



**HAL**  
open science

## Competition and the Two Margins of Privacy

David Bounie, Antoine Dubus, Patrick Waelbroeck

► **To cite this version:**

David Bounie, Antoine Dubus, Patrick Waelbroeck. Competition and the Two Margins of Privacy. 2024. hal-04427224v2

**HAL Id: hal-04427224**

**<https://hal.science/hal-04427224v2>**

Preprint submitted on 2 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Competition and the Two Margins of Privacy\*

David Bounie,<sup>†</sup> Antoine Dubus<sup>‡</sup> and Patrick Waelbroeck<sup>§</sup>

May 2, 2024

## Abstract

This article analyzes the relationship between privacy protection and market competition. We consider a model where firms collect data to price discriminate consumers in a competitive product market, and we distinguish two margins of privacy. Firms strategically choose the number of consumers on whom they collect data – the extensive margin of privacy – as well as the precision of information – the intensive margin of privacy. We show that policymakers can efficiently protect both margins of privacy and consumer surplus by safeguarding the intensive margin. Indeed, when both strategic variables are strategic complements, restricting the amount of information that firms have on each consumer (the intensive margin) also induces firms to collect data on fewer consumers, thereby protecting the extensive margin of privacy. This softens the intensity of competition but also reduces rent extraction by firms, and total consumer surplus increases. When both variables are strategic substitutes, protecting the intensive margin harms privacy at the extensive margin, but still increases consumer surplus.

---

\*We thank for useful remarks and comments members of the French Data Protection Agency (CNIL), and participants to the EARIE conference. Patrick Waelbroeck thanks for insightful discussions members of the Chair Values and Policies of Personal Information of Institut Mines Télécom, Paris. This research has also been conducted within the Chair "Digital Finance" under the aegis of the Risk Foundation, a joint initiative by Groupement des Cartes Bancaires CB, Caisse des Dépôts, INSEE, Télécom Paris and University of Paris 2 Panthéon – Assas.

<sup>†</sup>CREST, Télécom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France; [david.bounie@telecom-paris.fr](mailto:david.bounie@telecom-paris.fr).

<sup>‡</sup>Department of Management, Technology and Economics, ETH Zürich Leonhardstrasse 21 – 8092, Switzerland; [antoine1dubus@gmail.com](mailto:antoine1dubus@gmail.com).

<sup>§</sup>CREST, Télécom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France; [patrick.waelbroeck@telecom-paris.fr](mailto:patrick.waelbroeck@telecom-paris.fr).

# 1 Introduction

Privacy protection and competition are two central issues in the current digital economy. Data-driven services are developing rapidly, they have become integral to various digital markets and are essential for making informed decisions, enhancing user experience, and driving innovation. However, they have also raised concerns about privacy and the need for effective data protection regulations.

Privacy regulations play a crucial role in protecting the personal data of consumers, but can also have an impact on market competition in many contrasting ways. It may first encourage competition by leveling the playing field between firms that offer different levels of privacy protection when some of them use the data to compete aggressively (Niebel, 2021). Often, competition authorities have imposed mandated data sharing between competitors.<sup>1</sup> But it can also have unintended consequences such as decreasing competition in related technology markets by adding technical and administrative barriers to entry (Peukert et al., 2022). Moreover, data can allow firms to price discriminate consumers, thereby reducing their surplus.

Privacy protection and competition are therefore intertwined in the digital economy, and striking a balance is crucial to make sure that both competition and privacy rights are preserved. Yet, up to now, both regulations have been applied separately. In the Facebook/Whatsapp merger, the European Commission expressly stated that “any privacy-related concerns flowing from the increased concentration of data (...) do not fall within the scope of the EU competition law rules but within the scope of the EU data protection rules”. This stance was confirmed in early 2016 by Commissioner Vestager who declared in a speech that “privacy and competition concerns should be considered separately” (de Moncuit, 2018).

Recently, however, the need for co-operation and coordination across data protection agencies and competition authorities has been flagged by several reports. Experts argue that a strict separation between these two areas of law may result in suboptimal outcomes for consumer welfare.<sup>2</sup> For instance, many academics have

---

<sup>1</sup>See for instance the [Data Act](#) recently enacted in the European Union.

<sup>2</sup>As stated by Lina Khan, the Chair of the Federal Trade Commission, there is a “growing

argued that the acquisition of WhatsApp by Facebook and of Fitbit by Google have granted Facebook and Google better information on their users (reducing their privacy), thereby consolidating their dominant positions in their respective markets (Valletti and Zenger, 2019; Bourreau et al., 2020).

The objective of this article is to provide a framework to analyze the relationship between privacy and competition, and to study how protecting consumer privacy can harm or benefit consumer surplus. We consider a market where firms collect data to increase their competitive power. Firms are located at the extremities of a Hotelling line, and they strategically choose the number of consumers on whom they collect information and the number of data points that they collect.

This framework captures two dimensions of consumer privacy. On the one hand, privacy depends on the amount of information a firm has on individual consumers – namely the intensive margin of privacy. But privacy can also depend on whether information allows firms to identify more consumers – the extensive margin of privacy. In this model, firms use data to better identify the willingness to pay of consumers for their products, and a firm that collects more information about consumers will be able to charge them personalized prices, reducing their surplus. Yet, using consumer data to charge targeted prices will also increase the competitive pressure on the market, which benefits consumers (Thisse and Vives, 1988).

Using this framework, we show that when fewer data are collected on individuals – a protection of the intensive margin – firms charge lower prices and extract less surplus from consumers. In reaction firms collect data on consumers located closer to their location to maximize rent extraction. Hence, protecting privacy at the intensive margin also preserves the extensive margin of privacy, and we show that total consumer surplus also increases in this case.

On the contrary, regulators willing to protect the extensive margin of privacy have to balance increased privacy protection with a loss of consumer surplus. Indeed, when fewer consumers are identified by firms, the intensity of competition decreases and consumers pay higher prices. Hence, depending on the margin of

---

recognition that persistent commercial data collection implicates competition as well as privacy” (Statement of Chair Lina M. Khan Regarding the Report to Congress on Privacy and Security, October 1, 2021).

privacy considered, the objective of a data protection agency may be aligned or in conflict with the mission of a competition authority.

This article contributes to the economic literature and policy debates on two main points. First, we introduce a methodological distinction between the two margins of privacy, which is new in the literature. To this aim, we propose a flexible model in which two competing firms endogenously choose the number of consumers on whom they collect data and the amount of information that they collect on each consumer. This model allows us to analyze the relation between each margin of privacy and consumer surplus. Consumer privacy has extensively been analyzed in previous research, but most articles focus on one margin of privacy, providing incomplete analysis of data protection regulations and their interactions with competition law. For instance, in [Lee et al. \(2011\)](#), [Taylor and Wagman \(2014\)](#), [Belleflamme and Vergote \(2016\)](#), [Montes et al. \(2019\)](#), and [Braulín \(2023\)](#), firms charge personalized prices to consumers. Firms compete more fiercely when they have access to consumer information, and protecting consumer privacy softens market competition and reduces consumer surplus. [Liu and Serfes \(2004\)](#) analyze a framework in which firms can acquire information with exogenous precision. They show that consumer surplus can increase or decrease depending on the precision of information. Building on their model, we endogenize the collection of data by firms and we explicitly distinguish between both margins of privacy to identify their impact on competition and consumer surplus.

Secondly, our results provide new insights for policymakers willing to better understand the relation between privacy protection, competition, and consumer surplus. In particular, we show that, regardless of the relation between the strategic choice of the firms regarding the number of consumers on whom they collect information and the number of segments that they collect, protecting privacy at the intensive margin also benefits consumer surplus. This regulatory spillover highlights the importance of privacy protection at the intensive margin, which can be implemented through data minimization principles for instance. Moreover, we also identify a positive spillover from the protection of the intensive margin of privacy on the protection of the extensive margin of privacy. This result comes from the fact that, in our model, when a firm collects finer segments, it benefits from

identifying more consumers. Hence, both variables are strategic complements. When the choice of the number of consumers identified and of the number of segments collected are strategic substitutes, a data protection agency must balance the protection of privacy at the intensive and at the extensive margin. Overall, we call for greater collaboration between data protection agencies and competition authorities, as regulating the two margins of privacy may have unintended consequences on consumer surplus and lead to inefficient regulations.

The remainder of the article is organized as follows. We describe the model in Section 2. We characterize the equilibrium in Section 3. Section 4 provides an analysis of the relation between consumer data collection, price discrimination and consumer surplus. We extend the model and discuss our main assumptions in Section 5. Section 6 concludes.

## 2 Model

We consider a model of competition à la Hotelling in a product market.<sup>3</sup> Consumers are uniformly distributed on a unit line  $[0, 1]$ . They purchase one product from two competing firms that are located at the two extremities of the line.

Firms can collect information that partitions consumer demand into segments. They use this information to charge personalized prices to consumers on whom they have information, and set a homogeneous price for the remaining consumers.

### 2.1 Consumers

Consumers buy one product at a price  $p_1$  from Firm 1 located at 0, or at a price  $p_2$  from Firm 2 located at 1. Consumers located at  $x \in [0, 1]$  receive a utility  $V$  from purchasing the product, but incur a cost  $t > 0$  of consuming a product that does not perfectly fit their taste  $x$ . Therefore, buying from Firm 1 (resp. from Firm 2) incurs a cost  $tx$  (resp.  $t(1-x)$ ). Consumers choose the product that gives

---

<sup>3</sup>This simple model of horizontal differentiation is commonly used in the literature to analyze the impact of information on consumer surplus and on the profits of the firms (Thisse and Vives, 1988; Elliott et al., 2022; Ali et al., 2022).

the highest level of utility:<sup>4</sup>

$$u(x) = \begin{cases} V - p_1 - tx & \text{if buying from Firm 1,} \\ V - p_2 - t(1 - x) & \text{if buying from Firm 2.} \end{cases}$$

## 2.2 Firms

Firms strategically collect data that segments a share of the consumer demand into  $k$  segments. This information allows them to price discriminate these consumers. We describe the data collection strategy of each firm and the structure of information, then we characterize the resulting pricing strategy of each firm.<sup>5</sup>

### Extensive margin of privacy.

Firm 1 and Firm 2 respectively choose the shares of consumers  $[0, x_1]$  and  $[1 - x_2, 1]$  on whom they collect data. The shares  $x_1$  and  $x_2$  of identified consumers correspond to the extensive margin of privacy. Firm  $\theta$  ( $\theta = 1, 2$ ) then charges personalized prices on consumers on whom it has collected information, and a homogeneous price on the share  $1 - x_\theta$  of unidentified consumers.

Firm  $\theta$  chooses  $x_\theta$  according to two opposite effects of information on its profits. On the one hand, a larger share of identified consumers allows a firm to extract more surplus, which increases its profits. On the other hand, identifying consumers also increases competition because a firm has information on consumers that are closer to its competitor, and thus can lower prices for these consumers ([Thisse and Vives, 1988](#)). This competition effect lowers the profits of the firms. Hence, when choosing the share of identified consumers  $x_\theta$ , a firm will balance the rent extraction and the competition effects of information. We discuss in detail these information structures in [Section 5](#).

---

<sup>4</sup>We assume that the market is covered, so that all consumers buy at least one product from the firms. In terms of model primitives, this assumption requires that  $V \geq 2t$ .

<sup>5</sup>Note that in this model, consumers are passive and firms choose the amounts of data they collect according to a cost-benefit analysis. Hence, in the terms of [Dosis and Sand-Zantman \(2023\)](#), firms have ownership over the data generated by consumers. Other models consider optimal privacy regulations when consumers have control over their information and choose whether to disclose it ([Zogheib, 2023](#)).

### Intensive margin of privacy.

Firm 1 and Firm 2 collect endogenous amounts of information on consumers on  $[0, x_1]$  and  $[1 - x_2, 1]$  respectively.<sup>6</sup> Data allow Firm  $\theta$  to partition the demand of identified consumers into  $k_\theta$  segments of size  $\frac{x_\theta}{k_\theta}$ .<sup>7</sup> This ratio corresponds to the precision of information collected by Firm  $\theta$  which increases with  $k_\theta$  for a given  $x_\theta$ . Hence, the number of consumer segments  $k_1$  and  $k_2$  collected by each firm characterize the intensive margin of privacy.

A firm can then third-degree price discriminate consumers on whom it has information by charging different prices on different segments. For instance, when  $k_1 = 2$ , information is coarse, and Firm 1 can only distinguish whether consumers belong to  $[0, \frac{x_1}{2}]$  or to  $[\frac{x_1}{2}, x_1]$ . At the other extreme, when  $k_1$  converges to infinity, Firm 1 knows the exact location of each consumer on  $[0, x_1]$ , which corresponds to first-degree price discrimination.

Collecting more data allows firms to locate consumers more precisely. Considering firms that set prices on different segments of the consumer demand contrasts with the existing literature that assumes that firms can perfectly identify consumers. This allows us to analyze the effect of more precise information (finer segments) on profits and consumer surplus.

The  $k_\theta$  segments of size  $\frac{x_\theta}{k_\theta}$  form a partition illustrated in Figure 1. Both firms compete on the same unit line, but for clarity we represent the partitions of Firm 1 and Firm 2 on two separate lines.

---

<sup>6</sup>In practice, firms can collect such consumer information through various means, such as online cookies and pixels allowing them to observe the behavior of prospective buyers.

<sup>7</sup>Modeling information through a partition of the consumer demand was first introduced by Liu and Serfes (2004).



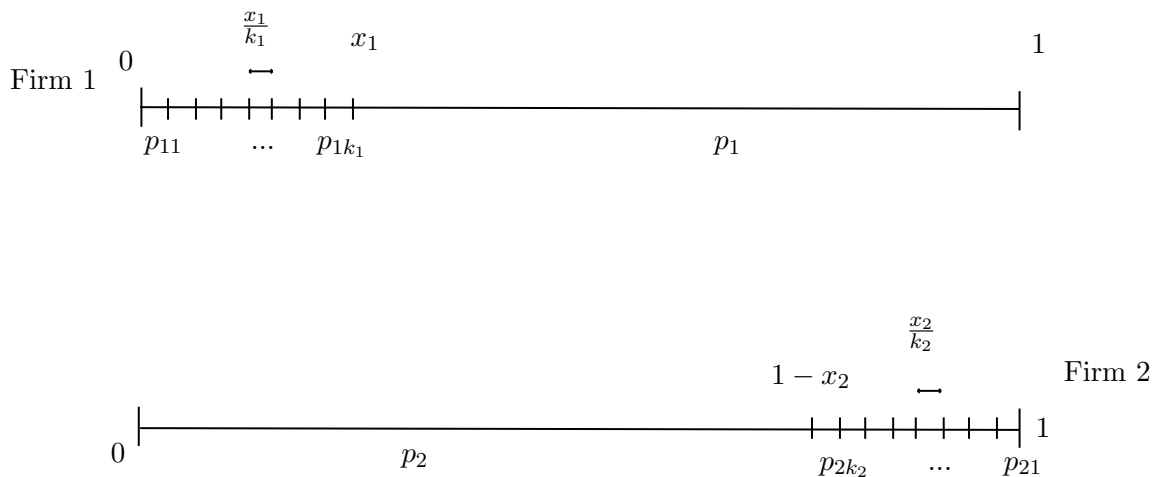


Figure 1: Data collection, Firms 1 and 2

### Cost to collect consumer data.

Firms incur a cost to collect consumer data which increases with the number of identified consumers and with the number of segments collected. The cost of collecting information encompasses various dimensions of the activity of the firms, such as installing trackers and online cookies allowing it to observe the behavior of web users, or storing and handling data to eventually build detailed consumer scores (see [Varian \(2018\)](#) for a detailed discussion on the structure of the costs to collect data).

We denote this cost by  $c(k_\theta, x_\theta)$ , which we assume to be multiplicative and quadratic in  $k_\theta$  and  $x_\theta$  and equal to  $c \cdot k_\theta^2 \cdot x_\theta^2$ ,  $c > 0$ . A firm will thus balance the benefits from collecting more segments on a large share of consumers and the cost to collect such data. Focusing on this functional form allows us to obtain closed-form solutions, but all our results hold with more general cost structures (see [Appendix A.2](#)).

### Profits.

Firms choose the shares  $x_1$  and  $x_2$  of consumers on which they will collect data and charge personalized prices, as well as the number of segments  $k_1$  and  $k_2$  collected on these consumers. We denote by  $d_{\theta i}$  the demand of Firm  $\theta$  on the  $i$ th segment that it price discriminates.  $d_{\theta i}$  depends on the size of the segment which is defined

by the number of data collected  $k_\theta$ : for Firm 1,  $d_{1i} = \frac{x_1}{k_1}$ , and for Firm 2  $d_{2i} = \frac{x_2}{k_2}$ . Additionally, Firm 1 reaches demand  $d_1$  in  $[x_1, 1]$  and Firm 2 reaches demand  $d_2$  in  $[0, 1 - x_2]$ . An informed Firm  $\theta$  maximizes the following profit function with respect to  $p_{\theta 1}, \dots, p_{\theta n}$ :<sup>8</sup>

$$\pi_\theta(p_{\theta i}, p_\theta) = \sum_{i=1}^{k_\theta} d_{\theta i} p_{\theta i} + p_\theta d_\theta - c(k_\theta, x_\theta), \quad (1)$$

We assume that firms set prices in two stages. First, Firm 1 and Firm 2 simultaneously set homogeneous prices  $p_1$  and  $p_2$  for the unidentified consumers in  $d_1$  and  $d_2$ . Secondly, Firm 1 sets a targeted price on each consumer segment on  $[0, x_1]$ , with  $p_{1i}$  being the price on the  $i$ th segment from the origin. Similarly, Firm 2 sets a targeted price on each consumer segment on  $[0, 1 - x_2]$ . Then consumers observe prices. When setting the competitive price  $p_\theta$ , Firm  $\theta$  already knows which consumers it price discriminates, and thus charges  $p_\theta$  accordingly.

Sequential pricing decision avoids the non-existence of Nash equilibrium in pure strategy, and allows an informed firm to charge consumers a higher price. This practice is common in the literature and is supported by managerial evidence. For instance, [Acquisti and Varian \(2005\)](#) use sequential pricing to analyze inter-temporal price discrimination with incomplete information on consumer demand. [Jentzsch et al. \(2013\)](#), [Belleflamme et al. \(2020\)](#) and [Dubus \(2024\)](#) also focus on sequential pricing where a higher personalized price is charged to identified consumers after a firm sets a uniform price.<sup>9</sup>

## 2.3 Timing

The data collection strategies of the firms take place in two separate stages.<sup>10</sup> Firms first choose the share of consumers on whom they collect information, and

---

<sup>8</sup>In our simple setting, firms use consumer data to charge targeted prices. Hence, we abstract from other uses of data such as improved matching or product personalization.

<sup>9</sup>Sequential pricing is also common in business practices (see also [Fudenberg and Villas-Boas \(2006\)](#)). Recently, Amazon has been accused of showing higher prices for Amazon Prime subscribers, who pay an annual fee for unlimited shipping services, than for non-subscribers ([Lawsuit alleges Amazon charges Prime members for "free" shipping, Consumer affairs, August 29 2017](#)). Thus Amazon first sets a uniform price, and then increases prices for high-value consumers who are better identified when they join the Prime program.

<sup>10</sup>We consider an alternative timing with simultaneous choice of the number of consumers identified and of the number of segments collected in Section 5.

then collect an endogenous number of segments on these consumers. This two-stage approach allows us to distinguish the two strategic choices of firms that correspond to the two margins of privacy.

The timing of the game is the following:

- Stage 1: Firm  $\theta$  chooses the share  $x_\theta$  of consumers on whom they collect information.
- Stage 2: Firm  $\theta$  collects  $k_\theta$  segments of consumers on the share  $x_\theta$  of identified consumers.
- Stage 3: Firm  $\theta$  sets prices  $p_\theta$  on the segment of unidentified consumers.
- Stage 4: Each firm charges personalized prices to identified consumers by setting  $p_{\theta i}$  ( $i \in [1, x_\theta(k_\theta)]$ ).

We analyze in Section 3 how a change in the intensity of competition impacts the data strategies of the firms, and subsequently the two margins of privacy. We then analyze in Section 4 how a change in  $x_\theta$  and  $k_\theta$  – the two margins of privacy – impacts market competition and consumer surplus.

### 3 Equilibrium Analysis

We solve the game backward. We first characterize the equilibrium prices in stages 3 and 4 when firms price discriminate  $x_1$  and  $x_2$  consumers respectively. We then analyze the data strategies of each firm.

#### Prices and demand

At the price setting stage, Firm  $\theta = 1, 2$  has collected data allowing to partition a share  $x_\theta$  of close-by consumers in  $k_\theta$  segments. Firm  $\theta$  sets a price  $p_{\theta i}$  for each segment of size  $\frac{x_\theta}{k_\theta}$ , and a unique price  $p_\theta$  on the rest of the unit line. Firm  $\theta$  sells to the  $x_\theta$  consumers on whom it has collected data, and has demand on each segment equal to  $d_{\theta i} = \frac{x_\theta}{k_\theta}$ . The corresponding prices are computed using the indifferent consumer located on the extremity of the segment,  $\frac{i x_\theta}{k_\theta}$ . Hence, personalized prices

satisfy the following relations in equilibrium:

$$p_{\theta i} = p_{-\theta} + t - 2t \frac{ix_{\theta}}{k_{\theta}}.$$

Let  $d_1$  denote the demand for Firm 1 (resp.  $d_2$  the demand for Firm 2) on the large segment  $[x_1, 1]$  where firms compete (resp. on  $[0, 1 - x_2]$ ).  $d_1$  is determined by the location of the indifferent consumer:  $\hat{x} = \frac{p_2 - p_1 + t}{2t}$  and  $d_1 = \hat{x} - x_1 = \frac{p_2 - p_1 + t}{2t} - x_1$  (resp.  $d_2 = \frac{p_1 - p_2 + t}{2t} - x_2$ ).

### ***Profits of the firms.***

Using these expressions for prices and demands, we can write the profits of the firms as:

$$\begin{aligned} \pi_{\theta} &= \sum_{i=1}^{k_{\theta}} d_{\theta i} p_{\theta i} + d_{\theta} p_{\theta} - c(k_{\theta}, x_{\theta}) \\ &= \sum_{i=1}^{k_{\theta}} \frac{x_{\theta}}{k_{\theta}} \left( p_{-\theta} + t - 2t \frac{ix_{\theta}}{k_{\theta}} \right) + \left( \frac{p_{-\theta} - p_{\theta} + t}{2t} - x_{\theta} \right) p_{\theta} - c k_{\theta}^2 x_{\theta}^2. \end{aligned}$$

These profits are equal to the sum of profits on each segment of size  $\frac{x_{\theta}}{k_{\theta}}$  where Firm  $\theta$  has collected information, and the competitive profits on the remaining segment where Firm  $\theta$  charges a homogeneous price.

### ***Prices and demands in equilibrium.***

We characterize the equilibrium pricing decision of the firms by applying first-order conditions on  $\pi_{\theta}$  with respect to  $p_{\theta}$ . By doing so, we can write prices in equilibrium:

$$p_{\theta} = t \left[ 1 - \frac{2}{3} x_{-\theta} - \frac{4}{3} x_{\theta} \right], \quad p_{\theta i} = 2t - \frac{4}{3} x_{-\theta} t - \frac{2}{3} x_{\theta} t - 2 \frac{itx_{\theta}}{k_{\theta}},$$

and we obtain the following demands in the competitive segment in equilibrium:

$$d_{\theta} = \frac{1}{2} - \frac{2}{3} x_{\theta} - \frac{1}{3} x_{-\theta}.$$

Plugging these expressions in the profits of the firms we obtain the profits of Firm  $\theta$

as the difference of its revenues with data  $r(x_\theta, k_\theta)$  and the cost to collect  $c(x_\theta, k_\theta)$ :

$$\begin{aligned}\pi_\theta^*(x_\theta, k_\theta) &= \sum_{i=1}^{k_\theta} \frac{x_\theta}{k_\theta} \left( p_{-\theta} + t - 2t \frac{i x_\theta}{k_\theta} \right) + \left( \frac{p_{-\theta} - p_\theta + t}{2t} - x_\theta \right) p_\theta - c k_\theta^2 x_\theta^2, \\ &= \underbrace{\frac{t}{2} - \frac{7}{9} x_\theta^2 t + \frac{2t}{9} x_{-\theta}^2 - \frac{4}{9} x_\theta x_{-\theta} t + \frac{2}{3} x_\theta t - \frac{2}{3} x_{-\theta} t - \frac{x_\theta^2 t}{k_\theta}}_{\text{Revenues with data } r(x_\theta, k_\theta)} - \underbrace{c k_\theta^2 x_\theta^2}_{\text{Cost to collect data } c(x_\theta, k_\theta)},\end{aligned}$$

The profits of the firms depend on the amounts of data that they collect  $k_1$  and  $k_2$ , and on the number of consumers on whom data is collected  $x_1$  and  $x_2$ .

### Number of segments collected

The number of segments collected by Firm  $\theta$  is found by maximizing its profits with respect to  $k_\theta$ . For Firm  $\theta$  we have:

$$\frac{\partial \pi_\theta(x_\theta, k_\theta)}{\partial k_\theta} = \frac{x_\theta^2 t}{k_\theta^2} - 2c k_\theta x_\theta^2 = 0 \implies k_\theta^* = \left( \frac{t}{2c} \right)^{1/3}$$

The number of segments collected in equilibrium increases with the transport cost  $t$ . As  $t$  increases, the potential for rent extraction of the firms increases, and they benefit from collecting more segments. Moreover, as expected,  $k_\theta^*$  decreases with the data collection cost. It is important to stress that  $k_\theta^*$  does not depend on the value of  $x_\theta$  when more consumers are identified ( $x_\theta$  increases). The benefits from increased rent extraction perfectly match the loss resulting from an increase in the total data collection cost. This is a property of our cost function, and we show that our results are robust to a general specification of the data collection cost in Appendix A.2.

Using the equilibrium values of  $k_\theta^*$ , we can rewrite the profits of Firm  $\theta$  as follows:

$$\pi_\theta^*(x_\theta) = \frac{t}{2} - \frac{7}{9} x_\theta^2 t + \frac{2t}{9} x_{-\theta}^2 - \frac{4}{9} x_\theta x_{-\theta} t + \frac{2}{3} x_\theta t - \frac{2}{3} x_{-\theta} t - x_\theta^2 t \left( \frac{2c}{t} \right)^{1/3} - c x_\theta^2 \left( \frac{t}{2c} \right)^{2/3}$$

### Number of consumers identified

Turning to the number of consumers identified in equilibrium, we solve for the optimal value of  $x_\theta^*$  by maximizing profits with respect to  $x_\theta$ . We can write the

best-response functions of the firms at this stage:

$$x_1^* = \frac{k_1^* t (3 - 2x_2^*)}{9ck_1^{*3} + (7k_1^* + 9)t}, \quad x_2^* = \frac{k_2^* t (3 - 2x_1^*)}{9ck_2^{*3} + (7k_2^* + 9)t}.$$

The number of consumers identified by the firms are strategic substitutes: when Firm 1 identifies more consumers, the number of consumers identified by Firm 2 as a best response decreases.

In turn, denoting  $\alpha = 21k_{-\theta}^* + 27$ , the number of consumers identified as functions of  $k_\theta$  and  $k_{-\theta}$  are equal to:

$$x_\theta^* = \frac{(5k_{-\theta}^* + 9)k_\theta^* t^2 + 9ck_{-\theta}^{*3} k_\theta^* t}{((15k_{-\theta}^* + 21)k_\theta^* + \alpha)t^2 + (c\alpha k_\theta^{*3} + 21ck_{-\theta}^{*3} k_\theta^* + 27ck_{-\theta}^{*3})t + 27c^2 k_{-\theta}^{*3} k_\theta^{*3}}.$$

Substituting the equilibrium number of segments collected in  $x_\theta^*$  we obtain the following equilibrium number of consumers identified:

$$x_\theta^* = \frac{1}{3} - \frac{27c}{5 \cdot 2^{4/3} \cdot c^{1/3} \cdot t^{2/3} + 21 \cdot 2^{5/3} \cdot c^{2/3} \cdot t^{1/3} + 81 \cdot c}.$$

Hence, the equilibrium number of consumers identified by each firm decreases with the data collection cost  $c$ .

## 4 Privacy Protection and Consumer Surplus

We now analyze how the protection of each margin of privacy by a regulator impacts competition and consumer surplus  $CS(x_1, x_2, k_1, k_2)$ . A regulation that protects the intensive margin of privacy sets a limit  $\bar{k}$  on the amount of data that firms can collect. Similarly, protecting the extensive margin of privacy sets a limit over the number of consumers  $\bar{x}$  on whom a firm can collect data. To analyze the impacts of such regulations on consumer surplus, we perform comparative statics on consumer surplus with respect to  $x_\theta$  and to  $k_\theta$ .

### 4.1 Protecting the Intensive Margin of Privacy

A data protection agency can protect privacy at the intensive margin by setting a limit  $\bar{k}$  over the number of segments that a firm can collect. This may also impact the incentives of firms to collect data on more consumers.

The impact of a limit on  $k_\theta$  on the choice of  $x_\theta$  depends on whether both variables are strategic complements or substitutes. Taking the cross derivative of the profit function we can write:

$$\frac{\partial^2 \pi_\theta(x_\theta, k_\theta)}{\partial k_\theta \partial x_\theta} = 2t \frac{x_\theta}{k_\theta^2} - 4ck_\theta x_\theta.$$

The cross derivative is positive for  $\bar{k} < k_\theta^* = \left(\frac{t}{2c}\right)^{1/3}$ . Thus, for any value of  $\bar{k}$  that limits the number of segments that firms can collect, reducing the number of segments collected to  $\bar{k}$  also lowers the value of  $x_\theta$ .

Protecting the intensive margin of privacy has then two opposite effects on consumer surplus. On the one hand, firms collect coarser segments which limits their ability to extract rent from consumers. On the other hand, fewer consumers are identified resulting in lower intensity of competition between firms. We show that the first effect dominates the second, and protecting the intensive margin of privacy increases consumer surplus.

**Proposition 1.**

*In our baseline model where  $x_\theta$  and  $k_\theta$  are strategic complements, when a regulator protects the intensive margin of privacy by reducing the number of consumer segments  $k_\theta$  collected by Firm  $\theta$ :*

- *Firms have incentives to collect data on fewer consumers, protecting their privacy at the extensive margin.*
- *Total consumer surplus increases.*

Proof: see Appendix A.3.

This proposition highlights the efficiency of regulations aiming at limiting the number of segments collected by firms, as doing so also protects their privacy at the extensive margin, as well as aggregate consumer surplus. Central to this result is the relationship between  $x_\theta$  and  $k_\theta$  as strategic complements, a fundamental characteristic of data which has been emphasized in recent research on data-driven network effects, underscoring the increasing returns of data collection.<sup>11</sup> In Section

---

<sup>11</sup>Note that the complementarity of variables  $x_\theta$  and  $k_\theta$  is robust to changing to cost function to an additive form such as:  $c(k_\theta, x_\theta) = c \cdot (k_\theta^\alpha + x_\theta^\beta)$  where  $\alpha$  and  $\beta$  are positive.

5 , we consider  $x_\theta$  and  $k_\theta$  as strategic substitutes and we show that it does not change our qualitative results regarding the impact of privacy protection at the intensive margin and consumers surplus.

## 4.2 Protecting the Extensive Margin of Privacy

A data protection agency may also protect privacy at the extensive margin by limiting the ability of firms to identify more consumers. We study in this section how a reduction in the number of consumers identified by firms may also impact the number of segments they collect depending on the complementarity between  $x_\theta$  and  $k_\theta$ . When both variables are strategic substitutes, protecting the extensive margin of privacy exerts a negative impact on the intensive margin by increasing the incentives of firms to collect finer segments.

On the contrary, when  $x_\theta$  and  $k_\theta$  are strategic complements a reduction of the value of  $x_\theta$  also reduces the incentives of the firms to collect data, protecting privacy at the intensive margin. In our setting, the number of segments collected in equilibrium does not vary with  $x_\theta$ , which is therefore the main relevant variable to analyze variations of surplus.

Suppose initially that Firm 1 has collected information on  $x_1$  consumers, and Firm 2 has information on  $x_2$  consumers. If the share of consumers identified by Firm 1 is reduced by  $\epsilon > 0$  so that  $[x_1 - \epsilon, x_1]$  are unidentified, consumer surplus varies according to two effects:

1. A rent extraction effect: Firm 1 charges a uniform price to consumers on  $[x_1 - \epsilon, x_1]$ , which increases their surplus.
2. A competitive effect: Firm 1 increases its price on  $[x_1 - \epsilon, 1]$ , which reduces the competitive pressure on Firm 2. In turn, Firm 2 increases its price, which has a negative effect on the surplus of consumers over the whole unit line.

Overall, the second effect always dominates the first, and aggregate consumer surplus decreases when fewer consumers are identified. Indeed, the rent extraction effect only increases surplus on the  $\epsilon$  consumers who are not identified anymore,



while the competitive effect affects the whole Hotelling line. Therefore, the extensive margin of privacy is deeply related to the competitive impact of information, which implies a necessary trade-off between the protection of privacy and consumer surplus.

Moreover, when the share of identified consumers varies, we have shown that the number of segments collected remains equal to  $k_\theta^*$ . Hence, segments are smaller when  $x_\theta$  decreases, allowing Firm  $\theta$  to better extract their surplus, which further reduces total consumer surplus.

**Proposition 2.**

*In our baseline model where  $x_\theta$  and  $k_\theta$  are strategic complements, a regulation protecting the extensive margin of privacy by reducing the number of consumers  $x_\theta$  on whom data is collected by Firm  $\theta$ :*

- *Reduces total consumer surplus.*
- *Increases the surplus of consumers on  $[x_\theta - \epsilon, 1]$ .*

Proof: See Appendix A.4.

Proposition 2 shows a general negative relation between privacy protection at the extensive margin and consumer surplus. When high-valuation consumers are identified, their surplus decreases. The intensity of competition increases on the rest of the line, which benefits low-valuation consumers. Hence, Proposition 2 highlights the difficulties to protect at the same time the extensive margin of privacy and consumer surplus.

In practical terms, regulatory authorities have the ability to protect the extensive margin of privacy by prohibiting the collection of data on certain groups of consumers. A natural illustration of such a policy is the collection of data on minors below a certain age. In the European Union, this age varies between thirteen and sixteen years old, and depends on the decision of national agencies.<sup>12</sup> In this respect, the recent move of WhatsApp lowering the minimal age required to use its services from 16 to 13 years old in the UK and in the European Union<sup>13</sup>

---

<sup>12</sup>Can personal data about children be collected?, European Commission, last accessed April 9, 2024.

<sup>13</sup>Anger from campaigners as WhatsApp lowers age limit to 13 in UK and EU, the Guardian, April 12 2024.

can be seen as a reduction of privacy at the extensive margin. Forbidding this change could be a way for regulators to protect this dimension of privacy among users, but may come at a cost in terms of surplus if it prevents WhatsApp from competing fiercely with other services.<sup>14</sup>

More generally, protecting privacy at the extensive margin may be a complicated task for regulators, in particular if firms must first collect data on consumers to infer that they do not have the right to do so.<sup>15</sup> This emphasizes the importance for regulators to protect privacy at the intensive margin, as when  $x_\theta$  and  $k_\theta$  are strategic complements, doing so also gives them a simple way to protect privacy at the extensive margin.

## 5 Extensions and Robustness

We present in this section one extension, namely the strategic substitutability between the number of consumers identified  $x_\theta^*$  and the number of segments collected  $k_\theta^*$ , and we discuss the robustness of the results related to the timing of the game, the information structures, and the ability of consumers to share their data with firms.

### 5.1 Strategic Substitutability

The positive interaction between the protection of privacy at the intensive and extensive margins results from the nature of  $x_\theta^*$  and  $k_\theta^*$ , which are strategic complements when the cost function to collect data is quadratic, as in our baseline model. We consider here the case where these variables are strategic substitutes. We provide conditions over the cost function for this substitutability to take place in Appendix A.2.

In this case, when a regulator enforces a protection of privacy at the intensive margin reducing the number of segments collected by Firm 1 and Firm 2, the best response of the firms is to collect segments on more consumers, so  $x_1$  and

---

<sup>14</sup>Of course, a fundamental assumption of our model is that the market is competitive. Otherwise, the competitive gains for consumers resulting from a reduction of privacy at the extensive margin do not take place.

<sup>15</sup>We thank an anonymous referee for highlighting this paradox.

$x_2$  increase. Conversely, by reducing the number of consumers on whom data is collected  $x_\theta$ , the regulator induces firms to collect finer segments, harming privacy at the intensive margin. This highlights a tension between the protection of each margin of privacy.

However, we show in Proposition 3 that considering  $x_\theta^*$  and  $k_\theta^*$  as strategic substitutes does not change our main result related to consumer surplus: a decrease in  $k_\theta$  will induce an increase in  $x_\theta$ , and both changes will increase consumer surplus. Conversely, a decrease in  $x_\theta$  will induce an increase in  $k_\theta$ , and both changes will reduce consumer surplus.

**Proposition 3.**

*Assume  $k_\theta$  and  $x_\theta$  are strategic substitutes:*

- *When a regulator protects the intensive margin of privacy by reducing the number of consumer segments  $k_\theta$  collected by Firm  $\theta$ :*
  - *Firms have incentives to collect data on more consumers, reducing their privacy at the extensive margin.*
  - *Consumer rent extraction is weaker, and the intensity of competition stronger, such that total consumer surplus increases.*
- *When a regulator protects the extensive margin of privacy by reducing the number of consumers  $x_\theta$  on whom data is collected by Firm  $\theta$ :*
  - *Firms have incentives to collect more segments, reducing consumer privacy at the intensive margin.*
  - *Consumer rent extraction is stronger, and the intensity of competition weaker, such that total consumer surplus decreases.*

*Proof.* Consider the expression of surplus given in Appendix A.3:

$$CS(x_1, x_2, k_1, k_2) = V + t[-\frac{5}{4} + \frac{17}{18}x_1^2 + \frac{17}{18}x_2^2 + x_1x_2] + \frac{1}{2} \frac{x_1^2 t}{k_1^2} + \frac{1}{2} \frac{x_2^2 t}{k_2^2}$$

Simple algebra allows us to see that, all things equal,  $CS(x_1, x_2, k_1, k_2)$  increases when  $k_\theta$  decreases and when  $x_\theta$  increases. Hence, even when variables are strategic substitutes, the protection of the intensive margin of privacy lowers  $k_\theta$  and

increases  $x_\theta$ , so that consumer surplus increases. A protection of the extensive margin of privacy reduces  $x_\theta$  and induces an increase of  $k_\theta$ , so that consumer surplus decreases.  $\square$

When  $k_\theta$  and  $x_\theta$  are strategic substitutes, a data protection agency must balance the protection of privacy at the extensive and intensive margins when setting a limitation over the value of  $k_\theta$  set by the firms, as doing so will give them incentives to collect data on a larger group of consumers, reducing privacy at the extensive margin.

## 5.2 Timing

In this section, we examine two aspects regarding the timing of the game: the order in which firms choose  $x_\theta$  and  $k_\theta$ , and the extent to which firms' data strategies are observable.

### **Simultaneous choice of $x_\theta$ and $k_\theta$ .**

In the baseline model, the data collection strategies of the firms take place in two separate stages (Stages 1 and 2): first, they choose the share of consumers on whom they collect information  $x_\theta$ , and secondly they collect an endogenous number of segments on these consumers  $k_\theta$ .

As we have shown in the analysis, the choice of  $k_\theta$  is independent of  $x_\theta$ , so that considering a sequential or simultaneous choice of  $x_\theta$  and  $k_\theta$  yields identical results. This property relies on the cost structure that we consider in the main analysis. In Appendix A.2, we analyze the impact of a simultaneous choice of  $x_\theta$  and  $k_\theta$  on market outcomes when considering alternative cost functions. We show that the relation between  $k_\theta$  and  $x_\theta$  remains unchanged when both variables are chosen simultaneously, but their equilibrium values may change. This implies that the conditions for  $k_\theta$  and  $x_\theta$  to be strategic complements or substitutes with general cost functions do not depend on the timing considered.

### **The data strategies of the firms are observable.**

In the baseline model, when firms compete in prices in stages 3 and 4, they know the number of consumers on whom their competitor has acquired information. This assumption allows the existence of subgame perfect equilibria at the price-competition stage, in particular when firms set prices  $p_1$  and  $p_2$ .

In practice, public knowledge over  $x_1$  and  $x_2$  can be obtained if firms engage in reverse engineering over the strategic choices of their competitor. For instance, a growing literature now analyzes the role of public adoption of machine learning algorithms as a device to commit to their strategies (Salcedo, 2015; Harrington Jr, 2022; Brown and MacKay, 2023). Hence, a firm that knows the technology of its competitor can infer its optimal number of consumers identified, and choose its data collection strategy accordingly.

### **5.3 Information Structures**

We model firms acquiring information on their closest consumers, on segments equal to  $[0, x_1]$  for Firm 1 and  $[1 - x_2, 1]$  for Firm 2. These information structures rule out the possibility for firms to acquire information on consumers located in the middle of the line instead of consumers located close to them. In turn, this implies that identified consumers are always closer to the respective firms' location than consumers who are unidentified. Moreover, as partitions do not overlap and  $x_1 < 1 - x_2$ , both firms cannot have information on the same consumers.

These assumptions are motivated by a recent literature analyzing which information structures are profitable when firms compete in a Hotelling duopoly setting. Indeed, Lee et al. (2011) show that, when firms have overlapping partitions allowing them to identify close-by consumers as well as those located in the middle of the line, firms may be better off by adopting a privacy-preserving policy preventing them to charge personalized prices to consumers that they both identify. Conversely, Dubus (2024) considers a firm located at 0 having access to information on consumers on  $[0, \frac{1}{2}]$ , while its competitor has information on the remaining consumers. In this model, firms do not have overlapping information partitions, and in equilibrium, firms reach the highest profits by using informa-

tion on the closest consumers, without charging personalized prices to consumers in the middle of the line. A similar effect highlighting the importance for firms of focusing on the closest consumers is also identified in the marketing literature (Iyer et al., 2005).

The rationale for this result is the following. Closest consumers are those for whom the benefits from rent extraction are the highest, as they have the strongest valuation for the product of the firm located close to them. Moreover, charging personalized prices to consumers located far away from a firm's location intensifies sharply competition between firms. Indeed, to sell its product to far-away consumers, a firm must lower drastically its price for these consumers, increasing the pressure on its competitor. If the competitor also has information on these consumers, it can lower its price to a point where it is not possible for a firm to poach consumers located far away.

To illustrate these effects, consider the symmetric case where Firm 1 and Firm 2 have information on consumers located respectively on  $[0, \hat{x}]$  and  $[1 - \hat{x}, 1]$ . There are three possible deviations for Firm 1 that would result in a different information structure than the one we consider: (i) Firm 1 can acquire information on a segment of consumers located in the middle of the line in  $[\hat{x}, 1 - \hat{x}]$  where it would capture all consumers; (ii) it can acquire information on a segment in the middle of the line in  $[\hat{x}, 1 - \hat{x}]$  where it would compete with Firm 2; (iii) it can acquire information on a segment of consumers on which Firm 2 also has information in  $[1 - \hat{x}, 1]$ .

The first two deviations are dominated by the following strategy:<sup>16</sup> instead of targeting consumers located far away, for which the potential for rent extraction is low, Firm 1 can achieve higher profits by targeting the closest consumers unidentified, located on  $[\hat{x}, \hat{x} + \frac{1}{k_1}]$ , from which it will be able to extract more surplus. In turn, this implies that the resulting partitions will be of the form  $[0, x_1]$  for Firm 1 and  $[1 - x_2, 1]$  for Firm 2.

In the third deviation, Firm 1 and Firm 2 will compete fiercely to sell their products to the consumers that they both identify. If these consumers are located close to Firm 2, Firm 1 will not be able to poach them. Hence, acquiring costly

---

<sup>16</sup>A formal analysis of these first two deviations is provided in Dubus (2024).

information on these consumers will not be profitable. [Liu and Serfes \(2004\)](#) have even shown that, when the market is segmented, poaching will be possible for the firms only for consumers located in the neighborhood of  $\frac{1}{2}$ . We have shown in our equilibrium analysis that, even when collecting data is costless, firms will not identify consumers located further away than  $\frac{1}{3}$  from their location, so that acquiring information to poach consumers close to the competitor is not a profitable strategy.

## 5.4 Consumers Share their Data

We have assumed that firms can choose the consumers on whom they collect data, and that consumers are passive. This assumption is relevant as the literature has emphasized the difficulties for consumers to hide from firms. For instance the literature on the privacy paradox documents that consumers are myopic when it comes to protecting their privacy, even if they claim to care about it ([Kokolakis, 2017](#)).

However, recent research has focused on consumer information-sharing strategies. [Ali et al. \(2022\)](#) and [Ali et al. \(2023\)](#) have shown for example that when consumers can choose to share their personal information, consumers located in the middle of the line prefer to share their data to both firms, while those located at the extremities decide not to do so. Conversely, [Montes et al. \(2019\)](#) analyze consumers who can hide from firms located at the extremities of a unit line. They also show that consumers who benefit the most from hiding are those located closest to a firm's location, while those in the middle of the line remain identified by firms. Extending this result to our model, we can hypothesize that the rent extraction effect would be weaker, but the competition effect of information would be stronger. We leave the question of strategic sharing of consumer data for future research.

## 6 Conclusion

We have analyzed a model where firms collect consumer data along two dimensions: they choose on how many consumers they collect information - the *extensive*

*margin of privacy* - and the amount of information they collect on these consumers - the *intensive margin of privacy*. Using this framework, we have explored the interaction between privacy protection and market competition. The main finding of this article is to show that a regulation that restricts the number of consumer segments collected by firms leads them to collect data on fewer consumers, and results in an increase in total consumer surplus; protecting the intensive margin preserves also the extensive margin of privacy. On the contrary, a regulation that exclusively safeguards the extensive margin of privacy by reducing the number of consumers from whom data is collected is detrimental to total consumer surplus.

These findings are particularly worth stressing in light of recent regulations that target the way firms collect and use consumer data, and how they impact competition in product markets. Competition authorities have indeed recently taken measures to intensify competition by increasing the amount of personal data accessible to firms.<sup>17</sup> For instance, Article 61 of the Digital Market Act establishes a data access right under which large digital platforms must grant their competitors access to ranking, query, click and view data.<sup>18</sup> Our results support the idea that such a policy has a positive impact on the intensity of product-market competition, which might increase the surplus of consumers. However, because more firms have access to consumer data, such a regulation reduces consumer privacy at the extensive margin, and the benefits of a stronger competitive pressure between firms should be cautiously balanced with the loss of privacy.

Moreover, in recent years, data protection agencies have also adopted measures to limit the amount of consumer data collected by firms, thereby protecting privacy at the intensive margin. Such a data minimization principle is enacted for instance in the [California Consumer Privacy Act](#) and the [Health Insurance Portability and Accountability Act](#) in the US, or in the European GDPR.<sup>19</sup> Our results suggest that limiting the number of consumer data that a firm can collect may give firms

---

<sup>17</sup>The Digital Market Act is the EU law to make digital markets fairer and more contestable. It provides conditions fostering data sharing between firms, in line with a theoretical literature arguing that mandatory data sharing can promote competition between firms (Prüfer and Schottmüller, 2021; Krämer and Schnurr, 2022).

<sup>18</sup>The Digital Markets Act, last accessed April 12, 2024.

<sup>19</sup>The data minimization principle is expressed in Article 5(1)(c) of the GDPR and Article 4(1)(c) of Regulation (EU) 2018/1725, which provide that personal data must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed".



incentives to collect data on fewer consumers. Hence, protecting the intensive margin may also protect privacy at the extensive margin.

Our results are particularly useful for policymakers, as, to the best of our knowledge, there is up to now no definition of welfare that includes privacy concerns. Hence, any policy that benefits one dimension while harming the other cannot be evaluated (and designed) in a systematic way. By highlighting a measure that protects at the same time both margins of privacy and consumer surplus, our results provide thus a useful first step towards an efficient protection of consumers in digital markets. In any other cases, it is essential that policymakers set up a co-regulation of digital markets protecting jointly consumer privacy and consumer surplus. The recent creation of the high-level group under article 40 of the Digital Markets Act,<sup>20</sup> including among others of the European Data Protection Supervisor and of the European Data Protection Board on the one hand, and of the European Competition Network and the Consumer Protection Cooperation Network on the other hand could be a useful task force to accomplish this goal. Similarly, such a task could be achieved by the Digital Regulation Cooperation Forum<sup>21</sup> and by the Digital Markets Unit<sup>22</sup> in the UK.

Finally, from an academic point of view, our analysis has primarily focused on data segmenting a unit line and allowing firms to price discriminate consumers. Alternatively, [Belleflamme et al. \(2020\)](#) consider competing firms having information on a common group of consumers, that they can identify with a certain probability. Such a setting allows both firms to identify the same group of consumers. Moreover, data may serve various other purposes, including online search and targeted advertising, as analyzed by [De Corniere and De Nijs \(2016\)](#) and [Bird and Neeman \(2023\)](#). The impact of privacy protection may differ depending on the specific use of data. Further research is needed to examine how the methodological distinction we have introduced between the intensive and extensive margins of privacy applies to these alternative contexts.

---

<sup>20</sup>The Digital Markets Act, last accessed April 12, 2024.

<sup>21</sup>The Digital Regulation Cooperation Forum, last accessed, April 12, 2024.

<sup>22</sup>Digital Markets Unit, April 7, 2021.

## References

- Acquisti, A. and H. R. Varian (2005). Conditioning prices on purchase history. *Marketing Science* 24(3), 367–381.
- Ali, S. N., G. Lewis, and S. Vasserman (2022, 07). Voluntary Disclosure and Personalized Pricing. *The Review of Economic Studies* 90(2), 538–571.
- Ali, S. N., G. Lewis, and S. Vasserman (2023). Consumer control and privacy policies. In *AEA Papers and Proceedings*, Volume 113, pp. 204–209. American Economic Association.
- Belleflamme, P., W. M. W. Lam, and W. Vergote (2020). Competitive imperfect price discrimination and market power. *Marketing Science* 39(5), 996–1015.
- Belleflamme, P. and W. Vergote (2016). Monopoly price discrimination and privacy: The hidden cost of hiding. *Economics Letters* 149, 141–144.
- Bird, D. and Z. Neeman (2023). The effect of privacy on market structure and prices. *The Journal of Law, Economics, and Organization*, ewad031.
- Bourreau, M., C. Caffarra, Z. Chen, C. Choe, G. S. Crawford, T. Duso, C. Genakos, P. Heidhues, M. Peitz, T. Rønne, et al. (2020). *Google/Fitbit will monetise health data and harm consumers*. Centre for Economic Policy Research.
- Braulin, F. C. (2023). The effects of personal information on competition: Consumer privacy and partial price discrimination. *International Journal of Industrial Organization* 87, 102923.
- Brown, Z. Y. and A. MacKay (2023). Competition in pricing algorithms. *American Economic Journal: Microeconomics* 15(2), 109–156.
- De Corniere, A. and R. De Nijs (2016). Online advertising and privacy. *The RAND Journal of Economics* 47(1), 48–72.
- de Moncuit, A. (2018). In which ways should privacy concerns serve as an element of the competition assessment. Technical report, Speech at the European Commission.
- Dosis, A. and W. Sand-Zantman (2023). The ownership of data. *The Journal of Law, Economics, and Organization* 39(3), 615–641.
- Dubus, A. (2024). Behavior-based algorithmic pricing. *Information Economics and Policy*, 101081.
- Elliott, M., A. Galeotti, A. Koh, and W. Li (2022). Market segmentation through information.
- Fudenberg, D. and J. M. Villas-Boas (2006). Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems* 1, 377–436.

- Harrington Jr, J. E. (2022). The effect of outsourcing pricing algorithms on market competition. *Management Science* 68(9), 6889–6906.
- Iyer, G., D. Soberman, and J. M. Villas-Boas (2005). The targeting of advertising. *Marketing Science* 24(3), 461–476.
- Jentzsch, N., G. Sapi, and I. Suleymanova (2013). Targeted pricing and customer data sharing among rivals. *International Journal of Industrial Organization* 31(2), 131–144.
- Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security* 64, 122–134.
- Krämer, J. and D. Schnurr (2022). Big data and digital markets contestability: Theory of harm and data access remedies. *Journal of Competition Law & Economics* 18(2), 255–322.
- Lee, D.-J., J.-H. Ahn, and Y. Bang (2011). Managing consumer privacy concerns in personalization: a strategic analysis of privacy protection. *Mis Quarterly*, 423–444.
- Liu, Q. and K. Serfes (2004). Quality of information and oligopolistic price discrimination. *Journal of Economics & Management Strategy* 13(4), 671–702.
- Montes, R., W. Sand-Zantman, and T. Valletti (2019). The value of personal information in online markets with endogenous privacy. *Management Science* 65(3), 1342–1362.
- Niebel, C. (2021). The impact of the general data protection regulation on innovation and the global political economy. *Computer Law & Security Review* 40, 105523.
- Peukert, C., S. Bechtold, M. Batikas, and K. Tobias (2022). Regulatory spillovers and data governance: Evidence from the gdpr. *Marketing Science* 2022.
- Prüfer, J. and C. Schottmüller (2021). Competing with big data. *The Journal of Industrial Economics* 69(4), 967–1008.
- Salcedo, B. (2015). Pricing algorithms and tacit collusion. *Manuscript, Pennsylvania State University*.
- Taylor, C. and L. Wagman (2014). Consumer privacy in oligopolistic markets: Winners, losers, and welfare. *International Journal of Industrial Organization* 34, 80–84.
- Thisse, J.-F. and X. Vives (1988). On the strategic choice of spatial price policy. *The American Economic Review*, 122–137.
- Valletti, T. M. and H. Zenger (2019). Increasing market power and merger control. *Competition Law & Policy Debate* 5(1), 26–35.
- Varian, H. (2018). Artificial intelligence, economics, and industrial organization. Technical report, National Bureau of Economic Research.
- Zogheib, J.-M. (2023). Competition, privacy, and multi-homing.

# A Appendix

## A.1 General Data Collection Costs

We discuss how the structure of the data collection cost impacts the relation between the number of segments collected by a firm and the share of consumers it identifies. We consider two general classes of cost functions that characterize general properties of data such as increasing and decreasing returns to scale, and multiplicative and additive costs with respect to  $k_\theta$  and  $x_\theta$ .

Note that we focus on cost functions that increase with  $k_\theta$  and  $x_\theta$ . This ignores effects of strong data complementarity that may reduce the cost when more consumers are identified.

We show how different cost structures impact our results. We have seen in Section 3 that the profit function of a firm can be written as:

$$\pi_\theta(k_\theta, x_\theta) = \frac{t}{2} - \frac{7}{9} x_\theta^2 t + \frac{2t}{9} x_{-\theta}^2 - \frac{4}{9} x_\theta x_{-\theta} t + \frac{2}{3} x_\theta t - \frac{2}{3} x_{-\theta} t - \frac{x_\theta^2 t}{k_\theta} - c(k_\theta, x_\theta)$$

Hence, noting  $f(x_\theta) = \frac{t}{2} - \frac{7}{9} x_\theta^2 t + \frac{2t}{9} x_{-\theta}^2 - \frac{4}{9} x_\theta x_{-\theta} t + \frac{2}{3} x_\theta t - \frac{2}{3} x_{-\theta} t$ , we can rewrite the profit function of a firm in the following way:

$$\pi_\theta(k_\theta, x_\theta) = f(x_\theta) - \frac{x_\theta^2 t}{k_\theta} - c(k_\theta, x_\theta),$$

where  $f(x_\theta)$  does not depend on  $k_\theta$ . Hence the first-degree derivative of the profits with respect to  $k_\theta$  can be written as:<sup>23</sup>

$$\frac{\partial \pi_\theta(k_\theta, x_\theta)}{\partial k_\theta} = \frac{x_\theta^2 t}{k_\theta^2} - \frac{\partial c(k_\theta, x_\theta)}{\partial k_\theta}.$$

FOCs imply that

$$x_\theta^2 t = k_\theta^2 \cdot \frac{\partial c(k_\theta, x_\theta)}{\partial k_\theta}.$$

### Multiplicative cost functions.

We first consider the class of multiplicative cost functions  $c(k_\theta, x_\theta) = c \cdot k_\theta^\alpha \cdot x_\theta^\beta$ .<sup>24</sup> Coefficient  $c$  corresponds to a productivity factor resulting for instance from different technologies used by an industry. We consider  $\alpha \geq 1$  so that second-order conditions hold. Such a specification characterizes both increasing and constant returns to scale w.r.t.  $k_\theta$ , and increasing and decreasing returns to scale w.r.t.  $x_\theta$ ,

<sup>23</sup>It is clear that FOCs can be applied as long as the cost function is convex w.r.t  $k_\theta$  or concave and such that  $\frac{\partial^2 c(k_\theta, x_\theta)}{\partial k_\theta^2} > -2 \frac{x_\theta^2 t}{k_\theta^3}$  with  $x_\theta$  constant.

<sup>24</sup>In this case, the second-order condition can be written  $c \cdot k_\theta^{\alpha+1} \cdot \alpha(\alpha - 1) > -2x_\theta^{2-\beta} t$ . This condition is clearly satisfied for  $\alpha \geq 1$ .

depending on the values of  $\alpha > 1$  and  $\beta > 0$ . Considering the FOCs, we have:

$$x_\theta^{2-\beta} t = c \cdot \alpha \cdot k_\theta^{\alpha+1}$$

$$k_\theta^* = \left( \frac{c \cdot \alpha}{t} \right)^{\frac{-1}{\alpha+1}} \cdot x_\theta^{\frac{2-\beta}{\alpha+1}}$$

Hence,  $k_\theta^*$  increases with  $x_\theta$  when  $\beta < 2$  and  $k_\theta^*$  decreases with  $x_\theta$  when  $\beta > 2$ . In the baseline model where  $c(\cdot)$  is proportional to  $x_\theta^2$ ,  $2 - \beta$  simplifies and  $k_\theta^*$  does not depend on  $x_\theta$ . Then,  $\beta = 2$  is the threshold value above which  $k_\theta$  and  $x_\theta$  are strategic substitutes and below which they are strategic complements.

### Additive cost functions.

We consider now the class of costs functions where the costs associated with  $x_\theta$  and  $k_\theta$  are additive:  $c(k_\theta, x_\theta) = c \cdot (k_\theta^\alpha + x_\theta^\beta)$ .<sup>25</sup> In this case the FOC w.r.t.  $k_\theta$  implies:

$$x_\theta^2 t = c \cdot \alpha \cdot k_\theta^{\alpha+1}$$

$$k_\theta^* = \left( \frac{c}{t} \right)^{\frac{-1}{\alpha+1}} \cdot x_\theta^{\frac{2}{\alpha+1}}$$

And  $k_\theta$  always increases with  $x_\theta$ .

### Market outcomes with general cost functions.

The main results of the analysis hold as consumer surplus increases with a higher level of privacy at the intensive margin and decreases with a higher level of privacy at the extensive margin.

Moreover, we identify another effect when  $k_\theta$  and  $x_\theta$  are strategic complements. When far away consumers are identified by firms, the profitability of data decreases, which lowers rent extraction and increases consumer surplus. Overall the two effects go in opposite directions, and their magnitude depends on the structure of the data collection cost.

## A.2 Simultaneous Choice of $x_\theta$ and $k_\theta$ with General Data Collection Costs

We consider here an alternative timing when firms choose simultaneously  $k_\theta$  and  $x_\theta$ . While we have seen that in the baseline model, the equilibrium values of  $k_\theta$  and  $x_\theta$  do not depend on the timing, it is not the case with general cost functions, which we analyze here.

We show how different cost structures impact our results. We have seen in

---

<sup>25</sup>In this case, the second-order condition can be written  $c \cdot k_\theta^{\alpha+1} \cdot \alpha(\alpha - 1) > -2x_\theta^2$ . This condition is clearly satisfied for  $\alpha \geq 1$ .

Section 3 that the profit function of a firm can be written:

$$\pi_\theta(k_\theta, x_\theta) = \frac{t}{2} - \frac{7}{9} x_\theta^2 t + \frac{2t}{9} x_{-\theta}^2 - \frac{4}{9} x_\theta x_{-\theta} t + \frac{2}{3} x_\theta t - \frac{2}{3} x_{-\theta} t - \frac{x_\theta^2 t}{k_\theta} - c(k_\theta, x_\theta)$$

The first-degree derivative of the profits with respect to  $k_\theta$  can be written as:

$$\frac{\partial \pi_\theta(k_\theta, x_\theta)}{\partial k_\theta} = \frac{x_\theta^2}{k_\theta^2} - \frac{\partial c(k_\theta, x_\theta)}{\partial k_\theta}.$$

FOCs imply that

$$x_\theta^2 = k_\theta^2 \cdot \frac{\partial c(k_\theta, x_\theta)}{\partial k_\theta}.$$

Hence, the relation between  $x_\theta^*$  and  $k_\theta^*$  does not change with the timing, as both variables must satisfy the first-degree derivative of the profits with respect to  $k_\theta$  regardless of whether  $k_\theta$  is chosen at the same time as or after  $x_\theta$ . Hence, the conditions for  $x_\theta^*$  and  $k_\theta^*$  to be strategic substitutes of complements in the case of multiplicative cost functions are identical to the sequential case: when  $c(k_\theta, x_\theta) = c \cdot k_\theta^\alpha \cdot x_\theta^\beta$ ,  $k_\theta^*$  increases with  $x_\theta$  when  $\beta < 2$  and  $k_\theta^*$  decreases with  $x_\theta$  when  $\beta > 2$ . Here again,  $\beta = 2$  is the threshold value above which  $k_\theta$  and  $x_\theta$  are strategic substitutes and below which they are strategic complements. In the case of additive cost functions  $c(k_\theta, x_\theta) = c \cdot (k_\theta^\alpha + x_\theta^\beta)$ , both variables are still strategic complements.

What will change when variables are chosen simultaneously are the equilibrium values  $x_\theta^*$  and  $k_\theta^*$ . Indeed, with simultaneity, these variables must also satisfy the first-order condition of the profits with respect to  $x_\theta$ , which can be written as:

$$\frac{\partial \pi_\theta(k_\theta, x_\theta)}{\partial x_\theta} = \frac{14tx_\theta}{9} - \frac{4tx_{-\theta}}{9} + \frac{2}{3} - \frac{2tx_\theta}{k_\theta} - \frac{\partial c(k_\theta, x_\theta)}{\partial x_\theta}.$$

### A.3 Proof of Proposition 1

We analyze the variation of consumer surplus with the number of segments collected by a firm.

## Consumer surplus

Consumer surplus when Firm 1 identifies  $x_1$  consumers with  $k_1$  segments and Firm 2 identifies  $x_2$  consumers with  $k_2$  segments is defined as follows.

$$\begin{aligned}
CS(x_1, x_2, k_1, k_2) &= \sum_{i=1}^{k_1} \left[ \int_0^{\frac{x_1}{k_1}} V - 2t \left[ 1 - \frac{1}{3}x_1 - \frac{2}{3}x_2 - \frac{ix_1}{k_1} \right] - txdx \right] \\
&+ \int_{x_1}^{\frac{1}{2} + \frac{x_1}{3} - \frac{x_2}{3}} V - t \left[ 1 - \frac{4}{3}x_1 - \frac{2}{3}x_2 \right] - txdx \\
&+ \int_{\frac{1}{2} + \frac{x_1}{3} - \frac{x_2}{3}}^{1-x_2} V - t \left[ 1 - \frac{2}{3}x_1 - \frac{4}{3}x_2 \right] - txdx \\
&+ \sum_{i=1}^{k_2} \left[ \int_0^{\frac{x_2}{k_2}} V - 2t \left[ 1 - \frac{1}{3}x_2 - \frac{2}{3}x_1 - \frac{ix_2}{k_2} \right] - txdx \right] \\
&= V + t \left[ -\frac{5}{4} + \frac{17}{18}x_1^2 + \frac{17}{18}x_2^2 + x_1x_2 \right] + \frac{1}{2} \frac{x_1^2 t}{k_1^2} + \frac{1}{2} \frac{x_2^2 t}{k_2^2}
\end{aligned} \tag{2}$$

The number of consumers identified as a function of the number of segments collected is equal to:

$$x_\theta = \frac{(5k_{-\theta} + 9)k_\theta t^2 + 9ck_{-\theta}^3 k_\theta t}{((15k_{-\theta} + 21)k_\theta + \alpha)t^2 + (c\alpha k_\theta^3 + 21ck_{-\theta}^3 k_\theta + 27ck_{-\theta}^3)t + 27c^2 k_{-\theta}^3 k_\theta^3}$$

Substituting these values in  $CS(x_1, x_2, k_1, k_2)$ , we obtain:

$$\begin{aligned}
CS(k_1, k_2) &= \left[ 324 (9c^2 k_1^3 k_2^3 + ct (k_1^3 (7k_2 + 9) + 7k_1 k_2^3 + 9k_2^3) + t^2 (k_1 (5k_2 + 7) + 7k_2 + 9)) \right]^{-1} \\
&\cdot \left[ 324V (9c^2 k_1^3 k_2^3 + ct (k_1^3 (7k_2 + 9) + 7k_1 k_2^3 + 9k_2^3) + t^2 (k_1 (5k_2 + 7) + 7k_2 + 9))^2 \right. \\
&- 405t (9c^2 k_1^3 k_2^3 + ct (k_1^3 (7k_2 + 9) + 7k_1 k_2^3 + 9k_2^3) + t^2 (k_1 (5k_2 + 7) + 7k_2 + 9))^2 \\
&+ 36k_1 k_2 t^3 (9ck_1^3 + (5k_1 + 9)t) (9ck_2^3 + (5k_2 + 9)t) + 34k_2^2 t^3 (9ck_1^3 + (5k_1 + 9)t)^2 \\
&+ 18t^3 (9ck_1^3 + (5k_1 + 9)t)^2 + 34k_1^2 t^3 (9ck_2^3 + (5k_2 + 9)t)^2 \\
&\left. + 18t^3 (9ck_2^3 + (5k_2 + 9)t)^2 \right]
\end{aligned}$$

We take the derivative of this expression with respect to  $k_1$  (without loss of generality):

$$\begin{aligned} \frac{\partial CS(k_1, k_2)}{\partial k_1} = & - \left[ t^3 (9ck_2^3 + (5k_2 + 9)t) \left( 81c^3 (18k_1^6 k_2^4 + (34k_1^2 + 27) k_1^2 k_2^6) \right. \right. \\ & + 9c^2 t \left( 2k_1^6 (k_2 + 3)(29k_2 - 6) + 6k_1^4 k_2^3 (77k_2 + 102) + 81k_1^3 k_2^4 \right. \\ & + 162k_1^2 k_2^3 (2k_2 + 3) - 153k_1 k_2^6 + 63k_2^6) \\ & + ct^2 \left( 2k_1^4 (k_2(650k_2 + 2079) + 1287) + 9k_1^3 (k_2 + 3)(29k_2 - 6) \right. \\ & + 27k_1^2 (5k_2 + 9)(7k_2 + 9) - 27k_1 k_2^3 (77k_2 + 102) - 9(k_2 - 126)k_2^3) \\ & \left. \left. - t^3 (k_1(k_2(650k_2 + 2079) + 1287) + 9(4k_2^2 + k_2 - 81)) \right) \right] \\ & \cdot \left[ 81(9c^2 k_1^3 k_2^3 + ct(k_1^3(7k_2 + 9) + 7k_1 k_2^3 + 9k_2^3) + t^2(k_1(5k_2 + 7) + 7k_2 + 9))^3 \right]^{-1} \end{aligned}$$

In turn, replacing  $k_1$  and  $k_2$  by their equilibrium value  $(\frac{t}{2c})^{1/3}$ , we obtain:

$$\frac{\partial CS(k_1, k_2)}{\partial k_1} \Big|_{k_1=k_1^*} = - \frac{4 \left( 32^{1/3} c \left( \frac{t}{c} \right)^{2/3} + 2t \right)}{9 \left( 2^{2/3} \left( \frac{t}{c} \right)^{1/3} + 3 \right)^3}$$

and this value is negative, so that reducing the number of segments collected below the market equilibrium increases consumer surplus.

## A.4 Proof of Proposition 2

We have seen that the number of segments collected in equilibrium does not depend on the number of consumers identified, and the derivative of  $CS(x_1, x_2, k_1, k_2)$  with respect to  $x_1$  is equal to

$$\frac{\partial CS}{\partial x_1} = \frac{17tx_1}{9} + tx_2 + \frac{x_1 t}{k_1^2},$$

which is always greater than zero.

A direct comparison of prices shows that consumers in segments  $\frac{ix_\theta}{k_\theta} \leq \frac{1}{2} - \frac{2x_\theta}{3} - \frac{x_\theta}{3}$  are charged a higher price when they are identified than without information.