Selling Strategic Information in Digital Competitive Markets*

David Bounie,[†] Antoine Dubus[‡] and Patrick Waelbroeck[§]

September 28, 2018

Abstract

This article investigates the strategies of a data broker when selling information to one or to two competing firms that can price-discriminate consumers. The data broker can strategically choose any segment of the consumer demand (the information structure) to sell to firms that implement third-degree price-discrimination. We show that data broker's equilibrium profits are maximized when (1) information identifies consumers with the highest willingness to pay; (2) consumers with a low willingness to pay remain unidentified; and (3) the data broker sells two symmetrical information structures. The data broker therefore strategically sells partial information on consumers to soften competition between firms. Extending the baseline model, we prove that these results hold under first-degree price-discrimination.

*We would like to thank Yann Balgobin, Marc Bourreau, Romain De Nijs, Alexis Larousse, Qihong Liu, Johannes Paha, Martin Quinn, Regis Renault, Konstantinos Serfes, Marvin Sirbu as well as participants at the MaCCI Annual Conference, the 67th Annual Meeting of the French Economic Association, the 35th Days of Applied Microeconomics (JMA), the Workshop of Industrial Organization in the Digital Economy, the CREST Workshop on Platforms and E-commerce, and the Paris Seminar on Digital Economics for useful remarks and comments. Antoine Dubus acknowledges financial support from the Chair Values and Policies of Personal Information of Institut Mines Télécom, Paris. Patrick Waelbroeck thanks for insightful discussions the members of the Chair Values and Policies of Personal Information of Institut Mines Télécom, Paris.

[†]i3/Telecom ParisTech/Economics and Social Sciences and CESP, 46 rue Barrault, 75634 Paris Cedex 13, France; david.bounie@telecom-paristech.fr.

[‡]i3/Telecom ParisTech/Economics and Social Sciences, 46 rue Barrault, 75634 Paris Cedex 13, France; antoine.dubus@telecom-paristech.fr.

[§]i3/Telecom ParisTech/Economics and Social Sciences and CESifo, 46 rue Barrault, 75634 Paris Cedex 13, France; patrick.waelbroeck@telecom-paristech.fr.

1 Introduction

Digital technologies are transforming modern economies, with far-reaching effects on productivity, employment, innovation, and growth (McAfee and Brynjolfsonn, 2012). In 2016, the digital economy accounted for USD 1.2 trillion in the United States, or 6.5 percent of GDP, according to the Commerce Department's Bureau of Economic Analysis. The sector that includes network infrastructure, e-commerce and digital media grew at an average annual rate of 5.6 percent per year from 2006 to 2016, compared to the 1.5 percent growth in the overall economy. A similar trend is also at work in China and the European Union.[1]

This growth rate depends essentially on consumer information, as digital companies such as Facebook, Apple, Amazon and Google, base their business models on personal data collection and store traces left by Internet users who visit their online websites. What is less known, however, is that these large companies share or acquire information from data brokers that also gather information about millions of people.[2] The recent Facebook scandal involving Cambridge Analytica has precisely revealed to the public the troubled relations between Facebook and data brokers.[3]

Data brokers collect all sorts of information on consumers from publicly available online and offline sources (such as names, addresses, revenues, loan default information, and registers). They are major actors in the data economy, as more than 4000 data brokers operate in a market valued around USD 156 billion per year [4]. In a study of nine data brokers from 2014,[5] the Federal Trade Commission found that data brokers have information "on almost every U.S. household and commercial transaction. [One] data broker's database has information on 1.4 billion consumer transactions and over 700 billion aggregated data elements; another data broker's database covers one trillion dollars in consumer transactions; and yet another data broker adds three billion new records each month to its databases."[6] Data brokers therefore possess considerable amounts of information that they can sell to help firms learn more about their customers to better target ads, tailor services, or price-discriminate consumers.

Competition between firms is thus influenced by how much consumer data firms can acquire from the data brokers. On the one hand, more information allows firms to better target consumers and price-discriminate, which increases their profits. On the other hand, more information means that firms will fight more fiercely for consumers that they have identified as belonging to their business segments. This increased competition decreases firms profit. Overall, there exists an economic trade-off between surplus extraction and increased competition. This paper analyzes this trade-off when a data broker strategically sells information in order to maximize its profits.

Understanding how the quantity of information available on a market influences competition is a central question in economics, dating back to Hayek's seminal work [7]. Data brokers add a

[1] European Commission, Digital Agenda for Europe

[2] Business Insider, Facebook is quietly buying information from data brokers about its users' offline lives, Dec. 30, 2016.

[3] Washington Post, Facebook, longtime friend of data brokers, becomes their stiffest competition, 29 March 2018.

[4] Pasquale (2015)

[5] Acxiom, CoreLogic, Datalogix, eBureau, ID Analytics, Intelius, PeekYou, Rapleaf, and Recorded Future.

[6] Federal Trade Commission, 2014, Data brokers: A Call for Transparency and Accountability.

[7] Hayek (1945)

strategic dimension to the question of information acquisition on markets. Indeed, the previous literature[8] has assumed that information was exogenously available on the market.

We argue in this article that information can also be endogenously provided by a data broker who strategically sets the quantity of information available to market participants. Thus data brokers can weaken or strengthen the intensity of competition on the product market by strategically determining the quantity of data available on the market. It is not necessarily optimal for the data broker to sell information on all available consumer segments, since doing so would reduce their profits and hence their willingness to pay for consumer data.

We build a model where a data broker can sell information that partitions the Hotelling unit line into segments of arbitrary sizes to one or two competing firms. In other words, the data broker has the choice to sell information on all available consumer segments, only some segments of information, or no information at all. The possibility to purchase only a subset of all available demand segments corresponds to observed marketing strategies where firms purchase information on groups of consumers who share specific characteristics (market segmentation). By doing so, firms can identify the demand segments that are likely to be the most profitable. Firms that acquire segments of information can set specific prices on each segment of the unit line.

Using this setting, we show that it is optimal for the data broker to sell segments that are located closest to firms, but to keep consumers located in the middle of the Hotelling line unidentified. This partition allows firms to extract surplus from consumers with the highest willingness to pay while keeping consumers with a low willingness to pay unidentified in order to soften competition between firms. Total consumer surplus is increased when firms are informed, despite the fact that some consumers are worse off due to price-discrimination.

We contribute to the existing literature on two points. First, we show that a strategic data broker can strengthen or weaken the intensity of competition between firms by deciding to leave consumers with low willingness to pay unidentified. This result is new as previous research only examine the cases where data-brokers only sell information to firms on all consumers or on none of them[9]. Liu and Serfes (2004) for instance study firm's incentives to acquire a technology that can target customer segments more or less precisely. In their approach, information is a partition of a mass of consumers distributed in different segments on a Hotelling unit line, and firms have the choice to acquire either information on all consumer segments or no information at all. Also, Braulin and Valletti (2016) study vertically differentiated products, for which consumers have hidden valuations. The data broker can sell firms information on these valuations. Montes, Sand-Zantman and Valletti (2018) consider information allowing firms competing à la Hotelling to first-degree price-discriminate consumers. The data broker sells either information on all consumers, or no information at all. Information is sold through a second-price auction mechanism with negative externalities (as in Jehiel and Moldovanu (2000)). They find that firms acquire all the segments of information.

The remainder of the article is organized as follows. In Section 2 we describe the model, and

[8] Radner et al. (1961), Vives (1984), Burke, Taylor and Wagman (2012)

[9] Following Radner et al. (1961), Vives (1984), Burke, Taylor and Wagman (2012)

in Section 3 we characterize the optimal structure of information. In Section 4, we provide the equilibrium of the game, and we conclude in Section 5.

2 Model set-up

We consider a game involving a data broker, two firms (noted $\theta = 1, 2$), and a mass of consumers uniformly distributed on a unit line $[0, 1]$. The data broker collects information about consumers who buy products from the competing firms at a cost that we normalize to zero. Firms can purchase information from the data broker to price-discriminate consumers.[10] In Section 4, we first analyze third-degree price-discrimination, then we extend the analysis to first-degree price-discrimination.

The two firms are located at 0 and 1 on the unit line and sell competing products to consumers. A consumer located at x derives a gross utility V from consuming the product, and faces a linear transportation cost with value $t > 0$. A consumer buys at most one unit of the product, and we assume that the market is fully covered, that is, all consumers buy the product.[11] Let p_1 and p_2 denote the prices set by Firm 1 and Firm 2, respectively. A consumer located at x receives the following utility:

$$\begin{cases} U(x) = V - tx - p_1, \text{ if he buys from Firm 1,} \\ \\ U(x) = V - t(1-x) - p_2, \text{ if he buys from Firm 2,} \\ \\ U(x) = 0, \text{ if he does not consume.} \end{cases} \tag{1}$$

In the following sections, we define the information structure, the profits of the data broker and of the firms, and the timing of the game.

Information structure

Firms know that consumers are uniformly distributed on the unit line, but absent further information, they are unable to identify consumers' locations. Therefore, firms do not know the degree to which consumers value their products and cannot price-discriminate them.[12] Firms can acquire an information structure from a monopolist data broker at cost w. The information structure consists of a partition of the unit line into n segments of arbitrary size. These segments are constructed by unions of elementary segments of size $\frac{1}{k}$, where k is an exogenous integer that can be interpreted as the quality of information. Although, the data broker can sell any such partition, it is useful to define a reference partition $\mathcal{P}_{ref}$, which includes k segments of size $\frac{1}{k}$. Liu and Serfes (2004) assume that the data broker can only sell (or not) the reference partition $\mathcal{P}_{ref}$ to competing firms.

[10] The marginal production costs are also normalized to zero.

[11] If the market is not covered, the competition effect that we identify is weakened, and new issues related to customer churn and customer acquisition arise.

[12] This assumption is also made by Braulin and Valletti (2016) and Montes, Sand-Zantman and Valletti (2018).

A major contribution of the present article is to demonstrate that the optimal partition sold by the data broker is not the reference partition $\mathcal{P}_{ref}$.

We introduce further notations. We denote $\mathcal{S}$ the set comprising the $k-1$ endpoints of the segments of size $\frac{1}{k}$: $\mathcal{S} = \{\frac{1}{k}, .., \frac{i}{k}, .., \frac{k-1}{k}\}$. Consider the mapping, i.e., a bijection, that associates to any subset $\{\frac{s_1}{k}, .., \frac{s_i}{k}, .., \frac{s_{n-1}}{k}\} \in \mathcal{S}$ a partition $\{[0, \frac{s_1}{k}], [\frac{s_1}{k}, \frac{s_2}{k}], .., [\frac{s_{n-1}}{k}, 1]\}$, where $s_1 < .. < s_i < .. < s_{n-1}$ are integers lower than k. We write $\mathbb{P}$ as the target set of the mapping: $M : \mathcal{S} \to \mathbb{P}$; this set comprises all possible partitions of the unit line generated by segments of size $\frac{1}{k}$. Thus, $\mathbb{P}$ is the sigma-field generated by the elementary segments of size $\frac{1}{k}$. In particular, $\mathcal{P}_{ref}$ and $[0, 1]$ are included in $\mathbb{P}$.

The data broker can sell any partition $\mathcal{P}$ of the set of partitions $\mathbb{P}$, for instance, a partition starting with one segment of size $\frac{1}{k}$, and another segment of size $\frac{2}{k}$.

A firm having information of the form $\{[0, \frac{s_1}{k}], [\frac{s_1}{k}, \frac{s_2}{k}], .., [\frac{s_{n-1}}{k}, 1]\}$ will be able to identify whether consumers belong to one of the segments of the set and charge them a corresponding price. Namely, the firm will charge consumers on $[0, \frac{s_1}{k}]$ price p_1, consumers on $[\frac{s_i}{k}, \frac{s_{i+1}}{k}]$ price p_{i+1}, and so forth for each segment.

Contrary to the existing literature, we allow the data broker to sell a partition different from $\mathcal{P}_{ref}$. In fact, it can sell any information structure belonging to $\mathbb{P}$. However, we rule out information structures that generate uncertainty over the location of the elementary segment of size $\frac{1}{k}$ to which a consumer belongs. As an illustration, suppose that $k = 8$ so that the finest partition consists of 8 segments of size $\frac{1}{8}$. Suppose also that the data broker sells a partition constructed from 3 segments in the following way. The first element of the partition includes segments 1 and 3 which have a size of $\frac{1}{8}$ and that are located at the extremities of the unit line. The second element of the partition is segment 2 of size $\frac{6}{8}$, located in the middle of the line. The information structure is therefore the partition $\{\{1, 3\}, 2\}$. Segments 1 and 3 are not connected and are therefore excluded from our analysis.

Strategies and timing

The data broker can sell any partition $\mathcal{P}_\theta$ to Firm θ. In fact, starting from any pairs of partitions, we will show that when the data broker decides to sell information to both firms, it will sell the same partition. We write the generic form of the profits for a partition as $\mathcal{P}$,[13] using the notation NI (resp. I) when a firm is not informed (resp. informed). Additionally, we denote whether a firm and its competitor are informed or not by the couple (A, B) where $A, B \in \{I, NI\}$. For instance, (I, NI) refers to a situation in which Firm θ is informed and Firm $-\theta$ is uninformed. For any information structure, we need to compute the profits for three possible configurations as $\pi^{NI,I}_{\mathcal{P},\theta} = \pi^{I,NI}_{\mathcal{P},\theta}$: $\{\pi^{NI,NI}_{\mathcal{P},\theta}, \pi^{I,NI}_{\mathcal{P},\theta}, \pi^{I,I}_{\mathcal{P},\theta}\}$.

Firms simultaneously set their prices on the unit line when they have no information or on each segment of the partition when they are informed. Each firm knows whether its competitor is informed, and the structure of the partition $\mathcal{P}_{-\theta}$.[14] Firms acquire information at a price that

[13] We drop the subscript θ when there is no confusion.

[14] This assumption is also standard in Braulin and Valletti (2016) and Montes, Sand-Zantman and Valletti (2018).

depends on the extent to which information increases their profits. This value of information varies according to whether the competitor purchases information. We consider the profits in equilibrium for any partition $\mathcal{P}_\theta$ of the unit line.

The data broker extracts all surplus from competing firms and maximizes the difference between the profits of an informed firm and those of an uninformed firm. The data broker profit function can be written as

$$\Pi = \begin{cases} \Pi_1 = w_1 = \max_{\mathcal{P}\in\mathbb{P}}\{\pi_{\mathcal{P},\theta}^{I,NI} - \pi_\theta^{NI,NI}\}, \\ \text{if the data broker sells information to only one firm,} \\ \\ \Pi_2 = 2w_2 = 2\max_{\mathcal{P}\in\mathbb{P}}\{\pi_{\mathcal{P},\theta}^{I,I} - \pi_{\mathcal{P},\theta}^{NI,I}\}, \\ \text{if the data broker sells information to both competitors.} \end{cases} \tag{2}$$

The partition proposed by the data broker depends on whether information is sold to one firm or to both firms. We define Π_1 as the maximum of the first part of Eq. (2), and Π_2 as the maximum of the second part of Eq. (2).

For any partition $\mathcal{P}$ composed of n segments, Firm θ maximizes its profits with respect to the prices on each segment, denoted by the vector $\mathbf{p}_\theta = (p_{\theta 1}, .., p_{\theta n}) \in \mathbb{R}_+^n$. The profit function of the firms can be written as follows:

$$\pi_{\mathcal{P},\theta} = \sum_{i=1}^{n} d_{\theta i}(\mathbf{p}_\theta, \mathbf{p}_{-\theta}) p_{\theta i}. \tag{3}$$

The timing of the game is the following:

- Stage 1: the data broker chooses the optimal partition, and whether to sell information to one firm or to two firms.
- Stage 2: firms compete, and they price-discriminate consumers if they acquire information.

3 Optimal information structure

Equilibrium prices charged to consumers and the profits of the firms in stage 2 depend, first, on the optimal partition sold by the data broker in stage 1, and second, on the data broker's strategy to serve either one or two firms in the market. As a consequence, the data broker has to calculate the prices of any possible information structure that can be sold to firms.

In this section, we prove in Theorems 1 and 2 that we can restrict the analysis to particular information structures that are optimal for the data broker. We first analyze the case where the data broker chooses to sell information to only one firm, i.e., the case of exclusive selling. Second, we characterize the optimal information structure when the data broker sells information to both firms. We find that the data broker sells a partition $\mathcal{P}(\mathbf{p}_1, \mathbf{p}_2)$ that identifies consumers close to the firm up to a cutoff point $\frac{j}{k}$, and that leaves consumers unidentified in the remaining segment. In Section 4, we calculate the number j^* of segments where consumers are identified in the optimal information structure; j^* will depend on the data broker's strategy (whether it sells information to one or to two firms). We finally discuss at the end of this section how information acquisition affects competition between firms.

Information is sold to only one firm

When information is sold exclusively to Firm 1 (without loss of generality), the profit-maximizing information structure for the data broker has the following features. Theorem 1 shows that the data broker sells information on all segments up to a point $\frac{j}{k}$, and leaves a large segment of unidentified consumers after that point. In the rest of the article, we refer to the consumers located on the j segments of size $\frac{1}{k}$ as the *identified consumers*; the remaining consumers located beyond the j segments of size $\frac{1}{k}$ are referred to as the *unidentified consumers.*

Firm 2 has no information and sets a unique price p_2 over the unit line. Firm 1 can identify consumers on each segment on the left (indexed by $i = 1, .., j$), of size $\frac{1}{k}$. Firm 1 can price-discriminate consumers and sets different prices on each segment, with p_{1i} being the price on the *ith* segment from the origin. Firm 1 sets price p_1 on the last segment.

THEOREM 1: *Let* $\boldsymbol{p}_1 \in \mathbb{R}_+^{j+1}$ *and* $p_2 \in \mathbb{R}_+$. *The profit-maximizing information structure* $\mathcal{P}^*(\boldsymbol{p}_1, p_2)$ *divides the unit line into two segments:*

- *The first segment (closest to the firm buying information) is partitioned into* j *segments of size* $\frac{1}{k}$.
- *Consumers in the second segment of size* $1 - \frac{j}{k}$ *are unidentified.*

Proof available upon request.

The proof proceeds in the following way. Consider any information structure. First, we show that the data broker finds it profitable to re-order segments and reduce their size to $\frac{1}{k}$ so that the firm has more information on consumers closest to its product. Second, the data broker can soften competition between firms by leaving a segment of unidentified consumers in the middle.

Theorem 1 makes an important contribution to the existing literature that assumes that the data broker either always sells all information segments to firms, or sells no information at all (Braulin and Valletti, 2016; Montes, Sand-Zantman and Valletti, 2018). We show that this assumption is questionable as selling all segments, i.e., the reference partition of the unit line, is not optimal.

The data broker sells information to both firms

When information is sold to both firms, the profit maximizing information structure for the data broker has the same features as the optimal partition described in Theorem 1. Theorem 2 first demonstrates that the data broker sells to each firm information on all segments up to a point, $\frac{j}{k}$ to Firm 1 and $\frac{j'}{k}$ to Firm 2. Then, it is established that in equilibrium, the data broker sells the same information structure to both firms, that is, $\frac{j}{k} = \frac{j'}{k}$. The remaining consumers are unidentified.

When information is sold to both firms, we rule out situations where firms compete and share demand segments at the extremities of the unit line. We assume that the data broker does not sell segments that would allow firms to poach consumers. We analyze the condition under which both firms have positive demands on a given segment $[\frac{s_i}{k}, \frac{s_{i+1}}{k}]$:

$$C_1: \quad \frac{s_i}{k} \leq \frac{p_2 + t}{2t} \qquad \text{and} \qquad \frac{p_2 + t}{2t} \leq \frac{2s_{i+1} - s_i}{k}$$

The first part of condition C_1 guarantees that there is positive demand for Firm 1, whereas the second part guarantees positive demand for Firm 2. Inequalities in condition C_1 are expressed as a function of p_2 without loss of generality.

Except for the segment in the middle of the line, we exclude segments located before $\frac{1}{2}$, where Firm 2 has positive demand (and similarly for Firm 1). Thus, we assume that $\frac{p_2+t}{2t} \geq \frac{2s_{i+1}-s_i}{k}$, which is achieved by setting $p_2 = 0$ in the previous inequality (the lowest possible value for p_2): $\frac{1}{2} \geq \frac{2s_{i+1}-s_i}{k}$.

ASSUMPTION 1: **(No consumer poaching condition)**

When the data broker sells a partition $\mathcal{P} = \{[0, \frac{s_1}{k}], .., [\frac{s_i}{k}, \frac{s_{i+1}}{k}], .., [\frac{s_{n-1}}{k}, 1]\}$ *to Firm 1 and* $\mathcal{P}' = \{[0, \frac{s'_{n'-1}}{k}], .., [\frac{s'_{i'+1}}{k}, \frac{s'_{i'}}{k}], .., [\frac{s'_1}{k}, 1]\}$ *to Firm 2, the segments verify:* $2\frac{s_{i+1}}{k} - \frac{s_i}{k} \leq \frac{1}{2}$ *and* $2\frac{s'_{i'+1}}{k} - \frac{s'_{i'}}{k} \leq \frac{1}{2}$ *for* $i = 0, .., n-2$, $i' = 0, .., n'-2$.[15]

Under Assumption 1, the optimal partition is similar to that found in the case of exclusive selling, i.e. when one firm acquires information. The optimal information structure has the following features.

THEOREM 2: *Under Assumption 1, the data broker sells to Firm 1 (resp. Firm 2) a partition with two different types of segments:*

a) *There are j (resp. j') segments of size $\frac{1}{k}$ on $[0, \frac{j}{k}]$ (on $[1-\frac{j'}{k}, 1]$ for Firm 2) where consumers are identified.*

b) *Consumers in the second segment of size $1 - \frac{j}{k}$ (resp. $1 - \frac{j'}{k}$) are unidentified.*

c) $j = j'$.

Proof available upon request.

The proof proceeds in a similar way as the proof of Theorem 1. We consider any partition satisfying Assumption 1. We show that the data broker always finds it more profitable to sell

[15] We note by convention that $s'_0 = s_0 = 0$.

segments of size $\frac{1}{k}$. Using the profit function in equilibrium, we then show that selling the same information structure to both firms is optimal, that is $\frac{j}{k} = \frac{j'}{k}$.

Thus, the data broker sells the same information structure to both competitors. This result differs from Belleflamme, Lam and Vergote (2017), where two firms compete in a market for a homogeneous product. Firms can acquire information on their customers to price-discriminate them. The authors show that firms do not acquire information with the same precision, and a data broker selling information will thus strategically lower the precision of information for one firm.

Competitive effects of information acquisition

We now interpret how information acquisition affects competition between firms. To do so, we analyze the impact of the acquisition of an additional segment to the optimal partition on the firms' respective profits and prices. Specifically, we compare the changes in prices and profits when Firm 1 acquires an optimal partition $\mathcal{P}$ with the last segment located at $\frac{j}{k}$, and when Firm 1 acquires $\mathcal{P}'$ with the last segment located at $\frac{j+1}{k}$. In the following discussion, Firm 2 remains uninformed.

Purchasing an additional segment will have several impacts on the profits of both firms:

a) Firm 1 price-discriminates consumers on $[\frac{j}{k}, \frac{j+1}{k}]$, which increases its profits.

b) Firm 1 lowers its price on $[\frac{j+1}{k}, 1]$, which increases the competitive pressure on Firm 2. In reaction to this increased competition, Firm 2 lowers its price on the whole unit line ($p_2' < p_2$). The competitive pressure on Firm 1 is increased throughout the unit line as the price charged by Firm 2 decreases, which has a negative impact on Firm 1's profits.

The optimal size of the segments where consumers are identified therefore depends on the two opposite effects of information acquisition on firm profits. Following Theorems 1 and 2, it is clear that selling all segments to competing firms is not optimal.

In the following section, we detail the resolution of the game by taking into account the optimal information structure established in Theorems 1 and 2. An informed firm can distinguish $j + 1$ segments.

4 Model resolution

In this section, we compute the equilibrium prices and profits of Firm 1 and 2 using the optimal partition described in Theorems 1 and 2. Then, we analyze whether the data broker sells information to one firm or to both competitors.

Characterization of the equilibrium

We characterize in this section the number of segments of information sold to firms when only one firm is informed and when both firms are informed. We then compare the profits of the data broker in the two cases, and we show that the data broker always sells information to both competitors in equilibrium.

The optimal number of segments

We first find the optimal values of j when one or both firms are informed, then we compare the profits in the two situations.[16]

LEMMA 1:

- *When one firm buys information, the data broker sets*

$$j_1^* = \frac{6k - 9}{14}.$$

- *When both firms buy information, the data broker sets*

$$j_2^* = \frac{6k - 9}{22}.$$

Proof available upon request.

Optimal choice of the data broker

From Lemma 1, we can finally calculate the optimal choice of the data broker by comparing its profits when it sells information to one firm or to both firms:

PROPOSITION 1: *Suppose Assumption 1, the data broker optimally sells information to both firms:*

$$\Pi_2^* \geq \Pi_1^*.$$

Proof available upon request.

5 Conclusion

Data brokers are major players in the Internet economy. They collect and process a vast amount of consumer data that they can choose to sell to firms for various objectives, including price-discrimination. Data brokers have a significant impact on market equilibria and social welfare. Indeed, a data broker can soften or increase the intensity of competition on the product market by choosing the amount and the quality of information that he sells to firms. We have argued that firms are in a prisoners dilemma situation since the data broker can use its strategic position

[16] For the proof of Lemma 1, we assume that j is defined over $\mathbb{R}$, and the resulting j chosen by the data broker is the integer part of j^*.

to extract their surplus. Moreover, consumers with a high willingness to pay suffer from increased price discrimination. These different elements point to the need for an increased scrutiny of the data broker's industry by privacy regulators and competition authorities.[17]

Understanding how new data strategies can impact competition on markets is a new promising field of research. We contribute to this literature by developing a model in which the data broker can choose among a large set of possible information structures to sell to firms. The optimal information structure segments consumers into two groups: on the one hand, consumers with the highest willingness to pay are identified, and, on the other hand, low-valuation consumers remain unidentified. This strategy allows the data broker to soften competition between firms, while still price discriminating consumers with a high willingness to pay.

Determining which consumer segments are sold is therefore important, and should be included in models that make the assumption that data brokers can sell all consumer segments or no information at all. Our model can be used to address two recent privacy issues. First, our results suggest that new privacy policies in the European Union (such as the general regulation on data protection) could increase consumer surplus. Stronger privacy protection in Europe means that firms now are able to distinguish only coarser consumers segments, which lowers the precision of information structures modeled by k. When k decreases, the share of unidentified consumers increases. Overall, consumer surplus increases with privacy protection regulation. Second, the share of identified consumers is higher when both firms are informed than when only one firm is informed. Thus, selling to more firms on the market leads to more price discrimination, but at the same time, increases competition. An important question arises: how does competition in the data brokerage industry affect data collection and the amount and quality of information sold on the product market. We address this question in a companion paper.[18]

References

Belleflamme, Paul, Wing Man Wynne Lam, and Wouter Vergote. 2017. "Price Discrimination and Dispersion under Asymmetric Profiling of Consumers."

Bounie, David, Antoine Dubus, and Patrick Waelbroeck. 2018. "Data brokers and data externalities."

Braulin, Francesco Clavorà, and Tommaso Valletti. 2016. "Selling customer information to competing firms." *Economics Letters*, 149: 10–14.

Burke, Jeremy M., Curtis R. Taylor, and Liad Wagman. 2012. "Information Acquisition in Competitive Markets: An Application to the US Mortgage Market." *American Economic Journal: Microeconomics*, 4(4): 65–106.

Crain, Matthew. 2018. "The limits of transparency: Data brokers and commodification." *New Media & Society*, 20(1): 88–104.

Hayek, Friedrich August. 1945. "The use of knowledge in society." *The American economic review*, 35(4): 519–530.

Jehiel, Philippe, and Benny Moldovanu. 2000. "Auctions with downstream interaction among buyers." *Rand journal of economics*, 768–791.

Liu, Qihong, and Konstantinos Serfes. 2004. "Quality of information and oligopolistic price discrimination." *Journal of Economics & Management Strategy*, 13(4): 671–702.

McAfee, Andrew, and Erik Brynjolfsonn. 2012. "Race Against the Machine." *MIT Sloan Management School of Management.*

[17] Regulators and legislators have recently analyzed the impacts of data brokers on markets (Crain, 2018).
[18] Bounie, Dubus and Waelbroeck (2018)

Montes, Rodrigo, Wilfried Sand-Zantman, and Tommaso Valletti. 2018. "The value of personal information in online markets with endogenous privacy." *Management Science.*

Pasquale, Frank. 2015. *The black box society: The secret algorithms that control money and information.*

Radner, Roy, et al. 1961. *The evaluation of information in organizations.* Management Science Research Group, University of California.

Vives, Xavier. 1984. "Duopoly information equilibrium: Cournot and Bertrand." *Journal of economic theory*, 34(1): 71–94.