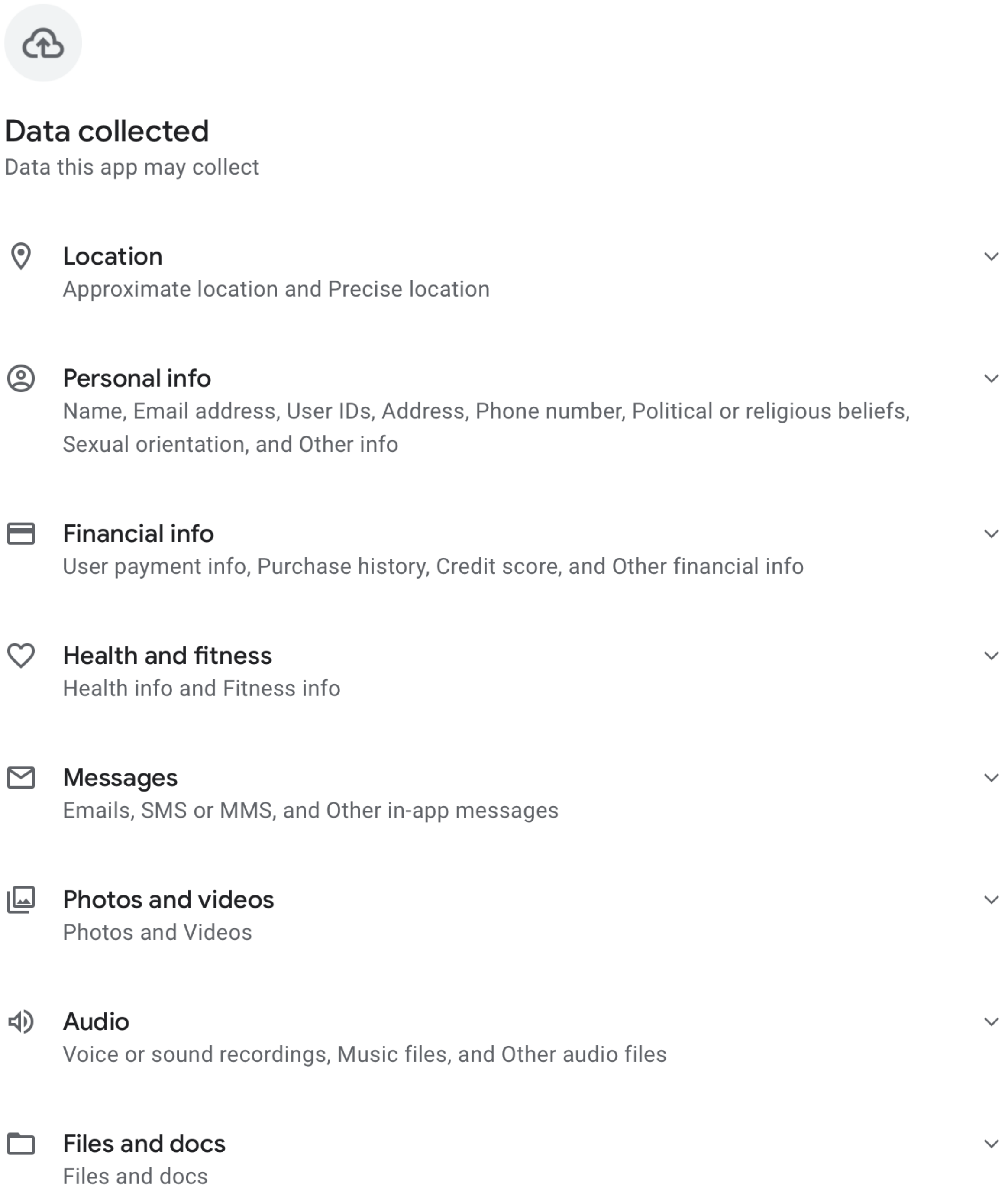


Figure A2: Data Increases Productivity and Personalizes Products

*Notes*: The figure illustrates the two main roles of data: first, data can be used by firms to develop new technology and boost productivity; second, data can be used by firms to tailor products to consumers' preferences. Related to the two main roles, the figure also shows how a regional privacy regulation like GDPR will affect US multinational digital firms and their customers. US multinational firms provide goods and services to both European customers and US customers (or, more accurately, customers from the rest of the world). Data is a byproduct of economic activities, and firms collect and analyze consumers' data to learn about their preferences and boost productivity. While consumers enjoy the advantages of personalized recommendations and enhanced service quality, they have concerns about sharing personal data with firms. After GDPR came into effect, EU privacy-conscious consumers can choose to share less data, but they do not internalize the positive externality of data sharing they have on others. Data is valuable for firms, and consumers' data sharing behaviors will be priced in by firms. Firms may adjust their business operations in each region according to the privacy regulations. They may shift away from the European market to other parts of the world where data access is more abundant.

## B.3 Instagram Data Safety Section on Google Play Store



Data collected

Data this app may collect

Location

Approximate location and Precise location

Personal info

Name, Email address, User IDs, Address, Phone number, Political or religious beliefs, Sexual orientation, and Other info

Financial info

User payment info, Purchase history, Credit score, and Other financial info

Health and fitness

Health info and Fitness info

Messages

Emails, SMS or MMS, and Other in-app messages

Photos and videos

Photos and Videos

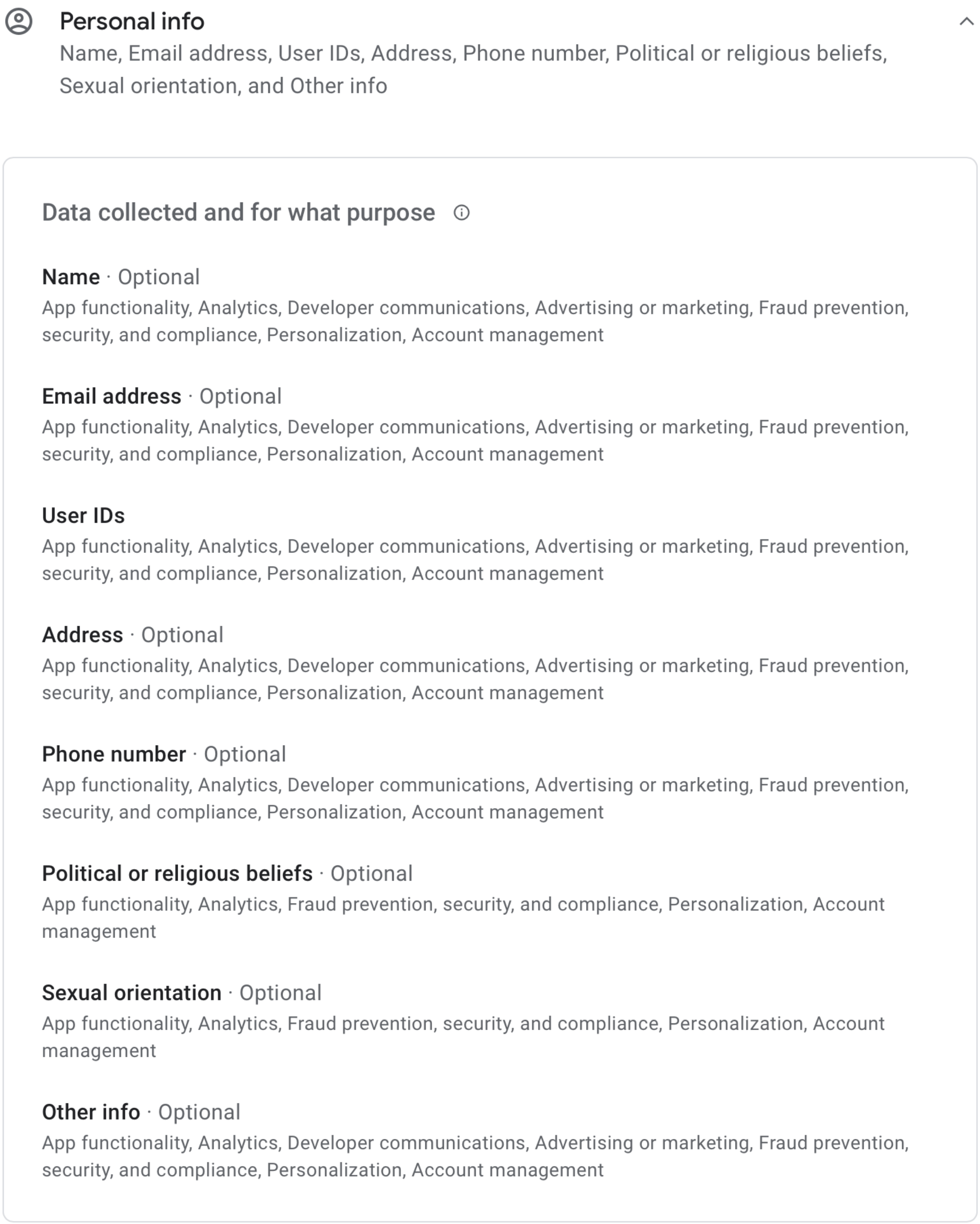
Audio

Voice or sound recordings, Music files, and Other audio files

Files and docs

Files and docs

Figure A3: Instagram: Data Collected



Personal info

Name, Email address, User IDs, Address, Phone number, Political or religious beliefs, Sexual orientation, and Other info

Data collected and for what purpose

**Name** · Optional

App functionality, Analytics, Developer communications, Advertising or marketing, Fraud prevention, security, and compliance, Personalization, Account management

**Email address** · Optional

App functionality, Analytics, Developer communications, Advertising or marketing, Fraud prevention, security, and compliance, Personalization, Account management

**User IDs**

App functionality, Analytics, Developer communications, Advertising or marketing, Fraud prevention, security, and compliance, Personalization, Account management

**Address** · Optional

App functionality, Analytics, Developer communications, Advertising or marketing, Fraud prevention, security, and compliance, Personalization, Account management

**Phone number** · Optional

App functionality, Analytics, Developer communications, Advertising or marketing, Fraud prevention, security, and compliance, Personalization, Account management

**Political or religious beliefs** · Optional

App functionality, Analytics, Fraud prevention, security, and compliance, Personalization, Account management

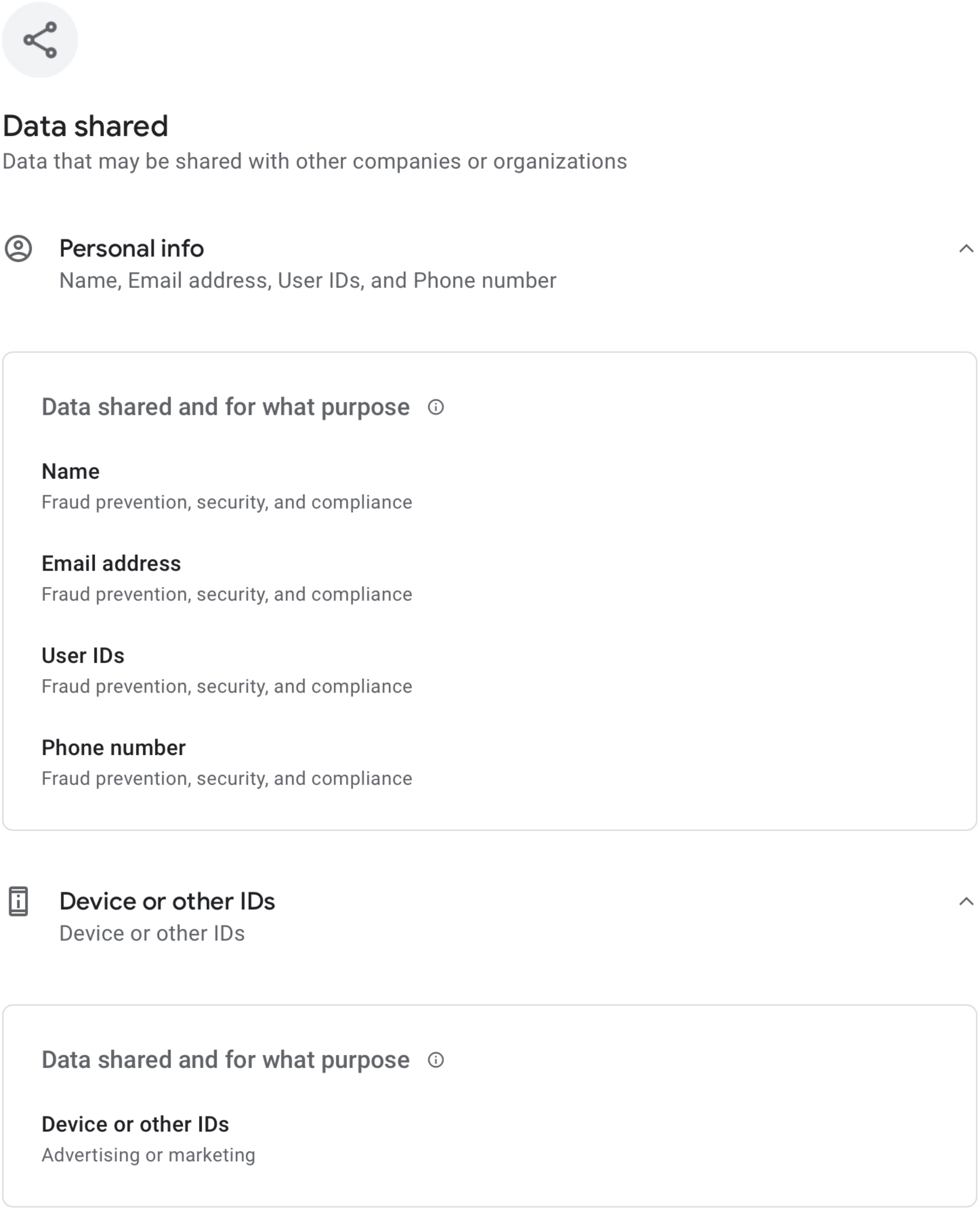
**Sexual orientation** · Optional

App functionality, Analytics, Fraud prevention, security, and compliance, Personalization, Account management

**Other info** · Optional

App functionality, Analytics, Developer communications, Advertising or marketing, Fraud prevention, security, and compliance, Personalization, Account management

Figure A4: Instagram: Data Collection Purpose



## Data shared

Data that may be shared with other companies or organizations

**Personal info**

Name, Email address, User IDs, and Phone number

Data shared and for what purpose

**Name**
Fraud prevention, security, and compliance

**Email address**
Fraud prevention, security, and compliance

**User IDs**
Fraud prevention, security, and compliance

**Phone number**
Fraud prevention, security, and compliance

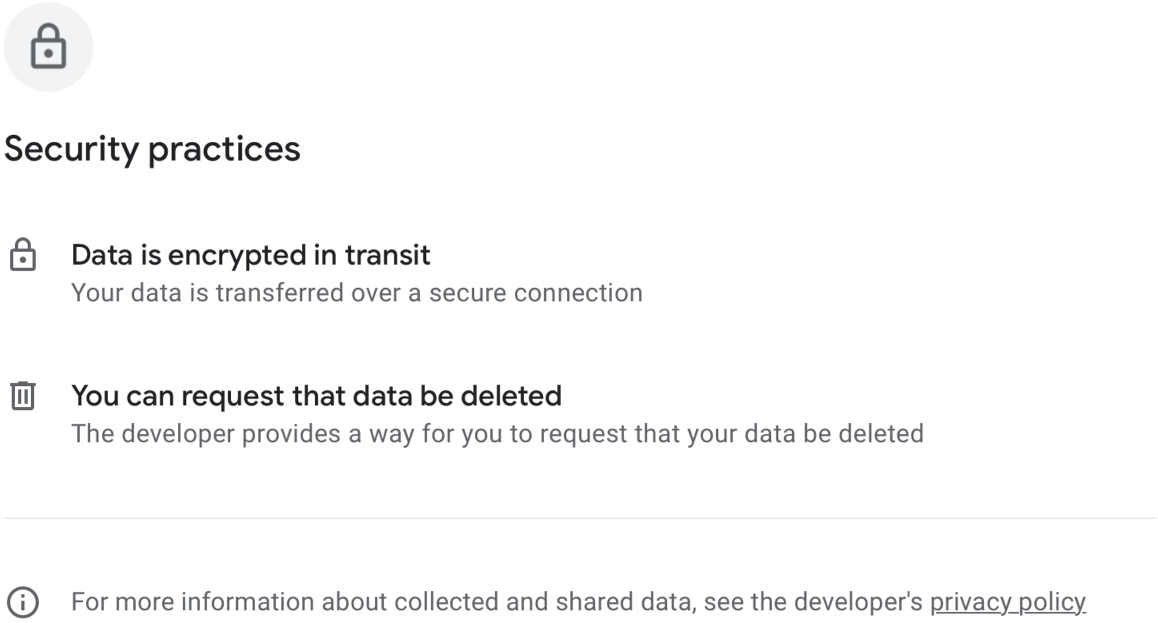
**Device or other IDs**

Device or other IDs

Data shared and for what purpose

**Device or other IDs**
Advertising or marketing

Figure A5: Instagram: Data Shared



Security practices

Data is encrypted in transit
Your data is transferred over a secure connection

You can request that data be deleted
The developer provides a way for you to request that your data be deleted

For more information about collected and shared data, see the developer's privacy policy

Figure A6: Instagram: Security Practice

# C Additional Tables

## C.1 Data-Intensive Patents

Table A1: Data-Intensive Patents from CPC Classification

*Notes:* In this table, I present the Cooperative Patent Classification categories that are classified as data-intensive.

| CPC Code | Description |
|---|---|
| G06F | Electric Digital Data Processing |
| G06N | Computing Arrangements Based On Specific Computational Models |
| G06Q | Information And Communication Technology [ICT] Specially Adapted For Administrative, Commercial, Financial, Managerial Or Supervisory Purposes; Systems Or Methods Specially Adapted For Administrative, Commercial, Financial, Managerial Or Supervisory Purposes, Not Otherwise Provided For |
| G06T | Image Data Processing Or Generation, In General |
| G06V | Image Or Video Recognition Or Understanding |
| G16 | Information And Communication Technology [ICT] Specially Adapted For Specific Application Fields |

## C.2 Sales by Region (Poisson Fixed-Effect Regression)

Table A2: Sales by Region (Poisson Fixed-Effect Regression)

*Notes:* In this table, I estimate a Poisson fixed effect regression on EU sales and sales from other regions of the world.

$$Y_{i,t} = \text{Poisson}\left(\alpha_t + \phi_i + \beta_1 \cdot \text{GDPR}_t \times \text{Data-Intensive}_i + \boldsymbol{\gamma}\boldsymbol{X}_{i,t} + \varepsilon_{i,t}\right)$$

$\alpha_t$ is the year fixed-effect, $\phi_i$ is the firm fixed-effect, and $\boldsymbol{X}_{i,t}$ is a vector of time-varying firm-level characteristics, including book to market ratio and total assets (one-period lagged). $\text{GDPR}_t$ is a binary variable that equals one if time $t$ is after GDPR's enactment date, May 2018. $\text{Data-Intensive}_i$ is a binary variable that equals one if firm $i$ is in the data-intensive category. I define data intensiveness in section 2.2.1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

| Dependent Variable: | EU Sales | Other Sales | US Sales |
|---|---|---|---|
| | (1) | (2) | (3) |
| GDPR Effective × Data-Intensive (binary) | -0.095*** | 0.117*** | 0.126*** |
| | (-41.981) | (128.294) | (107.424) |
| Controls | Yes | Yes | Yes |
| Controls × Data-Intensive (binary) | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| Firm FE | Yes | Yes | Yes |
| Observations | 8,202 | 8,295 | 8,037 |

## C.3 EU Segment and Firm-Level Profitability

Table A3: EU Segment and Firm-Level Profitability

*Notes:* In this table, I look at how the profitability of US multinational firms changes after GDPR' enactment. I run the following regression.

$$Y_{i,t} = \alpha_t + \phi_i + \beta_1 \cdot \text{GDPR}_t \times \text{Data-Intensive}_i + \boldsymbol{\gamma X}_{i,t} + \varepsilon_{i,t}$$

$\alpha_t$ is the year fixed-effect, $\phi_i$ is the firm fixed-effect, and $\boldsymbol{X}_{i,t}$ is a vector of time-varying firm-level characteristics, including book to market ratio and firm assets. $\text{GDPR}_t$ is a binary variable that equals one if time $t$ is after GDPR's enactment date, May 2018. $\text{Data-Intensive}_i$ is a binary variable that equals one if firm $i$ is in the data-intensive category. I focus on the firms with significant European market operations before 2018 and divide the sample into the more data-intensive group (above median) and the less data-intensive group (below median). I define data intensiveness in section 2.2.1. For the dependent variable, $Y_{i,t}$, I look into two measures of profitability, gross profit margin (GPM) and operating profit margin (OPM) in percentage points. The profitability measures are winterized at the 0.5% level on both ends. The standard errors are clustered at the industry level. t-statistics are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

| | EU GPM | EU OPM | Firm GPM | Firm OPM |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| GDPR Effective × Data-Intensive | -0.224 | 0.018 | 0.062 | 0.021 |
| | (-1.384) | (0.886) | (0.507) | (0.251) |
| Controls | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Firm FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.919 | 0.907 | 0.685 | 0.696 |
| Observations | 139 | 604 | 9,085 | 9,085 |

## C.4 Cross-Market Business Adjustment with Tech Controls

Table A4: Cross-Market Business Adjustment with Tech Controls

I employ a difference-in-differences design and study how US multinational firms respond when their access to EU consumers' data is restricted. Since most US firms report their geographical revenue compositions at an annual frequency, the observations of the sample used in this table are at the firm-year level. I run the following regression.

$$Y_{i,t} = \alpha_t + \phi_i + \beta_1 \cdot \text{GDPR}_t \times \text{Data-Intensive}_i + \boldsymbol{\gamma X}_{i,t} + \varepsilon_{i,t} \tag{103}$$

$\alpha_t$ is the year fixed effect, $\phi_i$ is the firm fixed effect, and $\boldsymbol{X}_{i,t}$ is a vector of time-varying firm-level characteristics, including book to market ratio and firm assets. $\text{GDPR}_t$ is a binary variable that equals one if time $t$ is after GDPR's enactment date, May 2018. $\text{Data-Intensive}_i$ is a binary variable that equals one if firm $i$ is in the data-intensive category. I focus on the firms with significant European market operations before 2018 and divide the sample into the more data-intensive group (above median) and the less data-intensive group (below median). I define data intensiveness in section 2.2.1. For the dependent variable, $Y_{i,t}$, I first look at the fraction of revenue generated from the European market by US firms in column (1). In column (2), I include one extra interaction term, GDPR Pass × Data-Intensive, which captures the time period between GDPR's passage and enactment. In columns (3) and (4), I look at total sales and sales scaled by total assets. The standard errors are clustered at the industry level. t-statistics are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

| Dependent Variable: | EU Sale Percentage | Sales/Assets | EU Sales/Assets |
|---|---|---|---|
| | (1) | (2) | (3) |
| GDPR Effective × Data-Intensive | -1.020*** | 0.019 | -1.717*** |
| | (-2.606) | (0.014) | (-3.415) |
| GDPR Effective × Tech Industry | -1.088* | 1.951 | -0.695 |
| | (-1.816) | (1.076) | (-1.177) |
| Controls | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| Firm FE | Yes | Yes | Yes |
| $R^2$ | 0.767 | 0.850 | 0.733 |
| Observations | 10,352 | 10,950 | 10,362 |

## C.5 Annual Purchase and Subscription Related Comments

Table A5: Annual Purchase and Subscription Related Comments

*Notes:* I employ a difference-in-differences design and study how limited access to data in the European market affects the number of comments related to purchase and subscriptions. The observations in this analysis are at the app-year level. In columns (1) and (2), I run the following regression:

$$\ln(1 + Y_{i,t}) = \alpha_t + \phi_i + \beta_1 \cdot \text{GDPR}_t \times \text{Target Advertising}_i + \varepsilon_{i,t}$$

where $Y_{i,t}$ is the total number of purchase related comments for app $i$ in year $t$. $\alpha_t$ is the year fixed effect, $\phi_i$ is the app fixed effect. $\text{GDPR}_t$ is a binary variable that equals one if time $t$ is after GDPR's enactment year, 2018. $\text{Target Advertising}_i$ is a binary variable that equals one if app $i$ collects user data for targeted advertising purposes. I analyze the reviews left by the EU and US users separately. In column (3), I run a triple difference regression:

$$\begin{aligned}\ln(1 + Y_{i,k,t}) =& \alpha_t + \phi_i + \psi_k + \beta_1^* \cdot \text{GDPR}_t \times \text{Target Advertising}_i \times \text{EU}_k \\ &+ \beta_2 \cdot \text{GDPR}_t \times \text{Target Advertising}_i + \beta_3 \cdot \text{GDPR}_t \times \text{EU}_k \\ &+ \beta_4 \cdot \text{Target Advertising}_i \times \text{EU}_k + \varepsilon_{i,k,t}\end{aligned}$$

where $Y_{i,k,t}$ is the total number of purchase related comments by users in region $k$ for app $i$ in year $t$. $\psi_k$ is the region (US or EU) fixed effect. $\text{EU}_k$ is an indicator variable that equals one if the reviews come from the EU users. The coefficient $\beta_1^*$ before the triple interaction term captures the differential change in the prevalence of paid services and subscriptions between the EU and US mobile app markets. t-statistics are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

| Dependent Variable: | EU Users | US Users | All |
|---|---|---|---|
| ln(1+Annual # of Purchase Comments) | (1) | (2) | (3) |
| GDPR Effective × Target Advertising × EU | | | 0.090** |
| | | | (2.383) |
| GDPR Effective × Target Advertising | 0.205*** | 0.117*** | 0.115*** |
| | (5.318) | (3.834) | (3.645) |
| GDPR Effective × EU | | | 0.030 |
| | | | (1.221) |
| Target Advertising × EU | | | -0.200*** |
| | | | (-3.667) |
| Year FE | Yes | Yes | Yes |
| App FE | Yes | Yes | Yes |
| Region FE | No | No | Yes |
| $R^2$ | 0.791 | 0.813 | 0.652 |
| Observations | 33,328 | 37,247 | 70,575 |

## C.6 Total Annual Reviews

Table A6: Total Annual Reviews

*Notes:* I employ a difference-in-differences design and study how limited access to data in the European market affects the total number of mobile app reviews. The observations in this analysis are at the app-year level. In columns (1) and (2), I run the following regression:

$$\ln(1 + Y_{i,t}) = \alpha_t + \phi_i + \beta_1 \cdot \text{GDPR}_t \times \text{Target Advertising}_i + \varepsilon_{i,t}$$

$Y_{i,t}$ is the total number of reviews for app $i$ in year $t$. $\alpha_t$ is the year fixed effect, $\phi_i$ is the app fixed effect. $\text{GDPR}_t$ is a binary variable that equals one if time $t$ is after GDPR's enactment year, 2018. $\text{Target Advertising}_i$ is a binary variable that equals one if app $i$ collects user data for targeted advertising purposes. I analyze the reviews left by the EU and US users separately. In column (3), I run a triple difference regression:

$$\begin{aligned}\ln(1 + Y_{i,k,t}) =& \alpha_t + \phi_i + \psi_k + \beta_1^* \cdot \text{GDPR}_t \times \text{Target Advertising}_i \times \text{EU}_k \\ &+ \beta_2 \cdot \text{GDPR}_t \times \text{Target Advertising}_i + \beta_3 \cdot \text{GDPR}_t \times \text{EU}_k \\ &+ \beta_4 \cdot \text{Target Advertising}_i \times \text{EU}_k + \varepsilon_{i,k,t}\end{aligned}$$

where $Y_{i,k,t}$ is the total number of reviews by users in region $k$ for app $i$ in year $t$. $\text{EU}_i$ is an indicator variable that equals one if the reviews come from the EU users. The coefficient $\beta_1^*$ before the triple interaction term captures the differential change in total reviews between the EU and US mobile app markets. t-statistics are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

| Dependent Variable: | EU Users | US Users | All |
|---|---|---|---|
| ln(1+Annual # of Reviews) | (1) | (2) | (3) |
| GDPR Effective × Target Advertising × EU | | | 0.067 |
| | | | (1.563) |
| GDPR Effective × Target Advertising | 0.033 | -0.014 | -0.025 |
| | (0.643) | (-0.327) | (-0.569) |
| GDPR Effective × EU | | | 0.209*** |
| | | | (6.942) |
| Target Advertising × EU | | | -0.047 |
| | | | (-0.632) |
| Year FE | Yes | Yes | Yes |
| App FE | Yes | Yes | Yes |
| Region FE | No | No | Yes |
| $R^2$ | 0.827 | 0.808 | 0.694 |
| Observations | 33,328 | 37,247 | 70,575 |

# D Skill Keywords

## D.1 AI Skills

Sentiment Analysis, Random Forests, Maximum Entropy Classifier, LDA, TensorFlow, Deep Learning, Classification Algorithms, Machine Learning, Libsvm, Latent Semantic Analysis, Backpropagation, Text Mining, Convolutional Neural Network, Geospatial Intelligence, Xgboost, Torch, NLP, Speech Recognition, Gradient Boosting, Neural Network, Long Short-Term Memory, Platfora, Latent Dirichlet Allocation, Nearest Neighbor, Reinforcement Learning, Neuroscience, Neural Nets, Recurrent Neural Network, Lasso, Pattern Recognition, Semi-Supervised Learning, Conditional Random Field, Natural Language Processing, Computer Vision, Artificial Intelligence, ND4J, Kernel Methods, Instance-Based Learning, Microsoft Cognitive Toolkit, Xgboost, Sentiment Classification, Long Short-Term Memory, LSTM, Libsvm, RNN, Word2Vec, MXNet, Caffe Deep Learning Framework, Autoencoders, MLPACK, Keras, Theano, Torch, Wabbit, Boosting, TensorFlow, Vowpal, Convolutional Neural Network, CNN, JUNG framework, OpenNLP, Natural Language Toolkit, NLTK, Unsupervised Learning, Dlib, Scikit-learn, Latent Semantic Analysis, Latent Dirichlet Allocation, Stochastic Gradient Descent, SGD, Dimensionality Reduction, Deep Learning, DBSCAN, Density-Based Spatial Clustering of Applications with Noise, AI ChatBot, Recommender Systems, Random Forests, Deeplearning4j, AdaBoost Algorithm, Support Vector Machines, SVM, Unstructured Information Management Architecture, Apache UIMA, Maximum Entropy Classifier, Pybrain, Computational Linguistics, Naive Bayes, H2O (software), WEKA, Clustering Algorithms, Matrix Factorization, Object Recognition, Classification Algorithms, Information Extraction, Image Recognition, Bayesian Networks, Supervised Learning, OpenCV, K-Means, Opinion Mining, Neural Networks, Support Vector Machine, Computer Vision, DBSCAN, Image Recognition, Mahout, Computational Linguistics, Object Recognition, Opinion Mining, Caffe Deep Learning Framework, Automatic Speech Recognition, Artificial Intelligence, Evolutionary Algorithm, Virtual Agents, Decision Trees, Predictive Models, Genetic Algorithm, Chatbot, OpenCV, Random Forest, Scikit-learn, Machine Translation, Elastic-Net, Keras, Ridge Regression, Image Processing, Big Data Analytics.

## D.2 Data Management Skills

Apache Hive, Information Retrieval, Data Management Platform, DMP, Data Collection, Data Warehousing, SQL Server, Data Visualization, Database Management, Data Governance, Data Transformation, Extensible Markup Language, XML, Data Validation, Data Architecture, Data Mapping, Oracle PL, SQL, Database Design, Data Integration, Teradata, Database Administration, BigTable, Data Security, Database Software, Data Integrity, File Management, Splunk, Relational DataBase Management System, Teradata DBA, Data Migration, Information Assurance, Enterprise Data Management, SSIS, Sybase, jQuery, Data Conversion, Data Acquisition, Master Data Management, Data Capture, Data Verification, MongoDB, Data Warehouse Processing, SAP HANA, Data Loss Prevention, Data Engineering, Database Schemas, Database Architecture, Data Documentation, Data Operations, Oracle Big Data, Domo, Data Manipulation, Data Management Platform, DMP, HyperText Markup Language, Data Access Object, DAO, Structured Query Reporter, SQR, Data Dictionary System, Data Entry, Data Quality, Data Collection, Information Systems, Information Security, Change data capture, Data Management, Data Governance, Data Encryption, Data Cleaning, Semi-Structured Data, Data Evaluation, Data Privacy, Dimensional and Relational Modeling, Data Loss Prevention, Data Operations, Relational Database Design, Database Programming, Information Systems Management, Database Tuning, Object Relational Mapping, Columnar Databases, Datastage, Data Taxonomy, Informatica Data Quality, Data Munging, Data Archiving, Warehouse Operations, Solaris, Data Modeling, Data Feed management, Data discovery, Exporting Large Datasets, Exporting Datasets, Database Performance, Designing Relational databases, Implementing Relational Databases, Designing and Implementing Relational Databases, Database Development, Data Production Process, Normalize Large Datasets, Normalize Datasets, Create Database, Develop Database, Data Onboarding, Data Sourcing, Data Purchase, Data Inventory, Cloud Security, Negotiating Data, Data Attorney, Data and Technology Attorney, Reliability Engineering, Reliability Engineer, Data Specialist, Enable Vast Data Analysis, Enable Data Analysis, Data Team, Capturing Data, Processing Data, Supporting Data, Error Free Data Sets, Error Free Datasets, Live Streams of Data, Data Accumulation, Kernel Level Development, Large Scale Systems, Hadoop, Distributed Computing, Multi Database Web Applications, Connect Software Packages to Internal and External Data, Explore Data Possibilities, Architect Complex Systems, Build Scalable Infrastructure for Data Analysis, Build Infrastructure for Data Analysis, Solutions for at Scale Data Exploration, Solutions for Data Exploration, Information Technology Security, Security Engineer, Security Architect.