Privacy and Competition*

David Bounie,†Antoine Dubus‡ and Patrick Waelbroeck§

March 15, 2022

Abstract

We investigate the relationship between consumer privacy, market competition and consumer surplus. Firms can collect consumer data that they use to price discriminate consumers in a competitive product market. Firms strategically choose which consumers they price discriminate, as well as the number of consumer data that they collect.

A data protection agency protecting consumer privacy by limiting the amount of data collected increases the intensity of competition and benefits consumers. We highlight the impossibility to protect at the same time aggregate and individual consumer surplus and privacy.

Very preliminary, do not circulate.

1 Introduction

Privacy and competitions are two central features of the digital economy. On the one hand, Big-tech companies such as Alphabet, Apple, Facebook, Amazon, and Microsoft are today the largest companies in the world with an aggregate

*We thank for useful remarks and comments. Antoine Dubus acknowledges financial support from the FNRS Grant PDR T.01.47.19. Patrick Waelbroeck thanks for insightful discussions the 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, La Banque Postale, la Caisse des Dépôts et Consignations, Telecom Paris and University of Paris 2 Panthéon - Assas.

†i3, Telecom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France; david.bounie@telecom-paris.fr.

‡Department of Management, Technology and Economics, ETH Zürich Leonhardstrasse 21 – 8092, Switzerland; antoine1dubus@gmail.com.

§i3, Telecom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France; patrick.waelbroeck@telecom-paris.fr.

market value of more than 5 Trillion USD.[1] They hold a dominant position in multiple sectors of the digital economy such as online search, e-commerce, and social networks, and are also active in multiple other markets (e.g. online advertising and payments). Their success is largely built upon the collection and use of huge amounts of consumer data. Data has become a competitive asset (Hagiu and Wright, 2020), and the ability of firms to collect and strategically use consumer data now shapes the competitiveness of all types of markets (Crémer et al., 2019). This has raised concerns among economists, and competition authorities (Furman et al., 2019; Scott Morton et al., 2019; Tirole, 2020).

On the other hand, given the importance of data for these companies, regulators are concerned with how personal data collection strategies threaten consumer privacy. A recent FT article[2] assesses that large data brokers have access to more than 3,000 data points on more than 200 million Americans, and the FTC recently fined Facebook $5 billion for not respecting consumer privacy,[3] while in France, Google was fined EUR 50 million for violating the European GDPR.[4] There is thus a pressing societal need to better understand the implications of large-scale data collection for consumers.

Privacy and competition are protected by different laws: personal data protection regulations such as the European General Data Protection Regulation (GDPR) and competition law. They have mainly been analyzed separately. [5] Recently competiton authorities have acknowledged the possibility that there could be a relationship between privacy and competition. A new line of reasoning has emerged following the German Competition Authority's decision to limit Facebook's collection of personal data.[6] If data provides a competitive advantage, the

[1] Stock Market Warning: 6 Mega Stocks Dominate S&P 500's $21.4 Trillion Cap; CCN, April 27, 2020.

[2] Data brokers: regulators try to rein in the 'privacy deathstars', Financial Times, January 8, 2019.

[3] FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook, Federal Trade Commission, July 24, 2019.

[4] The CNIL's restricted committee imposes a financial penalty of 50 Million euros against GOOGLE LLC, January 21, 2019.

[5] For instance in recent merger cases such Facebook /Whatsapp or Microsoft/Linkedin, competition authorities have acknowledge the role of consumer data but minimized its impact compared to efficiency gains due to better data processing tools and better algorithms.

[6] Bundeskartellamt prohibits Facebook from combining user data from different sources, last

protection of personal data may enhance competition between firms. Conversely, if competition between firms that use data intensively reduces their incentives to collect personal data, competition law may protect consumer privacy. Overall, as stated by Lina Khan, the Chair of the Federal Trade Commission, there is a "growing recognition that persistent commercial data collection implicates competition as well as privacy".[7]

In this article, we seek answers to the following two research questions: is there a relationship between privacy and competition? And if so, are privacy and competition friends or foes with respect to consumer surplus? The starting point of this article is that information has two opposite effects on consumer surplus. First, better information on consumers allows firms to determine more precisely their willingness to pay for products and this rent-extract effect of information reduces consumer surplus. Secondly, when different firms have information on the same consumers, they will compete fiercely for them, in turn, decreasing prices and increasing consumer surplus. [8] A model of privacy and competition should therefore include these two dimensions of information.

We analyze the relation between competition and privacy and we study whether there is a necessary tension between the protection of consumer privacy and consumer surplus. We build our analysis on a theoretical model of competition between digital companies collecting consumer data. Namely, we consider two firms competing in a product market, with the ability to collect consumer data that they use to price discriminate consumers. Firms strategically choose the number of consumers on whom they collect information and the number of data points that they collect.

Using this framework, we distinguish two dimensions of consumer privacy, which – we will show – are central to understand the relation between data protection and competition, but, to the best of our knowledge, are absent from previous

accessed October 1, 2021.

[7] Statement of Chair Lina M. Khan Regarding the Report to Congress on Privacy and Security, October 1, 2021.

[8] these two opposite effects of information are also well documented in the literature on advertisements, where ads can either be: persuasive and make consumers less price-sensitive, reducing their surplus; or informative and make firms compete for consumers, which increases their surplus.

literature and policy considerations. First, there is the extensive margin of privacy that depends on the number of consumers on whom a firm collects data: when firms collect information on a larger share of consumers, privacy decreases at the extensive margin. Policies aiming a protecting certain groups of consumers - such as children below 16 as enacted in the European GDPR - focus on this extensive margin of privacy. The second margin of privacy relates to the amounts of data collected on each consumer: privacy decreases at the intensive margin when more data are collected by a firm. The GDPR aims a protecting the intensive margin of privacy through a data minimization principle that limits the amounts of data firms can collect on consumers. This methodological distinction is an important contribution to the literature that usually focuses on one dimension of privacy. We will show that the relation between privacy and competition depends on the margin of privacy considered.

We show that an increase in the intensity of competition reduces the incentives of firms to identify consumers and to collect information on them. Hence the promotion of competition is essential to protect consumer privacy. This is a central contribution of this article that has important implication for policy makers and competition authorities. Conversely, we show that regulations protecting consumer privacy can increase or decrease consumer surplus depending on the margin of privacy. When fewer segments are collected (reduction of the intensive margin) firms extract less surplus from consumers and they charge lower prices. However, when fewer consumers are identified by firms (reduction of the extensive margin), the intensity of competition decreases and consumers pay higher prices.

The remainder of the article is organized as follows. We describe the model in Section 2. We analyze the relation between consumer data collection, price discrimination and consumer surplus in Section 3. Section 4 characterizes the impact of competition on the data strategies of the firms. We consider consumers' privacy behavior, general data collection costs, and regulatory instruments in Section 5, and section 6 concludes.

2 Model

We consider a model of competition à la Hotelling in a product market. Consumers are assumed to be 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, 0 and 1.

Firms can collect information that partitions a share of the consumer demand into segments. Firms then use this information to charge personalized prices to consumers on whom they have information, and set a homogeneous price on the remaining consumers. We use this simple model of horizontal differentiation to analyze the impact of information on consumer surplus and on the profits of the firms (Thisse and Vives, 1988).

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. A consumer located at $x \in [0, 1]$ receives a utility V from purchasing the product, but incurs a cost $t > 0$ of consuming a product that does not perfectly fit his 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 the highest level of utility:[9]

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

2.2 Firms

Firms strategically collect information that allows to price discriminate consumers. We describe the data collection strategy of each firm and the structure of information, then we characterize the resulting price structure of each firm.

[9] We assume that the market is covered, so that all consumers buy at least one product from the firms. This assumption is common in the literature. See for instance Bounie et al. (2021b) and Montes et al. (2019).

Share of identified consumers.

Firm 1 and Firm 2 respectively choose the share of consumer $[0, x_1]$ and $[1-x_2, 1]$ on whom they collect data, and to which we refer as the shares of consumers identified by a firm. Firms focus on close-by consumers who have the highest willingness to pay for their products (Iyer et al., 2005), but our model also accounts for the limit case where both firms collect information on all consumers.[10] Firm $\theta = 1, 2$ 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.

Firms can collect consumer information through various means, such as online cookies allowing them to observe the behavior of prospective buyers (Bergemann and Bonatti, 2015). They also have access to the purchase histories of their past customers, which they can use to build consumer scores, in the spirit of behavior-based price discrimination (Bonatti and Cisternas, 2020).

The data collected allows a firm to charge personalized prices to consumers, and Firm θ chooses x_θ 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 profit of the firms. Hence, when choosing the number of identified consumers x_θ, a firm will balance the rent extraction and the competition effects of information. Moreover, it will also account for a data collection cost which we introduce in the next section.

Consumer data collection.

Firm 1 and Firm 2 collect endogenous amounts of information on consumers on $[0, x_1]$ and $[1-x_2, 1]$ respectively. Data allows Firm θ to partition the demand of

[10] It is clear that if Firm 1 identifies consumer located at x_1, it also has interest to identify consumers on $x < x_1$. The same reasoning applies for Firm 2.

identified consumers into k_θ segments of size $\frac{x_\theta}{k_\theta}$.[11]

The fineness of the segments collected $\frac{x_\theta}{k_\theta}$ correspond to the precision of information collected by Firm θ, which increases with k_θ for a given x_θ.

A firm that has collected information can third-degree price-discriminate consumers by charging different prices on different segments. Collecting more data allows firms to locate consumers more precisely. 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 $[0, x_1]$.

The k_θ 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.

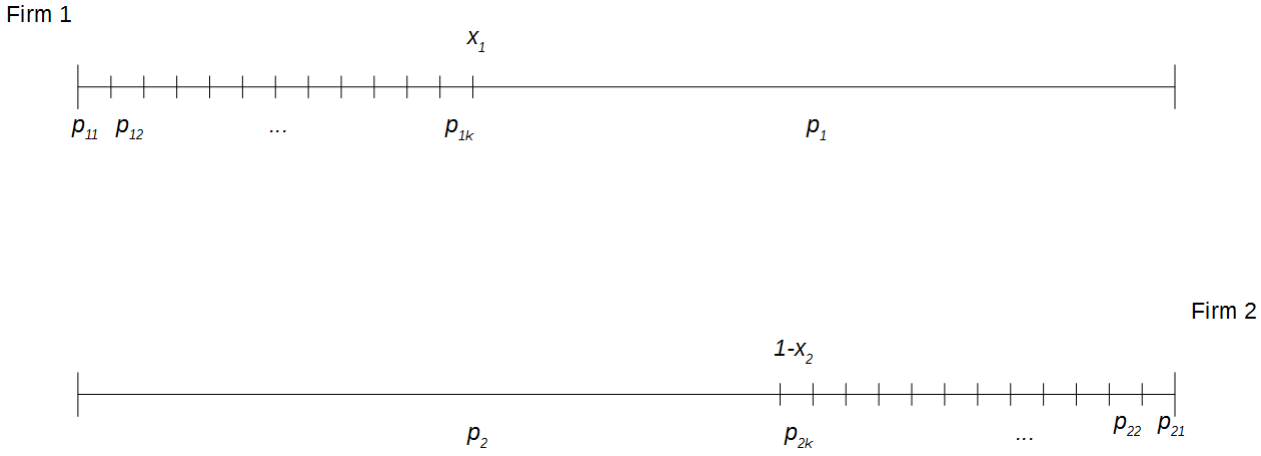

Figure 1: Data collection, Firms 1 and 2

todo mettre x sur k dans la figure

The two margins of privacy.

Our simple framework of competition with consumer data allows us to distinguish two dimensions of consumer privacy. On the one hand, consumer privacy is impacted at the extensive margin by the firms' strategic choices of the shares x_1 and

[11] Modeling information through a partition of the consumer demand was first introduced by Liu and Serfes (2004).

x_2 of identified consumers. On the other hand, the intensive margin of privacy also relates to the strategic choice of the number of consumer segments k_1 and k_2 collected by each firm.

Definition 1

- *The share x_θ of consumers identified by Firm θ corresponds to the extensive margin of privacy.*
- *The number k_θ of consumer segments collected by a firm corresponds to the intensive margin of privacy.*

Data protection laws such as the European GDPR distinguish the extensive and intensive margins of privacy. On the one hand, they aim at protecting groups of consumers such as children below 16, which corresponds to a limit over the share of consumers on whom firms can collect information. On the other hand, they also limit the volumes of data that firms can collect through data minimization principles.

This distinction is absent from the academic literature, which rather focuses on one margin of privacy and ignores the other, providing incomplete analysis of data protection regulations and their interactions with competition laws. Hence we make an important contribution to the literature by showing that the distinction of the two margins of privacy is essential to understand the relation between privacy protection and competition. We will show that protecting the intensive margin of privacy also protects consumer surplus while protecting the extensive margin of privacy lowers the intensity of competition between firms.

We characterize in the remaining of the article the two-way relationship between market competition and the two margins of privacy. We first analyze in Section 3 how a change in x_θ and k_θ – the two margins of privacy – impacts market competition and consumer surplus. We then analyze in Section 4 how a change in the intensity of competition changes the data strategies of the firms, impacting in turn the two margins of privacy.

Data collection cost.

There is a cost to collect consumer data, which increases with the number of identified consumers and with the number of segments collected. Let's denote by $c(k_\theta, x_\theta)$ this cost, which we assume to be multiplicative and quadratic in k_θ and x_θ and equal to $c \cdot k_\theta^2 \cdot x_\theta^2$, $c > 0$ in the baseline analysis. We will show that under this specification the number of segments collected in equilibrium does not depend on the shares of identified consumers. Hence this quadratic multiplicative cost allows us to analyze separately the impact of consumer identification and data collection on consumer surplus. We discuss alternative cost structures in Section 5.2, and we show that the relation between privacy and competition is not affected by the structure of $c(k_\theta, x_\theta)$.

The data collection cost encompasses various dimensions of data collection, such as storing and handling data or any other infrastructure-related costs.[12] A firm will thus balance the benefits from collecting more segments on a large share of consumers and the cost to collect such data.

Modelling Choices

We analyze a competitive market with strategic data collection to analyze the relation between competition and privacy. The Hotelling competition model is the simplest framework that captures two dimensions of consumer privacy. On the one hand, the extensive margin of privacy appears clearly through the strategic choice of the share of consumers identified by each firm. On the other hand, representing information that partitions consumer demand into segments of endogenous sizes – the intensive margin of privacy – also corresponds well to the marketing strategies of data-driven companies.

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

[12]Such cost structure is introduced by Bounie et al. (2020).

correspond to the two margins of privacy. Our results would hold under a different timing. For instance, in a companion article we consider competing data intermediaries who first collect information on the whole unit line, and then choose how many segments they sell to firms (Bounie et al., 2021a). We will see that the optimal shares of consumers identified by firms are identical under the different timings.

2.3 Profits and timing

Profits.

Firms choose the shares x_1 and x_2 of consumers on which they will collect data and charge personalized prices. We denote by $d_{\theta i}$ the demand of Firm θ on the ith segment of consumer that it price discriminates. $d_{\theta i}$ depends on the size of the segment which is defined by the number of data collected k_θ: for Firm 1, $d_{1i} = \frac{x_1}{k_1}$, and for Firm 2 $d_{2i} = \frac{x_2}{k_2}$. An informed Firm θ maximizes the following profit function with respect to $p_{\theta 1}, .., p_{\theta n}$:

$$\pi_\theta(p_{\theta i}, p_\theta) = \sum_{i=1}^{k_\theta} d_{\theta i} p_{\theta i} + p_\theta d_\theta, \tag{1}$$

We assume that firms set prices in two stages. First, Firm 1 and Firm 2 simultaneously set homogeneous prices p_1 and p_2 on the whole unit line. Secondly, Firm θ sets a personalized price on each consumer segment on $[0, x_\theta]$, with $p_{\theta i}$ being the price on the ith segment from the origin. Then consumers observe prices. When setting the competitive price p_θ, Firm θ already knows which consumers it price discriminate, and thus charges p_θ accordingly.

Sequential pricing decision avoids the non-existence of Nash equilibrium in pure strategies, 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 intertemporal price discrimination with incomplete information on consumer demand. Jentzsch et al. (2013), Lam et al. (2020) and Dubus (2021) also focus on sequential

pricing where a higher personalized price is charged to identified consumers after a firm sets a uniform price.

Sequential pricing is also common in business practices (see also Fudenberg and Villas-Boas (2006)). Recently, Amazon has been accused to show 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.

Timing.

The timing of the game is the following:

Stage 1: Firms choose the numbers x_1 and x_2 of consumers on whom they collect information.

Stage 2: Firm θ collects k_θ segments of consumers on the share x_θ of identified consumers.

Stage 3: Firms set prices p_1 and p_2 on the segment of unidentified consumers.

Stage 4: Each firm charges personalized prices to identified consumers by setting $p_{\theta i}$ $(i \in [1, x_\theta(k_\theta])$.

Equilibrium Concept.

Our focus is on subgame perfect equilibrium, and we will solve the game by backward induction. In particular, the number of consumer segments collected by each firm in Stage 2 will maximize their profits by balancing a better extraction of consumer surplus and higher data collection costs.

The values of x_θ are chosen as simultaneous best responses by the firms. Hence, taking the equilibrium prices and values of k_θ, our focus in the first stage of the game in on Pure Strategy Nash Equilibria. In particular, we require no assumption on commitment by the firms regarding their data collection strategies.

3 Privacy Protection and Consumer Surplus

We analyze how consumer surplus is impacted by a change of privacy at the extensive and intensive margins, using comparative statics with respect to x_θ and to k_θ.

3.1 Consumer surplus and the extensive margin of privacy

We analyze how the strategic choice of x_θ by Firm θ impacts competition and consumer surplus: $CS(x_1, x_2, k_1, k_2)$. We show that surplus increases when more consumers are targeted by a firm, which highlights a tension between privacy protection at the extensive margin and the protection of consumer surplus.

Suppose initially that Firm 1 has collected information on x_1 consumers, and Firm 2 has information on x_2 consumers. If Firm 1 identifies an additional element of consumer demand on $[x_1, x_1 + \epsilon]$, there are two effects on consumer surplus:

1. A rent extraction effect: Firm 1 price discriminates consumers on $[x_1, x_1+\epsilon]$, which reduces their surplus.

2. A competitive effect: Firm 1 lowers its price on $[x_1 + \epsilon, 1]$, which increases the competitive pressure on Firm 2. In turn, Firm 2 lowers its price, which has a positive effect on the surplus of consumer over the whole unit line.

Overall, the second effect always dominates the first, and aggregate consumer surplus increases when more consumers are identified. Indeed, the rent extraction effect only increases profits on the ϵ consumers who are newly identified, while the competitive effect affects the whole Hotelling line.

However, an increase in the number of identified consumers will impact consumers differently depending on whether they are newly identified or not. Newly identified consumers will pay a higher price than when they remain unidentified, and their surplus decreases. Consumers who pay a homogeneous price benefit from a more intense competition and their surplus increases when firms identify a larger share of consumers. Moreover, there are also consumers identified by a firm,

but for which the competitive impact of information dominates the rent extraction effect when a large share of consumers is identified.

We compare surplus without information, and surplus in equilibrium when firms identify their optimal shares of consumers. This allows us to characterize the relation between identification and consumer surplus in Proposition 1.

Proposition 1

Compared to a situation without information, the ability to identify x_θ consumers (extensive margin):

- *Increases total consumer surplus.*
- *Decreases the surplus of consumers on* $[0, \frac{1}{2} - \frac{2x_2}{3} - \frac{x_1}{3}] \cup [\frac{1}{2} + \frac{2x_1}{3} + \frac{x_2}{3}, 1]$ *pay a higher price.*
- *Increases the surplus of remaining consumers*

Proof: see Appendix A.1.

Proposition 1 highlights a general positive relation between identification and consumer surplus. A regulator thus faces a two-edged sword, as protecting consumer privacy reduces aggregate consumer surplus.

Moreover, consumers may benefit or suffer from identification depending on their location on the line. When a firm identifies more consumers, competition on the rest of the line increases, inducing a positive externality on other consumers. Hence, Proposition 1 highlights the difficulties for a regulator to protect consumer surplus: aggregate surplus increases with more identification, but individual consumers may suffer from price discrimination.

3.2 Consumer surplus and the intensive margin of privacy.

We now discuss the relation between the intensive margin of privacy and consumer surplus. More precise information increases the ability of a firm to price discriminate consumers, which reduces consumer surplus.

Proposition 2

When the number of consumer segments k_θ collected by Firm θ increases (intensive margin):

- *Aggregate consumer surplus decreases.*
- *The surplus of consumers identified by Firm θ decreases.*
- *The surplus of remaining consumers remains constant.*

Proof: see Appendix A.2.

Consumers who are price discriminated pay a higher price when a firm has more precise information about them. Contrary to x_θ, the number of segments k_θ collected by Firm θ does not have an impact on its competitor but only a negative rent extraction effect. Hence a better protection of consumer privacy at the intensive margin will also protect consumer surplus by limiting the ability of firms to charge high prices. For instance, data protection laws in Europe impose a data minimization principle, under which firms must collect a limited amount of data. This result contributes to the literature that usually finds a tension between privacy protection and the intensity of competition (Lee et al., 2011). Focusing on the intensive margin of privacy, we show that protecting consumer privacy also increases consumer surplus.

4 Competition and Privacy

We characterize the impact of competition on the data strategies of the firms by comparing market equilibrium when Firm 1 is a monopolist and when it competes with Firm 2. We show that promoting competition between firms allows to protect both margins of privacy.

We consider the data collection and pricing strategies of Firm 1 located at 0 in monopoly, which we compare with those of firms when they compete.

Proposition 3

- *Extensive margin of privacy: a monopolist identifies more consumers than competing firms.*
- *Intensive margin of privacy: a monopolist collects thinner consumer segments than competing firms.*

Proof: see Appendix A.3.

When a firm is a monopolist, there is no competitive effect of consumer identification, and it has interest to identify a larger share of consumers on the unit line. Moreover, the profitability of consumer data is higher than in competition, as information allows a better extraction of consumer surplus for a monopolist. It is clear that the two dimensions of privacy are negatively related to consumer surplus in a monopoly market. Consumers pay a higher price when they are price discriminated, and the ability of a firm to extract their surplus increases with the amount of data collected.

5 Extensions, Discussion and Robustness

5.1 Intensity of Competition and Consumer Privacy

We analyze in this section how the intensity of competition between firms – captured by transport cost t – impacts their incentives to identify consumers and to collect consumer data, impacting in turn the two margin of privacy. We show that when the intensity of competition decreases (t increases), consumer privacy decreases both at the intensive and extensive margins.

We first show in Proposition 4 that firms collect more consumer segments when t increases:

Proposition 4

A decrease of the intensity of competition increases the amount of consumer data collected by firms:

$$\frac{\partial k_\theta^*}{\partial t} > 0.$$

Proof: see Appendix A.4.

The lower the intensity of competition the higher the surplus extraction effect and the lower the competition effect due to information. A decrease of the intensity of competition increases the marginal value of information for the firms, which will thus collect more consumer data. This result is in line with recent empirical evidence showing that firms collect fewer consumer data in competitive environments (Kesler et al., 2017).

Moreover, the shares of consumers identified increase with t:

Proposition 5

The number of consumers identified by firms increases when the intensity of competition decreases:

$$\frac{\partial x_\theta^*}{\partial t} > 0.$$

Proof: see Appendix A.5.

The profitability of information decreases with the intensity of competition, and Proposition 5 highlights a positive relation between competition and consumer privacy protection at the extensive margin.

Hence our results emphasize the importance of protecting the competitiveness of markets to protect consumer privacy both at the intensive and extensive margin.

5.2 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 collection, such as increasing and decreasing returns to scale, and multiplicative and additive costs with respect to k_θ and x_θ. We show how different cost structures impact our results.

From the proof of Proposition 3 we can write the profit function of a firm the following way:

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

Hence the first degree derivative of the profits with respect to k_θ can be written as:[13]

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

FOCs imply that

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

Multiplicative cost functions.

We first consider the class of multiplicative cost functions $c(k_\theta, x_\theta) = c \cdot k_\theta^\alpha \cdot x_\theta^\beta$.[14] 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 specification characterizes both increasing and constant and returns to scale w.r.t. k_θ and increasing and decreasing returns to scale w.r.t. x_θ, depending on the values of $\alpha > 1$ and $\beta > 0$. Considering the FOCs, we have:

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

Hence, k_θ^* increases with x_θ when $\beta < 2$ and k_θ^* decreases with x_θ when $\beta > 2$. With the specification of our main analysis where $c(.)$ is proportional to x_θ^2, $2 - \beta$ simplifies and k_θ^* does not depend on x_θ. Hence $\beta = 2$ is the threshold value above which k_θ and x_θ are strategic substitutes and below which they are strategic complements.

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

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

Additive cost functions.

We consider now the class of costs functions where the costs associated with x_θ and k_θ are additive: $c(k_\theta, x_\theta) = c \cdot (k_\theta^\alpha + x_\theta^\beta)$.[15] In this case the FOC w.r.t. k_θ implies:

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

And k_θ always increases with x_θ.

Market outcomes with general cost functions.

The main results of the analysis hold:

- All things equal, consumer surplus increases with higher level of privacy at the intensive margin, and decreases with a higher level of privacy at the extensive margin.
- A more competitive market will protect both margins of consumer privacy by reducing the share of identified consumers and the number of consumer segments collected by firms.

Moreover, we identify another effect when k_θ and x_θ are strategic complements, resulting from the strategic choice of consumer segments collected by each firm. 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.

5.3 Privacy Protection and Competition

A data protection agency can set a limit $\bar{k}$ over the amount of consumer data collected by the firms. Such regulation aims at protecting consumer privacy by

[15] 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 $\alpha \geq 1$.

forcing firms to collect as little data as possible. A data minimization principle, aiming at limiting the amounts of data collected by firms, is enacted for instance in the European GDPR. We analyze the implications for market equilibrium of a change in the maximal amount of consumer data that firms can collect.

Proposition 6

- *Protecting the intensive margin of consumer privacy reduces privacy at the extensive margin when k_θ and x_θ are strategic substitutes.*
- *Protecting the intensive margin of consumer privacy also protects privacy at the extensive margin when k_θ and x_θ are strategic complements.*

When k_θ and x_θ are strategic complements, the marginal gain from collecting data decreases when fewer consumer segments are collected, and firms identify fewer consumers. The opposite holds when k_θ and x_θ are strategic substitutes, and this result highlights the need to better understand the costs that firms bear when collecting, treating and processing consumer data.

5.4 Consumers' Privacy Behavior

Consumers may anticipate the data strategies of the firms and hide to pay a lower price (Wathieu, 2006; Montes et al., 2019). We consider such strategic hiding in this section, which corresponds to the self-protection of their privacy by consumers. An identified consumer balances the benefits from hiding and paying a homogeneous cost, and from remaining identified and being charged a personalized price.

We use this framework to compare welfare when consumers protect their privacy with welfare when privacy is protected by a regulator. We will show that an efficient regulatory framework for privacy protection articulates self-protection by consumers with regulatory rules.todo

We highlight new effects of consumer's hiding on firms and consumers. We show that when consumers hide, firms identify consumers located further away

from their locations, and collect fewer consumers segments as the profitability of data decreases. Hence competition is higher, and rent extraction lower when consumers can hide. These results allow us to identify new types of externalities between consumers: when consumers hide, new consumers are identified, which reduces privacy at the extensive margin. However, as fewer segments are collected, privacy increases at the intensive margin.

Model with strategic hiding.

We consider a second unit line of size 1 (without loss of generality) where consumers are uniformly distributed, and that we denote by H. In market H firms are also located at the extremities of the line, and cannot collect consumer data for instance because they do not have cookies on these consumers. In the first stage of the game, consumers in the market with data collection, that we denote by I, can hide from firms and be charged the same homogeneous price as consumers in market H. For instance consumers can use privacy enhancing technologies such as Ghostery or private web browsers such as DuckDuckGo. There is a cost to this privacy-preserving behavior, such as an opportunity cost for not enjoying personalized recommendations. We denote by c_p such privacy cost, that we assume to be identical for all consumers for simplicity. When choosing whether to hide, consumers will balance the benefits of paying a lower homogeneous price with the cost of hiding.

This specification of consumer's privacy choice is standard in the literature on privacy (Belleflamme and Vergote, 2016; Montes et al., 2019).

Profits.

Firms makes profits on both markets. On market I, firms charge personalized prices $p_{\theta i}$ to a share of consumers who have not hidden. When consumer located at h_1 hides from Firm 1, it is clear by revealed preferences that consumers at $x < h_1$ also hide from Firm 1. The same applies for consumers located at $x > 1 - h_2$. When consumers located on $[0, h_1]$ hide from Firm 1 and consumers on $[1 - h_2, 1]$ hide from Firm 2. Hence each firm price discriminates a share of consumers equal

to $[h_\theta, x_\theta]$, and consumers on $[x_1, 1-x_2]$ are charged homogeneous prices $p_{\theta I}$. Firm θ then collects k_θ segments of size $x_\theta - h_\theta$.

On market H consumers are charged homogeneous prices $p_{\theta H}$. Consumer demand is composed of the uniform demand initially on market H, as well as of consumers on $[0, x_1] \cup [1 - x_2, 1]$ who migrated from market I. Denoting by $\tilde{x}$ the indifferent consumer on both markets, the overall profits of the firms can be written:

$$\pi_\theta(x_\theta, k_\theta) = \int_0^{h_\theta} p_{\theta H} dx + \int_{x_\theta}^{\tilde{x}} p_{\theta H} dx + \int_0^{\tilde{x}} p_{\theta H} dx + \sum_{i=1}^{k_\theta} \frac{x_\theta - h_\theta}{k_\theta} p_{\theta i}.$$

Timing.

The timing of the game with strategic hiding is as follows.

Stage 0: Consumers on market I choose whether to hide from firms or to have their data collected.

Stage 1: Firms choose the numbers x_1 and x_2 of consumers on whom they collect information.

Stage 2: Firm θ collects k_θ segments of consumers on the share $x_\theta - h_\theta$ of identified consumers.

Stage 3: Firms set prices p_{1I} and p_{2I} on the competitive segment of market I, and prices p_{1H} and p_{2H} on market H.

Stage 4: Each firm charges personalized prices to price discriminated consumers by setting $p_{\theta i}$.

Market equilibrium with privacy behavior.

We compare consumer surplus with consumer privacy behavior and surplus in the baseline model.

Proposition 7

Compared to the baseline model, when consumers can hide:

- *The intensity of competition is higher in the market with identified consumers.*
- *Consumer surplus is lower in the market with hidden consumers.*

Proof: see Appendix A.6.

Proposition 7 highlights opposite effects of privacy behaviors on consumers, resulting from two opposite externalities. In market I firms identify consumers located further away from their location when some consumers hide, which increases the intensity of competition. Firms have interest to identify more consumers and the values of x_1 and x_2 increase.

We also find a negative externality of hiding behaviors on other consumers in line with recent literature on privacy (Belleflamme and Vergote, 2016; Montes et al., 2019).[16] Consumers hiding are those with the highest willingness to pay for a firm's product, and firms increase prices in market H.

6 Conclusion

The relation between privacy and competition is today the source of intense debates between economists regarding the ability of competition authorities to protect consumer welfare. Our article contributes to these debates by emphasizing how the protection of consumer privacy and of consumer surplus cannot be achieved independently from each other. Data protection agencies and competition authorities must work together to better achieve data protection policies that protect consumer surplus.

We show that regulatory measures that aim at limiting data collection by firms will also increase consumer surplus. This can be achieved through regulations that

[16] Alternatively, Choi et al. (2019) and Acemoglu et al. (2019) develop models with data externalities, and show that too much data can be collected on consumers compared with a social optimum.

prohibit third party data collection, and through a tighter enforcement of illegal data acquisition. Practices of consumer information sharing between firms should also be limited as they will harm consumers through a better identification of their preferences.

Our results also suggest that regulators can protect at the same time consumer privacy and consumer surplus. Data protection agencies and competition authorities must carefully balance the loss in terms of surplus when protecting a group of consumers from having their data collected. In this respect, third parties selling information to firms such as data brokers can have a positive impact on consumer surplus, under the condition that they only sell relatively coarse information. New techniques of high precision price discrimination using big data analysis should raise concern regarding both the protection of consumer privacy, and the reduction of market competition.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2019). Too much data: Prices and inefficiencies in data markets. Technical report, National Bureau of Economic Research.

Acquisti, A. and H. R. Varian (2005). Conditioning prices on purchase history. *Marketing Science 24*(3), 367–381.

Belleflamme, P. and W. Vergote (2016). Monopoly price discrimination and privacy: The hidden cost of hiding. *Economics Letters 149*, 141–144.

Bergemann, D. and A. Bonatti (2015). Selling cookies. *American Economic Journal: Microeconomics 7*(3), 259–94.

Bonatti, A. and G. Cisternas (2020). Consumer scores and price discrimination. *The Review of Economic Studies 87*(2), 750–791.

Bounie, D., A. Dubus, and P. Waelbroeck (2020). Market for information and selling mechanisms. *Available at SSRN 3454193*.

Bounie, D., A. Dubus, and P. Waelbroeck (2021a). Competition and mergers with strategic data intermediaries. *Available at SSRN*.

Bounie, D., A. Dubus, and P. Waelbroeck (2021b). Selling strategic information in digital competitive markets. *The RAND Journal of Economics 52*(2), 283–313.

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.

Crémer, J., Y.-A. de Montjoye, and H. Schweitzer (2019). Competition policy for the digital era. *Report for the European Commission*.

Dubus, A. (2021). Behavior-based price discrimination with strategic customer targeting.

Fudenberg, D. and J. M. Villas-Boas (2006). Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems 1*, 377–436.

Furman, J., D. Coyle, A. Fletcher, D. McAules, and P. Marsden (2019). Unlocking digital competition: Report of the digital competition expert panel. *Report prepared for the Government of the United Kingdom, March*.

Hagiu, A. and J. Wright (2020). When data creates competitive advantage. *Harvard business review 98*(1), 94–101.

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.

Kesler, R., M. Kummer, and P. Schulte (2017). User data, market power and innovation in online markets: Evidence from the mobile app industry. Technical report, Working Paper.

Lam, W., P. Belleflamme, and W. Vergote (2020). Competitive imperfect price discrimination and market power. *Marketing Science*.

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.

Scott Morton, F., P. Bouvier, A. Ezrachi, B. Jullien, R. Katz, G. Kimmelman, A. Melamed, and J. Morgenstern (2019). Committee for the study of digital platforms: Market structure and antitrust subcommittee-report. *Chicago: Stigler Center for the Study of the Economy and the State, University of Chicago Booth School of Business*.

Thisse, J.-F. and X. Vives (1988). On the strategic choice of spatial price policy. *The American Economic Review*, 122–137.

Tirole, J. (2020). Competition and the industrial challenge for the digital age.

Wathieu, L. (2006). Marketing and privacy concerns.

A Appendix

A.1 Proof of Proposition 1

We analyze the variation of consumer surplus with the number of consumers price discriminated by a firm.

We first characterize the equilibrium prices in period 2 when firms price discriminates x_1 and x_2 consumers respectively.

Prices and demand

Firm $\theta = 1, 2$ sets a price $p_{\theta i2}$ for each segment of size $\frac{x_\theta}{k_\theta}$, and a unique price $p_{\theta 2}$ on the rest of the unit line. The demand for Firm θ on collected segments is $d_{\theta i} = \frac{x_\theta}{k_\theta}$. The corresponding prices are computed using the indifferent consumer located on the right extremity of the segment, $\frac{ix_\theta}{k_\theta}$. For Firm 1:

$$\begin{aligned} & V - t\frac{ix_1}{k_1} - p_{1i} = V - t(1 - \frac{ix_1}{k_1}) - p_2 \\ \implies & \frac{ix_1}{k_1} = \frac{p_2 - p_{1i} + t}{2t} \\ \implies & p_{1i} = p_2 + t - 2t\frac{ix_1}{k_1}. \end{aligned}$$

p_2 is the price set by Firm 2 on interval $[0, 1 - x_2]$ where it does not price discriminate consumers. Prices set by Firm 2 on segments in interval $[1 - x_2, 1]$ are:

$$p_{2i} = p_1 + t - 2t\frac{ix_2}{k_2}.$$

Let denote d_1 the demand for Firm 1 (resp. d_2 the demand for Firm 2) where firms compete. d_1 is determined by the indifferent consumer $\hat{x}$:

$V - t\hat{x} - p_1 = V - t(1 - \hat{x}) - p_2 \implies \hat{x} = \frac{p_2 - p_1 + t}{2t}$ and $d_1 = \hat{x} - x_1 = \frac{p_2 - p_{12} + t}{2t} - x_1$ (resp. $d_2 = 1 - x_2 - \frac{p_2 - p_1 + t}{2t}$).

Profits of the firms.

The profits of the firms are:

$$\begin{aligned} \pi_1 &= \sum_{i=1}^{k_1} d_{1i}p_{1i} + d_1 p_1 = \sum_{i=1}^{k_1} \frac{x_1}{k_1}(p_2 + t - 2t\frac{ix_1}{k_1}) + (\frac{p_2 - p_1 + t}{2t} - x_1)p_1, \\ \pi_2 &= \sum_{i=1}^{k_2} d_{2i}p_{2i} + d_2 p_2 = \sum_{i=1}^{k_2} \frac{x_2}{k_2}(p_1 + t - 2t\frac{ix_2}{k_2}) + (\frac{p_1 - p_2 + t}{2t} - x_2)p_2. \end{aligned}$$

Prices and demands in equilibrium.

We now compute the optimal prices and demands, using first-order conditions on π_θ with respect to p_θ. Prices in equilibrium are:

$$p_1 = t[1 - \frac{2}{3}x_2 - \frac{4}{3}x_1],$$
$$p_2 = t[1 - \frac{2}{3}x_1 - \frac{4}{3}x_2].$$

Replacing these values in the above demands and prices gives:

$$p_{1i} = 2t - \frac{4}{3}x_2 t - \frac{2}{3}x_1 t - 2\frac{itx_1}{k_1},$$
$$p_{2i} = 2t - \frac{4}{3}x_1 t - \frac{2}{3}x_2 t - 2\frac{itx_2}{k_2}.$$

and

$$d_1 = \frac{1}{2} - \frac{2}{3}x_1 - \frac{1}{3}x_2,$$
$$d_2 = \frac{4}{3}x_2 - \frac{1}{2} - \frac{1}{3}x_1.$$

Consumer surplus

Consumer surplus when Firm 1 identifies x_1 consumers and Firm 2 identifies x_2 consumers is defined as follows.

$$
\begin{aligned}
CS(x_1,x_2,k_1,k_2) = &\sum_{i=1}^{k_1}[\int_0^{\frac{x_1}{k_1}} V-2t[1-\frac{1}{3}x_1-\frac{2}{3}x_2-\frac{ix_1}{k_1}]-tx\mathrm{d}x] \\
&+\int_{x_1}^{\frac{1}{2}+\frac{x_1}{3}-\frac{x_2}{3}} V-t[1-\frac{4}{3}x_1-\frac{2}{3}x_2]-tx\mathrm{d}x \\
&+\int_{\frac{1}{2}+\frac{x_1}{3}-\frac{x_2}{3}}^{1-x_2} V-t[1-\frac{2}{3}x_1-\frac{4}{3}x_2]-tx\mathrm{d}x \\
&+\sum_{i=1}^{k_2}[\int_0^{\frac{x_2}{k_2}} V-2t[1-\frac{1}{3}x_2-\frac{2}{3}x_1-\frac{ix_2}{k_2}]-tx\mathrm{d}x]
\end{aligned}
$$

$$
\begin{aligned}
&=\sum_{i=1}^{k_1}\frac{x_1}{k_1}(V-2t[1-\frac{1}{3}x_1-\frac{2}{3}x_2-\frac{ix_1}{k_1}])-\frac{x_1^2t}{2k_1^2} \\
&+\sum_{i=1}^{k_2}\frac{x_2}{k_2}(V-2t[1-\frac{1}{3}x_2-\frac{2}{3}x_1-\frac{ix_2}{k_2}])-\frac{x_2^2t}{2k_2^2} \\
&+V[1-x_2-x_1]-[\frac{1}{2}-\frac{2x_1}{3}-\frac{x_2}{3}]t[1-\frac{4}{3}x_1-\frac{2}{3}x_2] \\
&-[\frac{1}{2}-\frac{2x_2}{3}-\frac{x_1}{3}]t[1-\frac{4}{3}x_2-\frac{2}{3}x_1]-t[\frac{1}{4}-\frac{1}{9}x_1x_2-\frac{7}{18}x_2^2-\frac{7}{18}x_1^2]
\end{aligned}
$$

$$
\begin{aligned}
&=x_1[V-2t[1-\frac{1}{3}x_1-\frac{2}{3}x_2]+\frac{x_1^2(k_1+1)t}{k_1}-\frac{x_1^2t}{2k_1^2} \\
&+x_2[V-2t[1-\frac{1}{3}x_2-\frac{2}{3}x_1]+\frac{x_2(k_2+1)t}{k_2}-\frac{x_2^2t}{2k_2^2} \\
&+V[1-x_2-x_1]+t[-\frac{5}{4}+\frac{1}{3}x_1+\frac{1}{3}x_2+\frac{5}{6}x_1^2+\frac{5}{6}x_2^2-2x_1x_2]
\end{aligned}
$$

$$
\begin{aligned}
&=V+t[-\frac{5}{4}+\frac{17}{18}x_1^2+\frac{17}{18}x_2^2+x_1x_2]+\frac{1}{2}\frac{x_1^2t}{k_1^2}+\frac{1}{2}\frac{x_2^2t}{k_2^2}
\end{aligned}
\tag{2}
$$

Aggregate consumer surplus and consumer privacy

We show that consumer surplus always increases with the number of consumers identified.

The first degree derivative of $CS(x_1,x_2,k_1,k_2)$ with respect to x_1 is

$$
\frac{\partial CS}{\partial x_1}=\frac{17tx_1}{9}+tx_2+\frac{x_1t}{k_1^2}
$$

which is always greater than zero.

Individual consumer surplus

We compare the prices paid by identified consumers when Firm 1 and Firm 2 choose x_1 and x_2 respectively with prices without information as in the standard Hotelling model: $p_1 = p_2 = t$.

Identified consumers pay prices equal to

$$p_{1i} = 2t - \frac{4}{3}x_2 t - \frac{2}{3}x_1 t - 2\frac{i x_1 t}{k_1},$$
$$p_{2i} = 2t - \frac{4}{3}x_1 t - \frac{2}{3}x_2 t - 2\frac{i x_2 t}{k_2}.$$

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

A.2 Proof of Proposition 2

It is straightforward that CS decreases with k_θ.

Moreover, it is clear that prices charged to identified consumers increase with the precision of information:

$$p_{1i} = 2t - \frac{4}{3}x_2 t - \frac{2}{3}x_1 t - 2\frac{i x_1 t}{k_1},$$
$$p_{2i} = 2t - \frac{4}{3}x_1 t - \frac{2}{3}x_2 t - 2\frac{i x_2 t}{k_2}.$$

Because an increase in k_θ only impacts the prices charged to consumers targeted by Firm θ, the surplus of unidentified consumers remains constant regardless of the number of segments collected by the firms.

A.3 Proof of Proposition 3

We focus on Firm 1 in monopoly without loss of generality.

The price charged to segment i of identified consumers is equal to:

$$p_{1i} = V - t\frac{x_1 i}{k_1}.$$

We assume the market covered, and the price charged to the remaining consumers is $p_1 = V - t$

The profit of Firm 1 identifying $x_1 \leq 1$ consumers and collecting k_1 segments is equal to:

$$\begin{aligned}\pi_1(x_1,k_1) &= \sum_{i=1}^{k_1}\frac{x_1}{k_1}(V - t\frac{x_1 i}{k_1}) + (V-t)(1-x_1) - ck_1^2x_1^2 \\ &= x_1V - tx_1^2 - t\frac{x_1^2}{k_1} + (V-t)(1-x_1) - ck_1^2x_1^2\end{aligned} \tag{3}$$

It is clear that with our specification of the data collection cost, k_1 is chosen independently of x_1. Focs with respect to x_1 yield:

$$x_1 = \frac{t}{2t + 2t\frac{1}{k_1} + 2ck_1^2}.$$

Data collection in competition.

we consider the profits the firms in competition:

$$\pi_1(x_1,k_1) = \sum_{i=1}^{k_1}\frac{x_1}{k_1}(p_2 + t - 2t\frac{ix_1}{k_1}) + (\frac{p_2 - p_1 + t}{2t} - x_1)p_1,$$
$$\pi_2(x_2,k_2) = \sum_{i=1}^{k_2}d_{2i}p_{2i} + d_2p_2 = \sum_{i=1}^{k_2}\frac{x_2}{k_2}(p_1 + t - 2t\frac{ix_2}{k_2}) + (\frac{p_1 - p_2 + t}{2t} - x_2)p_2.$$

Focs w.r.t. x_1 and x_2 give:

$$\pi_1(x_1,k_1)^* = \frac{t}{2} - \frac{7}{9}x_1^2t + \frac{2t}{9}x_2^2 - \frac{4}{9}x_1x_2t + \frac{2}{3}x_1t - \frac{2}{3}x_2t - \frac{x_1^2t}{k_1} - ck_1^2x_1^2,$$
$$\pi_2^*(x_2,k_2) = \frac{t}{2} - \frac{7}{9}x_2^2t + \frac{2t}{9}x_1^2 - \frac{4}{9}x_1x_2t + \frac{2}{3}x_2t - \frac{2}{3}x_1t - \frac{x_2^2t}{k_2} - ck_2^2x_2^2.$$

Focs w.r.t. x_1 and x_2 give:

$$x_1^* = \frac{(5k_2+9)k_1t^2 + 9ck_2^3k_1t}{((15k_2+21)k_1 + 21k_2 + 27)t^2 + (c(21k_2+27)k_1^3 + 21ck_2^3k_1 + 27ck_2^3)t + 27c^2k_2^3k_1^3}$$

and similarly for x_2.

It is immediate that data collection does not depend on x_θ in the competitive framework either, and competition does not change the number of segments collected by the firms with this cost function.

Straightforward comparison of x_1 in monopoly and in competition shows that more consumers are identified when Firm 1 is a monopolist.

A.4 Proof of Proposition 4

The number of segments collected by a firm is determined by considering the derivative of its profits with respect to k_θ.

Consider Firm 1, as we have:

$$\frac{\partial \pi_1(x_1, k_1)}{\partial k_1} = \frac{x_1^2 t}{k_1^2} - 2ck_1x_1^2.$$

$$k_1^* = \left(\frac{t}{2c}\right)^{1/3}$$

It is clear that data collection increases with t.

A.5 Proof of Proposition 5

Going back to our expression of x_1^* in competition and replacing k_1 and k_2 by their equilibrium values we obtain:

$$x_\theta^* = \frac{5 \cdot 2^{4/3} \cdot t + 27 \cdot 2^{2/3} \cdot c^{1/3} \cdot t^{2/3}}{15 \cdot 2^{4/3} \cdot t + 63 \cdot 2^{5/3} \cdot c^{1/3} \cdot t^{2/3} + 243 \cdot c^{2/3} \cdot t^{1/3}}$$

which clearly increases with t.

A.6 Proof of Proposition 7

We provide the functions of the location of the last hidden consumer, and prices on both markets. Consumers hide when the price they pay is higher than the homogeneous price plus a privacy cost c_p.

Prices set by Firm θ on the ith segment closest to its location is:

$$p_{\theta i} = p_{-\theta H} + t - 2t(h_\theta + i\frac{x_\theta - h_\theta}{k_\theta}).$$

Which decreases with i, and only consumers at the extremities consider hiding. $p_{\theta H}$ is found by maximizing firms' profits on the anonymous market:

$$\pi_{\theta H} = \int_0^{h_\theta} p_{\theta H} dx + \int_0^{x_\theta} p_{\theta H} dx + 2\int_{x_\theta}^{\frac{p_{-\theta H} - p_{\theta H} + t}{2t}} p_{\theta H} dx$$

We find:

$$p_{\theta H}^* = t + \frac{2h_\theta t}{3} + \frac{h_{-\theta} t}{3} - \frac{2x_\theta t}{3} - \frac{x_{-\theta} t}{3}$$

$$p_{\theta i} = 2t + \frac{h_\theta t}{3} + \frac{2h_{-\theta} t}{3} - \frac{x_\theta t}{3} - \frac{2x_{-\theta} t}{3} - 2t(h_\theta + i\frac{x_\theta - h_\theta}{k_\theta}).$$

consumers hide if

$$p_{\theta i} = 2t + \frac{h_\theta t}{3} + \frac{2h_{-\theta} t}{3} - \frac{x_\theta t}{3} - \frac{2x_{-\theta} t}{3} - 2th_\theta \geq c_p + p_{\theta H} = c_p + t + \frac{2h_\theta t}{3} + \frac{h_{-\theta} t}{3} - \frac{2x_\theta t}{3} - \frac{x_{-\theta} t}{3}$$

$$t + \frac{h_{-\theta} t}{3} + \frac{x_\theta t}{3} - 2th_\theta \geq c_p + \frac{h_\theta t}{3} + \frac{x_{-\theta} t}{3}$$

$$h_\theta = \frac{1}{7}(3 + h_{-\theta} + x_\theta - 3\frac{c_p}{t} - x_{-\theta})$$

Profits are:

$$\begin{aligned}
\pi_1^* &= \sum_{i=1}^{k_1} \frac{x_1 - h_1}{k_1}[2t + \frac{h_1 t}{3} + \frac{2h_2 t}{3} - \frac{x_1 t}{3} - \frac{2x_2 t}{3} - 2t(h_1 + i\frac{x_1 - h_1}{k_1})] + \pi_{1H}(x_1) \\
&= t(1 - \frac{2h_1}{3} + \frac{h_1^2}{9} - \frac{2h_1^2}{k_1} + \frac{2h_2}{3} + \frac{h_2^2}{9} - h_1 h_2 \frac{2}{9} + x_1 h_1(\frac{4}{3} + \frac{4}{k_1}) \\
&+ x_2 \frac{2h_1}{3} - x_1 x_2 \frac{4}{9} - x_1^2(\frac{13}{9} + \frac{2}{k_\theta}) - \frac{x_2^2}{9}) - ck_1^2 x_1^2
\end{aligned}$$

FOCs on $\pi_{\theta H}$ with respect to x_θ gives us the optimal pricing strategy of Firm θ. We substitute this value in the expression of h_θ.

The last identified consumers are located at:

$$x_1 = \frac{(k_1(h_1(26k_2 + 36) - 4h_2 k_2 - 12h_2) + h_1(78k_2 + 108))t^2 + (18ch_1 k_1 k_2^3 + 54ch_1 k_2^3)t}{(k_1(55k_2 + 78) + 78k_2 + 108)t^2 + (39ck_1 k_2^3 + 54ck_2^3 + ck_1^3(39k_2 + 54))t + 27c^2 k_1^3 k_2^3}$$

Which increases with h_1 for $k_1, k_2 > 1$.