Strategic information and competition in digital markets

Antoine Dubus

université PARIS-SACLAY

NNT : 2018SACLT037

Thèse de doctorat

TELECOM Paris

IP PARIS

Strategic Information and Competition in Digital Markets

Thèse de doctorat de l'Université Paris-Saclay
préparée à Télécom Paris

École doctorale n°578 Sciences de l'Homme et de la Société (SHS)
Spécialité de doctorat : Sciences Sociales et Humanités

Thèse présentée et soutenue à Paris, le 30 septembre 2019, par

ANTOINE DUBUS

Composition du Jury :

Andreea Cosnita-Langlais Maître de conférence habilitée à diriger des recherches, Université Paris Nanterre	Présidente
Doh-Shin Jeon Professeur, Toulouse School of Economics	Rapporteur
Patrick Legros Professeur, Université libre de Bruxelles et Northeastern University	Examinateur
David Bounie Professeur, Télécom Paris	Directeur de thèse
Patrick Waelbroeck Professeur, Télécom Paris	Co-directeur de thèse

Contents

Introduction

Information plays a key role in modern economies. Since the seminal work of Hayek (1945), its importance in policy decision making has been widely acknowledged by scholars and politics. Information is also at the heart of our understanding of markets. Major contributions by Akerlof (1978), Rothschild and Stiglitz (1978), and Samuelson (1985) have emphasized how information asymmetry between market participants can cause important harms to uninformed actors. For instance, consider a consumer who wants to buy a mobile phone. The mobile phone can be either of a good type or bad type. Consumers cannot distinguish the types of a mobile phone before buying it. However, phone sellers know the quality of the phone that they sell, this is an example of information asymmetry. There is a risk for consumers who might pay a high price for a bad type phone, and thus consumers will be ready to buy mobile phones only at a low price. Firms selling mobile phones will not have interest to propose good type mobile phone, as they will not be sold at a high enough price, and only low quality mobile phone will remain on the market. This is known as adverse selection, which can lead to the disappearance of markets.

Digital markets have given information an even more essential role. Digital technologies have enabled firms to collect huge amounts of information on consumers. Information on consumers' private characteristics and preferences is then used by data driven companies to propose consumers targeted advertisement and prices, but also personalized services and goods. The most emblematic of these companies is probably Google, which collects every information possible on consumers and sells to firms the possibility to show up on the first page of their web browser when potential buyers enter a related query. Google uses information to match as precisely as possible consumers who are the most eager to buy a product with the company that sells it. Firms are willing to pay for this service as it allows them to reach more consumers, thus to be more competitive and to make more profits.

Recently, data brokers have specialized in extracting and treating as much information on as many consumers possible. Consumer information allows them to reduce information asymmetry on consumers' characteristics, that is, to know as much as possible about them. Data brokers sell consumer data to firms willing to personalize their advertising campaigns or their products depending on the preferences of consumers. There is thus a market for consumer information: information is supplied by data brokers, and purchased by firms willing to acquire consumer data. Consumer data are now economic goods, whose monetary values depend on their quality, and on the supply and demand of the market.

Little is known about the functioning of the market for consumer information. Data brokers operate outside the scope of regulators by keeping a well-hidden secret on the nature of their activities. What are their strategies regarding how much information they collect and sell to competing firms? How do these strategies change when competition between data brokers is more intense? How do they react to a change of regulation on data protection? Finally, how does the data brokerage activity affect markets where firms acquire information? Meanwhile, the fact that companies can buy and sell consumer information raises several questions on the risks involved for consumers. Will consumer welfare be lowered due to the collection and sale of their information? Can consumers protect their privacy from data brokers? Can a regulator influence how much consumer information a data broker collects and sells?

This thesis contributes to the theoretical literature on markets for consumer information in three points. First, it analyzes the strategies of data brokers regarding how much consumer information they collect, and how they sell it to competing firms. Secondly, it examines the effects of competition between data brokers on consumers and on firms that buy information. Thirdly, it proposes recommendations to data protection agencies and competition authorities to regulate competition between data brokers. By providing insights on how data brokers operate and how they affect consumers and firms, this thesis can help policy makers to better answer the issues raised by the emergence of a market for consumer information.

The data brokerage industry

We describe in this section the data brokerage industry. The description builds on the FTC report on data brokers[1], as well as on an interview of a specialist of the data brokerage industry that was conducted in April 2019.[2]

Data brokers are companies specialized in collecting, treating, and selling consumer information to firms. Data brokers appeared in the mid-sixties, and have grown fast since then, to such extent that a major data broker such as LiveRamp Holdings, Inc. (formerly known as Acxiom) is now valued around 3.59 billion $.

Data brokers collect information from various sources. They combine offline data - such as public records of marriages, divorces, voting - with online data that they mostly collect using cookies. For instance, data brokers have had an open access for years to the personal information of Facebooks users.[3] As Crain (2018) emphasizes, data brokers also obtain a substantial part of their information from other data brokers. They merge data bases in order to obtain thinner segments of the population, to such extent that, for instance, the FTC states in its report that "one of the nine data brokers [that they consider] has 3000 data segments for nearly every U.S. consumer."

Data brokers sell information to companies for various purposes. From a marketing perspective, information can be used for instance to send solicitation e-mails, or advertise for specific products to specific consumer segments.

> "Traditionally, what comes to mind is the basics of marketing information, where [data brokers] will make available to companies who want to carry out marketing operations files to put them in contact and hope to market themselves, to develop. [Firms acquire

[1] Data brokers: a Call for transparency and accountability; FTC, May 27 2014.
[2] For privacy issues the name of the respondent is not provided in this thesis. The full interview is available upon request.
[3] Facebook, longtime friend of data brokers, becomes their stiffest competition; Washington Post, March 29, 2018.

> information] in a logic of marketing and improvement of their profits.[...]
>
> Then data brokers also operate around the notion of risk, sometimes called credit bureau, with three major actors in the U.S., that are Equifax, Experian and Transunion. Credit scoring can be either marketed as a proportion of the customer to spend, or as the level of risk that the customer presents with respect to its creditworthiness."

Information can thus be used to verify the identity of consumers, or to detect fraud. Banks use information for credit scoring purposes, that is, to adapt credit allocation and credit rates to the ability of clients to reimburse a loan. Data brokers thus interact with banks, media, and in the end with any company willing to know more about consumers.

In many cases, consumers benefit from firms being more informed. They are proposed advertising and discounts adapted to their preferences; more information on the markets usually means more competition, which lowers the prices of products; less fraud and a better allocation of credit improve the functioning of the market which shall also benefit consumers.

Nevertheless, in other ways, consumers can also be harmed by the activity of data brokers. For instance, an error in credit allocation prevents a liable consumer to access a legitimate credit. Discrimination may also arise as consumers can be targeted depending on their ethnicity, gender, or age. These issues are exacerbated by security risks, that are inherent to the massive storage of data. The recent breach of Equifax, a major data broker specialized in credit scoring is eloquent:[4] more than a 143 Million U.S. citizens – almost half of the population – have had their personal information stolen during this breach. However, the data brokerage industry is characterized by an unusual and worrying opacity, and information collection is largely done without the knowledge and informed consent of consumers.

Several questions arise regarding a regulation of the data brokerage industry. Policy makers are concerned by data brokers as they affect consumer welfare and privacy. Also, data brokers have recently faced a wave of mergers. In 2014,

[4] The Equifax Data Breach; FTC, September 18, 2019.

for instance, Oracle, a major technology company, has acquired two large data brokers: Bluekai and Datalogix.[5] This increase of concentration in the industry raises concerns about the market power of these companies. The impacts that a higher concentration of data brokers will have on how much consumer information they collect and sell have not been studied in the economic literature. The purpose of this thesis is thus to unveil some of the mechanisms of information acquisition and selling of data brokers.

Outline

The thesis is composed of four chapters building on a homogeneous model in which data brokers sell consumer information to competing firms for price discrimination purposes. The first three chapters are co-authored with David Bounie and Patrick Waelbroeck. The last chapter is single authored.

Selling strategic information: a critical review of the literature

In the first chapter, we review the literature on third parties selling informational goods. We identify two strategic dimensions used by a third party when selling to firms information that has a competitive effect on the market. First, the literature shows that a third party that sells information can also lower the precision of information in order to soften competition between firms. Such practice increases the value of information. Secondly, they can decide to sell information to only a subset of firms, and choose which firms will dominate the market. Thus these third parties have a strong impact on competition on the market. In the remaining of the thesis, we focus on data brokers selling consumers information to competing firms.

Selling strategic information in digital competitive markets

The second chapter analyzes a monopolist data broker selling consumer information to competing firms. We relax a classical assumption of the theoretical

[5] AdExchanger, How Datalogix Made Oracle's BlueKai Acquisition Even Smarter; Adexchanger, January 19, 2015.

literature, that assumes that the data broker either sells all information, or does not sell information at all. We show that this assumption is highly restrictive: the data broker will sell information to firms on consumers with the highest willingness to pay for their product. Consumers with the lowest willingness to pay will remain unidentified, even though the data broker has information on them. This information allows firms to extract surplus from consumers who have a high valuation of their product. Keeping a share of consumers unidentified softens competition between firms. We argue that this new theoretical result is fundamental to understand the impacts of data brokers on consumers and firms that acquire information. By choosing which consumer segment to sell to competing firms, data brokers can soften competition on the market, which increases the value of information. The following chapters build on this result and analyze its implications for competing firms and consumer welfare.

Selling mechanisms and the market for information

In the third chapter, we focus on how different mechanisms of sales impact the strategy of a data broker regarding how much consumers information it collects and sells. We compare equilibrium in the following cases: the data broker sells information through a take it or leave it offer, a sequential negotiation, and finally an auction. We show that the more the data intermediary collects information, the lower consumer surplus. Take it or leave it offers, maximize consumer surplus and minimize information collection, but is the least profitable mechanism for the intermediary. Thus, selling mechanisms can be used as a regulatory tool by data protection agencies and competition authorities to minimize information collection and selling, and to maximize consumer surplus.

Collecting and selling consumer information: the two faces of data brokers

The fourth chapter analyzes a competitive data brokerage industry. Namely, we study the effects of competition between data brokers on how much consumer information they collect and sell to competing firms. We argue that, contrary to the current regulatory practice, both data protection and competition regulation should be taken into account together. Indeed, we show that competition between data brokers may increase the amount of consumer information that they collect and sell. This will positively impact consumers as competition on the product market is increased. Where data protection authorities are concerned, competition between data brokers has a negative impact on consumers as more consumer data are collected and sold to firms. However, competition authorities prefer data brokers to compete as consumers end up paying a lower price.

CHAPTER 1

Information Goods and Strategic Gatekeepers

1.1 Introduction

Information plays a key role in modern economies. Since the seminal work of Hayek (1945), scholars and policy makers have recognized that information greatly influences how markets function. Major contributions by Akerlof (1978), Rothschild and Stiglitz (1978), and Samuelson (1985) have emphasized how information asymmetry between market participants can cause important harms to uninformed actors, leading to market inefficiencies, and ultimately to market failures. By showing how most of classical economic equilibrium fall when markets display imperfect information, they have put the role of information at the core of economic research.

Information also plays a key role in the functioning of firms. From a business oriented perspective, firms face many informational problems (Holmstrom and Tirole, 1989). They have to manage internal information flows for decision making purposes. The manager-employee relationship is compromised by adverse selection and moral hazard issues (Spence, 1978). Moreover, employers have to allocate employees with tasks fitting their personal ability (Canidio and Legros, 2017). Firms also have to distinguish themselves from competitors when consumers have imperfect information on the reliability and the quality of their products, which can lead to reputation issues (Shapiro, 1982). They face uncertain market characteristics such as demand or consumer willingness to pay (Vives, 1984). Overall, investing in information systems is a market share winning strategy (Weill, 1992).

Big data has given information an even more central role. McAfee et al. (2012) show that data driven companies are 5% more productive and 6% more profitable than their competitors. The big data revolution affects all sectors of the economy and captures the attention of regulators and politics who do not want to be left behind (Kaisler et al., 2013). Economic analysis is also deeply impacted by this increasing ubiquity of information (Einav and Levin, 2014).

A recent literature has focused on the effect of big data infomediaries on competition and surplus.[1] However, we are still lacking a comprehensive review of the literature on the strategic use of information and innovation on markets. Yet, the economic literature has recognized the role of strategic information selling on markets and the spectacular growth of the digital economy has attracted more and more researchers to this issue. Information can be strategically sold by data brokers or exogenously by third party agents. In particular, Bergemann and Bonatti (2019) review the literature on markets for information, with a special emphasize on how the mechanisms through which information is sold affects market equilibria. In our paper complements their research as we focus on intermediaries who strategically decide to which firm they sell information, and which information they sell.

Information economics challenges traditional views of competition, welfare, barriers to entry, and regulation (Crémer et al., 2019). Information is long known to have important effects on competition in product markets. Early articles such as Thisse and Vives (1988), Ulph and Vulkan (2000) or Encaoua and Hollander (2007), highlight two effects of consumer information on competition. On the one hand, firms who can better target their consumers can better extract their surplus - this will be referred to as the rent-extraction effect, increasing market power. On the other hand, competing firms who both have information

[1] See for instance Belleflamme et al. (2017), Montes et al. (2018) or Bounie et al. (2018).

on consumers will compete more fiercely for them - this will be referred to as the competition-effect.

Protection of personal data also impacts the innovation process (Goldfarb and Tucker (2012), Lefouili et al. (2017)). Understanding how data protection can or not hinder innovation is of crucial importance in the data economy, where economic performance and innovation rely on data acquisition. The topic is of course of great interest to regulators and policy makers who will have to balance data protection, competition and innovation policies.

We provide a comprehensive review of the economic literature that analyzes how data intermediaries shape competition and innovation processes. These actors have been referred to as infomediaries, information gatekeepers, or more recently, data brokers.

We consider third party selling information to firms. Studies such as Ulph and Vulkan (2000), Stole (2007), Thisse and Vives (1988), Liu and Serfes (2004) or Taylor and Wagman (2014) where information is exogenously given on a market are thus excluded from this survey. We also rule out studies where firms collect information themselves, for instance in models of behavior based price discrimination (Fudenberg and Villas-Boas, 2006; Esteves, 2010). Finally we do not consider models where firms share information such as Vives (1984), Liu and Serfes (2006), Shy and Stenbacka (2013), Jentzsch et al. (2013) and Liu and Serfes (2013).

The structure of the article is the following. We analyze how authors address the issue of which information quality a data seller will propose to which firms. The literature shows that information sellers can strategically lower the quality of information that they sell to firms.[2] We identify three arguments in favor of such practices.

First, information increases competition on the product market, lowers the profits of the firms, and thus lowers their willingness to pay for information.

[2] This effect recalls the one in Deneckere and Preston McAfee (1996), where firms voluntarily lower the quality of the good they sell, which in particular, can be Paretto improving.

Information sellers can decide to sell voluntarily distorted information to only part of the firms in order to moderate competition on the market, and increase their profits. Such effect is also observed in the case of an innovator selling licenses to competing firms.

Secondly, we identify a parallel of the Arrow effect in the case of a firm selling information: the price of information already reveals part of its content. Firms thus face a dilemma when selling information, since setting the highest price naturally decreases the value of information. Lowering the quality of information is a way for an information seller to hide part of its content, and increase the value of information.

Thirdly, lowering information quality can be used by an uninformed principal to break information asymmetries.

Finally, we discuss how decision makers can regulate these strategies of information selling on markets.

We will use indifferently the terms information distortion, voluntary lowering or degradation of information quality, precision or accuracy, or to add noise to a signal.

The remaining of the paper is organized as follow. Section 1.2 analyzes by third party who develop selling strategies to modulate competition on a downstream market. In Section 1.3 we focus on the arrow effect, and how information and innovation sellers can correct it. Finally we review in Section 1.4 the strategies of sellers to correct for information asymmetry. Section 1.5 concludes.

1.2 Shaping of downstream competition

The role of strategic information intermediaries is well identified in the literature. These actors can shape the intensity of competition in downstream markets by limiting the amount of information sold to competing firms. We review in this section the literature analyzing the relation between information and competition on the sides of the seller and of the buyer of information.

Villas-Boas (1994) provides an early illustration a mechanism that shapes competition by addressing the question of whether competing firms should share the same marketing agency. Indeed, the effect of a marketing campaign that two competing firms have at the same time might be reduced. The author characterizes situations in which when the agency chooses to contract with only one firm in the market, and keep its competitor excluded. Such an equilibrium may occur because dealing with both competitors would lower the value of the agency's services as they compete more fiercely.

Baye and Morgan (2001) also emphasize the role of information gatekeepers on market competition by considering how they can break local monopolies. Gatekeepers can open new markets to firms which increases competition between them.

Turning to third parties selling information for decision making, Sarvary and Parker (1997) consider two firms selling information to consumers who are heterogeneous in their willingness to pay for the quality of information. Consumers can acquire information from any sellers, and they can merge information from different sources to get more precise information. The authors show that the strategies of information sellers depend on the degree of complementarity or substitutability between them, that is on the intensity of competition. When information sellers propose complementary products, price competition will be soft, and buyers will acquire information from several information sellers. When information sellers propose substitute information, they will compete more fiercely, prices will be low, and buyers will acquire information from one seller only.[3] Christen and Sarvary (2007) empirically support these results.

Xiang and Sarvary (2013) in an extension of Sarvary and Parker (1997) and

[3] Recent contributions propose related models of competition between data brokers. Gu et al. (2018) study the incentives for data brokers to share information together, and they show that information sharing occurs when sellers have substitute data sets. The rationale behind their result is that information sharing is a way for data brokers to avoid price competition.

Bimpikis et al. (2019) show that when information buyers see their actions as strategic complements, the data broker sells the most accurate information to all customers. When buyers view their actions as strategic substitutes the provider maximizes her profits by either restricting the overall supply of the information product, or distorting its content by offering information of inferior quality. To sum-up, information can increase competition between buyers, and the data broker takes this competitive effect of information into account when choosing which signal to sell to which firm. If information increases competition between firms, he will sell partial information so that the profits of the firms are not decreased too much, and thus the value of information remains high.

We analyze in the remaining of the section the literature about innovation licensing and information selling. As both innovations and information have a competitive effect on the downstream market, we identify similar results in both literatures, that we summarize in the following sections. When it is not specified the term information also refers to licensing.

1.2.1 Selling information to a subset of firms

Information can increase competition between information buyers, and thus lower their willingness to pay for information. We analyze cases where third parties selling information strategically limit the access to information (to a subset of firms), so that competition is not increased too much on the market. Thus the value of information remains high for firms that can buy it, and information sellers can charge firms a high price. In order to understand the effect of competition on the downstream market on a sellers strategies, we first consider a benchmark case where information has no competitive effect on downstream competition. Then we analyze how the literature takes this effect into account, allowing the seller not to sell information to a subset of firms to lower competition.

Katz and Shapiro (1986), Kamien (1992), and Rey and Salant (2012) study a monopolist specialized in innovation licensing to competing firms. In their model, the willingness to pay of a firm for the license depends on whether its competitors acquired it. This is because a firm that acquires an innovation becomes more competitive on the market, which lowers the profits of its competitors. Thus, by selling licenses to competing firms, part of the competitive gain is lost for the firms, as they also face more informed competitors. The authors find that the innovator optimally restricts the number of licenses on the market, so that competition is not intensified too much, and the value of the licenses does not decrease.

Finally, Iyer and Soberman (2000) study a sophisticated model of information selling to add value to a product. They consider two types of information. First, innovations can allow a firm to increase the value of the product for loyal customers. In this case ,the innovator sells licenses to all firms on the market. Secondly, innovations can increase the value for all customers and thus facilitate consumer poaching. In this case, the innovators sells information to a subset of firms only.

Information can also be used to price discriminate consumers. Braulin and Valletti (2016) propose a model in which a data broker sells consumer information to two vertically differentiated competitors. Montes et al. (2018) propose a model of information selling to competing firms for price discrimination purposes. In both approaches, information is perfect and allows firms to identify consumers' willingness to pay for their product. Information allows firms to better extract surplus from consumers, as they know their willingness to pay for a product, but it also increases competition between firms as they set their prices more aggressively. They show that a data broker will sell information to only one of the competitors in order to soften the competitive effect of increased information on the market. This in turns increases the value of information.

The literature shows clearly how information sellers can influence competition on the downstream market by controlling the access to information. However they ignore the possibility for information sellers to voluntarily lower the quality of information. We consider in the next section such a possibility and how an information seller can use it to lower competition between firms.

1.2.2 Lowering the quality of information and innovations

How a data broker can strategically choose information so that consumer surplus is extracted without competition being too much intensified on the product market? Bounie et al. (2018), and Belleflamme et al. (2017) provide answers to this question by considering a data broker selling information to competing firms for price discrimination purposes.

Belleflamme et al. (2017), show that a data broker will voluntarily lower the quality of information that he sells to firms for price discrimination consumers. Bounie et al. (2018) show that the data broker sells information on consumers with the highest willingness to pay, which allows firms to extract more consumer surplus. Information on low-valuation consumers is not sold to firms in order to soften competition. In other words, it is not optimal for the data broker to sell information on all consumers, as doing so would reduce the profits of the firms, and hence their willingness to pay for consumer information.

The literature on innovation also study similar mechanisms. Kastl et al. (2018) consider a consulting company selling to a firm information about the structure of its internal organization. Information is modelled as a set of signals and likelihood functions mapping states of nature to signals. The provider can produce any information structure at no cost. The authors allow the information provider to voluntarily supply imperfect information to competing firms. The rationale behind such practice is similar to the literature above where information and innovations are sold in competitive markets: as companies are more efficient with better information, they increase their production, which lowers the price of their product and consequently their profits, and thus lowers their willingness to pay for information. The sale of imperfect information thus occurs in equilibrium.

Turning to information in financial markets, García and Sangiorgi (2011) analyze the strategies of a monopolist information provider selling information of endogenous quality to competing investors. As the investing strategy of an agent has a negative externality on the profits of its competitor, and thus lowers the value of information, they find that the information provider sells either to all agents very imprecise information, or to a small number of agents information of the highest quality.

Cespa (2008) focuses on information selling at two successive stages. Information acquired in stage 1 by a firm can be partially reused in stage 2. Thus, the data broker will lower the quality of information in stage 1 so that the value of information in stage 2 is higher. Thus a firm competes inter-temporally with herself, and moderates self-competition by lowering the quality of information.

In a search model, Lizzeri (1999) focuses on intermediaries selling to buyers information on a product quality. They find that the intermediary will voluntarily lower competition between sellers by two means: first, by keeping one of the buyers uninformed, secondly, by selling partial information to firms.

1.3 The Arrow effect

The arrow paradox has long been identified in the literature on innovation and licensing (Arrow, 1972). It states that when a technological transfer occurs, an innovator needs to provide information ex ante about the license to the buyer, which directly lowers the value of the innovation. This effect can be translated into the context of financial markets: the price of an information product already contains some information for the buyers, and thus inherently lowers its value. Authors have studied how an information seller can counter this effect, either by using specific mechanisms of sale, of by blurring information.

Anton and Yao (2004) consider whether a firm should reveal information about its invention, and take the risk of being imitated, or keep it hidden but not being able to commercialize it. They show that small inventions are not imitated, and thus an innovator can license them. Important innovations, however will be kept secret by innovators when property rights are weak, because licensing them would expose them to imitation by competing firms.

In another paper, Anton and Yao (2002) study how much an inventor should reveal about its invention. They show that revealing partial information allows the buyer to have an accurate estimation of the value of the invention, while keeping enough hidden information for the inventor to sell.

Turning to financial markets Admati and Pfleiderer (1986) model information as signals on different states of nature. A monopolist information seller sells information to competing firms. The main takeaway of their paper is that the price of information reveals part of its content, and thus naturally lowers its value. Thus, if the information seller gives the genuine price of information, its value naturally decreases. A way to circumvent this effect is to sell artificially noisy versions of information. On top of this mechanism, and so that market prices are not affected by having symmetrically informed companies, the information seller will sell information with different degrees of noise to competing firms. By doing so, the seller can extract all the surplus generated by information from buyers

Admati and Pfleiderer (1990) show that this issue can be avoided by using a mechanism of indirect sales where an information seller creates a mutual fund, and buyers do not observe information but pay for a response. Contrary to direct sales, adding noise to information is never optimal even though they allow for such strategy in their model.

1.4 Payoff uncertainty

A seller can lower the precision of an information good to circumvent uncertainties that he faces when selling information. We consider here two types of uncertainty that are analyzed in the literature. First, an information or innovation seller can sell partial lower the quality of information or innovation when he faces an information asymmetry. Lowering the quality of information can force the information buyer to reveal part of its private information. Secondly, an information seller can have dissonant risk aversion with an information buyer. By selling partial information, an information seller can force an information buyer to take decisions that fit with more risk averse utilities.

Breaking information asymmetry

In Arora and Fosfuri (2005), Hörner and Skrzypacz (2016) and Bergemann et al. (2018), a data seller proposes information to agents that have heterogeneous valuations for information. The data seller thus faces an information asymmetry. Arora and Fosfuri (2005) find that in equilibrium, the seller offers similar information to everyone. Hörner and Skrzypacz (2016) show that by selling information at successive periods, the data seller can correct this information asymmetry. The data broker will split information in different selling periods in order to force firms to reveal their willingness to pay. Bergemann et al. (2018) propose an alternative mechanism where the data seller will propose a menu with information of different precision. Proposing a menu with information of various quality allows to extract more rent from consumers.[4]

Dissonant risk aversion

Weber and Croson (2004) focus on a dissonance between the utilities of an information seller and an information buyer, when information is used for a risky investment. Both the seller and the buyer are asymmetric in their utilities, and in particular in their aversion to risks. The buyers pays an ex post price that depends on the payoff of the investment. The firm is willing to invest in some assets, but that the data sellers finds too risky. The information seller can sell noisy information to the firms to avoid her to invest in such asset.

[4] This result is similar to Deneckere and Preston McAfee (1996).

1.5 Conclusion

We have reviewed the literature on third parties selling information and licenses to firms. This comprehensive view of the literature identifies two strategies used by third parties to sell their good.

First, third parties can exclude part of the firms from acquiring information or innovation. They deal with certain firms, while preventing other firms from acquiring the good. Firms that acquire information or innovation thus obtain a competitive advantage over excluded firms. This practice is worrying from the point of view of a competition authority as it leads to the emergence of dominant actors on the market. By preventing exclusive contracts, a regulator can restore fair competition on a market.

Secondly, intermediaries can shape the intensity competition on product markets by selling only partial information in order to extract more surplus from firms. Innovators can adopt similar strategies, for instance by improving only partially the efficiency of a firm so that the level of competition on the market is not too high. Competition authorities should also be concerned with such practice, that result in a lower competition on product markets.

Finally, as consumers have access to new tools and to new ways to hide their personal information from data intermediaries, further research should focus on how consumer empowerment affects information selling. On the on hand, data brokers selling consumer information to firms can increase consumer surplus by increasing competition on the market (Choi et al., 2019). On the other hand, the lack of transparency of the data brokerage industry raises concern about whether consumer empowerment is enough to protect privacy (Crain, 2018; Jann and Schottmüller, 2018). In particular, there is a need to integrate data protection regulation and competition policy.

CHAPTER 2

Selling Strategic Information in Digital Competitive Markets

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 combine any segment of the consumer demand in an information structure that is sold to firms that implement third-degree price discrimination. We show that the data broker (1) sells information on consumers who have the highest willingness to pay; (2) keeps consumers with low willingness to pay unidentified. The data broker strategically chooses to withhold information on consumer demand in order to soften competition between firms. Moreover, we prove that these results hold under first-degree price discrimination, which is a limit case of our model when firms have access to perfect information.

2.1 Introduction

The digital economy is driven by consumer information, what analysts have called 'the new oil' of the twenty first century.[1] Digital giants such as Facebook, Apple, Amazon and Google, base their business models on traces left by Internet users who visit their online websites. In a race to information dominance, these large companies also acquire information from data brokers that gather information about millions of people.[2]

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 (Pasquale (2015)). In a study of nine data brokers from 2014,[3] 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."[4] 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 information firms can acquire from data brokers. On the one hand, more information

[1] The Economist, The world's most valuable resource is no longer oil, but data, May 6, 2017.

[2] The recent Facebook scandal involving Cambridge Analytica has precisely revealed to the public the troubled relations between Facebook and data brokers (Washington Post, Facebook, longtime friend of data brokers, becomes their stiffest competition, March 29, 2018; Business Insider, Facebook is quietly buying information from data brokers about its users' offline lives, December 30, 2016.)

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

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

allows firms to better target consumers and price discriminate. They can extract more consumer surplus, 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 lowers the profits of the firms. Overall, there is an economic trade-off between surplus extraction and increased competition. This article analyzes this trade-off when a data broker strategically combines consumer segments 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 (Hayek, 1945). The emergence of data brokers adds a strategic dimension to the literature that assumes that information is exogenously available on the market.[5] Braulin and Valletti (2016) study vertically differentiated products, for which consumers have hidden valuations. The data broker can sell to firms information on these valuations. Montes et al. (2018) consider information allowing competing firms to first-degree price discriminate consumers. In both articles, the data broker sells either information on all consumers, or no information at all.

We build a model where a data broker can sell information that partitions consumer demand into segments of arbitrary sizes to one or to two competing firms. The data broker can strategically sell consumer segments of information to firms competing on the product market, and can weaken or strengthen the intensity of competition by determining the quantity of information available on the market. In other words, the data broker has the choice to sell information on all available consumer segments, on a subset of consumer segments, or no information at all. By acquiring information from the data broker, firms can identify the most profitable consumer segments, on which they set specific prices.

Using this setting, we show that the data broker sells information on con-

[5] See for instance Radner et al. (1961), Vives (1984), Thisse and Vives (1988), Burke et al. (2012), Shy and Stenbacka (2016), Kim et al. (2018).

sumers with the highest willingness to pay, which allows firms to extract more consumer surplus. Information on low-valuation consumers is not sold to firms in order to soften competition. In other words, it is not optimal for the data broker to sell all consumer segments, as doing so would reduce the profits of the firms, and hence their willingness to pay for consumer information.

This paper contributes to the fast growing literature on customer information acquisition by allowing a data broker to sell any combination of consumer segments. Data brokers can strengthen or weaken competition between firms by choosing the amount of consumer information that they sell. Thus, the strategies of data brokers can have conflicting policy implications for competition authorities and data protection agencies. On the one hand, competition authorities could encourage industry practices that increase information and competition on the market. On the other hand, more information available on the market allows firms to extract more consumer surplus, which can harm consumers. Data protection agencies could be wary of such practices.

The remainder of the article is organized as follows. In Section 2.2 we describe the model, and in Section 2.3 we characterize the optimal structure of information. In Section 2.4, we provide the equilibrium of the game, and we discuss the effects of information acquisition on welfare. We conclude in Section 2.5.

2.2 Model set-up

The model involves 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.[6]

The two firms are located at 0 and 1 on the unit line and sell competing

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

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. 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:

$$\begin{cases} U(x) = V - tx - p_1, \text{ if he buys from Firm 1,} \\ \\ U(x) = V - t(1-x) - p_2, \text{ if he buys from Firm 2,} \\ \\ U(x) = 0, \text{ if he does not consume.} \end{cases} \tag{2.1}$$

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.

2.2.1 Information structure

Firms know that consumers are uniformly distributed on the unit line, but without further information, they are unable to identify their locations. Therefore, firms do not know the degree to which consumers value their products and cannot price discriminate them.[7]

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 or the precision 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 $\frac{1}{k}$.

Figure 4.1 illustrates the reference partition that includes all segments of size $\frac{1}{k}$. All existing models in the literature assume that the data broker can only

[7] This assumption is also made by Braulin and Valletti (2016) and Montes et al. (2018).

sell the reference partition $\mathcal{P}_{ref}$ to competing firms, or no information at all. A major contribution of the present article is to demonstrate that the optimal partition sold by the data broker is not the reference partition $\mathcal{P}_{ref}$.

Fig. 2.1: 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 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 includes 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 $\emptyset$ are included in $\mathbb{P}$.

The data broker can sell any partition $\mathcal{P}$ in the set of partitions $\mathbb{P}$: for instance, a partition starting with one segment of size $\frac{1}{k}$, and another segment of size $\frac{2}{k}$, and so on, as illustrated in Figure 4.4.

Fig. 2.2: Example of a partition of the unit line

A firm that has information $\{[0, \frac{s_1}{k}], [\frac{s_1}{k}, \frac{s_2}{k}], .., [\frac{s_{n-1}}{k}, 1]\}$ 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 a price p_1 on $[0, \frac{s_1}{k}]$, a price p_{i+1} on $[\frac{s_i}{k}, \frac{s_{i+1}}{k}]$, and so forth for each segment. Firms thus practice third-degree price discrimination. We show in Section 4 that first-degree price discrimination is a limit case of our model when $k \to \infty$.

Finally, to keep the analysis as simple as possible, we rule out elements of the partition $\mathcal{P}$ that consist of several disjoint intervals, and that add uncertainty on the location of consumers, such as $[\frac{s_i}{k}, \frac{s_{i+1}}{k}] \cup [\frac{s_{i'}}{k}, \frac{s_{i'+1}}{k}]$ (with $i' > i + 1$).

2.2.2 Strategies and timing

We present the strategies and profits of the data broker and of firms, and then the timing of the game.

The data broker maximizes its profits by choosing a pair of partitions noted $\mathcal{P}_1, \mathcal{P}_2$ in the set $\mathbb{P}$ that are respectively proposed to Firm 1 and Firm 2. These partitions, or information structures, can be potentially different for Firm 1 and Firm 2. Finding the optimal partitions is a complex optimization problem given that the cardinality of $\mathbb{P}$ can be very large, and that we do not impose restrictions on the total number k of segments of the reference partition.

We denote whether a firm and its competitor are informed (I) or uninformed (NI) by the couple (x, y) where $x, y \in \{I, NI\}$, and (I, NI) refers to a situation in which Firm θ is informed and Firm $-\theta$ is uninformed. We note $\pi_\theta^{x,y}(\mathcal{P}_1, \mathcal{P}_2)$ the profit of Firm θ with information $\mathcal{P}_1$ whereas Firm $-\theta$ has information $\mathcal{P}_2$ in situation (x, y). For instance, $\pi_1^{I,I}(\mathcal{P}_1, \mathcal{P}_2)$, is the profit of Firm 1 when both firms are informed.[8] For any information structure, we need to compute the profits in four possible configurations: $\{\pi_\theta^{NI,NI}, \pi_\theta^{I,NI}, \pi_\theta^{NI,I}, \pi_\theta^{I,I}\}$.

The data broker decides to sell information to one firm only or to both firms. In both cases, information is sold through an auction mechanism with negative externalities as in Jehiel and Moldovanu (2000). Before the auction takes place, the data broker proposes partition $\mathcal{P}_1$ to Firm 1 and partition $\mathcal{P}_2$ to Firm 2. When the data broker sells information to only one firm, we assume that it is Firm 1, without loss of generality.

[8] To simplify notations, we will drop arguments $\mathcal{P}_1, \mathcal{P}_2$ when there is no confusion.

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

$$\Pi = \begin{cases} \Pi_1(\mathcal{P}_1, \mathcal{P}_2) = \pi_1^{I,NI}(\mathcal{P}_1, \emptyset) - \pi_1^{NI,I}(\emptyset, \mathcal{P}_2) \\ \text{if the data broker sells information to Firm 1,} \\ \\ \Pi_2(\mathcal{P}_1, \mathcal{P}_2) = \pi_1^{I,I}(\mathcal{P}_1, \mathcal{P}_2) - \pi_1^{NI,I}(\emptyset, \mathcal{P}_2) + \pi_2^{I,I}(\mathcal{P}_2, \mathcal{P}_1) - \pi_2^{NI,I}(\emptyset, \mathcal{P}_1) \\ \text{if the data broker sells information to both competitors.} \end{cases} \tag{2.2}$$

The first part of Eq. (3.3.3), Π_1, is the profit of the data broker when selling partitions $\mathcal{P}_1$ to Firm 1 only; the second part of Eq. (3.3.3), Π_2, is the profit of the data broker when selling partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ to Firm 1 and Firm 2 respectively.[9]

In order to compute the profits of the firms, we need to compute demand and prices on each consumer segment. When a firm has no information, it sets a uniform price on the whole interval $[0, 1]$. However, when a firm has a partition $\mathcal{P}_1$, it sets a price on each segment of the partition. There are two types of segments to analyze: segments on which both firms have a strictly positive demand, and segments on which a firm is a monopolist. We assume that Firm θ sets prices in two stages.[10] First, it sets prices on segments where it shares consumer demand with its competitor. Then, on segments where it is a monopolist, it sets a monopoly price, constrained by the price proposed by its competitor. Each firm knows whether its competitor is informed, as well as the structure of the partition acquired by its competitor.

For any partition $\mathcal{P}$ composed of n segments ($n \leq k$), Firm θ maximizes

[9] We check the implementability constraint: $\pi_\theta^{I,NI} - (\pi_\theta^{I,I} - \pi_\theta^{NI,I}) \geq \pi_\theta^{NI,NI}$, which is always verified in equilibrium.

[10] Introducing a sequential pricing decision avoids the non-existence of Nash equilibrium in pure strategies, and is supported by managerial practices (Fudenberg and Villas-Boas, 2006).

its profit 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 Firm θ when both firms are informed is given by:

$$\pi_\theta^{I,I} = \sum_{i=1}^{n} d_{\theta i}(\mathbf{p}_\theta, \mathbf{p}_{-\theta}) p_{\theta i}, \tag{2.3}$$

where $d_{\theta i}(.)$ is the demand of Firm θ on segment i. We define $\pi_\theta^{I,NI}, \pi_\theta^{NI,I}$ in a similar way.

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 set prices on the competitive segments.
- Stage 3: firms price discriminate consumers on the segments where they have monopoly power.

2.3 Optimal information structure

Equilibrium prices charged to consumers and profits of the firms in stages 2 and 3 depend, first, on the optimal partition sold by the data broker in stage 1, and second, on the strategy of the data broker 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. 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 that identifies consumers close to the firm up to a cutoff point, and that leaves consumers unidentified in the remaining segment. In Section 2.4, we determine the number of segments where consumers are identified in the optimal information structure, which depends on whether the data broker sells information to one or to both firms. We finally discuss at the end of this section how information acquisition affects competition between firms.

2.3.1 Information is sold to only one firm

When information is only sold to Firm 1 (without loss of generality), Theorem 1 shows that the data broker sells information on all segments up to a point $\frac{j}{k}$, and does not sell information after that point. In the remainder 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*.

THEOREM 1: *The data broker sells to Firm 1 a partition that divides the unit line into two intervals:*

- The first interval consists of j segments of size $\frac{1}{k}$ on $[0, \frac{j}{k}]$ where consumers are identified.

- Consumers in the second interval of size $1 - \frac{j}{k}$ are unidentified.

Proof: See Appendix A.1.

The proof proceeds in the following way. Consider any information structure. The data broker maximizes the profit of Firm 1 with respect to $\mathcal{P}_1$ when Firm 2 is not informed. When Firm 1 does not purchase information, the data broker determines its least favorable outside option. It is intuitively clear that the worst case scenario for Firm 1 when it is not informed is when Firm 2 has the most informative partition, that is when the data broker proposes the reference partition to Firm 2. When choosing $\mathcal{P}_1$, we first show that the data broker finds profitable to re-order segments and reduce their size to $\frac{1}{k}$ so that Firm 1 has more information on consumers closest to its location. Secondly, the data broker can soften competition between firms by leaving a segment of unidentified consumers in the middle of the line.

Figure 3.2 illustrates Theorem 1. The data broker sells partition $\mathcal{P}_1$ to Firm 1 and does not sell information to Firm 2. Therefore, Firm 2 sets a uniform price p_2 on the whole unit line. Firm 1 sets price p_1 on the segment of unidentified consumers. Firm 1 can identify consumers on each segment on the left (indexed by $i = 1, .., j$), of size $\frac{1}{k}$. Firm 1 price discriminates consumers and sets different prices on each segment, with p_{1i} being the price on the ith segment from the origin.

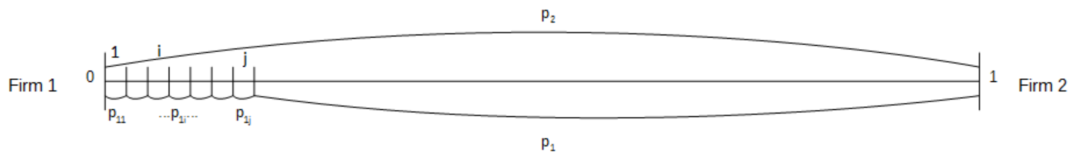

Fig. 2.3: Selling information to one firm: Firm 1 informed

Theorem 1 makes an important contribution to the existing literature that assumes that the data broker either always sells all consumer segments to firms, or sells no information at all (Braulin and Valletti, 2016; Montes et al., 2018). We show that selling the reference partition (with information on all consumer segments) is not optimal. The existing literature thus overestimates the effects of data brokers on the prices paid by consumers. Indeed, when firms have information on each consumer, they compete more intensively, resulting in lower prices. For instance in Baye and Morgan (2001), firms end up competing à la Bertrand, making zero profits in equilibrium. Our model shows on the contrary that a data broker has always incentives to soften competition.

2.3.2 *The data broker sells information to both firms*

In this section we analyze a situation where the data broker sells information to both firms. We show that the optimal partition is similar to Theorem 1, and that it has the following features. Theorem 2 first demonstrates that the data broker sells to each firm all segments up to a point $\frac{j}{k}$ to Firm 1, and $\frac{j'}{k}$ to Firm 2, where j' is defined as the number of segments starting from point 1. Theorem

2 then shows that the data broker sells the same information structure to both firms, that is $\frac{j}{k} = \frac{j'}{k}$.

In order to simplify the analysis, we make the following assumption that rules out situations where firms compete and share demand segments at the extremities of the unit line:

ASSUMPTION 1: *The data broker proposes partitions in which there is only one segment on which firms compete.*

Assumption 1 greatly simplifies the resolution and is not too restrictive for the following reasons. First, partitions that are ruled out by Assumption 1 are those that increase the competitive pressure on both firms, and which do not increase profits for the data broker. Secondly, in Theorem 1 the optimal partition when the data broker sells information to only one firm satisfies Assumption 1. Thirdly, we will show that Assumption 1 is not binding in equilibrium when the data broker sells information to both firms.

We now state Theorem 2:

THEOREM 2: *Suppose that Assumption 1 holds.*

(a) The data broker sells to Firm 1 (resp. Firm 2) a partition that divides the unit line into two intervals:

- *The first interval consists of 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.*
- *Consumers in the second interval of size $1 - \frac{j}{k}$ (resp. $1 - \frac{j'}{k}$) are unidentified.*

(b) Partitions sold to Firm 1 and Firm 2 are symmetric, and $j = j'$.

Proof: See Appendix A.2.

Theorem 2 generalizes the results of Theorem 1 when the data broker sells information to two firms: the data broker does not sell segments in the middle of the Hotelling line in order to reduce competition between firms, and extract more surplus. The proof proceeds in the following way. We consider any partition satisfying Assumption 1. We show that the data broker always finds it more profitable to sell 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 $\frac{j}{k} = \frac{j'}{k}$.

Figure 4.6 illustrates Theorem 2. Firm 1 (resp. Firm 2) sets a unique price p_1 (resp. p_2) on $[\frac{j}{k}, 1]$ (resp. $[0, 1 - \frac{j'}{k}]$). Firm 1 (resp. Firm 2) is informed on each segment of size $\frac{1}{k}$ closest to its location until $\frac{j}{k}$ (resp. $1 - \frac{j'}{k}$), and sets prices p_{1i} (resp. p_{2i}).

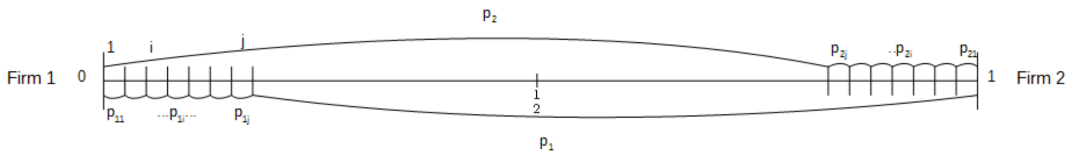

Fig. 2.4: Selling information to both firms

2.3.3 *Competitive effects of information acquisition*

We now interpret how information acquisition affects competition between firms. Consider the case where the data broker sells information to Firm 1 only.[11] We analyze the impact of the acquisition of an additional segment to the optimal partition on prices and profits of firms. 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 $\frac{j+1}{k}$.

Purchasing an additional segment has two effects on the profits of both firms:

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

[11] A similar reasoning applies when the data broker sells information to two firms.

b) A competitive effect; 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 the profits of Firm 1.

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

2.4 Model resolution

In this section, we solve the game by backward induction. We compute the equilibrium prices and profits of Firm 1 and 2 in stages 2 and 3, 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.

2.4.1 Stages 2 and 3: firms set prices

We denote by $d_{\theta i}$ the demand of Firm θ on the ith segment. By convention, d_θ is the demand on the last segment. An informed Firm θ maximizes the following profit function with respect to $p_{\theta 1}, .., p_{\theta j}$, and p_θ:

$$\pi_\theta = \sum_{i=1}^{j} d_{\theta i} p_{\theta i} + p_\theta d_\theta. \tag{2.4}$$

When $j = 0$,[12] the firm does not distinguish any consumer on the unit line, and sets a uniform price as in the Hotelling model. An uninformed Firm θ maximizes $\pi_\theta = p_\theta d_\theta$ with respect to p_θ.

[12] By convention, $\sum_{i=1}^{0} d_{\theta i} p_{\theta i} = 0$.

Theorems 1 and 2 show that the optimal partition sold by a data broker is not the reference partition.[13] The data broker only sells segments of size $\frac{1}{k}$ that are located closest to Firm θ. This partition allows firms to better extract surplus from consumers with the highest willingness to pay while softening competition between firms by keeping consumers with low willingness to pay unidentified.

Using Theorems 1 and 2, we characterize the sub-game perfect equilibria for the optimal structure of information by backward induction. There are three cases to consider. In the first case, firms have no information. In the second case, the data broker sells information to only one firm. In the third case, the data broker sells information to both firms.

2.4.1.1 *The data broker does not sell information*

Firms have no information on consumers and compete in the standard Hotelling framework. Firm θ sets $p_\theta = t$ in equilibrium, and the resulting demand is $d_\theta = \frac{p_{-\theta} - p_\theta + t}{2t}$. The profits of Firm θ are $\pi_\theta = \frac{t}{2}$.

2.4.1.2 *The data broker sells information to one firm*

Without loss of generality, we assume that only Firm 1 is informed. Firm 1 can distinguish $j+1$ segments of consumer demand, with j being an integer lower than k. Firm 1 price discriminates by setting a price for each segment p_{1i}. Firm 2 has no information, and sets a uniform price p_2.

Firm 1 maximizes $\pi_1 = \sum_{i=1}^{j} d_{1i}p_{1i} + p_1 d_1$ first with respect to p_1, then with respect to p_{1i} for $i = 1, .., j$. Firm 2 maximizes $\pi_2 = p_2 d_2$ with respect to p_2.

Profits maximization leads to the prices given in Lemma 1 that we will use to compute the profits of the data broker in Lemma 3.

LEMMA 1: *The market equilibrium when the data broker chooses a partition of j segments of size $\frac{1}{k}$ on $[0, \frac{j}{k}]$ and one segment of unidentified consumers on $[\frac{j}{k}, 1]$ is as follows:*

[13] Thus, Montes et al. (2018) is a special case of our model when the data broker cannot recombine consumers segments, and only sells the reference partition.

- *Firm 1 captures all demand on each segment* $i = 1, .., j$*, and:*

$$p_{1i} = 2t[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k}].$$

- *Firms compete on the segment of unidentified consumers, and the prices are:*

$$p_1 = t[1 - \frac{4}{3}\frac{j}{k}], \quad \text{and} \quad p_2 = t[1 - \frac{2}{3}\frac{j}{k}].$$

Proof: See Appendix B.

Lemma 1 shows how prices and profits vary with j. When j increases, the competitive pressure on the market increases. As a result, prices decrease both on the competitive segment and on the segments where consumers are identified. Firm 2 suffers from more information on the market, since more information reduces the uniform price p_2. For low values of j, Firm 1 benefits from more information and extracts more consumer surplus. Profits reach a maximum and then decrease due to increased competition.

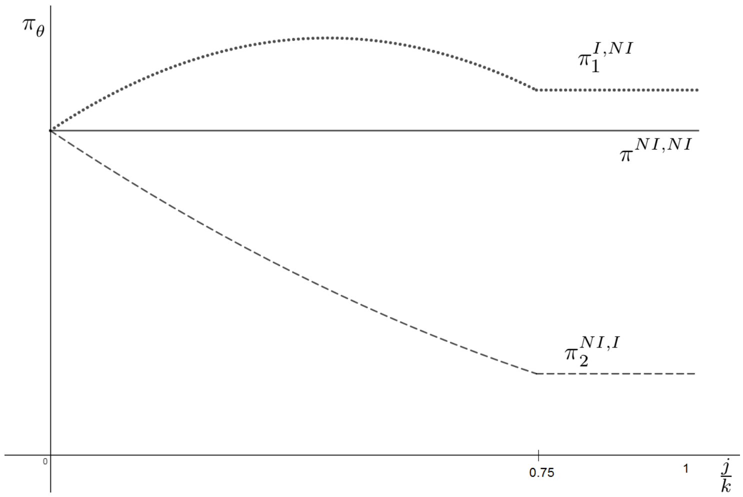

Fig. 2.5: Profits of the informed firm ($\pi_1^{I,NI}$) and of the uninformed firm ($\pi_2^{NI,I}$) when the data broker sells information only to Firm 1 ($t = 1$ and $k = 200$).

Figure 2.5 displays how the profits of the firms change with respect to $\frac{j}{k}$ when Firm 1 acquires j segments of size $\frac{1}{k}$ on $[0, \frac{j}{k}]$, and Firm 2 remains uninformed (the formulas are given in Appendix B). We observe that the profits of Firm 1 follow an inverted U-shaped curve on $[0, \frac{3}{4}]$: more information increases the profits of Firm 1 when the surplus extraction effect dominates the competition effect. The profits reach a maximum and then decrease in a second phase. At this point, more information leads to more competition, which dominates the extraction of consumer surplus and thus reduces the profits of Firm 1. Firm 2 is always harmed when Firm 1 acquires information and its profits always decrease with j. Comparing the profits of the firms with information to those obtained in the Hotelling case, we see that the profits of Firm 1 (resp. Firm 2) are always higher (resp. lower) than the profits of the firms without information. On $[\frac{3}{4}, 1]$, more information does not change prices set by Firm 1, and acquiring information on these consumers does not change the profits of the firms.

2.4.1.3 *The data broker sells information to both firms*

We have characterized in Theorem 2 the optimal information structure that the data broker sells to both firms. Firm 1 identifies j segments, $\{[\frac{i-1}{k}, \frac{i}{k}]\}$ with $i = 1, .., j$ and $j \in \mathbb{N}^*$, and Firm 2 identifies the segments $\{[1 - \frac{i}{k}, 1 - \frac{i-1}{k}]\}$. This leaves a segment of unidentified consumers in the middle of the line $[0, 1]$ where both firms compete. At the extremities of the unit line, both firms price discriminate identified consumers, as described in Figure 4.6 in Section 2.3.2.

Lemma 2 gives the equilibrium prices that we will use to compute the profits of the data broker in Lemma 3.

LEMMA 2: *The equilibrium when both firms are informed is characterized by the following properties*:

- For each segment $i = 1, .., j$:

$$p_{\theta i} = 2t[1 - \frac{j}{k} - \frac{i}{k}].$$

- For the segment of size $1 - \frac{j}{k}$, where firms compete:

$$p_\theta = t[1 - 2\frac{j}{k}].$$

Proof: See Appendix C.

According to Lemma 2, prices p_1 and p_2 set by firms on the segment were they compete decrease with j. Prices for identified consumers $p_{\theta i}$ also decrease with j. More information increases competition between firms, which reduces the prices that they set. Figure 2.6 illustrates this effect when j increases from 0 to 1.

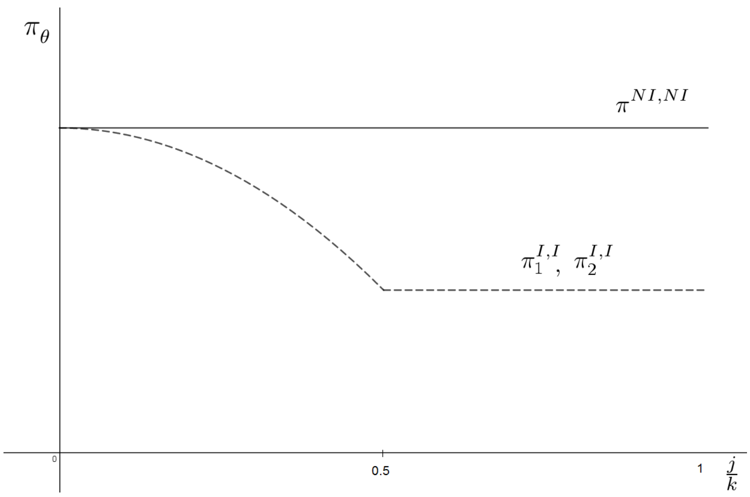

Fig. 2.6: Profits of the firms ($\pi_1^{I,I}$ and $\pi_2^{I,I}$) when the data broker sells information to both firms (for $t = 1$ and $k = 200$).

Figure 2.6 displays the profits of the firms when they are informed. On the horizontal axis, $\frac{j}{k}$ is the limit between the identified and unidentified consumers (the formulas are given in Appendix C). $\pi^{NI,NI}$ denotes the profits of the firms in the standard Hotelling model. When both firms acquire information, their profits always decrease with j, and reach a minimum when the data broker sells information on all segments of size $\frac{1}{k}$ on $[0, \frac{1}{2})$. Beyond $\frac{1}{2}$, more information does not change the profits of the firms.

2.4.2 Stage 1: profits of the data broker

The data broker can choose among the set of allowable partitions that we have proved to be optimal. The data broker compares the three different outcomes analyzed in stages 2 and 3: selling no information, selling information to only one firm or selling information to both competitors. When no information is sold, the data broker makes no profits, and we refer to this case as the outside option. In the remainder of the article, all lemmas and propositions are stated under Assumption 1.

Using Lemma 1 and Lemma 2, we compute the profits of the data broker with respect to j, first when only one firm is informed and, secondly when both firms are informed. Using Theorems 1 and 2, profits are straightforward to compute, following the mechanism explained in Section 2.2.2, and are given in Lemma 3.

LEMMA 3: *The profits of the data broker are as follows:*

- *When the data broker sells information to only one firm:*

$$\begin{aligned}\Pi_1(j) = w_1(j) &= \pi^{I,NI}(j,\emptyset) - \pi^{NI,I}(\emptyset,\mathcal{P}_{ref}) \\ &= \frac{3t}{8} + \frac{2jt}{3k} - \frac{t}{4k} - \frac{7j^2t}{9k^2} - \frac{jt}{k^2} - \frac{t}{8k^2}.\end{aligned}$$

- *When the data broker sells information to both competitors:*

$$\begin{aligned}\Pi_2(j) = 2w_2(j) &= 2[\pi^{I,I}(j) - \pi^{NI,I}(j)] \\ &= 2[\frac{2jt}{3k} - \frac{11j^2t}{9k^2} - \frac{jt}{k^2}].\end{aligned}$$

2.4.3 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 one firm in equilibrium.

Using Lemma 3, we first determine the optimal values of j when one or both firms are informed, then we compare the profits of the firms in the two situations.[14]

PROPOSITION 1:

- *(a) When one firm buys information, the data broker sets:*

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

- *(b) When both firms buy information, the data broker sets:*

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

- *(c) The data broker sells information to only one firm.*

Proof: See Appendix D.

Proposition 1 (a) and 1 (b) show that the optimal number of segments j_1^* and j_2^* sold to firms is less than k. In other words, the optimal partition is not the reference partition used in the existing literature, and the data broker does not sell information on all consumer segments. Moreover, the total amount of information sold on the market is larger when the data broker sells information to two firms: $2j_2^* > j_1^*$. Finally, $j_2^* < \frac{1}{2}$, which means that the constraints imposed by Assumption 1 are not binding in equilibrium.

2.4.4 Welfare analysis

Total surplus remains constant compared to the standard Hotelling case where firms are uninformed. Firms lose surplus that is captured by the data broker and by consumers. Even though total consumer surplus increases, some consumers win and other consumers lose. We define $\Delta CS(k)$ as the difference between

[14] For the proof of Proposition 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^*.

consumer surplus when the data broker sells partition with information precision k and consumer surplus in the Hotelling model (i.e. without information).

2.4.4.1 Profits of the firms in equilibrium

Using the optimal values found in Proposition 1, we can compare the profits of the firms in equilibrium:

PROPOSITION 2: *The profits of the firms verify the following property:*

$$\pi^{*I,NI} \geq \pi^{*NI,NI} \geq \pi^{*I,I} \geq \pi^{*NI,I}.$$

Proof: See Appendix E.

Proposition 2 confirms that the firms that acquire information face a prisoner's dilemma. As illustrated in Stole (2007), the profits of both firms when they are informed ($\pi^{*I,I}$) are lower than those of the Hotelling model when both firms are uninformed ($\pi^{*NI,NI}$). In other words, competing firms choose to acquire information even though more information increases competition on the market, as being the only uninformed firm would induce even lower profits $\pi^{*NI,I}$.

2.4.4.2 Consumer surplus

We show that consumer surplus increases compared with the standard Hotelling model, and that the change in consumer surplus decreases with information precision k. These results are stated in Proposition 3.

PROPOSITION 3:

- *(a) For a given k, consumer surplus is higher than in the Hotelling model:*

$$\Delta CS(k) > 0.$$

- *(b) Consumer surplus decreases with k:*

$$\frac{\partial CS(k)}{\partial k} < 0.$$

Proof: The proof is straightforward and available upon request.

Proposition 3 (a) shows that consumer surplus increases compared to the Hotelling model without information; however, some consumers gain and other lose when firms acquire information. Suppose that Firm 1 is informed and Firm 2 is uninformed as described in Proposition 1. It is straightforward to show that, on the one hand, identified consumers on $[0, \frac{5k+3}{14k}]$ pay a higher price than in the Hotelling model: $p_{\theta i}^{I,NI} \geq p_{\theta}^{NI,NI}$. They lose surplus from better price discrimination. On the other hand, consumers located on $[\frac{5k+3}{14k}, 1]$ pay a lower price than in the Hotelling model, $p_{\theta i}^{I,NI}, p_{\theta}^{I,NI}, p_{\theta}^{NI,I} \leq p_{\theta}^{NI,NI}$, and benefit from increased competition between firms, resulting from more information on the market. Some of these consumers, located on $[\frac{5k+3}{14k}, \frac{6k-9}{14k}]$ are identified. The others, located on $[\frac{6k-9}{14k}, 1]$, are unidentified.

Information acquisition by a firm has a positive effect on consumer surplus. Due to increased competition, unidentified consumers located furthest away from the informed firm benefit from lower prices even though that firm extracts more surplus from identified consumers. Consumers located closest to the informed firm are identified, they pay a higher price and suffer from price discrimination.

Proposition 3 (b) states that consumers suffer from higher information precision k: $\frac{\partial CS(k)}{\partial k<0}$. Moreover, the share of identified consumers, $\frac{6k-9}{14k}$, increases with k. As the data broker has more precise consumer information, the share of identified consumers increases.

2.4.4.3 First-degree price discrimination

We analyze in this section how first-degree price discrimination impacts the strategies of the data broker. Indeed, digital technologies allow firms to better classify and target consumers, which increases the precision of information. It is important to understand how increasing information precision ($k \to \infty$) affects our results.

We show that the model with first-degree price discrimination is a special case of the model with third-degree price discrimination when $k \to +\infty$. Similarly to Theorem 1, the data broker sells to one firm the following information structure: a share of consumers is fully identified, and low-valuation consumers are unidentified.

PROPOSITION 4:

- *When firms first-degree price discriminate, the data broker sells only to Firm 1 (without loss of generality) an information structure characterized by the following partition:*
  - *on* $[0, \frac{3}{7}]$*, consumers are identified.*
  - *on* $[\frac{3}{7}, 1]$*, consumers are unidentified.*
- *When* $k \to +\infty$*, the equilibrium partition under third-degree price discrimination converges to the partition under first-degree price discrimination.*

Proof: See Appendix F.

From Proposition 4, it is straightforward to show that the profits of the firms and consumer surplus under third-degree price discrimination converge to their corresponding values under first-degree price discrimination: $\pi^{I,NI} \xrightarrow{k\to\infty} \frac{9}{14}t > \pi^{NI,NI}$ and $\pi^{NI,I} \xrightarrow{k\to\infty} \frac{25}{98}t < \pi^{NI,NI}$. Additionally, when Firm 1 is informed (without loss of generality), consumers on $[\frac{5}{14}, 1]$ benefit from lower prices.

Let CS^* denote consumer surplus under first-degree price discrimination, Corollary 1 shows that, as firms acquire more precise information ($k \to +\infty$), consumer surplus under third-degree price discrimination converges to CS^*.

COROLLARY 1:

Consumer surplus under third-degree price discrimination converges to CS^**.*

2.5 Conclusion

Understanding how data brokers impact competition on markets is a new promising field of research. We contribute to this literature by developing a model in which a data broker can choose among a large set of possible information structures to sell to firms. The optimal information structure divides consumers into two groups: consumers with the highest willingness to pay are identified while low-valuation consumers remain unidentified. The data broker strategically chooses the amount of information sold to market competitors by balancing consumer surplus extraction and competition effects.

Our results have several policy and managerial implications. First, selling information on more consumers increases the competitive pressure on the market, which could be desirable from the perspective of a competition authority. Thus, as the market develops, additional data brokers could increase this competitive pressure. However, the data brokerage industry faces a wave of consolidation and data sharing agreements.[15] For instance, Bluekai and Datalogix were acquired by Oracle,[16] and Equifax and FICO have agreed to share information on consumer financial data.[17] Two questions should be at least further investigated: how is competition on the product market affected when the number of

[15] Regulators and legislators have recently analyzed the impacts of data brokers on markets (Crain, 2018).

[16] AdExchanger, How Datalogix Made Oracle's BlueKai Acquisition Even Smarter, January 19, 2015.

[17] The Wall Street Journal, Equifax, FICO Team Up to Sell Consumer Data to Banks, March 27, 2019.

data brokers increases? Conversely, how is competition on the market affected by a merger between data brokers?

Secondly, with the development of big data, machine learning and artificial intelligence, data brokers can identify consumers almost perfectly, which increases the quality of information sold to firms. We have shown that moving from third-degree price discrimination to first-degree price discrimination, that is, when information becomes more precise, consumer surplus decreases. Countering this trend, new regulations have been implemented in different parts of the world. For instance, the General Data Protection Regulation in the European Union creates new ways to protect consumers through opt-in, right to be forgotten, data minimization and privacy by design. Further research should analyze the effects of these new data regulations on the strategies of data brokers.

Thirdly, selling information on more consumers increases consumer surplus due to the competition effect that exercises a downward pressure on prices. However, more precise information reduces consumer surplus through a stronger surplus extraction effect. Thus, there could be a conflict between competition authorities that see competition enhancing data sales with a keen eye, and data protection agencies concerned about consumer targeting. Again, further research should investigate how data collection and selling strategies are related, and how they impact consumer surplus.

Finally, we discuss the managerial implications of our results. We have assumed that the data broker sells information through an auction mechanism with negative externality: if a firm does not purchase information from the data broker, the other firm will. However, other selling mechanisms may well be used. For instance, Bergemann et al. (2018) consider an information seller, or equivalently a data broker, who does not know the willingness to pay of firms, and therefore, proposes a menu of information structures with different precisions. As the data brokerage industry is expanding, several questions arise.

Which selling mechanism will be chosen by the industry? How do different selling mechanisms change the strategies of data brokers?

2.6 Appendix

Proofs of Theorems 1 and 2

In Appendix A.1, we show that the data broker optimally sells a partition that divides the unit line into two intervals. The first interval identifies the closest consumers to a firm and is partitioned in j segments of size $\frac{1}{k}$. The second interval is of size $1 - \frac{j}{k}$ and leaves the other consumers unidentified. We first establish this claim when the data broker sells information to only one firm, and second when it sells information to both firms.[18]

Proof of Theorem 1: the data broker sells information to one firm

The data broker can choose any partition in the sigma-field $\mathbb{P}$ generated by the elementary segments of size $\frac{1}{k}$, to sell to Firm 1 (without loss of generality). There are three types of segments to consider:

- Segments A, where Firm 1 is in constrained monopoly;
- Segments B, where Firms 1 and 2 compete.
- Segments C, where Firms 1 makes zero profit.

We proceed in three steps. In step 1 we analyze type A segments. We show that it is optimal to sell a partition where type A segments are of size $\frac{1}{k}$. In step 2, we show that all segments of type A are located closest to Firm 1. In step 3 we analyze segments of type B and we show that it is always more profitable to sell a union of such segments. Therefore, there is only one segment of type B, located furthest away from Firm 1, and of size $1 - \frac{j}{k}$ (with j an integer, $j \leq k$).

[18] All along the proofs, we refer to Liu and Serfes (2004) who prove the continuity and concavity of the profit functions with third-degree price discrimination.

Finally, we can discard segments of type C because information on consumers on these segments does not increase profits.

Step 1: We analyze segments of type A where Firm 1 is in constrained monopoly, and show that reducing the size of segments to $\frac{1}{k}$ is optimal.

Consider any segment $I = [\frac{i}{k}, \frac{i+l}{k}]$ of type A with l, i integers verifying $i + l \leq k$ and $l \geq 2$, such that Firm 1 is in constrained monopoly on this segment. We show that dividing this segment into two sub-segments increases the profits of Firm 1. Figure 4.7 shows on the left panel a partition with segment I of type A, and on the right, a finer partition including segments I_1 and I_2, also of type A. We compare profits in both situations and show that the finer segmentation is more profitable for Firm 1. We write $\pi_1^A(\mathcal{P})$ and $\pi_1^{AA}(\mathcal{P}')$ the profits of Firm 1 on I with partitions $\mathcal{P}$ and on I_1 and I_2 with partition $\mathcal{P}'$.

Fig. 2.7: Step 1: segments of type A

To prove this claim, we establish that the profit of Firm 1 is higher with a finer partition $\mathcal{P}'$ with two segments : $I_1 = [\frac{i}{k}, \frac{i+1}{k}]$ and $I_2 = [\frac{i+1}{k}, \frac{i+l}{k}]$ than with a coarser partition $\mathcal{P}$ with I.

First, profits with the coarser partition is: $\pi_1^A(\mathcal{P}) = p_{1i}d_1 = p_{1i}\frac{l}{k}$. The demand is $\frac{l}{k}$ as Firm 1 gets all consumers by assumption; p_{1i} is such that the indifferent consumer x is located at $\frac{i+l}{k}$:

$$V - tx - p_{1i} = V - t(1-x) - p_2 \implies x = \frac{p_2 - p_{1i} + t}{2t} = \frac{i+l}{k} \implies p_{1i} = p_2 + t - 2t\frac{i+l}{k},$$

with p_2 the price charged by (uninformed) Firm 2. This price is only affected by strategic interactions on the segments where firms compete, and therefore does not depend on the pricing strategy of Firm 1 on type A segments.

We write the profit function for any p_2, replacing p_{1i} and d_1:

$$\pi_1^A(\mathcal{P}) = \frac{l}{k}(t + p_2 - \frac{2(l+i)t}{k}).$$

Secondly, using a similar argument, we show that the profit on $I_1 \cup I_2$ with partition $\mathcal{P}'$ is:

$$\pi_1^{AA}(\mathcal{P}') = \frac{1}{k}(t + p_2 - \frac{2(1+i)t}{k}) + \frac{l-1}{k}(t + p_2 - \frac{2(l+i)t}{k}).$$

Comparing $\mathcal{P}$ and $\mathcal{P}'$ shows that the profit of Firm 1 using the finer partition increases by $\frac{2t}{k^2}(l-1)$, which establishes the claim.

By repeating the previous argument, it is easy to show that the data broker will sell a partition of size $\frac{l}{k}$ with l segments of equal size $\frac{1}{k}$.

Step 2: We show that all segments of type A are closest to Firm 1 (located at 0 ***on the unit line by convention).***

Going from left to right on the Hotelling line, look for the first time where a type B interval, $J = [\frac{i}{k}; \frac{i+l}{k}]$ of length $\frac{l}{k}$, is followed by an interval $I_1 = [\frac{i+l}{k}, \frac{i+l+1}{k}]$ of type A, shown to be of size $\frac{1}{k}$ in step 1. Consider a reordering of the overall interval $J \cup I_1 = [\frac{i}{k}, \frac{i+l+1}{k}]$ in two intervals $I_1' = [\frac{i}{k}; \frac{i+1}{k}]$ and $J' = [\frac{i+1}{k}, \frac{i+l+1}{k}]$. We show in this step that such a transformation increases the profits of Firm 1.

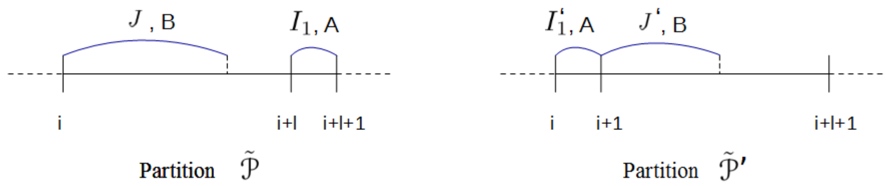

Fig. 2.8: Step 2: relative position of type A and type B segments

The two cases are shown in Figure 2.8 and correspond respectively to the partitions $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{P}}'$. The curved line represents the demand of Firm 1, which does not cover type B segments. In partition $\tilde{\mathcal{P}}$, a segment of type B of size $\frac{l}{k}$, J, is followed by a segment of type A of size $\frac{1}{k}$, I_1. We show that segments of type A are always located closest to Firm 1 by proving that it is always optimal to change partition starting with segments of type B with a partition starting with segments of type A like in partition $\tilde{\mathcal{P}}'$. To show this claim, we compare the profits of the informed firm with J, I_1 under partition $\tilde{\mathcal{P}}$ and with I_1', J' under partition $\tilde{\mathcal{P}}'$, and we show that the latter is always higher than the former. The other segments of the partition remain unchanged.

To compare the profits of the informed firm under both partition, we first characterize type B segments. Segment J of type B is non null (has a size greater than $\frac{1}{k}$), if the following restrictions imposed by the structure of the model are met: respectively positive demand and the existence of competition on segments of type B. In order to characterize type A and type B segments, it is useful to consider the following inequality:

$$\begin{aligned} &\forall \;\; i, l \in \mathbb{N} \;\; s.t. \;\; 0 \leq i \leq k-1 \;\; \text{and} \;\; 1 \leq l \leq k-i-1, \\ &\frac{i}{k} \leq \frac{\tilde{p}_2 + t}{2t} \qquad \text{and} \qquad \frac{\tilde{p}_2 + t}{2t} - \frac{l}{k} \leq \frac{i+l}{k}. \end{aligned} \tag{2.5}$$

In particular, we use the relation that Eq. 2.5 draws between price $\tilde{p}_2$ and segments endpoint $\frac{i}{k}$ and $\frac{i+l}{k}$ to compare the profits of Firm 1 with $\tilde{\mathcal{P}}'$ and with

$\tilde{\mathcal{P}}$.

Without loss of generality, we rewrite the notation of type A and B segments. Segments of type A are of size $\frac{1}{k}$ and are located at $\frac{u_i-1}{k}$, and segments of type B, are located at $\frac{s_i}{k}$ and are of size $\frac{l_i}{k}$.[19] There are $h \in \mathbb{N}$ segments of type A, of size $\frac{1}{k}$, where prices are noted $\tilde{p}_{1i}^A$. On each of these segments, the demand is $\frac{1}{k}$. There are $n \in \mathbb{N}$ segments of type B, where prices are noted $\tilde{p}_{1i}^B$. We find the demand for Firm 1 on these segments using the location of the indifferent consumer:

$$d_{1i} = x - \frac{s_i}{k} = \frac{\tilde{p}_2 - \tilde{p}_{1i}^B + t}{2t} - \frac{s_i}{k}.$$

We can rewrite profits of Firm 1 as the sum of two terms. The first term represents the profits on segments of type A. The second term represents the profits on segments of type B.

$$\pi_1(\tilde{\mathcal{P}}) = \sum_{i=1}^{h} \tilde{p}_{1i}^A \frac{1}{k} + \sum_{i=1}^{n} \tilde{p}_{1i}^B [\frac{\tilde{p}_2 - \tilde{p}_{1i}^B + t}{2t} - \frac{s_i}{k}].$$

Profits of Firm 2 are generated on segments of type B only, where the demand for Firm 2 is:

$$d_{2i} = \frac{s_i + l_i}{k} - x = \frac{\tilde{p}_{1i}^B - \tilde{p}_2 - t}{2t} + \frac{s_i + l_i}{k}.$$

Profits of Firm 2 can be written therefore as:

$$\pi_2(\tilde{\mathcal{P}}) = \sum_{i=1}^{n} \tilde{p}_2 [\frac{\tilde{p}_{1i}^B - \tilde{p}_2 - t}{2t} + \frac{s_i + l_i}{k}]. \tag{2.6}$$

Firm 1 maximizes profits $\pi_1(\tilde{\mathcal{P}})$ with respect to $\tilde{p}_{1i}^A$ and $\tilde{p}_{1i}^B$, and Firm 2 maximizes $\pi_2(\tilde{\mathcal{P}})$ with respect to $\tilde{p}_2$, both profits are strictly concave.

Equilibrium prices are:

[19] With u_i and s_i integers below k. See Section 4.2.2.2.

$$
\begin{aligned}
\tilde{p}_{1i}^{A} &= t + \tilde{p}_2 - 2\frac{u_i t}{k} \\
\tilde{p}_{1i}^{B} &= \frac{\tilde{p}_2 + t}{2} - \frac{s_i t}{k} = \frac{t}{3} + \frac{2t}{3n}[\sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}]] - \frac{s_i t}{k} \\
\tilde{p}_2 &= -\frac{t}{3} + \frac{4t}{3n}\sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}].
\end{aligned} \tag{2.7}
$$

We can now compare profits with $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{P}}'$. When we move segments of type B from the left of segments of type A to the right of segment of type A, it is important to check that Firm 1 is still competing with Firm 2 on each segment of type B, and that Firm 1 is still in constrained monopoly on segments of type A. The second condition is met by the fact that price $\tilde{p}_2$ is higher in $\tilde{\mathcal{P}}'$ than in $\tilde{\mathcal{P}}$. The first condition is guaranteed by Eq. 2.5: $\frac{\tilde{p}_2+t}{2t} - \frac{l_i}{k} \leq \frac{s_i+l_i}{k}$ for some segments located at s_i of size l_i. By abuse of notation, let s_i denote the segment located at $[\frac{s_i}{k}, \frac{s_i+l_i}{k}]$, which corresponds to segments of type B that satisfy these condition. Let $\tilde{s}_i$ denote the m segments ($m \in [0, n-1]$) of type B with partition $\tilde{\mathcal{P}}$ located at $[\frac{\tilde{s}_i}{k}, \frac{\tilde{s}_i+\tilde{l}_i}{k}]$ that do not meet these conditions, and therefore are type A segments with partition $\tilde{\mathcal{P}}'$.

Noting $\tilde{p}_2'$ and $\tilde{p}_{1i}^{B'}$ the prices with $\tilde{\mathcal{P}}'$, we have:

$$
\begin{aligned}
\tilde{p}_2' &= \frac{4t}{3(n-m)}[-\frac{n}{4} + \sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}] + \frac{m}{4} + \frac{1}{2k} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}] \\
&= \tilde{p}_2 + \frac{4t}{3(n-m)}[\frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}],
\end{aligned}
$$

for segments of type B where inequalities in Eq. 2.5 hold:

$$
\tilde{p}_{1i}^{B'} = \tilde{p}_{1i} + \frac{1}{2}\frac{4t}{3(n-m)}[\frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}],
$$

for segments of type B where inequalities in Eq. 2.5 do not hold:

$$
\tilde{p}_{1i}^{B'} = \tilde{p}_{1i} + \frac{1}{2}\frac{4t}{3(n-m)}[\frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}] - \frac{t}{k}.
$$

Let us compare the profits between $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{P}}'$. To compare profits that result by reordering J, I_1 into I_1', J', that is, by moving the segment located at $\frac{i+l}{k}$ to $\frac{i}{k}$ (A to B), we proceed in two steps. First we show that the profits of Firm 1 on $[\frac{i}{k}, \frac{i+l+1}{k}]$ are higher with $\tilde{\mathcal{P}}'$ than with $\tilde{\mathcal{P}}$, and that $\tilde{p}_2$ increases as well; and secondly we show that the profits of Firm 1 on type B segments are higher with $\tilde{\mathcal{P}}'$ than with $\tilde{\mathcal{P}}$.

First we show that the profits of Firm 1 increase on $[\frac{i}{k}, \frac{i+l+1}{k}]$, that is, we show that $\Delta\pi_1 = \pi_1(\tilde{\mathcal{P}}') - \pi_1(\tilde{\mathcal{P}}) \geq 0$:

$$\begin{aligned}\Delta\pi_1 =& \pi_1(\tilde{\mathcal{P}}') - \pi_1(\tilde{\mathcal{P}}) \\ =& \frac{1}{k}[\tilde{p}_2' - 2\frac{it}{k} - \tilde{p}_2 + 2\frac{i+l}{k}t] \\ &+ \tilde{p}_{1i}^{B'}[\frac{\tilde{p}_2' - \tilde{p}_{1i}^{B'} + t}{2t} - \frac{i+1}{k}] - \tilde{p}_{1i}^{B}[\frac{\tilde{p}_2 - \tilde{p}_{1i}^{B} + t}{2t} - \frac{i}{k}].\end{aligned}$$

By definition, $\tilde{s}_i$ verifies the inequalities in Eq. 2.5, thus $\frac{\tilde{s}_i}{k} \leq \frac{\tilde{p}_2+t}{2t}$, which allows us to establish that $\frac{4t}{3(n-m)}[\frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^{m} \frac{\tilde{s}_i}{2k}] \geq \frac{2t}{3nk}$. It is then immediate to show that:

$$\Delta\pi_1 \geq \frac{t}{k}[1 - \frac{1}{3n}][\frac{2}{k}\frac{3nl+1}{3n-1} - \frac{\tilde{p}_2}{2t} - \frac{1}{2} - \frac{1}{6nk} + \frac{i}{k} + \frac{1}{2k}].$$

Also, by assumption, firms compete on $J = [\frac{i}{k}, \frac{i+l}{k}]$ with $\tilde{\mathcal{P}}$, which implies that inequalities in Eq. 2.5 hold, and in particular, $\frac{\tilde{p}_2+t}{4t} - \frac{i}{2k} \leq \frac{l}{k}$.

Thus:

$$\Delta\pi_1 \geq \frac{t}{k}[1 - \frac{1}{3n}][\frac{2}{k}\frac{3nl+1}{3n-1} - \frac{2l}{k} - \frac{1}{6nk} + \frac{1}{2k}] \geq 0.$$

Profits on segment $[\frac{i}{k}, \frac{i+l+1}{k}]$ are higher with $\tilde{\mathcal{P}}'$ than with $\tilde{\mathcal{P}}$.

Second we consider the profits of Firm 1 on the rest of the unit line. We write the reaction functions for the profits on each type of segments, knowing that $\tilde{p}_2' \geq \tilde{p}_2$.

For segments of type A:

$$\frac{\partial}{\partial \tilde{p}_2}\pi_{1i}^{A} = \frac{\partial}{\partial \tilde{p}_2}(\frac{1}{k}[t + \tilde{p}_2 - 2\frac{u_i t}{k}]) = \frac{1}{k},$$

which means that a higher $\tilde{p}_2$ increases the profits.

For segments of type B:

$$\frac{\partial}{\partial \tilde{p}_2}\pi_{1i}^B = \frac{\partial}{\partial \tilde{p}_2}(p_{1i}[\frac{\tilde{p}_2 - \tilde{p}_{1i}^B + t}{2t} - \frac{s_i}{k}]) = \frac{\partial}{\partial \tilde{p}_2}(\frac{1}{2t}[\frac{\tilde{p}_2 + t}{2} - \frac{s_i t}{k}]^2) = \frac{1}{2t}[\frac{\tilde{p}_2 + t}{2} - \frac{s_i t}{k}],$$

which is greater than 0 as $\frac{\tilde{p}_2+t}{2} - \frac{s_i t}{k}$ is the expression of the demand on this segment, which is positive under Eq. 2.5.

Thus for any segment, the profits of Firm 1 increase with $\tilde{\mathcal{P}}'$ compared to $\tilde{\mathcal{P}}$.

Intermediary result 1: *By iteration, we conclude that type A segments are always at the left of type B segments.*

Step 3: We now analyze segments of type B where firms compete. Starting from any partition with at least two segments of type B, we show that it is always more profitable to sell a coarser partition.

As there are only two possible types of segments (A and B) and that we have shown that segments of type A are the closest to the firms, segment B is therefore further away from the firm. We prove the claim of step 3 by showing that if Firm 1 has a partition of two segments where it competes with Firm 2, a coarser partition produces a higher profits. We compute the profits of the firm on all the segments where firms compete, and compare the two situations described below with partition $\hat{\mathcal{P}}$ and partition $\hat{\mathcal{P}}'$.

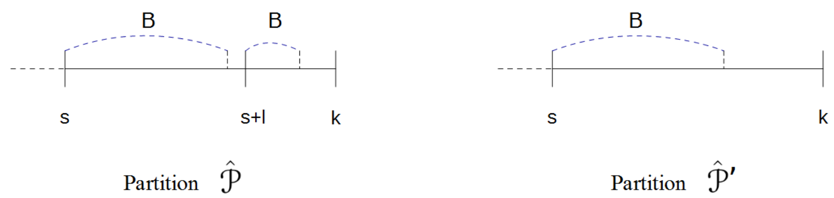

Fig. 2.9: Step 3: demands of Firm 1 on segments of type B (dashed line)

Figure 2.9 depicts partition $\hat{\mathcal{P}}$ on the left panel, and partition $\hat{\mathcal{P}}'$ on the right panel (on each segment the dashed line represents the demand for Firm

1). Partition $\hat{\mathcal{P}}$ divides the interval $[\frac{i}{k}, 1]$ in two segments $[\frac{i}{k}, \frac{i+l}{k}]$ and $[\frac{i+l}{k}, 1]$, whereas $\hat{\mathcal{P}}'$ only includes segment $[\frac{i}{k}, 1]$. We compare the profits of the firm on the segments where firms compete and we show that $\hat{\mathcal{P}}'$ induces higher profits for Firm 1. There are three types of segments to consider:

1. segments of type A that with partition $\hat{\mathcal{P}}$ that remain of type A with partition $\hat{\mathcal{P}}'$.
2. segments of type B with partition $\hat{\mathcal{P}}$ that are of type A with partition $\hat{\mathcal{P}}'$.
3. segments of type B with partition $\hat{\mathcal{P}}$ that remain of type B with partition $\hat{\mathcal{P}}'$.

1. Profits always increase on segments that are of type A with partitions $\hat{\mathcal{P}}$ and $\hat{\mathcal{P}}'$. Indeed, we will show that $\hat{p}_2'$ with partition $\hat{\mathcal{P}}'$ is higher than $\hat{p}_2$ with partition $\hat{\mathcal{P}}$, and thus the profits of Firm 1 on type A segments increase.

2. There are m segments which were type B in partition $\hat{\mathcal{P}}$ are no longer necessarily of type B in partition $\hat{\mathcal{P}}$ (and are therefore of type A).

3. There are $n+1-m$ segments of type B with partition $\hat{\mathcal{P}}$ that remain of type B with partition $\hat{\mathcal{P}}'$. We compute prices and profits on these $n+1+m$ segments.

We proved in step 2 that prices can be written as:

$$\begin{aligned}
\hat{p}_2 &= -\frac{t}{3} + \frac{4t}{3(n+1)} \sum_{i=1}^{n+1} [\frac{s_i}{2k} + \frac{l_i}{k}], \\
\hat{p}_{1i}^B &= \frac{\hat{p}_2 + t}{2} - \frac{s_i t}{k} \\
&= \frac{t}{3} + \frac{2t}{3(n+1)} \sum_{i=1}^{n+1} [\frac{s_i}{2k} + \frac{l_i}{k}] - \frac{s_i t}{k}.
\end{aligned}$$

Let $\hat{p}_{1s}^B$ and $\hat{p}_{1s+l}^B$ be the prices on the last two segments when the partition is $\hat{\mathcal{P}}$.

$$\begin{aligned}
\hat{p}_{1s}^B &= \frac{\hat{p}_2 + t}{2} - \frac{st}{k}, \\
\hat{p}_{1s+l}^B &= \frac{\hat{p}_2 + t}{2} - \frac{s+l}{k} t,
\end{aligned}$$

$\hat{p}_2'$ is the price set by Firm 2 with partition $\hat{\mathcal{P}}'$, and $\hat{p}_{1s}^{B'}$ is the price set by Firm 1 on the last segment of partition $\hat{\mathcal{P}}'$.

Inequalities in Eq. 2.5 might not hold as price $\hat{p}_2$ varies depending on the partition acquired by Firm 1. This implies that segments which are of type B with partition $\hat{\mathcal{P}}$ are then of type A with partition $\hat{\mathcal{P}}'$. This is due to the fact that the coarser the partition, the higher $\hat{p}_2$. We note $\tilde{s}_i$ the m segments where it is the case. We then have:

$$\begin{aligned}
\hat{p}_2' &= \frac{4t}{3(n-m)}[-\frac{n-m}{4} + \sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}] - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}] \\
&= \frac{4t}{3(n-m)}[-\frac{n+1}{4} + \sum_{i=1}^{n+1}[\frac{s_i}{2k} + \frac{l_i}{k}] + \frac{m+1}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k} - \frac{s+l}{2k}] \\
&= \hat{p}_2 + \frac{4t}{3(n-m)}[\frac{3(m+1)\hat{p}_2}{4t} + \frac{m+1}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k} - \frac{s+l}{2k}] \\
&\geq \hat{p}_2 + \frac{4t}{3(n-m)}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}], \\
\hat{p}_{1s}^{B'} &= \frac{\hat{p}_2 + t}{2} - \frac{st}{k},
\end{aligned}$$

$$\begin{aligned}
\pi_1(\hat{\mathcal{P}}) &= \sum_{i=1, s_i \neq \tilde{s}_i}^{n} p_{1i}[\frac{\hat{p}_2 + t}{4t} - \frac{s_i}{2k}] + \sum_{i=1}^{m}\hat{p}_{1i}^{B}[\frac{\hat{p}_2 + t}{4t} - \frac{\tilde{s}_i}{2k}] + \hat{p}_{1s+l}^{B}[\frac{\hat{p}_2 + t}{4t} - \frac{s+l}{2k}] \\
\pi_1(\hat{\mathcal{P}}') &= \sum_{i=1, s_i \neq \tilde{s}_i}^{n} \hat{p}_{1i}^{B'}[\frac{\hat{p}_2' + t}{4t} - \frac{s_i}{2k}] + \sum_{i=1}^{m}\frac{\tilde{l}_i}{k}[\hat{p}_2' + t - 2t\frac{\tilde{s}_i + \tilde{l}_i}{k}].
\end{aligned}$$

We compare the profits of Firm 1 in both cases in order to show that $\hat{\mathcal{P}}'$ induces higher profits:

$$
\begin{aligned}
\Delta\pi_1 &= \pi_1(\hat{\mathcal{P}}') - \pi_1(\hat{\mathcal{P}}) \\
&= \sum_{i=1, s_i \neq \tilde{s}_i}^{n} \hat{p}_{1i}^{B'} [\frac{\hat{p}_2' + t}{4t} - \frac{s_i}{2k}] - \sum_{i=1, s_i \neq \tilde{s}_i}^{n} \hat{p}_{1i}^{B} [\frac{\hat{p}_2 + t}{4t} - \frac{s_i}{2k}] \\
&+ \sum_{i=1}^{m} \frac{\tilde{l}_i}{k} [\hat{p}_2' + t - 2t\frac{\tilde{s}_i + \tilde{l}_i}{k}] - \sum_{i=1}^{m} \hat{p}_{1i}^{B} [\frac{\hat{p}_2 + t}{4t} - \frac{\tilde{s}_i}{2k}] - \hat{p}_{1s+l}^{B} [\frac{\hat{p}_2 + t}{4t} - \frac{s+l}{2k}] \\
&= \frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^{n} [\frac{\hat{p}_2' + t}{2t} - \frac{s_i}{k}]^2 - \frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^{n} [\frac{\hat{p}_2 + t}{2t} - \frac{s_i}{k}]^2 \\
&+ \frac{t}{2} \sum_{i=1}^{m} \frac{\tilde{l}_i}{k} [2\frac{\hat{p}_2' + t}{t} - 4\frac{\tilde{s}_i + \tilde{l}_i}{k}] - \frac{t}{2} \sum_{i=1}^{m} [\frac{\hat{p}_2 + t}{2t} - \frac{\tilde{s}_i}{2k}]^2 - \frac{t}{2} [\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k}]^2 .
\end{aligned}
$$

We consider the terms separately. First,

$$
\begin{aligned}
& \frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^{n} [\frac{\hat{p}_2' + t}{2t} - \frac{s_i}{k}]^2 - \frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^{n} [\frac{\hat{p}_2 + t}{2t} - \frac{s_i}{k}]^2 \\
=& \frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^{n} [[\frac{2}{3(n-m)} [\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}]]^2 \\
& + [\frac{\hat{p}_2 + t}{2t} - \frac{s_i}{k}][\frac{4}{3(n-m)} [\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}]]] \\
\geq & \frac{t}{2} [\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k}] \frac{4}{3} [\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}].
\end{aligned}
$$

Second, on segments of type B with partition $\hat{\mathcal{P}}$ that are of type A with partition $\hat{\mathcal{P}}'$:

$$
\frac{t}{2} \sum_{i=1}^{m} \frac{\tilde{l}_i}{k} [2\frac{\hat{p}_2' + t}{t} - 4\frac{\tilde{s}_i + \tilde{l}_i}{k}] - \frac{t}{2} \sum_{i=1}^{m} [\frac{\hat{p}_2 + t}{2t} - \frac{\tilde{s}_i}{2k}]^2 .
$$

On these m segments, inequalities in Eq. 2.5 hold for price $\hat{p}_2'$ but not for $\hat{p}_2$. Thus we can rank prices according to $\tilde{s}_i$ and $\tilde{l}_i$:

$$
\frac{\tilde{s}_i + \tilde{l}_i}{k} \geq \frac{\hat{p}_2 + t}{2t} - \frac{\tilde{l}_i}{k} \quad \text{and} \quad \frac{\hat{p}_2' + t}{2t} - \frac{\tilde{l}_i}{k} \geq \frac{\tilde{s}_i + \tilde{l}_i}{k}.
$$

thus:

$$
2\frac{\tilde{l}_i}{k} \geq \frac{\hat{p}_2 + t}{2t} - \frac{\tilde{s}_i}{k} \quad \text{and} \quad \frac{\hat{p}_2' + t}{2t} - 2\frac{\tilde{l}_i}{k} \geq \frac{\tilde{s}_i}{k}.
$$

By replacing $\tilde{s}_i$ by its upper bound value and then $\tilde{l}_i$ by its lower bound value we obtain:

$$\frac{t}{2}\sum_{i=1}^{m}\frac{\tilde{l}_i}{k}[2\frac{\hat{p}_2'+t}{t}-4\frac{\tilde{s}_i+\tilde{l}_i}{k}]-\frac{t}{2}\sum_{i=1}^{m}[\frac{\hat{p}_2+t}{2t}-\frac{\tilde{s}_i}{2k}]^2\geq 0.$$

Getting back to the profits difference, we obtain:

$$\begin{aligned}\Delta\pi_1 &\geq \frac{t}{2}[\frac{\hat{p}_2+t}{2t}-\frac{s+l}{k}]\frac{4}{3}[\frac{3}{4t}\hat{p}_2+\frac{m\hat{p}_2}{2t}+\frac{1}{4}-\frac{s+l}{2k}]-\frac{t}{2}[\frac{\hat{p}_2+t}{2t}-\frac{s+l}{k}]^2\\ &\geq \frac{t}{2}[\frac{\hat{p}_2+t}{2t}-\frac{s+l}{k}][\frac{\hat{p}_2}{2t}+\frac{s+l}{3k}-\frac{1}{6}].\end{aligned} \tag{2.8}$$

The first bracket of Equation 2.8 is positive given Eq. 2.5. The second bracket is positive if $\frac{\hat{p}_2}{2t}+\frac{s+l}{3k}\geq\frac{1}{6}$. A necessary condition for this result to hold is $\hat{p}_2\geq\frac{1}{6}$. We now show that $\hat{p}_2\geq\frac{t}{2}$

We show in Equation 4.12 that $\hat{p}_2=-\frac{t}{3}+\frac{4t}{3(n+1)}\sum_{i=1}^{n+1}[\frac{s_i}{2k}+\frac{l_i}{k}]$. We now show that p_2 is minimal when the data broker sells the reference partition $\mathcal{P}_{ref}$ to Firm 1, which consists of segments of size $\frac{1}{k}$. Indeed, it is immediate to see that, p_2 always decreases when $\mathcal{P}$ becomes finer. It is thus immediate that p_2 is minimal with the reference partition and $p_2\geq\frac{t}{2}$[20]. And as this price is greater than $\frac{1}{6}$, the second bracket of Equation 2.8 is positive. This proves that $\Delta\pi_1\geq 0$.

We have just established that it is always more profitable for the data broker to sell a partition with one segment of type B than to sell a partition with several segments of type B.

The profits of Firm 1 are minimized when Firm 2 acquires $\mathcal{P}_{ref}$.

This claim is straightforward to establish, as we have shown in step 3 that the price set by an uninformed Firm is minimized when its competitor acquires the reference partition. Thus, demand and profit are also minimized for this partition and the data broker sells $\mathcal{P}_{ref}$ to Firm 2.

Conclusion

[20] As shown in Liu and Serfes (2004).

These three steps prove that the optimal partition includes two intervals, as illustrated in Figure 3.2. The first interval is composed of j segments of size $\frac{1}{k}$ located at $[0, \frac{j}{k}]$, and the second interval is composed of unidentified consumers, and is located at $[\frac{j}{k}, 1]$.

Proof of Theorem 2: the data broker sells symmetric information to both firms

Part a: optimal information structure when the data broker sells information to both firms

We prove that the partition described in Theorem 2 is optimal when information is sold to both firms. For each firm, the partition divides the unit line into two intervals. The first interval identifies the closest consumers to a firm and is partitioned in j segments of size $\frac{1}{k}$. The second interval is of size $1 - \frac{j}{k}$ and leaves unidentified the other consumers.

Three types of segments are defined as before:

- Segments A, where Firm θ is in constrained monopoly;
- Segments B, where Firms 1 and 2 compete;
- Segments C, where Firm θ gets no demand.

We use Assumption 1 that states that the unit line is composed of one interval where firms compete, located at the middle of the line. As we will show, the optimal partition under this assumption is similar to the optimal partition when the data broker sells information to one firm.

Inequalities in Eq. 2.5 characterize segments $[\frac{s_i}{k}, \frac{s_{i+1}}{k}]$ where both firms have positive demand:

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

The first part of Eq. 2.5 guarantees that there is positive demand for Firm 1, whereas the second part guarantees positive demand for Firm 2. Inequalities in Eq. 2.5 are expressed as a function of p_2 without loss of generality. We use Eq. 2.5 to characterize type A and type B segments, in order to compute the profits of the firms.

The profits of the data broker when it sells information to both firms is the difference between the profits of the firms when they are informed and their outside option, where they do not have information, but their competitor is informed:

$$\Pi_2 = (\pi_1^{I,I}(\mathcal{P}_1, \mathcal{P}_2) - \pi_1^{NI,I}(\emptyset, \mathcal{P}_2)) + (\pi_2^{I,I}(\mathcal{P}_1, \mathcal{P}_2) - \pi_2^{NI,I}(\emptyset, \mathcal{P}_1)).$$

Firm θ buys a partition composed of segments of type A and one segment of type B. To show that a partition in which type A segments are of size $\frac{1}{k}$ is optimal, we prove that 1) such a partition maximizes $\pi_{\mathcal{P},\theta}^{I,I}$ and 2) such a partition does not change $\pi_{\mathcal{P},\theta}^{NI,I}$.

1) *A partition which maximizes* $\pi_{\mathcal{P},\theta}^{I,I}$ *is necessarily composed of type A segments of size* $\frac{1}{k}$.

The proof of this claim is similar to step 1 of the proof of Theorem 1 in Appendix A.1 the price of the competing firm $-\theta$ does not change when Firm θ gets more precise information on type A segments, and the profits of Firm θ increase as it can target more precisely consumers with this information.

2) *Changing from a partition with type A segments of arbitrary size to a partition where type A segments are of size* $\frac{1}{k}$ *does not change* $\pi_{\mathcal{P},\theta}^{NI,I}$.

It is immediate to show that the profit of the uninformed firm does not depend on the fineness of type A segments. As a result, Π_2 is maximized when segments of type A are of size $\frac{1}{k}$.

We conclude that the optimal partition is composed of two intervals, sold to each firm. For Firm 1, the first interval is partitioned in j segments of size $\frac{1}{k}$, and is located at $[0, \frac{j}{k}]$. Consumers are unidentified on the second interval of size $1-\frac{j}{k}$ located at $[\frac{j}{k},1]$. For Firm 2, the first interval is partitioned in j' segments of size $\frac{1}{k}$, and is located at $[1-\frac{j'}{k},1]$. Consumers are unidentified on the second interval of size $1-\frac{j'}{k}$ located at $[0,1-\frac{j'}{k}]$.

Part b: the data broker sells symmetric information to both firms

We show now that selling symmetric information is optimal for the data broker, that is, in equilibrium $j=j'$.

We compute prices and profits in equilibrium when both firms are informed with the optimal partition found above.

Firm 1 is a monopolist on the j segments of size $\frac{1}{k}$ in $[0,\frac{j}{k}]$ and Firm 2 has information on $[1-\frac{j'}{k},1]$. On $[\frac{j}{k},1]$ Firm 1 sets a unique price p_1 and gets demand d_1, similarly on $[0,1-\frac{j'}{k}]$ Firm 2 sets a unique price p_2 and gets demand d_2.

We write in step 1 prices and demands, in step 2 we give the profits, and solve for prices and profits in equilibrium in step 3.

Step 1: prices and demands.

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

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

p_2 is the price set by Firm 2 on interval $[0,\frac{j'}{k}]$ where it cannot identify consumers. Prices set by Firm 2 on segments in interval $[\frac{j'}{k},1]$ are:

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

Let denote d_1 the demand for Firm 1 (resp. d_2 the demand for Firm 2) where firms compete. d_1 is found in a similar way as when information is sold to one firm, which gives us $d_1 = \frac{p_2-p_1+t}{2t} - \frac{j}{k}$ (resp. $d_2 = 1 - \frac{j'}{k} - \frac{p_2-p_1+t}{2t}$).

Step 2: profits of the firms.

The profits of the firms are:

$$\pi_1 = \sum_{i=1}^{j} d_{1i}p_{1i} + d_1p_1 = \sum_{i=1}^{j} \frac{1}{k}(p_2 + t - 2t\frac{i}{k}) + (\frac{p_2 - p_1 + t}{2t} - \frac{j}{k})p_1,$$
$$\pi_2 = \sum_{i=1}^{j'} d_{2i}p_{2i} + d_2p_2 = \sum_{i=1}^{j} \frac{1}{k}(p_1 + t - 2t\frac{i}{k}) + (\frac{p_1 - p_2 + t}{2t} - \frac{j'}{k})p_2.$$

Step 3: prices, demands and profits 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:

$$p_1 = t[1 - \frac{2}{3}\frac{j'}{k} - \frac{4}{3}\frac{j}{k}],$$
$$p_2 = t[1 - \frac{2}{3}\frac{j}{k} - \frac{4}{3}\frac{j'}{k}].$$

Replacing these values in the above demands and prices gives:

$$p_{1i} = 2t - \frac{4}{3}\frac{j't}{k} - \frac{2}{3}\frac{jt}{k} - 2\frac{it}{k},$$
$$p_{2i} = 2t - \frac{4}{3}\frac{jt}{k} - \frac{2}{3}\frac{j't}{k} - 2\frac{it}{k}.$$

and

$$d_1 = \frac{1}{2} - \frac{2}{3}\frac{j}{k} - \frac{1}{3}\frac{j'}{k},$$
$$d_2 = \frac{4}{3}\frac{j'}{k} - \frac{1}{2} - \frac{1}{3}\frac{j}{k}.$$

Profits are:

$$\pi_1^* = \sum_{i=1}^{j} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k} - \frac{2}{3}\frac{j'}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{j}{k} - \frac{1}{3}\frac{j'}{k})t[1 - \frac{2}{3}\frac{j'}{k} - \frac{4}{3}\frac{j}{k}]$$
$$= \frac{t}{2} - \frac{7}{9}\frac{j^2 t}{k^2} + \frac{2}{9}\frac{j'^2 t}{k^2} - \frac{4}{9}\frac{jj't}{k^2} + \frac{2}{3}\frac{jt}{k} - \frac{2}{3}\frac{j't}{k} - \frac{jt}{k^2}.$$

$$\pi_2^* = \sum_{i=1}^{j'} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j'}{k} - \frac{2}{3}\frac{j}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{j'}{k} - \frac{1}{3}\frac{j}{k})t[1 - \frac{2}{3}\frac{j}{k} - \frac{4}{3}\frac{j'}{k}]$$
$$= \frac{t}{2} - \frac{7}{9}\frac{j'^2 t}{k^2} + \frac{2}{9}\frac{j^2 t}{k^2} - \frac{4}{9}\frac{jj't}{k^2} + \frac{2}{3}\frac{j't}{k} - \frac{2}{3}\frac{jt}{k} - \frac{j't}{k^2}.$$

The data broker maximizes the following profit function:

$$\Pi_2(j,j') = (\pi_1^{I,I}(j,j') - \pi_1^{NI,I}(\emptyset,j')) + (\pi_2^{I,I}(j,j') - \pi_2^{NI,I}(\emptyset,j))$$
$$= -\frac{7}{9}\frac{j'^2 t}{k^2} - \frac{4}{9}\frac{jj't}{k^2} + \frac{2}{3}\frac{j't}{k} - \frac{j't}{k^2} - \frac{7}{9}\frac{j^2 t}{k^2} - \frac{4}{9}\frac{jj't}{k^2} + \frac{2}{3}\frac{jt}{k} - \frac{jt}{k^2}.$$

At this stage, straightforward FOCs with respect to j and j' confirm that, in equilibrium, $j = j'$. The fact that the solution is a maximum is directly found using the determinant of the Hessian matrix.

Proof of Lemma 1

We compute prices and profits in equilibrium when information is sold to one firm. Without loss of generality we consider the situation where Firm 1 is informed only. We consider the optimal partition found in Appendix A.1.

Firm 1 owns a partition of $[0, \frac{j}{k}]$ that includes j segments of size $\frac{1}{k}$, and has no information on consumers on $[\frac{j}{k}, 1]$. Again, firms face three types of segments, A, B, and C defined in Appendix A.1.

We write in step 1 prices and demands, in step 2 we give the profits, and solve for prices and profits in equilibrium in step 3.

Step 1: prices and demands.

Type A segments are of size $\frac{1}{k}$, and the last one is located at $\frac{j-1}{k}$. Firm 1 sets a price p_{1i} for each segment $i = 1,..,j$ and where it is in constrained monopoly: $d_{1i} = \frac{1}{k}$. Prices on each segment are determined by the indifferent consumer of each segment located at its right extremity, $\frac{i}{k}$:[21]

$V - t\frac{i}{k} - p_{1i} = V - t(1-\frac{i}{k}) - p_2 \implies \frac{i}{k} = \frac{p_2 - p_{1i}+t}{2t} \implies p_{1i} = p_2 + t - 2t\frac{i}{k}$.

The rest of the unit line is a type B segment. Firm 1 sets a price p_1 and competes with Firm 2. Firm 2 sets a unique price p_2 for all consumers on the segment $[0,1]$. We note d_1 the demand for Firm 1 on this segment, which is determined by the indifferent consumer:

$V - tx - p_1 = V - t(1-x) - p_2 \implies x = \frac{p_2 - p_1 + t}{2t}$ and $d_1 = x - \frac{j}{k} = \frac{p_2 - p_1 + t}{2t} - \frac{j}{k}$.

Firm 2 sets p_2 and the demand, d_2, is found similarly to d_1, and $d_2 = 1 - \frac{p_2 - p_1 + t}{2t} = \frac{p_1 - p_2 + t}{2t}$.

Step 2: profits.

The profits of both firms can be written as follows:

$$\pi_1 = \sum_{i=1}^{j} d_{1i}p_{1i} + d_1 p_1 = \sum_{i=1}^{j} \frac{1}{k}(p_2 + t - 2t\frac{i}{k}) + (\frac{p_2 - p_1 + t}{2t} - \frac{j}{k})p_1,$$
$$\pi_2 = d_2 p_2 = \frac{p_1 - p_2 + t}{2t} p_2.$$

Step 3: prices, demands and profits in equilibrium.

We solve prices and profits in equilibrium. First order conditions on π_θ with respect to p_θ give us $p_1 = t[1 - \frac{4}{3}\frac{j}{k}]$ and $p_2 = t[1 - \frac{2}{3}\frac{j}{k}]$. By replacing these values in profits and demands we deduce that: $p_{1i} = 2t[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k}]$, $d_1 = \frac{1}{2} - \frac{2}{3}\frac{j}{k}$ and $d_2 = \frac{1}{2} - \frac{1}{3}\frac{j}{k}$.

Profits are:[22]

[21] Assume it is not the case. Then, either p_{1i} is higher and the indifferent consumer is at the left of $\frac{i}{k}$, which is in contradiction with the fact that we deal with type A segments, or p_{1i} is lower and as the demand remain constant, the profits are not maximized.

[22] For $p_{1i} \geq 0 \implies \frac{j}{k} \leq \frac{3}{4}$. Profits are equal whatever $\frac{j}{k} \geq \frac{3}{4}$.

$$\begin{aligned}\pi_1^* &= \sum_{i=1}^{j} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k}] + \frac{t}{2}(1 - \frac{4}{3}\frac{j}{k})^2 \\ &= \frac{t}{2} + \frac{2jt}{3k} - \frac{7t}{9}\frac{j^2}{k^2} - \frac{tj}{k^2} \\ \pi_2^* &= \frac{t}{2} + \frac{2t}{9}\frac{j^2}{k^2} - \frac{2}{3}\frac{jt}{k}.\end{aligned} \tag{2.9}$$

Proof of Lemma 2

We compute prices and profits in equilibrium when both firms are symmetrically informed, with the optimal partition found in Appendix A.2.

Firm 1 is a monopolist on the j segments of size $\frac{1}{k}$ in $[0, \frac{j}{k}]$ and Firm 2 has symmetric information, composed of j segments of size $\frac{1}{k}$ on $[1-\frac{j}{k}, 1]$. On $[\frac{j}{k}, 1]$ Firm 1 sets a unique price p_1 and gets demand d_1, similarly on $[0, 1-\frac{j}{k}]$ Firm 2 sets a unique price p_2 and gets demand d_2.

We do not go through the computation of prices and demand which are already described in Appendix A.2, and we directly give prices and profits in equilibrium.

Prices in equilibrium are $p_1 = p_2 = t[1 - 2\frac{j}{k}]$, $p_{\theta i} = 2t[1 - \frac{j}{k} - \frac{i}{k}]$ and $d_\theta = \frac{1}{2} - \frac{j}{k}$.

Profits are:[23]

$$\begin{aligned}\pi_\theta^* &= \sum_{i=1}^{j} \frac{2t}{k}[1 - \frac{i}{k} - \frac{j}{k}] + \frac{1}{2}(1 - 2\frac{j}{k})^2 t \\ &= \frac{t}{2} - \frac{j^2}{k^2}t - \frac{jt}{k^2}.\end{aligned}$$

Proof of Proposition 1

In this section we assume that j is continuous. The optimal value of j will be the integer closest to the optimum found in the continuous case.

[23] For $\frac{j}{k} < \frac{1}{2}$. Profits are equal as soon as $\frac{j}{k} > \frac{1}{2}$.

Using the profits from Lemma 3, we determine the optimal size j_1^* of the segments of type A when the data broker only sells information to Firm 1, by maximizing profits with respect to j. When the data broker sells information to both firms, we determine the optimal number j_2^* of type A segments in a similar way. We then compare the maximized profits of the data broker to find whether it sells information to one or to both firms in equilibrium.

1) Optimal partition when the data broker sells information to one firm.

The profits of the data broker when it sells to one firm are:[24]

$$\begin{aligned}\Pi_1(j) = w_1(j) &= \pi^{I,NI}(j,\emptyset) - \pi^{NI,I}(\emptyset,\mathcal{P}_{ref}) \\ &= \frac{3t}{8} + \frac{2jt}{3k} - \frac{t}{4k} - \frac{7j^2t}{9k^2} - \frac{jt}{k^2} - \frac{t}{8k^2}.\end{aligned}$$

FOC on j leads to the following maximizing value: $j_1^* = \frac{6k-9}{14}$ and:

$$\Pi_1^* = \frac{29t}{56} - \frac{19t}{28k} + \frac{11t}{56k^2}.$$

2) Optimal partition when the data broker sells information to both firms.

We maximize the profit function with respect to the j segments sold to Firm 1 and Firm 2. The profits of the data broker when both firms are informed are:

$$\Pi_2(j) = 2w_2 = 2[\frac{2jt}{3k} - \frac{11j^2t}{9k^2} - \frac{jt}{k^2}].$$

FOC on j leads to $j_2^* = \frac{6k-9}{22}$ and:

$$\Pi_2^* = \frac{2t}{11} - \frac{6t}{11k} + \frac{9t}{22k^2}.$$

3) Data broker's selling strategy in equilibrium.

[24] The expression of $\pi^{NI,I}(\emptyset,\mathcal{P}_{ref})$ is provided in Liu and Serfes (2004).

We compare the profits of the data broker when it sells information to one firm or to both firms. The difference between the two profits is:

$$\Pi_1(j_1^*) - \Pi_2(j_2^*) = \frac{(207k^2 - 82k - 131)t}{616k^2}.$$

which is positive for any $k \geq 2$.

Proof of Proposition 2

The profits of the firms without information are similar to the standard Hotelling model and are:

$$\pi^{*NI,NI} = \frac{t}{2}.$$

The profits of the firms depending on whether they acquire information can be derived immediately by replacing j by its optimal values in the different scenarios. We obtain:

$$\pi^{*I,NI}(j_1^*, \emptyset) = \frac{(29k^2 - 38k + 11)t}{56k^2}.$$

$$\pi^{*I,I}(j_2^*, j_2^*) = \frac{(206k^2 - 24k + 117)t}{484k^2}.$$

$$\pi^{NI,I}(\emptyset, \mathcal{P}_{ref}) = \frac{(k^2 + 2k + 1)t}{8k^2}.$$

Proposition 2 stems immediately from the above results.

Proof of Proposition 4

We generalize the results to first-degree price discrimination, and show that profits and the optimal structure correspond to the limit of the profits under third-degree price discrimination, when $k \to \infty$.

We prove that the optimal structure when firms first-degree price discriminate is identical to the structure when firms third-degree price discriminate. We first characterize the information structure under first-degree price discrimination, then we determine the optimal partition.

When a firm first-degree price discriminates, for instance on a segment $[\frac{l_1}{k}, \frac{l_2}{k}]$ with $l_1 \leq l_2$ integers lower than k, two types of segments are defined. On type A' segments, the firm sets a personalized price for each consumer, here $[\frac{l_1}{k}, \frac{l_2}{k}]$. On type B' segments, the firm sets a homogeneous price on each segment, here a price p_1 on $[0, \frac{l_1}{k}]$ and a price p_2 on $[\frac{l_2}{k}, 1]$. If there are n segments of type B', then the firm sets n prices $p_1, .., p_n$, one on each of these segments.

The optimal partition is composed of two intervals: on $[0, l]$ ($l \in [0, 1]$) consumers are perfectly identified, and on $[l, 1]$, consumers are unidentified. The proof of this result is not detailed here, as it is similar to the proof of Theorem 1 in Appendix A.1.

Step 1: Profits under third-degree price discrimination converge to profits under first-degree price discrimination

It remains to show that on the first interval $[0, l]$, profits under third-degree price discrimination converge to profits under first-degree price discrimination when $k \to \infty$, and to find the optimal size of these segments.

First we write the profits of Firm 1 under first-degree price discrimination, then we show that when $k \to \infty$, profits under third-degree price discrimination converge to profits under first-degree price discrimination. In the next section, we find the optimal length of the segment of identified consumers under first-degree price discrimination.

Profits of Firm 1 under first-degree price discrimination.

Let l denote the size of the interval of identified consumers under first-degree price discrimination. We want to compare profits for identical partitions, that is for which $l = \lim_{k\to\infty} \frac{j(k)}{k}$. Under first-degree price discrimination, Firm 1 sets personalized prices on $[0, l]$, and a single price on $[l, 1]$. Firm 2 sets a single price on the unit line: $p_2 = t - \frac{2}{3}l$ (similarly to Lemma 1).

$$\pi_1^{FD} = \int_0^l p_1(x)\mathrm{d}x + \frac{t}{2}(1 - \frac{4}{3}l)^2.$$

$p_1(x)$ verifies $V - tx - p_1(x) = V - t(1-x) - p_2 \implies p_1(x) = 2t[1 - x - \frac{1}{3}l]$.

We thus have:

$$\pi_1^{FD} = \int_0^l 2t[1 - x - \frac{1}{3}l]\mathrm{d}x + \frac{t}{2}(1 - \frac{4}{3}l)^2.$$

Third-degree price discrimination profits converge to first-degree price discrimination profits.

Starting from Equation 3.1, we want to prove that the sum $\sum_{i=1}^{lk} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j(k)}{k}]$ converges to profits of first-degree price discrimination when $k \to \infty$, that is:

$$\lim_{k\to\infty} \sum_{i=1}^{lk} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j(k)}{k}] = \int_0^l 2t[1 - x - \frac{1}{3}l]\mathrm{d}x.$$

Let $f(i) = \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j(k)}{k}]$. It is immediate to see that f is decreasing and continuous on $[0, \infty[$, we can thus write: $\int_{i-1}^{i} f(z)\mathrm{d}z \geq f(i) \geq \int_i^{i+1} f(z)\mathrm{d}z$.

Summing each term from 1 to lk we get: $\int_0^{lk} f(z)\mathrm{d}z \geq \sum_{i=1}^{lk} f(i) \geq \int_1^{lk+1} f(z)\mathrm{d}z$.

We have $\int_1^{lk+1} f(z)\mathrm{d}z = \int_0^{lk} f(z)\mathrm{d}z + \int_{lk}^{lk+1} f(z)\mathrm{d}z - \int_0^1 f(z)\mathrm{d}z$.

$$\begin{cases} \lim_{k\to\infty} \int_{lk}^{lk+1} f(z)\mathrm{d}z = \lim_{k\to\infty} \int_{lk}^{lk+1} \frac{2t}{k}[1 - \frac{z}{k} - \frac{1}{3}\frac{j(k)}{k}]\mathrm{d}z = 0. \\ \\ \lim_{k\to\infty} \int_0^1 f(z)\mathrm{d}z = \lim_{k\to\infty} \int_0^1 \frac{2t}{k}[1 - \frac{z}{k} - \frac{1}{3}\frac{j(k)}{k}]\mathrm{d}z = 0. \end{cases}$$

Thus we have: $\lim_{k\to\infty} \int_0^{lk} f(z)\mathrm{d}z \geq \lim_{k\to\infty} \sum_{i=1}^{lk} f(i) \geq \lim_{k\to\infty} \int_0^j f(z)\mathrm{d}z$.

By the sandwich theorem we have : $\lim_{k\to\infty} \sum_{i=1}^{lk} f(i) = \lim_{k\to\infty} \int_0^{lk} f(z)\mathrm{d}z = \int_0^l 2t[1 - x - \frac{1}{3}l]\mathrm{d}x$ the last equality is immediate by substitution. Profits under third-degree price discrimination converge to profits under first-degree price discrimination when $k \to \infty$ (thus when quality $\frac{1}{k} \to 0$).

It is straightforward to establish the same result when the data broker sells information to both firms.

Step 2: Optimal size of the interval of identified consumers.

We compute the profits of Firm 1 when the data broker does not sell information to Firm 2, and Firm 1 has information that allows it to first-degree price discriminate. We find the following profits:

$$\pi_1^{FD;I,NI} = \int_0^l 2t[1 - x - \frac{l}{3}]\mathrm{d}x + \frac{t}{2}(1 - \frac{4}{3}l)^2 = \frac{t}{2} + \frac{2lt}{3} - \frac{7}{9}l^2 t.$$

The profits of Firm 1 when only Firm 2 is informed with the reference partition are, similarly to the third-degree price discrimination case:

$$\pi_1^{FD;NI,I} = \frac{t}{8}.$$

The profits of the data broker are then: $\Pi_2 = \frac{3t}{8} + \frac{2}{3}lt - \frac{7}{9}l^2 t$, maximized with $l^* = \frac{3}{7}$.

CHAPTER 3

Selling Mechanisms and the Market for Consumer Information

Abstract

We investigate the strategies of a data intermediary selling consumer information to firms for price discrimination purpose. We analyze how the mechanism through which information is sold influences how much consumer data is collected and sold by an intermediary, and how it impacts consumer surplus. We consider three selling mechanisms: take it or leave it offers, sequential bargaining, and auctions. We show that the more the data intermediary collects information, the lower consumer surplus. Consumer surplus is maximized in take it or leave it offers, which is the least profitable mechanism for the intermediary. Our result show that selling mechanisms can be used as a regulatory tool by data protection agencies and competition authorities to limit consumer information collection and increase consumer surplus.

3.1 Introduction

Since the seminal works of Hayek (1945) and Marschak (1974), scholars and policy makers have acknowledged that information greatly enhances the efficiency of markets. A new market for information is emerging: data intermediaries supply consumer information that is purchased by firms to improve their business practices. Companies of a new type - data intermediaries - have specialized in collecting data from different sources, and selling customized datasets to firms.[1] The emergence of this new market for data intermediation raises several policy concerns.

First, data intermediaries have become major actors of the economy, up to a point where, in 2014, the market for consumer data was valued around USD 156 billion per year (Pasquale, 2015). Recent scandals of data breaches and violation of consumer privacy have revealed the huge amount of information possessed by data intermediaries.[2] For instance, in a study of nine data brokers from 2014, 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." (Federal Trade Commission, 2014, Data brokers: A Call for Transparency and Accountability). The sheer volume of personal data collected by these data intermediaries can raise concerns for data protection agencies. Indeed, new regulations try to limit the amount of personal data collected by data intermediaries. The California Consumer Privacy Act provides a detailed list of safeguards to protect personal data. Similarly, a (personal) data minimization principle is enacted in the Health

[1] For instance, data brokers such as Equifax or Transunion, sell specific consumer segments to firms willing to personalize their advertising campaigns.

[2] Huge data breach reveals hundreds of millions of emails and passwords from across the Internet.

Insurance Portability and Accountability Act in the US, and in the EU General Data Protection Regulation in Europe.

Secondly, market practices have revealed that data intermediaries have a significant role in shaping competition, which can cause important harms to other companies and to consumer welfare. For instance, Facebook offered companies such as Netflix, Lyft, or Airbnb special access to data, while denying its access to other companies such as Vine.[3] There is a risk that more precise consumer information could lead to more consumer surplus extraction and to increased market power in the data intermediaries' industry.[4] There thus is a pressing need to analyze the strategies of data intermediaries.

The emergence of data intermediaries raises two questions that have been studied in the literature: what is the impact of the selling mechanism on market equilibrium, and for a given selling mechanism, how to sell consumer information. Economists have for long acknowledged that the selling mechanism plays a central role on the organization of markets (Riley and Zeckhauser, 1983). In traditional markets, the economic literature has clearly shown that equilibrium prices and quantities of a product change with the selling mechanism used in the market. This literature has mainly focused on take it or leave it offers (Binmore et al., 1986), auctions (Vickrey, 1961; Jehiel and Moldovanu, 2000), and sequential bargaining (Rubinstein, 1982; Sobel and Takahashi, 1983). Recent trends of literature reconsider the question of how to sell a good, especially on digital markets and online platforms. They study similar selling mechanisms. Bajari et al. (2008) compares auction with negotiations when selling customized products. Jindal and Newberry (2018) study in which case it is optimal for a seller to use bargaining or fixed price to sell a good. Backus et al. (2018a), Backus et al. (2018b) and Backus et al. (2019) empirically analyze patterns of bargaining on e-bay and how bargaining environments are affected by informa-

[3] Facebook gave Lyft and others special access to user data; engadget, May 12th, 2018.

[4] See the recent debate on a potential breakup of major data intermediaries. (Is Big Tech Too Big Or Not Big Enough?; Forbes, June 20th, 2019.; Warren Wants To Break Up Amazon, Facebook, Google; Forbes, March 8th, 2019.)

tion asymmetries, bargaining power, and private characteristics on the buyer and on the seller's side. Milgrom and Tadelis (2018) study how machine learning is used to improve mechanism design, in particular to set reserve prices optimally in auctions.

For a given selling mechanism, a recent literature has analyzed the selling strategies of a data intermediary. Montes et al. (2018) consider information allowing competing firms to first-degree price discriminate consumers. Bounie et al. (2018) endogenize the amount of information sold to firms, and show that it is not optimal for a data intermediary to sell all consumer segments to firms. Bergemann et al. (2018) studies a situation of information asymmetries where the willingness to pay of the buyer for information is unknown to the data intermdiary. The data intermediary proposes a menu of signals to information buyers. However, the existing literature on data intermediary only focuses on a single selling mechanism: Montes et al. (2018), Bounie et al. (2018) only consider auction; Bergemann et al. (2018) focuses on take it or leave it offers.

In this article, we merge the literature on selling mechanisms with the recent literature on data intermediaries who sell information to firms competing on the product market. We compare the strategies of a data intermediary under three selling mechanisms: a take it or leave it offer, a sequential bargaining, and an auction. We contribute to the existing literature on two points.

We first show that the three selling mechanisms share common properties, what we will call, independent data contracts. The key element that will determine how much information will be collected and sold in equilibrium is the threat for a firm of being uninformed. They will determine the willingness to pay for information, and thus the price of information. This conclusion could not be reached in models where the data intermediary is assumed to sell no information at all, or all consumer segments.[5]

Secondly, by comparing different selling mechanisms, we show that the take

[5] See for instance Montes et al. (2018).

it or leave it offer maximizes consumer surplus, and minimizes data collection, but that the data intermediary would prefer auctions or sequential bargaining that lead to lower consumer surplus and more data collection. We discuss the regulatory implications in the conclusion.

The reminder of the article is organized as follows. We explicit the model in Section 3.2. Section 3.3 describes the three selling mechanisms. We solve the game in Sections 3.4 and 3.5. We discuss alternative selling mechanisms in Section 3.6. Section 3.7 concludes.

3.2 Model

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. The data intermediary collects and sells data on consumer segments. An informed firm can set a price on each consumer segment. An uninformed firm cannot distinguish consumer segments and sets a single price on the entire line.

3.2.1 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:[6]

[6] 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. (2018) or Montes et al. (2018).

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

This simple model of horizontal differentiation can be used to analyze the impact of information acquisition on the profits of firms (Thisse and Vives, 1988).

3.2.2 Data intermediary

The data intermediary collects information on consumers that allows firms to distinguish k consumer segments on the unit line. The data intermediary can decide to sell all segments collected or only a subset of these segments. We will show that the data intermediary never sells all available consumer segments.[7]

3.2.2.1 Collecting data

The data intermediary collects k consumer segments at a cost $c(k)$. The cost of collecting information encompasses various dimensions of the activity of the data intermediary, such as installing trackers, or storing and handling data. Collecting more information by increasing the number of segments allows a firm to locate consumers more precisely, and thus increases the value of information. For instance, when $k = 2$, the information is coarse, and firms can only distinguish whether consumers belong to $[0, \frac{1}{2}]$ or to $[\frac{1}{2}, 1]$. At the other extreme, when k converges to infinity, the data broker knows the exact location of each consumer. Thus, $\frac{1}{k}$ can be interpreted as the precision of the information collected by the data intermediary. The k segments of size $\frac{1}{k}$ form a partition $\mathcal{P}$, illustrated in Figure 4.1.

[7] Previous research has assumed that the data intermediary sells all available information (Braulin and Valletti, 2016; Montes et al., 2018). We show that this assumption is not valid.

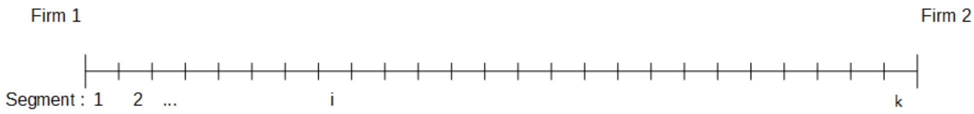

Fig. 3.1: Partition $\mathcal{P}$

3.2.2.2 Selling information

To present our argument in the simplest way, we assume that the data intermediary only sells information to Firm 1[8] using one of the three following selling mechanisms: a take it or leave it offer, a sequential bargaining, and an auction.

The data intermediary can potentially sell any subset of segments collected in the partition depicted in Figure 4.1. It is easy to understand that selling all consumer segments is not optimal for the data intermediary. On the one hand thinner segments in the partition allow a firm to extract more surplus from consumers. On the other hand selling more consumer segments also increases competition because Firm 1 has information on consumers that are closer to Firm 2, and can poach them (Thisse and Vives, 1988). For instance, if the data intermediary sells all consumer segments, Firm 1 can set a price on the consumer segment that is the closest to Firm 2.

Thus, an optimal partition must balance the competition and surplus extraction effects. Consider partition $\mathcal{P}_1$ represented in Figure 3.2. Partition $\mathcal{P}_1$ divides the unit line into two intervals: the first interval consists of j_1 segments of size $\frac{1}{k}$ on $[0, \frac{j_1}{k}]$ where consumers are identified so that Firm 1 can price discriminate them. The data intermediary does not sell information on consumers in the second interval of size $1 - \frac{j_1}{k}$, who remain unidentified, and firms charge a uniform price on this second interval. The number of segments of identified consumers j_1 depends on the total number of segments on the market k. We

[8] Selling information to both firms is in general not optimal because it increases the competitive pressure on the product market (Braulin and Valletti, 2016; Montes et al., 2018; Bounie et al., 2018), and thus lowers the profits of the data intermediary, who extracts part of the surplus of the firms.

denote by $j_1(k)$ the number of segments as a function of k. Any optimal partition must be similar to partition $\mathcal{P}_1$, and the optimization problem for the data intermediary boils down to choosing the number of segments $j_1(k)$ in partition $\mathcal{P}_1$.[9]

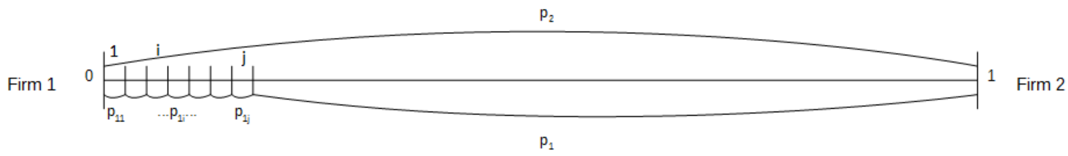

Fig. 3.2: Selling partition $\mathcal{P}_1$ to Firm 1

3.2.3 Firms

Without information, firms only know that consumers are uniformly distributed on the unit line. When Firm 1 acquires $j_1(k)$ segments of information, it can price discriminate consumers on these segments. Firm 1 sets prices in two stages.[10] First Firm 1 sets price p_1 on the segment where it competes directly with Firm 2 (the competitive segment). Secondly, Firm 1 sets a price on each segment where it is in a monopoly position, with p_{1i} being the price on the ith segment from the origin. Firm 2 is uninformed but knows the price p_1 set by Firm 1 on the competitive segment, and sets a price p_2 on the whole unit line.

We denote by $d_{\theta i}$ the demand of Firm θ on the ith segment.[11] Firm 1 is informed and maximizes the following profit function with respect to $p_{11}, .., p_{1j_1}, p_1$:

$$\pi_1 = \sum_{i=1}^{j_1+1} d_{1i} p_{1i} = \sum_{i=1}^{j_1} \frac{1}{k} p_{1i} + d_1 p_1.$$

Firm 2 is uninformed and maximizes $\pi_2 = d_2 p_2$ with respect to p_2.

[9] See Bounie et al. (2018) for a more detailed discussion.
[10] Sequential pricing decision avoids the non existence of Nash equilibrium in pure strategies, and is supported by managerial practices (see for instance, Fudenberg and Villas-Boas (2006)).
[11] The marginal production costs are also normalized to zero.

3.2.4 Timing

We summarize the timing of the game. The data intermediary first collects data and sells the partition $\mathcal{P}_1$ to Firm 1. Then Firms 1 and 2 set prices on segments where they compete. Finally Firm 1 sets prices on the monopolistic segments.

- Stage 1: the data intermediary collects data on k consumer segments.
- Stage 2: the data intermediary sells information partition $\mathcal{P}_1$ by choosing the number of segments $j_1(k)$ to include in the partition.
- Stage 3: firms set prices p_1 and p_2 on the competitive segments.
- Stage 4: Firm 1 price discriminates consumers where it is in a monopoly position by setting p_{1i}, $i \in [1, j_1(k)]$.

The game is solved by backward induction. In stage 4, Firm 1 sets prices $p_{11}, .., p_{1j_1}$ on segments where it is in a monopoly position. In stage 3, Firm 1 and Firm 2 set prices p_1 and p_2 on the competitive segments. In stage 2, we characterize the strategies of the data intermediary regarding how much consumer information to sell to Firm 1 in Section 3.4. In stage 1, we determine how much data the data intermediary collects information in equilibrium in Section 3.5. The strategies of the firms and of the data intermediary critically depend on the way information is sold, i.e. the selling mechanism, which influences the willingness to pay of the firms for information.

3.3 Selling mechanisms

We analyze three mechanisms that have been extensively studied in the literature for final goods, that we apply in the context of intermediary information goods: information can be sold through a take it or leave it offer, a sequential bargaining and an auction. First, under the take it or leave it selling mechanism, the data intermediary proposes an information partition to Firm 1. After the offer is made, there is no possibility for the data intermediary to sell information to Firm 2, even if Firm 1 refuses the offer. This approach has been studied for instance by Binmore et al. (1986). The second mechanism, the sequential bargaining, allows the data intermediary to propose information to Firm 2 if Firm 1 declines the offer, and so on until one of the firms acquires information. This type of dynamic games has been studied for instance by Rubinstein (1982) or Sobel and Takahashi (1983).[12] Thirdly, the data intermediary can auction information partition to firms. Firms bid for information partitions that can be different for Firm 1 and Firm 2. The firm with the highest bid wins the auction. Jehiel and Moldovanu (2000) analyze this type of auctions.

The three selling mechanisms have a major impact on the strategies of the data intermediary and the value of information. We compute for each selling mechanism what a firm is ready to pay for information, and determine its outside option if it does not purchase information. We will show in Section 3.4 that the three selling mechanisms have an outside option that is independent from the information sold to Firm 1. In the remainder of this section, we define the outside option for the three selling mechanisms, as the data intermediary can propose information to Firm 2 when Firm 1 does not acquire information. In other words, the outside option can be used as a threat by the data intermediary to extract more surplus from Firm 1.

We introduce further notations. We denote by $\pi_1(j_1)$ the profit of Firm 1 when it has information on the j_1 consumer segments closest to its location (Firm 2 is uninformed). In the take it or leave it format, if Firm 1 declines the offer, Firm 2 is not informed either, and both firms are uninformed. In this case, they set a single price on the unit line and make profits π. In the sequential bargaining and auctions format, Firm 2 has information when Firm 1 is uninformed. Let $\bar{\pi}_1(j_2)$ denote the profit of Firm 1 when Firm 2 has information on the j_2 consumer segments closest to its location.

[12] They consider a bilateral monopoly with a time discount factor.

It is also useful to define a data contract as a couple (j_1, j_2) where j_1 is the information proposed to Firm 1, and j_2 is the information sold to Firm 2 if Firm 1 does not acquire information, which can include the empty set, for instance in the take it or leave it offer.

Definition 1: **(Data contract)**. A data contract is a couple (j_1, j_2).

We will show in Definition 2 and Theorem 7 that the three selling mechanisms belong to a specific class of data contracts.

3.3.1 Take it or leave it

The data intermediary proposes information to Firm 1 that accepts or declines the offer. If Firm 1 declines the offer, the data intermediary does not propose information to Firm 2, and both Firm 1 and Firm 2 remain uninformed. This selling mechanism rules out the possibility for the data intermediary to renegotiate if no selling agreement is found, contrary to the sequential bargaining format that we analyze in Section 3.3.2.[13]

The data intermediary makes an offer to Firm 1 that consists of an information partition j_1^{tol}, and a price of information p_{tol}. Firm 1 can either accept the offer and make profits $\pi_1(j_1^{tol}) - p_{tol}$, or reject the offer and make profits π. The data contract is therefore $(j_1^{tol}, \emptyset)$. Thus, the willingness to pay of Firm 1 for information is $\pi_1(j_1^{tol}) - \pi$. The data intermediary sets the price of information to:

$$p_{tol}(j_1^{tol}) = \pi_1(j_1^{tol}) - \pi.$$

3.3.2 Sequential bargaining

Under the sequential bargaining mechanism, the data intermediary proposes information to each firm sequentially, in an infinite bargaining game. There is

[13] The take it or leave it format includes in fact many such mechanisms where there is no possibility for renegotiation, including Nash bargaining and menu pricing.

no discount factor and the game stops when one firm acquires information. At each stage, the data intermediary proposes information j_θ^{seq} to Firm θ and no information to Firm $-\theta$.

Firm 1 can acquire information j_1^{seq} and make profits $\pi_1(j_1^{seq})$, or decline the offer, and the data intermediary proposes information j_2^{seq} to Firm 2. If Firm 2 acquires information, the profits of Firm 1 are $\bar{\pi}_1(j_2^{seq})$. If Firm 2 declines the offer, the two previous stages are repeated. The data contract is therefore (j_1^{seq}, j_2^{seq}).

To compute the value of information under the sequential bargaining format, we characterize a stationary equilibrium of this game where Firm 1 is making profit $\pi_1(j_1^{seq})$ if it accepts the offer, but makes profits $\bar{\pi}_1(j_2^{seq})$ if it declines the offer and Firm 2 purchases information. It is important to stress that when Firm 1 declines the offer of the data intermediary, it will compete with Firm 2 that is proposed the symmetric partition (j_2^{seq} is the symmetric of j_1^{seq}). We show in Appendix 3.8 that the data intermediary sets the price of information to:

$$p_{seq} = \pi_1(j_1^{seq}) - \bar{\pi}_1(j_2^{seq}).$$

3.3.3 *Auction*

Under the auction mechanism, firms bid and the highest bidder wins the auction. The data intermediary can design a data contract to select Firm 1 as the wining bidder. When Firm 1 wins the auction, it makes profits $\pi_1(j_1^a) - p_a$, where p_a is the value of its bid. To maximize the threat on Firm 1, the data intermediary proposes k segment of information to Firm 2, so that the profit of Firm 1 without information is $\bar{\pi}_1(k)$, which is the lowest possible value. The data contract is therefore (j_1^a, k). The data intermediary sets the price of information to:

$$p_a(j_1^a, k) = \pi_1(j_1^a) - \bar{\pi}_1(k)$$

3.4 Number of segments sold in equilibrium

In this section, we characterize the number of consumer segments sold to Firm 1 for each selling mechanism. We first establish that for a given k, the number of consumer segments sold by the data intermediary is the same for the three selling mechanisms (Proposition 1). We then show that the take it or leave it, the sequential bargaining and the auction mechanisms belong to a class of data contracts that we refer to as independent data contracts. These contracts have the property that the information proposed to Firm 2 is independent from the information proposed to Firm 1. Theorem 7 generalizes Proposition 1 for independent data contracts.

3.4.1 Number of segments sold in equilibrium

We characterize in Proposition 1 the number of consumer segments sold to Firm 1 in equilibrium under the take it or leave it, sequential bargaining and auction mechanisms.

Proposition 1:

The number of consumer segments sold in equilibrium is:

$$j_1^{nb*}(k) = j_1^{seq*}(k) = j_1^{a*}(k) = \frac{6k-9}{14}.$$

Proof: see Appendix 3.8.

The proof of Proposition 1 is based on the fact that the data intermediary optimizes j_1 and j_2 independently. In other words, the information proposed to Firm 1 (j_1) is independent from the information proposed to Firm 2 (j_2) if Firm 1 does not acquire information. It is the case for the take it or leave it and the auction mechanisms. Under the take it or leave it format, Firm 1 has no information when it declines the offer of the data intermediary, and thus its outside option is independent with the information structure proposed by the data intermediary to Firm 1. Under the auction format, when Firm 1 loses the auction, it has no information but Firm 2 has information on all consumer segments. Thus, the outside option of Firm 1 that is affected by the partition proposed to Firm 2 is independent from the partition proposed to Firm 1. Under sequential bargaining, at each stage of the process, the firm who declines the offer has no information, even though the competitor can acquire information at the following stage. Here again, the outside option of Firm 1 is independent from the information structure proposed by the data intermediary to Firm 1. Regardless of the selling mechanism, when the outside option does not depend on j_1, the data intermediary simply maximizes the profit of Firm 1 with respect to j_1.

3.4.2 *Independent data contracts*

Using the intuition developed in the previous section, we can generalize Proposition 1 to a specific class of data contracts. These independent data contracts have the property that the information sold to Firm 1 (j_1) is independent from the information proposed to Firm 2 (j_2) if Firm 1 does not acquire information. Theorem 7 shows that, for a given amount of data collected k, selling mechanisms characterized by independent data contracts lead to the same number of consumer segments sold to Firm 1 (j_1^*).

Let (j_1, j_2) be the data contract proposed to Firm 1.

Definition 2: **(Independent data contract)**

A data contract (j_1, j_2) is independent if the data intermediary maximizes profits by choosing j_1 and j_2 separately.

Definition 2 includes a large set of selling mechanisms such as various forms of Nash and infinite sequential bargaining with discount factors, but also the three selling mechanisms of the article. For instance, under a Nash bargaining selling mechanism, the data intermediary maximizes with respect to j_1 a share of the joint profits with Firm 1, and does not propose information to Firm 2 if the negotiation breaks down. Also, infinite sequential bargaining with discount factors alternate offers to Firm 1 and to Firm 2 independently. However, there are mechanisms that do not satisfy Definition 2. For instance, the data intermediary can propose a symmetric partition to Firm 1, then to Firm 2 if Firm 1 declines the offer. The information structure proposed to Firm 1 appears in its outside option: $p_{alt} = \pi_1(j_1^{alt}) - \bar{\pi}_1(j_1^{alt})$. Thus, the number of segments chosen by the data intermediary affects both the profit of Firm 1 and its outside option, violating Definition 2.

Theorem 7 shows that for a given k, all selling mechanisms satisfying Definition 2 lead to the same number of consumer segments sold by the data intermediary.

Theorem 1:

Consider s and s', two selling mechanisms that satisfy Definition 2:

$$\forall\ k, \quad j_1^{s*}(k) = j_1^{s'*}(k).$$

Theorem 7 has theoretical and practical implications. First, Theorem 7 provides a first attempt to characterize data contracts based on their theoretical properties. Other dimensions of interest include the length of the data contract, exclusive sales, renegotiation conditions, or quantity discount.

Secondly, under independent data contracts, the data intermediary maximizes the profits of Firm 1. Thus, the joint profits of the data intermediary and Firm 1 are maximized. This collusive behaviour favors Firm 1 on the market, at the expense of Firm 2. This is not necessarily the case under other types of contracts. For instance under second-price auctions the data intermediary maximizes the willingness to pay of the second highest bidder, and the interest of Firm 1 and the data intermediary are not aligned.

Thirdly, Theorem 7 offers a convenient criteria to assess the impact of a selling mechanism on the amount of information sold on the market. Two selling mechanisms that belong to the class of data contracts of Theorem 7 will always lead to the same number of consumer segment sold to Firm 1. Thus a competition authority can analyze the properties of the data contract to determine if an action is required to limit the amount of information sold on a market.

We have shown in this part that the number of consumer segments sold to Firm 1 does not vary with the selling mechanism. In the next part, we will analyze how the amount of data collected varies with different selling mechanisms.

3.5 Collecting data in equilibrium

In this section we analyze how the profits of the data intermediary vary with the number of consumer segments collected (k) for the three selling mechanisms. The amount of data collected depends on the value of information, which is determined by the outside option, and thus varies according to the selling mechanism. We show that even though the data intermediary sells the same information structure to firms under the different selling mechanisms, the number of segments collected in the first stage of the game is not necessarily the same.[14]

The profit of the data intermediary $\Pi \in \{\Pi_{tol}, \Pi_{seq}, \Pi_a\}$ is given by the price of information $p \in \{p_{tol}, p_{seq}, p_a\}$, net of the cost of data collection $c(k)$:[15] $\Pi(k) = p(k) - c(k)$.

We have established in Proposition 1 that the number of segments sold by the data intermediary in the second stage of the model is the same for the three selling mechanisms: $j_1^*(k) = \frac{6k-9}{14}$. Thus, selling mechanisms will only impact the strategies of the data intermediary through the number of consumer segments collected k. Indeed, different selling mechanisms will lead to different

[14] We assume that the cost of collecting data does not depend on the selling mechanism.

[15] We make the assumption that Π_{net} is concave and reaches a unique maximum on $\mathbb{R}^+$. See Appendix 3.8 for a mathematical expression of this assumption.

prices for information, and thus to different amount of data collected by the data intermediary.

Proposition 2 compares the number of segments collected by the data intermediary and consumer surplus under the three selling mechanisms.

Proposition 2:

The number of consumer segments collected k and consumer surplus CS are inversely correlated:

$$k_{seq} > k_a > k_{tol}, \quad and \quad CS_{tol} > CS_a > CS_{seq}.$$

Proof: see Appendix 3.8.

Proposition 2 shows that the number of consumer segments collected is minimized under the take it or leave it mechanism. The optimal level of data collected depends on the marginal gain from increasing information precision. This marginal gain is the lowest in the take it or leave it offer since the outside option of the firm does not depend on any partition proposed by the data broker. Thus, the surplus extraction effect is the less intense in this mechanism, and consumer surplus is maximized.[16]

More information on the market leads to lower consumer surplus. This result sharply contrasts with the existing literature that argues that more information leads to higher consumer surplus due to the competitive effect of information (Thisse and Vives, 1988; Stole, 2007).

We show in Proposition 3 that the data intermediary chooses the auction mechanism, and that the take it or leave it is the least profitable selling mechanism.

Proposition 3:

[16] The marginal gain is higher in the auction mechanism since the data intermediary threatens the highest bidder with the harshest partition, $\mathcal{P}$, which includes all consumer segments. The marginal gain is highest under the sequential selling mechanism, since increasing precision in that case increases the most the threat of the outside option.

The profits of the data intermediary are maximized under auctions and minimized under the take it or leave it mechanism:

$$\Pi_a > \Pi_{seq} > \Pi_{tol}.$$

Proof: see Appendix 3.8.

Under the auction mechanism, the data intermediary can maximize the value of the threat of the outside option, and maximizes the willingness to pay of Firm 1. On the contrary, under the take it or leave it mechanism, both firms are uninformed when a firm rejects the offer of the data intermediary, resulting in a lower willingness to pay of firms for information.

Proposition 3, is relevant for regulators. The data intermediary chooses the auction mechanism that maximizes its profits among the three mechanisms that we propose in this article. Thus a data intermediary will never choose the take it or leave it mechanism. However, Proposition 2 shows that a competition authority, concerned with consumer surplus, and a data protection agency, concerned with the amount of consumer data collected, would choose the take it or leave it format. Enforcing specific selling mechanisms is a simple and powerful tool for regulators.

3.6 Extension: alternative selling mechanisms

We analyze an alternative selling mechanism in which the data intermediary proposes symmetric partitions to both firms (symmetric offers, indexed by *sym*). Such selling mechanism therefore does not verify Definition 2. We show that the main results of Sections 3.4 and 3.5 hold under this alternative selling mechanism. Finally we show that another class of selling mechanisms, namely second-price auctions, are equivalent to symmetric offers, by proving that the data intermediary proposes symmetric partitions to both bidders.

In the symmetric offers mechanism, the data intermediary proposes a partition j_1^{sym} to Firm 1. If Firm 1 declines the offer, a symmetric partition is proposed to Firm 2. Such a mechanism can be enforced by a competition authority to guarantee a level playing field. The price of information p_{sym} can be written as follows: $p_{sym} = \pi_1(j_1^{sym}) - \bar{\pi}_1(j_1^{sym})$. The data contract does not satisfy Definition 2 as j_1^{sym} appears in the outside option of Firm 1. The data intermediary will take this negative effect of j_1^{sym} on the profits of Firm 1 when it declines the offer.

Proposition 4:

The equilibrium with the symmetric offers mechanism has the following properties:

$$(a) \quad j_1^{sym*} = \frac{4k-3}{6}.$$

$$(b) \quad \Pi_a > \Pi_{sym} > \Pi_{seq} > \Pi_{tol}$$

$$(c) \quad k_{seq} > k_a > k_{sym} > k_{tol}$$

$$(d) \quad CS_{sym} > CS_{tol} > CS_a > CS_{seq}.$$

Proof: see Appendix 3.8.

First, the take it or leave it mechanism still minimizes the number of consumer segments collected, so that a data protection agency would prefers it to any other selling mechanism. Secondly, the data intermediary still chooses the first price auction mechanism as it leads to the highest willingness to pay of Firm 1. Thus there is still a tension between private and public regulatory motives. Thirdly, consumer surplus is now maximized in the symmetric offers mechanism. Thus there is a new tradeoff between data protection agencies and competition authorities. On the one hand, a data protection agency chooses the take it or leave it mechanism that minimizes personal data collected. On the other hand, a competition authority prefers the symmetric offers mechanism that maximizes consumer surplus.

The main results of the article also hold for second-price auctions, where the winning bidder pays the second highest bid (indexed by a_2). In the second-price auction, the data intermediary sells to Firm 1 an information structure characterized by $j_1^{a_2}$ but pays the price corresponding to the partition $j_2^{a_2}$ that Firm 2 is willing to bid. We show in Proposition 5 that a second-price auction is equivalent to the symmetric offers mechanism, and lead to the same number of consumer segments collected and sold at the equilibrium for a given k (we drop reference to k when there is no confusion).

Proposition 5:

$$j_1^{sym*} = j_1^{a_2*}.$$

Proof: See Appendix 3.8.

To prove Proposition 5, we show that the partition proposed to firms in second price auctions are symmetric. Since the symmetric equilibrium is unique, Proposition 5 is established. Consider the second-price auction. The winning bidder has more information than the second highest bidder. The data intermediary can increase the willingness to pay of the losing bidder, by increasing the number of consumer segments proposed in $j_2^{a_2*}$. Repeating this reasoning, the equilibrium is reached when the two partitions are symmetric: $j_1^{a_2*} = j_2^{a_2*}$.

To sum-up, we have identified another class of selling mechanisms, where partitions proposed to both firms are perfectly correlated, and as a matter of fact symmetric, under which our main results hold. It remains to show that Propositions 4 and 5 also hold for a broader set of classes. This is likely to be true given the fact that first price auctions is the selling mechanism that extract the most surplus from the firm who purchase information.

3.7 Conclusion

With the rise of digital giants such as Facebook, Apple, Google and Amazon, access to data and information is now central for competition policy in the digital era. As Crémer et al. (2019) emphasize, data create a high barrier to entry on a market, which encourages the emergence of dominant actors. This strategic role of data has led the FTC and the European Commission to increase their scrutiny on the activity of web giants, with concerns of potential anticompetitive practices.[17] It is important to take access to data into account, but even more important is the *differentiated* access to consumer data. Indeed, we have shown that data intermediaries can influence competition on product markets by selling different information to firms.

Our results suggest that the amount of data collected is a key strategic asset in the data strategies of firms in the digital era. Recent legislations such as the European GDPR impose a data minimization principle. Data intermediaries are growing fast, collecting any type of information on huge masses of consumers[18]. Moreover, data breaches are becoming more and more frequent, and it is important to know how much personal information data intermediaries collect to understand the strategies of large data intermediaries with respect to collecting and selling consumer information.

Policy makers could feel uneasy about leaving the market for information unregulated. Data intermediaries will choose a selling mechanism that maximizes the amount of data collected, which can raise privacy concerns, and also, that extracts more consumer surplus, which raises concerns for competition authorities. Moreover, as different selling mechanisms lead to different amounts of data collected and at the same time increase or decrease consumer surplus extraction, a tradeoff can arise between privacy concerns and the competitive effect of information. Thus, enforcing specific selling mechanisms is a simple and powerful tool for regulators to guarantee data minimization principles, or limit the market power of data intermediaries.

Our main results indicate that the design of the market for information is of most importance for the outcomes of data strategies of market participants.

[17] Congress, Enforcement Agencies Target Tech; Google, Facebook and Apple could face US antitrust probes as regulators divide up tech territory; If you want to know what a US tech crackdown may look like, check out what Europe did.

[18] Data brokers: regulators try to rein in the 'privacy deathstars'.

We have shown that different classes of selling mechanisms lead to the same tradeoff between private incentives and regulatory requirements. Other dimensions include resolving information asymmetries between sellers and buyers of information (Anton and Yao, 2002; Bergemann and Bonatti, 2015). Further research should explore the implications of market design on data strategies.

3.8 Appendix

Mathematical interpretation of Assumption 1

The cost function is defined such that:

$$\begin{cases} \frac{\partial^2[p(k)-c(k)]}{\partial k^2} < 0 \text{ and } \exists! \;\; k^* \;\; s.t. \;\; \frac{\partial[p(k)-c(k)]}{\partial k} = 0 \\ \frac{\partial^2[p(k)-c(k)]}{\partial k^2} < 0 \text{ and } \exists! \;\; k^* \;\; s.t. \;\; \frac{\partial[p(k)-c(k)]}{\partial k} = 0 \\ \exists! \;\; k^* \;\; s.t. \;\; \frac{\partial \Pi}{\partial k} = 0 \text{ and } \Pi(k^*) \geq 0 \\ c(0) = 0 \end{cases}$$

This technical hypothesis is common in the literature. It allows profits to be maximized in a unique point, which is usually true for linear cost functions.

Proof of optimal prices in sequential bargaining

We propose a candidate equilibrium policy function. We show that $p_{seq} = \pi_1(j_1^{seq}) - \bar{\pi}_1(j_2^{seq})$ is an SPE. As only the data intermediary has a non-binary choice, uniqueness will result naturally.

We write V_1 the value function of Firm 1 in stage 1 to determine its willingness to pay:

$$\begin{cases} V_1 + \pi_1(j_1^{seq}) - p_{seq} \;\; \text{if Firm 1 accepts the offer,} \\ \bar{\pi}_1(j_2^{seq}) \;\; \text{if Firm 1 declines the offer and Firm 2 accepts the offer,} \\ V_1 \;\; \text{if Firm 2 declines the offer.} \end{cases}$$

Thus, the overall value of Firm 1 is:

$$V_1 + \pi_1(j_1^{seq}) - p_{seq} - \bar{\pi}_1(j_2^{seq}) - V_1 = \pi_1(j_1^{seq}) - p_{seq} - \bar{\pi}_1(j_2^{seq})$$

Thus:

$$p_{seq} = \pi_1(j_1^{seq}) - \bar{\pi}_1(j_2^{seq})$$

The data intermediary has no interest in deviating from this value, as lowering p_{seq} would decrease its profits, and increasing p_{seq} would have Firm 1 rejecting the offer. Thus $p_{seq} = \pi_1(j_1^{seq}) - \bar{\pi}_1(j_2^{seq})$ is the unique SPE of this game.

Proof of Proposition 1

We prove that the optimal partition in equilibrium does not depend on the selling mechanism.

The data intermediary profit functions in the different timings are:

$$p_a(\mathcal{P}_1, \mathcal{P}_2) = \pi_1^{I,NI}(\mathcal{P}_1, \emptyset) - \pi_1^{NI,I}(\emptyset, \mathcal{P}_{ref})$$

$$p_{tol} = \pi_1^{I,NI}(\mathcal{P}_1, \emptyset) - \pi_1^{NI,NI}$$

$$p_{seq} = \pi_1^{I,NI}(\mathcal{P}_1, \emptyset) - \pi_1^{NI,I}(\emptyset, \mathcal{P}_2)$$

It is immediate to see that in each mechanism, the data intermediary chooses $\mathcal{P}_1$ in order to maximize the profits of Firm 1. Thus, the optimal information structure in equilibrium $\mathcal{P}_1^*$ does not depend on the selling mechanism.

Prices and demands on the unit line are identical to Bounie et al. (2018) and can be written as follow:

$p_1 = t[1 - \frac{4}{3}\frac{j}{k}]$; $p_{1i} = 2t[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k}]$; $d_1 = \frac{1}{2} - \frac{2}{3}\frac{j}{k}$.

Profits are:[19]

$$\pi_1^* = \sum_{i=1}^{j} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k}] + \frac{t}{2}(1 - \frac{4}{3}\frac{j}{k})^2 \tag{3.1}$$

Thus, first order conditions on π_1 gives us

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

Proof of Proposition 2

Data collection

We compare the first derivative of the profits of the data intermediary in the different mechanisms in order to compare the optimal precisions in equilibrium.

$$\frac{\partial p_a^*}{\partial k} = \frac{(19k - 11)t}{28k^3},$$

$$\frac{\partial p_{tol}^*}{\partial k} = \frac{(6k - 9)t}{14k^3},$$

$$\frac{\partial p_{seq}^*}{\partial k} = \frac{(72k - 45)t}{98k^3}.$$

Comparing the derivatives gives us:

$$\frac{\partial p_{seq}^*}{\partial k} > \frac{\partial p_a^*}{\partial k} > \frac{\partial p_{tol}^*}{\partial k}.$$

From the convexity of the cost function, it is straightforward that:

[19] For $p_{1i} \geq 0 \implies \frac{j}{k} \leq \frac{3}{4}$. Profits are equal whatever $\frac{j}{k} \geq \frac{3}{4}$.

$$k_{seq} > k_a > k_{tol}$$

Consumer surplus

Prices when the data intermediaries sells j segments of information to Firm 1 are given in Bounie et al. (2018) and are as follow:

- *Firm 1 captures all demand on each segment $i = 1, .., j$, and:*

$$p_{1i} = 2t[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k}].$$

- *Firms compete on the segment of unidentified consumers, and the prices are:*

$$p_1 = t[1 - \frac{4}{3}\frac{j}{k}], \quad \text{and} \quad p_2 = t[1 - \frac{2}{3}\frac{j}{k}].$$

We need to compute demands in order to find consumer surplus. On the j segments of size $\frac{1}{k}$ where Firm 1 has information, it is a monopolist and demand is $\frac{1}{k}$ on each segment.

On the segment of unidentified consumers, where firms compete, the indifferent consumer is characterized by

$$\tilde{x} = \frac{p_2 - p_1 + t}{2t} + \frac{j}{k} \implies \tilde{x} = \frac{4}{3}\frac{j}{k}$$

As $j^* = \frac{6k-9}{14}$, $\tilde{x}^* = \frac{4k-12}{7k}$.

We can write consumer surplus in equilibrium:

$$
\begin{aligned}
CS(k) &= \sum_{i=1}^{j^*}[\int_0^{\frac{1}{k}} V - 2t[1-\frac{1}{3}\frac{j}{k}] + \frac{t}{k} + \frac{it}{k} - tx\mathrm{d}x] \\
&+ \int_{\frac{j^*}{k}}^{\frac{4j^*}{3k}} V - t[1-\frac{4}{3}\frac{j^*}{k}] - tx\mathrm{d}x + \int_0^{1-\frac{4j^*}{3k}} V - t[1-\frac{2}{3}\frac{j^*}{k}] - tx\mathrm{d}x \\
&= \sum_{i=0}^{j^*-1}\frac{1}{k}[V - 2t[1-\frac{1}{3}\frac{j^*}{k}] + \frac{t}{k} + \frac{it}{k}] - \frac{j^*}{2k^2} \\
&+ \frac{j^*}{3k}[V - t + \frac{4}{3}\frac{j^*t}{k}] - \frac{t}{2}[\frac{16}{9}\frac{j^{*2}}{k^2} - \frac{j^{*2}}{k^2}] \\
&+ [1-\frac{4j^*}{3k}][V - t + \frac{2}{3}\frac{j^*t}{k}] - \frac{t}{2}[1-\frac{4}{3}\frac{j^*}{k}]^2 \\
&= \frac{j^*}{k}[V - 2t[1-\frac{1}{3}\frac{j^*}{k}] + \frac{t}{k}] + \frac{j^*(j^*+1)t}{k^2} - \frac{j^*}{2k^2} \\
&+ V[1-\frac{j^*}{k}] - \frac{3t}{2} + 3t\frac{j^*}{k} - \frac{31t}{18}\frac{j^{*2}}{k^2} \\
&= \frac{j^*}{k}V - \frac{2tj^*}{k} + \frac{5}{3}\frac{j^{*2}t}{k^2} + \frac{5j^*t}{2k^2} \\
&+ V[1-\frac{j^*}{k}] - \frac{3t}{2} + 3t\frac{j^*}{k} - \frac{31t}{18}\frac{j^{*2}}{k^2} \\
&= V - \frac{3t}{2} + \frac{j^*t}{k} + \frac{5j^*t}{2k^2} - \frac{j^{*2}t}{18k^2} \\
&= -\frac{(424k^2 - 180k + 639)t - 392Vk^2}{392k^2}
\end{aligned} \tag{3.2}
$$

Consider now the first degree derivative of consumer surplus with respect to k:

$$\frac{\partial CS(k)}{\partial k} = -\frac{(90k - 639)t}{196k^3}$$

This is always negative for $k \geq 5$, and thus consumer surplus decreases with information precision.

Proof of Proposition 3

We compare the profits of the data intermediary in the different selling mechanisms. The profits of the firms depending on the information structure are provided in Bounie et al. (2018):

$$\pi^{NI,NI} = \frac{t}{2}.$$

$$\pi^{I,NI}(j_1^*, \emptyset) = \frac{(18k^2 - 12k + 9)t}{28k^2}.$$

$$\pi^{NI,I}(\emptyset, \mathcal{P}_{ref}) = \frac{(k^2 + 2k + 1)t}{8k^2}.$$

$$\pi^{NI,I}(\emptyset, j_1^*) = \frac{(25k^2 + 30k + 9)t}{98k^2}.$$

Profits are found directly from these values:

$$p_a^* = \pi^{I,NI}(j_1^*, \emptyset) - \pi^{NI,I}(\emptyset, \mathcal{P}_{ref}) = \frac{(29k^2 - 38k + 11)t}{56k^2}$$

$$p_{tol}^* = \pi^{I,NI}(j_1^*, \emptyset) - \pi^{NI,NI} = \frac{(4k^2 - 12k + 9)t}{28k^2}$$

$$p_{seq} = \pi^{I,NI}(j_1^*, \emptyset) - \pi^{NI,I}(\emptyset, j_1^*) = \frac{(76k^2 - 144k + 45)t}{196k^2}$$

Direct comparison of the profits provide the ranking of Proposition 2.

Proof of identical structures in alternative mechanisms

We prove that information structures are identical in both alternative mechanisms. In the first alternative mechanism, the price of information can be written

$$p_{alt} = \pi_1(j_1^{alt}) - \bar{\pi}_1(j_1^{alt}).$$

In the auction mechanism, the willingness to pay of firms when the data intermediary proposes information j_1^{alt} to Firm 1 and j_2^{alt} to Firm 2 are:

$$\begin{cases} \pi_1(j_1^{alt}) - \bar{\pi}_1(j_2^{alt}), \\ \pi_2(j_2^{alt}) - \bar{\pi}_2(j_1^{alt}) \end{cases}$$

We show that in equilibrium $j_1^{alt} = j_2^{alt}$ and thus that the optimization problem is identical to the one in the first mechanism.

Assume $j_1^{alt} > j_2^{alt}$ (the other case is solved similarly). Then it is straightforward that:

- $\pi_1 > \pi_2$,
- π_2 increases when j_1^{alt} decreases.

Thus the data intermediary choses values of j_1^{alt} as low as possible, under the constraint that $j_1^{alt} \geq j_2^{alt}$; that is, the data intermediary chooses $j_1^{alt} = j_2^{alt}$.

FOC on p_{alt} with respect to j_1^{alt} gives us:

$$p_{alt}^* = \frac{4t}{9} - \frac{2t}{3k} + \frac{t}{9k^2}$$

$$\frac{\partial p_{alt}^*}{\partial k} = \frac{(6k-2)t}{9k^3}$$

The ranking of profits and optimal data collection is then straightforward.

CHAPTER 4

Collecting and Selling Consumer Information: the Two Faces of Data Brokers

Abstract

We investigate how competition between data brokers affects their strategies of *collecting* and *selling* personal information to competing firms for price discrimination purposes. Two dimensions of competition between data brokers are considered: their market shares, and the size of the market on which they compete. We show that when competing data brokers are symmetric in terms of market share *i.* they collect less information than in monopoly; *ii.* they sell information on more consumers than under monopoly, resulting in a higher competition between firms. However, when the market shares of data brokers are asymmetric, as the size of the competitive market increases *i.* only one data broker sells information; *ii.* this data broker collects more information than in monopoly; *iii.* the share of identified consumers is larger than in monopoly. We argue that these findings are central for privacy and competition laws as we identify an inverted U-shape relationship between competition and data collection: data collection is the lowest when competition is the strongest between data brokers; it reaches an intermediate level under monopoly, and is the highest under constrained monopoly.

4.1 *Introduction*

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 (Hayek, 1945).[1] This topic has recently been revisited with the development of data brokers who collect and sell consumer information to firms.

The data brokerage industry is characterized by a high degree of competition. In 2015, there were more than 4,000 data brokers (Pasquale, 2015). Data brokers have two main activities. First, they collect personal information with different degrees of precision on various consumer markets. The precision or granularity of personal information allows data brokers to create unique consumer segments. They sell these consumer segments to companies that want to better target or price-discriminate consumers. Data brokers can increase their profitability by selling more precise segments to firms. The activities of collecting and selling consumer information are related, as the more data a data broker collects, and the higher the precision of information, the more firms are willing to pay for information.

In this paper, we study how competition between data brokers affects their strategies regarding how much consumer information they collect and sell, and we show that data protection and competition laws may be conflicted with one another.

To date, the economic literature has mostly ignored the strategic role of data brokers on product markets, as well as the relation between strategies of data collection and information selling. A first strand of the literature mainly focuses on data collection, and analyzes how firms choose the amount of personal information that they collect. Data collection is a black box, and information is considered as any other costly input whose only special feature is to raise privacy concerns. For instance, Bloch and Demange (2018) look at the effects of taxation on business strategies related to consumer information collection. In

[1] See also Radner et al. (1961), Vives (1984), Thisse and Vives (1988), Burke et al. (2012).

a competitive set-up, Casadesus-Masanell and Hervas-Drane (2015) and Gal-Or et al. (2018) show how consumer privacy concerns can be used as an asset by data driven firms. In these papers the sale of consumer information does not influence the strategic choice of the amount of data collected.

A second strand of the literature focuses on third parties selling information to competing firms. Sarvary and Parker (1997) and Xiang and Sarvary (2013) show that when information increases competition between firms, a data broker will sell partial information to firms so that the profits of the firms are not decreased too much, and thus the value of information remains high. Their results are empirically supported by Christen and Sarvary (2007). Braulin and Valletti (2016) and Montes et al. (2018) consider data brokers selling information of exogenous quality, and Bounie et al. (2018)[2] study the strategies of data brokers related to how much information they sell to firms. They show that data brokers do not find it profitable to sell all information available to competing firms, as selling all information would increase competition and lower how much firms are willing to pay for consumer information. The aforementioned studies do not focus on the data collection strategies of data brokers, and also rule out the role of competition between data brokers.[3]

The novelty of this paper is to integrate strategies of information collection and selling in a model where two data brokers compete. Two competing data brokers collect and sell information on consumers' willingness to pay for a product. Data brokers sell information represented by a partition of the consumer demand in segments of variable length. Data brokers can sell any information partition, that is, any combination of consumer segments to only one firm or to both competing firms. Data brokers endogenously choose the precision of their information as well as which partition they sell to firms. They then compete to sell information to firms.

[2] This working paper is a previous version of Chapter 2 of the thesis. In this initial version, information is sold through a take it or leave mechanism, in which case it is optimal for the data broker to sell information to both firms.

[3] Gu et al. (2018) look at information sharing between competing data brokers. They do not however focus on the amount of information collected and sold by data brokers in equilibrium.

We show that when competing data brokers are symmetric in terms of market share, they collect less data than in monopoly, and they sell information on more consumers than in monopoly. Thus, the goals of privacy laws and competition laws are aligned since less data are collected and since information is sold on more consumer competition increases on the market. However, when there is a dominant data brokers in terms of market shares, as the size of the competitive market increases only the dominant data broker sells information. Moreover, the dominant data broker collects more information than in monopoly, and he sells information on more consumers than in monopoly. Thus we find an inverted U-shape relationship between competition and data collection: data collection is minimized when competition is the strongest, it reaches an intermediate level under monopoly, and is the highest when data brokers compete, and one data broker is bigger than the other. Overall, consumer surplus always increases when data brokers compete.

The remainder of this article is organized as follow. In Section 4.2 we describe the model. We solve the game when two data brokers compete in Section 4.3. We discuss consumer surplus in Section 4.4. Section 4.5 concludes.

4.2 Description of the model

There are three types of agents: a mass 1 of consumers, two firms noted $\theta \in \{1, 2\}$ competing on a product market, and two data brokers noted $\gamma \in \{1, 2\}$. 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. Without information firms set a single price for their product. They can purchase information from data broker γ who collects data on consumer segments. An information partition segments the unit line into k_γ identical intervals. We interpret k_γ as the amount of information collected by data broker γ. Firms who purchase information can set a price on each consumer segment.

Consumers are divided into three sub-markets. On sub-markets 1 and 2 of size m_1 and m_2, data brokers 1 and 2 are in monopoly positions; we therefore implicitly assume that data-broker 1 [data broker 2] is the only one to have information on consumers in sub-market 1 [sub-market 2]. However, on sub-market 3 of size l, both data brokers 1 and 2 have information on consumers, and compete to sell information to firms. We allow the model to handle several cases such as $m_1 = 0$, $m_2 = 0$, $m_1 = m_2$, and $l > 0$.

4.2.1 Consumers

Consumers are uniformly located on a unit line $[0, 1]$. They buy one product at a price p_1 from Firm 1 located at 0, or p_2 from Firm 2 located at 1.[4] A consumer located at $x \in [0, 1]$ derives a utility V from purchasing the product. He incurs a transportation cost $t > 0$ so that buying from Firm 1 (resp. from Firm 2), has a total cost tx (resp. $t(1 - x)$). Consumers purchase the product for which they have the highest utility.

To summarize, a consumer located at x has a utility function defined by:

$$u(x) = \begin{cases} V - p_1 - tx, & \text{if he buys from Firm 1,} \\ \\ V - p_2 - t(1 - x), & \text{if he buys from Firm 2.} \end{cases} \tag{4.1}$$

4.2.2 Data brokers

Data brokers collects information on consumers that allows firms to distinguish k consumer segments on the unit line. The data intermediary can decide to sell all segments collected or only a subset of these segments. We will show that the data intermediary never sells all available consumer segments.[5]

[4] We assume that the market is covered. This assumption is common in the literature. See for instance Bounie et al. (2018) or Montes et al. (2018).

[5] Previous research has assumed that the data intermediary sells all available information (Braulin and Valletti, 2016; Montes et al., 2018). We show that this assumption is not valid.

4.2.2.1 Collecting information

Data broker $\gamma = 1, 2$ collects information of precision $\frac{1}{k_\gamma}$ on consumers, at a cost $c(k_\gamma)$. Data broker γ collects a partition consisting of k_γ segments of size $\frac{1}{k_\gamma}$ which we refer to as the reference partition: $\mathcal{P}_{\gamma,ref}$. Figure 4.1 illustrates $\mathcal{P}_{\gamma,ref}$.

Fig. 4.1: Reference partition $\mathcal{P}_{\gamma,ref}$

We assume that the number of segments that data broker γ collects on the Hotelling line can always be written $k_\gamma = 2^\kappa$, with κ an integer. This assumption has the following implication: suppose that there are two partitions $\mathcal{P}$ and $\mathcal{P}'$ with intervals of respective sizes $\frac{1}{k'}$ and $\frac{1}{k}$ with $k' \leq k$. All elements of the sigma field generated by $\mathcal{P}$ also belong to the sigma field generated by $\mathcal{P}'$. In other words there cannot be any intersections between the segments of both partitions. This assumption greatly increases the tractability of the model.

Figure 4.2 illustrates the case with $\kappa' = 2$ and $\kappa = 3$:

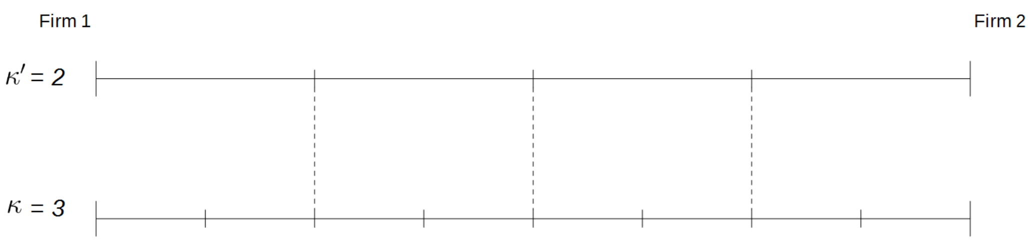

Fig. 4.2: Information collection

Each data broker γ is in a monopoly position in a sub-market of mass m_γ and data brokers compete on a sub-market of size l. We assume that the precision

k_γ chosen by data broker γ is identical on both sub-markets.[6]

Assumption 1:

For each data broker γ, the reference partitions in sub-markets m_γ and l are identical.

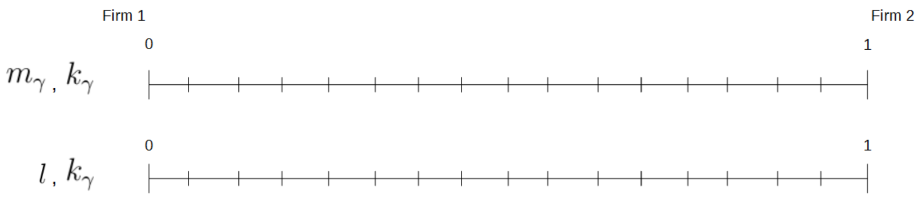

Fig. 4.3: Information collection of data broker γ

Figure 4.3 shows information partitions on two sub-markets. On the top part, data broker γ is in monopoly position, and has an information partition that divides the Hotelling line in k_γ segments of equal size. On the bottom part, both data brokers compete for the sub-market of size l. Assumption 1 implies that data brokers collect the same amount of information on consumers, independently from whether consumers belong to m_γ or to l. Moreover, even if a data broker does not sell information on the competitive sub-market l, he incurs a per-user cost $c(k_\gamma)$ to collect information.

Data broker γ has a total market share of $\sigma_\gamma = m_\gamma + l$. On its market share σ_γ, data broker γ segments consumer demand in k_γ intervals of size $\frac{1}{k_\gamma}$ at a cost $\sigma_\gamma c(k_\gamma)$, where k_γ characterizes the amount of information collected.[7]

4.2.2.2 Selling information partitions

Data brokers sell information partitions which do not necessarily include all data collected. As a matter of fact, we will show that it is not optimal for data

[6] This assumption is not essential to prove our main results since the crucial element of the analysis is the fact that data brokers are differentiated. Our results would carry on with different precisions on different sub-markets.

[7] Function c is detailed in Appendix 4.6.

brokers to sell all consumer segments.

As we have argued in the introduction, it is not optimal to sell all consumer segments. On the one hand, more information allows firms to better target consumers and price discriminate. They can extract more consumer surplus, 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 lowers the profits of the firms. Overall, there is an economic trade-off between surplus extraction and increased competition. By only selling a subset of the intervals, data brokers can soften competition on the product market.

An information partition consists of a partition of the unit line into $n_\gamma \leq k_\gamma$ segments of arbitrary sizes. These segments are constructed by unions of elementary segments of size $\frac{1}{k_\gamma}$, with k_γ the number segments in the reference partition.

Fig. 4.4: Example of a partition of the unit line

For instance, data broker γ can sell a partition starting with one segment of size $\frac{1}{k}$, and another segment of size $\frac{2}{k}$, and so on, as illustrated in Figure 4.4.

4.2.2.3 Pricing information

Data broker γ can sell information separately on m_γ and l. Data brokers thus choose an information partition as well as the price of information on the two sub-markets: m_γ, l. Let $\mathcal{P}^\theta_{m_\gamma}$ (resp. $\mathcal{P}^\theta_{l,\gamma}$) denote the information partition sold by data broker γ on m_γ (resp. on l) to Firm θ. On m_γ, data broker γ has the choice to sell information to one firm only, or to both firms. We focus on the case where data broker γ sells information to both firms.

Price of information in monopolistic markets

On m_γ, data broker γ can sell any partition $\mathcal{P}^\theta_{m_\gamma}$ to Firm θ. We note $\mathcal{P}^\theta_{m_\gamma} = 0$ when Firm θ does not acquire information. Each data broker knows the information partition of its competitor, and adjusts its price and information partition accordingly.

On m_γ, data broker γ sells information through a take it or leave it offer, at the highest price that firms are willing to pay. This price, w_{m_γ}, corresponds to the difference of the profits of a firm when it is informed $\pi_\theta(\mathcal{P}^\theta_{m_\gamma}, \mathcal{P}^{-\theta}_{m_\gamma}, k_\gamma)$, and its outside option, which is remaining uninformed whereas its competitor is informed, $\pi_\theta(0, \mathcal{P}^{-\theta}_{m_\gamma}, k_\gamma)$.

The price of information can be written as:

$$w^\theta_{m_\gamma}(\mathcal{P}^\theta_{m_\gamma}, \mathcal{P}^{-\theta}_{m_\gamma}, k_\gamma) = \pi_\theta(\mathcal{P}^\theta_{m_\gamma}, \mathcal{P}^{-\theta}_{m_\gamma}, k_\gamma) - \pi_\theta(0, \mathcal{P}^{-\theta}_{m_\gamma}, k_\gamma)$$

Thus, the revenue of data broker γ is:

$$\Pi_{m_\gamma}(\mathcal{P}^1_{m_\gamma}, \mathcal{P}^2_{m_\gamma}, k_\gamma) = w^1_{m_\gamma}(\mathcal{P}^1_{m_\gamma}, \mathcal{P}^2_{m_\gamma}, k_\gamma) + w^2_{m_\gamma}(\mathcal{P}^2_{m_\gamma}, \mathcal{P}^1_{m_\gamma}, k_\gamma)$$

Price of information in the competitive market

Data brokers compete on sub-market l. Data broker γ sells partition $\mathcal{P}^\theta_{l,\gamma}$ to Firm θ. The price of information is defined by the difference of the profits of Firm θ with this partition, $\pi_\theta(\mathcal{P}^\theta_{l,\gamma}, \mathcal{P}^{-\theta}_{l,\gamma}, k_\gamma)$, and its outside option, which is buying information $\mathcal{P}^\theta_{l,-\gamma}$ from data broker $-\gamma$, $\pi_\theta(\mathcal{P}^\theta_{l,-\gamma}, \mathcal{P}^{-\theta}_{l,\gamma}, k_{-\gamma})$. We note w_l the price of information on consumers where data brokers compete:

$$w^\theta_{l,\gamma}(\mathcal{P}^\theta_{l,\gamma}, \mathcal{P}^{-\theta}_{l,\gamma}, k_\gamma) = max\{0, \pi_\theta(\mathcal{P}^\theta_{l,\gamma}, \mathcal{P}^{-\theta}_{l,\gamma}, k_\gamma) - \pi_\theta(\mathcal{P}^\theta_{l,-\gamma}, \mathcal{P}^{-\theta}_{l,\gamma}, k_{-\gamma})\}. \quad (4.2)$$

The profit function of data broker γ is:

$$\Pi_{l,\gamma}(\mathcal{P}^1_{l,\gamma}, \mathcal{P}^2_{l,\gamma}, k_\gamma) = w^1_{l,\gamma}(\mathcal{P}^1_{l,\gamma}, \mathcal{P}^2_{l,\gamma}, k_\gamma) + w^2_{l,\gamma}(\mathcal{P}^2_{l,\gamma}, \mathcal{P}^1_{l,\gamma}, k_\gamma). \qquad (4.3)$$

4.2.3 Firms

Without information, firms only know that consumers are uniformly distributed on the unit line. When a firm acquires an information partition $\mathcal{P}_\gamma$, it knows which interval of this partition a consumer belongs to. Firms simultaneously set their prices on each segment of the unit line where they have information. Firm θ sets prices in two stages.[8] First she sets prices on segments where she shares consumer demand with its competitors. Then, on segments where she is a monopolist, she sets a monopoly price. Each firm knows whether its competitor is informed, and the partition $\mathcal{P}^{-\theta}$.[9]

We denote by $d_{\theta i}$ the demand of Firm θ on the ith segment. An informed Firm θ maximizes the following profit function with respect to $p_{\theta 1}, .., p_{\theta n}$:

$$\pi_\theta = \sum_{i=1}^{n} d_{\theta i} p_{\theta i} \qquad (4.4)$$

4.2.4 Timing

The timing of the game is the following:

- Stage 1: data broker γ collects information k_γ.
- Stage 2: data broker γ sets $\mathcal{P}^\theta_{m_\gamma}$ and $\mathcal{P}^\theta_{l,\gamma}$, and sells information at prices w_{m_γ} and $w_{l,\gamma}$.

[8] Making a sequential pricing decision avoids the non-existence of Nash equilibrium in pure strategies, and is supported by managerial practices (see for instance, Fudenberg and Villas-Boas (2006)).
[9] This assumption is also standard in Braulin and Valletti (2016) and Montes et al. (2018).

- Stage 3: firms set prices on the competitive segments.
- Stage 4: firms price discriminate consumers on the segments where they have monopoly power.

We solve the game by backward induction on the competitive sub-market l in Section 4.3.

4.3 Collecting and selling information

We analyze the strategies of data collection and information selling of data brokers on monopoly sub-markets m_1 and m_2, and on the competitive market l. We solve the game by backward induction. We prove in Theorem 3 that the optimal information partition sold by data broker γ is composed of a partition of j segments of size $\frac{1}{k_\gamma}$ closest to firms, and a unique segment of unidentified consumers on the rest of the line.

4.3.1 Stages 3 and 4: profits of the firms in equilibrium

When no information is sold, the profit of the firms is $\pi = \frac{t}{2}$.

When both firms have information partitions $\mathcal{P}^{\theta}_{l,\gamma}$, the profit of Firm θ is:

$$\pi_\theta(\mathcal{P}^{\theta}_{l,\gamma}, \mathcal{P}^{-\theta}_{l,\gamma}, k_\gamma)$$

The equilibrium profits are detailed in Appendix 4.6.

4.3.2 Stage 2: Selling information

We analyze in this section the strategies of competing data brokers selling information to competing firms. We first provide the optimal information structure sold by data brokers. We then compare the equilibrium under competition with the one when data brokers are monopolists.

4.3.2.1 Selling information in monopoly

A monopolist data broker can sell any partition belonging to the sigma field generated by $\mathcal{P}_{ref}$, which is potentially a huge set as we have no upper bound on k. We show that we can reduce the dimensionality of the problem to solve for the profit functions of the data broker and of firms. Indeed, given the equilibrium prices of Stage 3, the optimal information structure is a partition that consists of elementary intervals of the reference partition up to the jth segment, and then a last segment of unidentified consumers.

In the remaining of the paper, we assume that data brokers do not sell segments that would allow firms to poach consumers. Assumption 2 is detailed in Appendix 4.6. It allows us to avoid considering situations where information that increases a lot competition between firms is sold. Such information is clearly not profitable for the data broker, thus we rule them out to simplify the resolution.

We describe the optimal information structure sold to firms by a monopolist data broker in Theorem 2:

Theorem 2:

Under Assumption 2, on the sub-market m, a monopolist data broker sells to Firm 1 (resp. Firm 2) a partition with two different types of segments:

a) *There are* j_1 *(resp.* j_2*) segments of size* $\frac{1}{k}$ *on* $[0, \frac{j_1}{k}]$ *(on* $[1 - \frac{j_2}{k}, 1]$ *for Firm 2) where consumers are identified.*

b) *Consumers in the second segment of size* $1 - \frac{j_1}{k}$ *(resp.* $1 - \frac{j_2}{k}$*) are unidentified.*

c) $j_1 = j_2$.

The proof is given in Bounie et al. (2018).

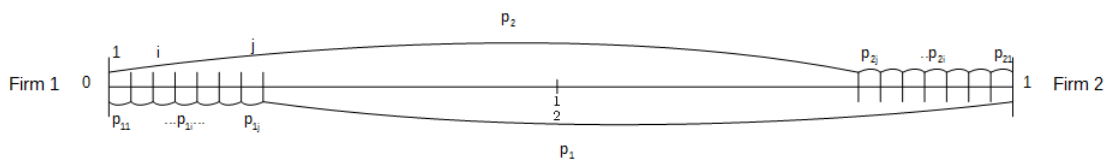

Fig. 4.5: Optimal information structure in monopoly

Theorem 2 implies that a monopolist data broker sells to firms information on consumers who have the highest valuation for their product. Consumers with the lowest valuation are kept unidentified.

When a firm acquires information of the form $\{[0,\frac{1}{k}],[\frac{1}{k},\frac{2}{k}],..,[\frac{i}{k},\frac{i+1}{k}],..,[\frac{j}{k},1]\}$, it can set a different price p_{i+1} on each segment $[\frac{i}{k},\frac{i+1}{k}]$. We denote by $d_{\theta i}$ the demand of Firm θ on the ith segment. An informed Firm θ maximizes the following profit function with respect to $p_{\theta 1},..,p_{\theta j}$, and p_θ:

$$\pi_\theta = \sum_{i=1}^{j} \frac{1}{k} p_{\theta i} + p_\theta d_\theta. \tag{4.5}$$

The problem of choosing the optimal structure of information by the data broker is complex to solve, since it requires to compare firms' profits for each possible partition of the unit line. However, we have shown in Theorem 2 that it can be simplified to the choice of one variable j. The expressions of the profit functions depending on j are detailed in Appendix 4.6.

The optimal information structure in equilibrium on m is :

$$\frac{j_1^*}{k} = \frac{j_2^*}{k} = \frac{3}{11} - \frac{9}{22k}.$$

The price of information is

$$w_m(k) = \frac{t}{11} - \frac{3t}{11k} + \frac{9t}{44k^2}.$$

The price of information constitutes the revenue of the data broker, which will drive its information collection decision.

4.3.2.2 Selling information under competition

Data brokers can potentially sell any partition belonging to the sigma field generated by $\mathcal{P}_{ref}$. We will show that the optimal partition is composed of elementary intervals of the reference partition up to the jth segment, and then a last segment of unidentified consumers.

We assume without loss of generality that $k_1 \geq k_2$.

Proposition 6:

On sub-market l where data brokers compete:

- Only data broker 1 sells information.

- Both firms acquire information.

Proof: see Appendix 4.6.

Proposition 6 states that in markets where data brokers compete, both firms are informed in equilibrium. The rationale behind Proposition 6 is the following. Firms acquire information from data broker 1 as it sells information with the highest precision. Thus, any information partition that data broker 2 can propose to firms, data broker 1 can offer too. As firms can acquire information from only one data broker, they choose to buy information that allows them to best extract consumer surplus, which is the one sold by data broker 1. Data broker 1 necessarily sells information to both firms, as if he were selling information to one firm only, data broker 2 would sell information to the uninformed firm.

Data broker 1 sells to firms information partitions $\mathcal{P}^1_{l,1}$ and $\mathcal{P}^2_{l,1}$ which maximize its profit. Information partitions and price in equilibrium must maximize the profit of the firms, so that firms do not prefer to buy partition $\mathcal{P}^{\theta}_{l,2}$ from data broker 2:

$$\pi_\theta(\mathcal{P}^{\theta}_{l,1}, \mathcal{P}^{-\theta}_{l,1}, k_1) - w^{\theta}_{l,1}(\mathcal{P}^{\theta}_{l,1}, \mathcal{P}^{-\theta}_{l,1}, k_1) \geq \pi_\theta(\mathcal{P}^{\theta}_{l,2}, \mathcal{P}^{-\theta}_{l,1}, k_2) \tag{4.6}$$

In equilibrium, the constraint is binding, and the price of information is:

$$w^{\theta}_{l,1}(\mathcal{P}^{\theta}_{l,1}, \mathcal{P}^{-\theta}_{l,1}, k_1) = \pi_\theta(\mathcal{P}^{\theta}_{l,1}, \mathcal{P}^{-\theta}_{l,1}, k_1) - \pi_\theta(\mathcal{P}^{\theta}_{l,2}, \mathcal{P}^{-\theta}_{l,1}, k_2).$$

Thus, data broker 1 acts as a monopolist, constrained by a limit price, that leaves its competitor out of the market.

We describe the optimal information partition sold to firms by data brokers in Theorem 3:

Theorem 3:

On sub-market l where data brokers compete, data broker 1 sells to Firm 1 (resp. Firm 2) a partition with two different types of segments:

a) *There are j_1 (resp. j_2) segments of size $\frac{1}{k_1}$ on $[0, \frac{j_1}{k_1}]$ (on $[1-\frac{j_2}{k_1}, 1]$ for Firm 2) where consumers are identified.*

b) *Consumers in the second segment of size $1-\frac{j_1}{k_1}$ (resp. $1-\frac{j_2}{k_1}$) are unidentified.*

c) *Partitions sold to Firm 1 and Firm 2 are symmetric and $j_1 = j_2$.*

Proof: See Appendix 4.6.

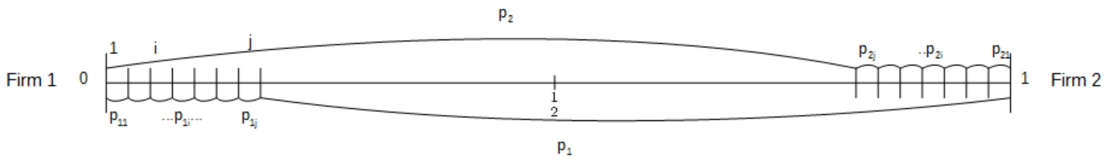

Fig. 4.6: Optimal information selling in duopoly

Theorem 3 implies that data brokers sell to firms information on consumers who have the highest valuation for their product. Consumers with the lowest valuation are kept unidentified. This optimal information partition is similar to the one on sub-markets m_γ where data brokers are monopolists.[10] However, more information will be sold in competition, as shown in Proposition 7.

Even though the problem of choosing the optimal partition by the data broker is complex to solve, since it requires to compare the profits of the firms for each possible partition of the unit line, we have shown in Theorem 3 that it can be simplified to the choice of one variable $j_{l,1}$. We give the optimal information structure when data broker compete:

Proposition 7:

On sub-market l where data brokers compete:

- when $k_1 > k_2 > 0$ (without loss of generality)
  - data broker 1 sells to both firms an information partition characterized by:
$$\frac{j^*_{l,1}}{k_1} = \frac{1}{3} - \frac{1}{9k_2} - \frac{7}{18k_1}.$$
  - data broker 1 sets a price $w_l(k_1)$ increasing in k_1.
  - data broker 2 does not sell information.
- when $k_1 = k_2 > 0$, data brokers compete à la Bertrand and they sell at a

[10] This information partition is given in Bounie et al. (2018).

zero price, information characterized by:

$$\frac{j^*_{l,\gamma}}{k_\gamma} = \frac{1}{3} - \frac{1}{2k_\gamma}.$$

Proof: See Appendix 4.6.

Proposition 7 shows that the data broker who has more precise information than its competitor ($k_1 > k_2$), will be the only one selling information, and the data broker with the lowest information precision will not sell information on the competitive market.

Proposition 7 also characterizes the number of consumer segments j that data broker 1 sells to firms when it faces competition. Parameter j can be interpreted as the strength of competition between firms. Indeed, competition increases with the information available in the market as discussed by Thisse and Vives (1988), Ulph and Vulkan (2000), and Stole (2007). In order to assess the impact of competition between data brokers on competition on the downstream market, we compare in the next section the value of j in equilibrium when data brokers compete with its value when data brokers are monopolists.

Firms buy information from data brokers to price discriminate consumers with precision k_1 on sub-markets m_1 and l,[11] and with precision k_2 on sub-market m_2.

When a firm acquires information of the form $\{[0, \frac{1}{k}], [\frac{1}{k}, \frac{2}{k}], .., [\frac{i}{k}, \frac{i+1}{k}], .., [\frac{j}{k}, 1]\}$, it can set a different price p_{i+1} on each segment $[\frac{i}{k}, \frac{i+1}{k}]$. We denote by $d_{\theta i}$ the demand of Firm θ on the ith segment. An informed Firm θ maximizes the following profit function with respect to $p_{\theta 1}, .., p_{\theta j}$, and p_θ:

$$\pi_\theta = \sum_{i=1}^{j} \frac{1}{k} p_{\theta i} + p_\theta d_\theta. \tag{4.7}$$

The expressions of the profits of the firms are detailed in Appendix 4.6.

[11] by Assumption 1.

4.3.2.3 Competing gatekeeper effect

We compare the selling strategies of data brokers when they are in a monopoly position (sub-market m_γ) or when they compete (sub-market l). To this aim, we define the share of identified consumers, which indicates how much consumer information a data broker sells.

Definition 3:

The share of identified consumers $s(j,k)$ is given by:

$$s(j,k) \equiv 2\frac{j}{k}.$$

When data brokers compete, there are two dimensions of competition to consider. When they compete à la Bertrand ($m_1 = m_2$, which implies $k_1 = k_2$), they make zero profit on sub-market l, and firms are indifferent between purchasing from data broker 1 and data broker 2. We note $s_0(k_1)$[12] (with $k_1 = k_2$) the share of identified consumers under Bertrand competition. When data brokers are asymmetric in terms of market shares ($m_1 > m_2$, which implies $k_1 > k_2$), only data broker 1 sells information. We note $s_1(k_1, k_2)$ the share of identified consumers sold by data broker 1.

In order to understand the effects of competition between data brokers on the amount of consumer information that they sell to firms, we now define $\Delta s_0(k_1) = s_0(k_1) - s(k_1)$ and $\Delta s_1(k_1, k_2) = s_1(k_1, k_2) - s(k_1, k_2)$, the incremental shares of identified consumers due to competition respectively under Bertrand and asymmetric competition. In each competitive set-up, they correspond to the difference between the share of identified consumers under competition (on the sub-market of size l) and the share of identified consumers under monopoly.

Lemma 1:

[12] When there is no confusion, we drop subscript j, since j^* only depends on k ($s(k) = s(j^*, k)$).

$$\Delta s_1(k_1, k_2) \underset{k_1 \to^+ k_2}{\longrightarrow} \Delta s_0(k_1)$$

Proof: See Appendix 4.6.

Lemma 1 states that the market share under asymmetric competition converges to the share under Bertrand competition as k_1 tends to k_2. Thus we only need to define the incremental share under asymmetric competition since $\Delta s_0(k_1)$ is a limit case when $k_1 = k_2$.

The share of identified consumers always increases with competition between data brokers, as stated in Proposition 8.

Proposition 8:

The incremental share of consumers identified due to competition is always positive:

$$\Delta s_1(k_1, k_2) \geq 0$$

Proof: Immediate by substituting $\frac{j^*_{l,1}}{k_1}$ and $\frac{j^*_{m_1}}{k_1}$ in $\Delta s(k_1, k_2)$.

Proposition 8 shows that data brokers sell more consumer information to firms when they compete. This is the competing-gatekeeper effect which increases competition in the product market. Proposition 8 is a central result of this paper, and is new in the literature. Previous models do not allow the data broker to choose the amount of information that he sells to firms (it is usually assumed that he will either sell all information or no information at-all). By allowing data brokers to sell information strategically to firms, we prove that competition between data brokers increases the amount of consumer information on the market, which increases consumer surplus.

We have focused on the selling strategy of the data brokers with a given precision k. We now consider the strategic choice of information precision by data brokers, how this choice is impacted by competition between data brokers, and how a change in information precision affects competition and consumer surplus on the product market.

4.3.3 Stage 1: Collecting information with competing data brokers

We characterize in this section the collecting strategies chosen by data brokers in equilibrium. We then analyze how information collection is affected by the intensity of competition between data brokers. In order to do so, we focus on two dimensions of competition: the market shares of data brokers, and the size of the market on which they compete. Then, we study how increasing the intensity of competition between data brokers changes the amount of data that they collect k_γ.

4.3.3.1 Profits of the data brokers in equilibrium

The profits of the data brokers on the competitive sub-market l can be written $[2w_{l,\gamma}(k_\gamma) - c(k_\gamma)]$. The total profits of the data brokers can thus be written:

$$\Pi_\gamma(k_\gamma) = m_\gamma[2w_{m_\gamma}(k_\gamma) - c(k_\gamma)] + l[2w_{l,\gamma}(k_\gamma) - c(k_\gamma)]. \tag{4.8}$$

The first term of this equation corresponds to the profit of data broker γ on the monopolistic market. Information is sold at a price $w_{m_\gamma}(k_\gamma)$, to both firms, and thus payoff is $2w_{m_\gamma}(k_\gamma)$. Information collection costs $c(k_\gamma)$, which gives net profits $2w_{m_\gamma}(k_\gamma) - c(k_\gamma)$. This value is weighted by the size of this monopolistic sub-market m_γ.

The second term corresponds to the profits of data broker γ on the competitive market, and is constructed similarly. $w_{l,\gamma}(k_\gamma)$ is the price of information and l the size of this market.

In equilibrium, the data broker with the highest m_γ sets the highest k_γ.

If $m_1 > m_2$:

$$\begin{aligned} \Pi_1 &= m_1[2w_{m_1}(k_1) - c(k_1)] + l[2w_{l,1}(k_1) - c(k_1)] \\ \Pi_2 &= m_2[2w_{m_2}(k_2) - c(k_2)] - lc(k_2) \end{aligned} \tag{4.9}$$

If $m_1 = m_2 > 0$:

$$\Pi_\gamma = m_\gamma[2w_{m_\gamma}(k_\gamma) - c(k_\gamma)] - lc(k_\gamma).$$

By assumption, each data broker has a unique optimal k_γ which depends on the structure of the market. We detail the property of k_γ in the following section.

4.3.3.2 Rent-extraction effect

We now analyze whether competition between data brokers increases or decreases the amount of personal data that they collect. To do so, we need to keep in mind that data broker γ operates on two markets: sub-market m_γ and sub-market l. Recall that the two sub-markets are related by Assumption 1 that states that a given data broker collects the same amount of information on both m_γ and l.[13] Thus, it is important to keep track of the relative size of these two sub-markets. Moreover, by observing the primitive l, data protection authorities can determine which scenarios on data collection are likely to occur. Proposition 9 analyzes how the amount of data collected by data brokers changes with the size of sub-market l.

Proposition 9:

- If $m_1 > m_2$: $k_1^* > k_2^*$ and
  - k_1^* increases with l.
  - k_2^* decreases with l.
- If $m_1 = m_2$: $k_1^* = k_2^*$ and
  - $k_1^* = k_2^*$ decrease with l.

Proof: See Appendix 4.6.

[13] This is not a strong assumption, as our result would hold under weaker conditions on the link between the two sub-markets.

When data broker 1 has a bigger market share than data broker 2 ($m_1 > m_2$), he collects more data than its competitor, and in addition, the amount of data collected increases with the size of the competitive sub-market l. This is due to the fact that collecting information becomes more profitable for the dominant data broker. This rent-extraction effect decreases consumer surplus. It is worth noting that competition between data brokers leads to a snowballing effect identified by Begenau et al. (2018): large firms have more market power and collect more information while small firms collect less information. When both data brokers have similar monopoly sub-markets ($m_1 = m_2$), competition between data brokers is symmetric, and they collect less personal data since they make zero profit on l, regardless of the amount of data that they collect.

These results are related to the question, relatively new in the literature, about how competition between firms affects the amount of data that they collect. In particular, our results nuance the findings of the empirical study of Kesler et al. (2017) that finds that competition between firms (apps) is negatively correlated with the amount of consumer information that they collect on Google Playstore. Indeed, there are two dimensions of competition to consider: the difference in market shares between data brokers on the monopoly markets, and the size of the competitive sub-market l. On the one hand, when $m_1 = m_2$, the intensity of competition is the strongest, which leads to a Bertrand outcome. We have shown that, in that case, data brokers collect less information because the marginal value of information is smaller. On the other hand, when $m_1 > m_2$, only one data broker sells information, and collects more information when the size of the competitive sub-market l increases. This challenges the results of Kesler et al. (2017) who only use the number of apps on Google Playstore as a measure of competition.

To summarize, the intensity of competition can be captured by the difference of market shares between data brokers (m_1 and m_2), and by l, the size of the market where data brokers compete. Both parameters need to be taken into account when assessing the effects of competition between data brokers on the amount of data that they collect.

4.4 Consumer surplus

We consider in this section the effect of competition between data brokers on consumer surplus. We focus on sub-market l where data brokers compete.

Proposition 10:

On sub-market l where data brokers compete, for $k_{l,\gamma}, k_{m_\gamma} > 4$, competition between data brokers always increases consumer surplus.

Proof: see Appendix 4.6.

Competition between data brokers has two opposite effects on consumer surplus. On the one hand, we have shown in Proposition 8 that when data brokers compete, they sell information that allows firms to identify a larger share of consumers. This is the competing-gatekeeper effect that increases consumer surplus, due to more intense competition on the product market. Indeed, on sub-market l where data brokers compete, the last segment of consumers sold to firms is located further away from the extremities than under monopoly. On the other hand, as stated in Proposition 9, either data brokers are symmetric in terms of market shares, and they collect less information when l increases; or data brokers are asymmetric in terms of market shares, and the amount of information collected by data broker 1 increases with the size of l. The competing-gatekeeper effect increases consumer surplus, while the rent-extraction effect decreases consumer surplus. Overall, Proposition 10 shows that the competing-gatekeeper effect always dominates the rent-extraction effect when considering consumer surplus. Consumers on sub-market l always benefit from competition between data brokers.

The effect of competition between data brokers on total consumer surplus is ambiguous. Consumers benefit from competition on l and on m_2 where information precision is lowered by competition, but they suffer from additional rent extraction on sub-market m_1 when data brokers are asymmetric. The effect of competition between data brokers thus depends on the relative sizes of the sub-markets.

4.5 Discussion

We summarize our three main results. First, we show in Proposition 8 that competing data brokers sell more consumer information to firms than in their monopoly market. Second, we also show that data broker 1 collects more personal data on the competitive sub-market compared to a situation in which data broker 1 is in a monopoly position on l. Third, consumer surplus always increases when data brokers compete.

The relation between the intensity of competition between data brokers and how much consumer information they collect is more and more scrutinized by protection agencies and competition authorities. On the one hand, recent legislations such as the European GDPR impose a data minimization principle. Data brokers are growing fast, collecting any type of information on huge masses of consumers[14], understanding how competition between data brokers changes under the GDPR is of crucial importance. Moreover, data breaches are becoming more and more frequent, and it is important to know how much personal information data brokers collect. On the other hand, competition authorities are more concerned by how much information is sold in the downstream market since more information usually means more competition between firms. However, the European Commission has overlooked the role of consumer information when assessing the effects of a merger between firms.[15] Since the acquisition of

[14] Data brokers: regulators try to rein in the 'privacy deathstars'.
[15] Data in Eu Merger Control.

Linkedin by Microsoft[16], practitioners have started to realize that merging data sets can have different consequences than merging companies, and that data brokers can affect the intensity of competition between downstream firms.[17]

This paper has analyzed the effects of competition between data brokers on how much data they collect and sell to downstream firms competing on the product market (the two faces of data brokers). We have argued that these two faces of data brokers are highly entangled. Thus, considering each dimension of a data broker separately can be misleading . This supports the idea that data protection authorities should closely work hand in hand with competition authorities.

Finally, our paper demonstrates that it is essential to assess the market conditions in which data brokers operate. When data brokers are symmetric in terms of market shares, increasing the intensity of competition between them has positive outcomes for consumers: data brokers collect less information and the downstream market is more competitive. However, when there is an asymmetry in market size between the two data brokers, only one data broker sells information, and he collects and sells more data. Our results call again for a better integration between data protection agencies and competition authorities.

4.6 Appendix

Firms' profits

We assume without loss of generality that $k_1 \geq k_2$. By definition, Firms' profits are:

[16] Microsoft officially closes its $26.2B acquisition of LinkeIn.
[17] Facebook Used People's Data to Favor Certain Partners and Punish Rivals, Documents Show.

$$\begin{cases} \pi_{1,2} = [\pi(j_1^{\theta}, j_1^{-\theta}, k_1) - w_{m_1}^{\theta}(j_1^{\theta}, j_{-\theta}^{2}, k_1)]m_1 \\ \text{from acquiring information from DB 1 on } m_1 \\ +[\pi(j_2^{\theta}, j_2^{-\theta}, k_2) - w_{m_2}^{\theta}(j_2^{\theta}, j_2^{-\theta}, k_2)]m_2 \\ \text{from acquiring information from DB 2 } m_2 \\ +[\pi(j_{l,1}^{\theta}, j_{l,1}^{-\theta}, k_1) - w_{l,1}^{\theta}(j_{l,1}^{\theta}, j_{l,1}^{-\theta}, k_1)]l \\ \text{from acquiring information from DB 1 on } l \\ +(1 - m_1 - m_2 - l)\pi \\ \text{on consumers who remain hidden} \end{cases} \tag{4.10}$$

Where 1) and 2) are the profits of firms when they acquire information from respectively data broker 1 and 2, on sub-markets where they are monopolists. 3) corresponds to firms profit on sub-market l where data brokers compete. On 4) firms do not acquire information and compete in the classical Hotelling way.

In the limit case where $k_\gamma = \infty$, firms perfectly recognize consumers on whom they have information on m_γ and l and price discriminate them at the first degree.

Properties of the cost function

The cost function is defined such that:

$$\begin{cases} \frac{\partial^2[2w_{m_\gamma}(k_\gamma) - c(k_\gamma)]}{\partial k_\gamma^2} < 0 \text{ and } \exists! \ k_\gamma^* \ \ s.t. \ \frac{\partial[2w_{m_\gamma}(k_\gamma) - c(k_\gamma)]}{\partial k_\gamma} = 0 \\ \frac{\partial^2[[2w_{l,\gamma}(k_\gamma) - c(k_\gamma)]}{\partial k_\gamma^2} < 0 \text{ and } \exists! \ k_\gamma^* \ \ s.t. \ \frac{\partial[[2w_{l,\gamma}(k_\gamma) - c(k_\gamma)]}{\partial k_\gamma} = 0 \\ \exists! \ k_\gamma^* \ \ s.t. \ \frac{\partial \Pi_\gamma}{\partial k_\gamma} = 0 \text{ and } \Pi_\gamma(k_\gamma^*) \geq 0 \\ c(0) = 0 \end{cases}$$

This technical hypothesis is common in the literature. It allows profits to be maximized in a unique point, which is usually true for linear cost functions.

No consumer poaching condition

Assumption 2: **(No consumer poaching condition)**

When data broker γ sells a partition $\mathcal{P} = \{[0, \frac{s_1}{k_\gamma}], .., [\frac{s_i}{k_\gamma}, \frac{s_{i+1}}{k_\gamma}], .., [\frac{s_{n-1}}{k_\gamma}, 1]\}$ to Firm 1 and $\mathcal{P}' = \{[0, \frac{s'_{n'-1}}{k_\gamma}], .., [\frac{s'_{i'+1}}{k_\gamma}, \frac{s'_{i'}}{k_\gamma}], .., [\frac{s'_1}{k_\gamma}, 1]\}$ to Firm 2, the segments verify: $2\frac{s_{i+1}}{k_\gamma} - \frac{s_i}{k_\gamma} \leq \frac{1}{2}$ and $2\frac{s'_{i'+1}}{k_\gamma} - \frac{s'_{i'}}{k_\gamma} \leq \frac{1}{2}$ for $i = 0, .., n-2$, $i' = 0, .., n'-2$.[18]

Notations

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

A firm having information of the form $\{[0, \frac{s_1}{k_\gamma}], [\frac{s_1}{k_\gamma}, \frac{s_2}{k_\gamma}], .., [\frac{s_{n-1}}{k_\gamma}, 1]\}$ 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 $[0, \frac{s_1}{k_\gamma}]$ price p_1, consumers on $[\frac{s_i}{k_\gamma}, \frac{s_{i+1}}{k_\gamma}]$ price p_{i+1}, and so forth for each segment.

We allow data brokers to sell a partition different from $\mathcal{P}_{\gamma,ref}$. In fact, it can sell any information partition belonging to $\mathbb{P}_\gamma$. However, we rule out information partitions that generate uncertainty over the location of the elementary segment of size $\frac{1}{k_\gamma}$ to which a consumer belongs. As an illustration, suppose that $k_\gamma = 8$ so that the finest partition consists of 8 segments of size $\frac{1}{8}$. Suppose also that data broker γ sells a partition consisting of 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

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

information partition is therefore the partition $\{\{1,3\},2\}$. Segments 1 and 3 are not connected and are therefore excluded from our analysis.

Mathematical Appendix

Firms' profits

- Segments A, where Firm 1 is in constrained monopoly;
- Segments B, where Firms 1 and 2 compete.

We can write profits of Firm 1 as the sum of two terms. The first term represents the profits on segments of type A where prices are denoted by p'_{1i}. The second term represents the profits on segments of type B, where prices are denoted by p_{1i}.

Without loss of generality, we rewrite the notation of type A and B segments. Segments of type A are of size $\frac{1}{k}$ and are located at $\frac{u_i-1}{k}$, and segments of type B, are located at $\frac{s_i}{k}$ and are of size $\frac{l_i}{k}$.[19] There are $n \in \mathbb{N}$ segments of type A, of size $\frac{1}{k}$. On each of these segments, the demand is $\frac{1}{k}$. There are $n' \in \mathbb{N}$ segments of type B. We find the demand for Firm 1 on these segments using the location of the indifferent consumer:

$$d_{1i} = x - \frac{s_i}{k} = \frac{p_2 - p_{1i} + t}{2t} - \frac{s_i}{k}$$

We can rewrite the profits as:

$$\pi_1(\tilde{\mathcal{P}}) = \sum_{i=1}^{n} p'_{1i}\frac{1}{k} + \sum_{i=1}^{n'} p_{1i}[\frac{p_2 - p_{1i} + t}{2t} - \frac{s_i}{k}]$$

Profits of Firm 2 are generated on segments of type B, where the demand for Firm 2 is

$$d_{2i} = \frac{s_i + l_i}{k} - x = \frac{p_{1i} - p_2 - t}{2t} + \frac{s_i + l_i}{k}$$

[19] With u_i and s_i integers below k. See Section 4.2.2.2.

Profits of Firm 2 can be written therefore as

$$\pi_2(\tilde{\mathcal{P}}) = \sum_{i=1}^{n'} p_2[\frac{p_{1i} - p_2 - t}{2t} + \frac{s_i + l_i}{k}] \tag{4.11}$$

Firm 1 maximizes profits $\pi_1(\mathcal{P})$ with respect to p_{1i} and p'_{1i}, and Firm 2 maximizes $\pi_2(\mathcal{P})$ with respect to p_2, both profits are strictly concave.

Equilibrium prices are:

$$\begin{aligned} p'_{1i} &= t + p_2 - 2\frac{u_i t}{k} \\ p_{1i} &= \frac{p_2 + t}{2} - \frac{s_i t}{k} = \frac{t}{3} + \frac{2t}{3n'}[\sum_{i=1}^{n'}[\frac{s_i}{2k} + \frac{l_i}{k}]] - \frac{s_i t}{k} \\ p_2 &= -\frac{t}{3} + \frac{4t}{3n'}\sum_{i=1}^{n'}[\frac{s_i}{2k} + \frac{l_i}{k}] \end{aligned} \tag{4.12}$$

Proof of Proposition 6

We assume that $k_1 \geq k_2$. We show that in equilibrium data broker 1 sells information to both firms. Firms can either buy no information, or information from one of the data brokers. The information acquisition game can be described by the Nash table:

		Firm 1		
		NI	I_{DB1}	I_{DB2}
Firm 2	NI	(π_1^{NI}, π_2^{NI})	$(\pi_1^{I_{DB1}}, \pi_2^{NI})$	$(\pi_1^{I_{DB2}}, \pi_2^{NI})$
	I_{DB1}	$(\pi_1^{NI}, \pi_2^{I_{DB1}})$	$(\pi_1^{I_{DB1}}, \pi_2^{I_{DB1}})$	$(\pi_1^{I_{DB2}}, \pi_2^{I_{DB1}})$
	I_{DB2}	$(\pi_1^{NI}, \pi_2^{I_{DB2}})$	$(\pi_1^{I_{DB1}}, \pi_2^{I_{DB2}})$	$(\pi_1^{I_{DB2}}, \pi_2^{I_{DB2}})$

Both firms are eventually informed. Assume the opposite. Then, only one firm acquires information from one data broker. Then the other data broker makes zero profit. It is thus profitable for him to sell information to the uninformed firm.

In this situation, both data broker have interest to propose information to both firms. Firms acquire information from the data broker who maximizes their profits. Necessarily, firms acquire information from data broker 1 since any partition that data broker 2 can propose, data broker 1 can propose too.

Proof of Theorem 3

We prove that the partition described in Theorem 3 is optimal for data brokers on sub-market l, where they compete. For each firm, the partition divides the unit line into two segments. The first segment identifies the closest consumers to a firm and is partitioned in $j_{l,\gamma}$ segments of size $\frac{1}{k_\gamma}$. The second segment is of size $1 - \frac{j_{l,\gamma}}{k_\gamma}$ and leaves unidentified the other consumers.

Without loss of generality, we take here $k_1 \geq k_2$

Three types of segments are defined:

- Segments A, where Firm θ is in constrained monopoly;
- Segment B, where firms 1 and 2 compete;
- Segments C, where Firm θ gets no demand.

We use Assumption 2 to show that the unit line is composed of one type B segment where firms compete, located at the middle of the line, and segments where firms are monopolists, located close to them.

On l where data brokers compete, data broker 1 sells to firm θ an information partition $\mathcal{P}_1^\theta$ at the highest price that firms are ready to pay, which corresponds now to the difference of profits firms make between the situation where they buy information, and their outside option, which is buying information from data broker 2:

$$w_{m_1}^\theta(\mathcal{P}_1^\theta, \mathcal{P}_1^{-\theta}, k_1) = \pi_\theta(\mathcal{P}_1^\theta, \mathcal{P}_1^{-\theta}, k_1) - \pi_\theta(\mathcal{P}_2^\theta, \mathcal{P}_1^{-\theta}, k_2).$$

Data broker 2 (without loss of generality), with the lowest precision of information will propose an information partition that maximizes the profits of each firm.

$$\Pi_2(\mathcal{P}_2^\theta, k_2) = max_{\mathcal{P}_2^\theta, \mathcal{P}_2^{-\theta}}\{\pi_\theta(\mathcal{P}_2^\theta, \mathcal{P}_2^{-\theta}, k_2)\}. \tag{4.13}$$

Firm θ buys a partition composed of segments of type A and one segment of type B, as it is given by Assumption 2. We show that a partition in which type A segments are of size $\frac{1}{k_\gamma}$ maximizes $\pi_\theta(\mathcal{P}_\gamma^\theta, \mathcal{P}_\gamma^{-\theta}, k_\gamma)$.

A partition which maximizes $\pi_\theta(\mathcal{P}_\gamma, k_\gamma)$ is necessarily composed of type A segments of size $\frac{1}{k_\gamma}$.

We analyze segments of type A where Firm 1 is in constrained monopoly, and we show that reducing the size of segments to $\frac{1}{k_\gamma}$ is optimal.

Consider any segment $[\frac{i}{k_\gamma}, \frac{i+h}{k_\gamma}]$ with h, i integers verifying $i + h \leq k_\gamma$ and $h \geq 2$, such that Firm 1 is in constrained monopoly on this segment. We show that selling a finer partition of this segment increases the profits of Firm 1. To prove this claim, we establish that Firm 1 profits is higher with a finer partition $\mathcal{P}'$ with two segments : $[\frac{i}{k_\gamma}, \frac{i+1}{k_\gamma}]$ and $[\frac{i+1}{k_\gamma}, \frac{i+h}{k_\gamma}]$ than with a coarser partition $\mathcal{P}$ with one segment $[\frac{i}{k_\gamma}, \frac{i+h}{k_\gamma}]$.

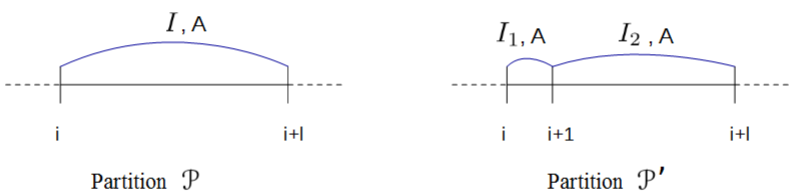

Fig. 4.7: Step 1: segments of type A

Figure 4.7 shows on the left panel a partition with a coarse segment of type A, and on the right, finer segments of type A. We compare profits in both situations and show that the finer segmentation is more profitable for Firm 1. We write $\pi_1^A(\mathcal{P})$ and $\pi_1^{AA}(\mathcal{P}')$ the profits of Firm 1 on $[\frac{i}{k_\gamma}, \frac{i+h}{k_\gamma}]$ for respectively partitions $\mathcal{P}$ and $\mathcal{P}'$.

First, profits with the coarser partition is: $\pi_1^A(\mathcal{P}) = p_{1i}d_1 = p_{1i}\frac{h}{k_\gamma}$. The demand is $\frac{l}{k_\gamma}$ as Firm 1 gets all consumers by assumption; p_{1i} is such that the indifferent consumer x is located at $\frac{i+h}{k_\gamma}$:

$$V - tx - p_{1i} = V - t(1-x) - p_2 \implies x = \frac{p_2 - p_{1i} + t}{2t} = \frac{i+h}{k_\gamma} \implies p_{1i} = p_2 + t - 2t\frac{i+h}{k_\gamma},$$

with p_2 the price charged by (uninformed) Firm 2. This price is only affected by strategic interactions on the segments where firms compete, and therefore does not depend on the pricing strategy of Firm 1 on type A segments.

We write the profit function for any p_2, replacing p_{1i} and d_1:

$$\pi_1^A(\mathcal{P}) = \frac{h}{k_\gamma}(t + p_2 - \frac{2(h+i)t}{k_\gamma})$$

Secondly, using a similar argument, we show that the profit on $[i, i+h]$ with partition $\mathcal{P}'$ is:

$$\pi_1^{AA}(\mathcal{P}') = \frac{1}{k_\gamma}(t + p_2 - \frac{2(1+i)t}{k_\gamma}) + \frac{h-1}{k_\gamma}(t + p_2 - \frac{2(h+i)t}{k_\gamma})$$

Comparing $\mathcal{P}$ and $\mathcal{P}'$ shows that the Firm 1 with the finer partition increases by $\frac{2t}{k_\gamma^2}(h-1)$, which establishes the claim.

By repeating the previous argument, it is easy to show that the data broker will sell a partition of size $\frac{h}{k_\gamma}$ with h segments of equal size $\frac{1}{k_\gamma}$.

Assumption 2 implies that, even when only one firm is informed, the unit line is divided in type A and one type B segments. It is immediate to show that the profit of the uninformed firm does not depend on the fineness of type

A segments. As a result, $\Pi_\gamma(\mathcal{P}_\gamma, k_\gamma)$ is maximized when segments of type A are of size $\frac{1}{k_\gamma}$.

We deduce that the optimal partition for both data brokers is composed of two segments, sold to each firm. For Firm 1, the first segment is partitioned in j segments of size $\frac{1}{k_\gamma}$, and is located at $[0, \frac{j}{k_\gamma}]$. The second segment is of size $1 - \frac{j}{k_\gamma}$, located at $[\frac{j}{k_\gamma}, 1]$ and is composed of unidentified consumers. For Firm 2, the first segment is partitioned in j' segments of size $\frac{1}{k_\gamma}$, and is located at $[1 - \frac{j'}{k_\gamma}, 1]$. The second segment is of size $1 - \frac{j'}{k_\gamma}$, located at $[0, 1 - \frac{j'}{k_\gamma}]$ and is composed of unidentified consumers.

The profit of Firm θ with information j_γ^θ when its competitor is informed with $j_\gamma^{-\theta}$ is:

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

Firms' profits with the optimal information partition

When the data broker sells the optimal information partition, firms profits are:

$$\begin{aligned}\pi_1^* &= \sum_{i=1}^{j_1} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{j_1}{k} - \frac{1}{3}\frac{j_2}{k})t[1 - \frac{2}{3}\frac{j_2}{k} - \frac{4}{3}\frac{j_1}{k}] \\ &= \frac{t}{2} - \frac{7}{9}\frac{j_1^2 t}{k^2} + \frac{2}{9}\frac{j_2^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_1 t}{k} - \frac{2}{3}\frac{j_2 t}{k} - \frac{j_1 t}{k^2}\end{aligned}$$

$$\begin{aligned}\pi_2^* &= \sum_{i=1}^{j_2} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{j_2}{k} - \frac{1}{3}\frac{j_1}{k})t[1 - \frac{2}{3}\frac{j_1}{k} - \frac{4}{3}\frac{j_2}{k}] \\ &= \frac{t}{2} - \frac{7}{9}\frac{j_2^2 t}{k^2} + \frac{2}{9}\frac{j_1^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_2 t}{k} - \frac{2}{3}\frac{j_1 t}{k} - \frac{j_2 t}{k^2}\end{aligned}$$

The proof is given in Bounie et al. (2018).

Proof of Proposition 7

Optimal information partition

Data broker 1 maximizes the sum of the difference $\pi_1(j^1_{l,1}, j^2_{l,1}) - \pi_1(j^1_{l,2}, j^2_{l,1}) + \pi_2(j^2_{l,1}, j^1_{l,1}) - \pi_2(j^2_{l,2}, j^1_{l,1})$ with respect with $j^1_{l,1}$ and $j^2_{l,1}$. Data broker 2 maximizes firms profits $\pi_1(j^1_{l,2}, j^2_{l,1})$ and $\pi_2(j^2_{l,2}, j^1_{l,1})$.

Thus data broker 1 maximizes

$$\pi_1(j^1_{l,1}, j^2_{l,1}) - \pi_1(j^1_{l,2}, j^2_{l,1}) + \pi_2(j^2_{l,1}, j^1_{l,1}) - \pi_2(j^2_{l,2}, j^1_{l,1}) = \frac{(7k_2k_1(j^2_{l,2})^2 + (4k_2j^1_{l,1} - 6k_2 + 9)k_1j^2_{l,2} + 7k_2k_1(j^1_{l,2})^2 + (4k_2j^2_{l,1} - 6k_2 + 9)k_1j^1_{l,2})t}{9k_2k_1} + \frac{((-7k_2(j^2_{l,1})^2 + (6k_2 - 8k_2j^1_{l,1})j^2_{l,1} - 7k_2(j^1_{l,1})^2 + 6k_2j^1_{l,1})k_1 - 9k_2j^2_{l,1} - 9k_2j^1_{l,1})t}{9k_2k_1} \tag{4.14}$$

with respect with $j^1_{l,1}$ and $j^2_{l,1}$.

And data broker 2 maximizes

$$\pi_\theta(j^\theta_{l,2}, j^{-\theta}_{l,1}) = \frac{t}{2} - \frac{7}{9}\frac{(j^\theta_{l,2})^2t}{k^2} + \frac{2}{9}\frac{(j^{-\theta}_{l,1})^2t}{k^2} - \frac{4}{9}\frac{j^\theta_{l,2}j^{-\theta}_{l,1}t}{k^2} + \frac{2}{3}\frac{j^\theta_{l,2}t}{k} - \frac{2}{3}\frac{j^{-\theta}_{l,1}t}{k} - \frac{j^\theta_{l,2}t}{k^2}$$

with respect with $j^1_{l,2}$ and $j^2_{l,2}$.

FOC on $j^1_{l,1}$, $j^2_{l,1}$, $j^1_{l,2}$ and $j^2_{l,2}$ give respectively in equilibrium:

$j^{1*}_{l,1} = j^{2*}_{l,1} = \frac{1}{3} - \frac{1}{9k_2} - \frac{7}{18k_1}$

$j^{1*}_{l,2} = j^{2*}_{l,2} = \frac{1}{3} - \frac{11}{18k_2} + \frac{1}{9k_1}$

Price of information

We substitute the values of $j^{1*}_{l,1}$, $j^{2*}_{l,1}$ and $j^{1*}_{l,2}$ in $\pi_1(j^1_{l,1}, j^2_{l,1}) - \pi_1(j^1_{l,2}, j^2_{l,1})$. The price of information is

$$w_l(k_2) = 0$$

$$w_l(k_1) = [\pi_1(j^1_{l,1}, j^2_{l,1}) - \pi_1(j^1_{l,2}, j^2_{l,1})] \tag{4.15}$$
$$= \frac{((12k_2 - 11)k_1^2 + (4k_2 - 12k_2^2)k_1 + 7k_2^2)t}{36k_2^2k_1^2}$$

which clearly increases in k_1.

Proof of Prop 9

The data broker collects more information when selling information to both firms

Data brokers collect more information when sell it to both firms than when they sell it to one firm only.

When selling information to both competitors, a data broker has revenue

$$2w_{m_\gamma}(k_\gamma) = \frac{2t}{11} - \frac{6t}{11k} + \frac{9t}{22k^2}.$$

$$\frac{\partial w_{m_\gamma}(k_\gamma)}{\partial k_\gamma} = \frac{6t}{11k^2} - \frac{9t}{11k^3}$$

and

$$\frac{\partial^2 w_{m_\gamma}(k_\gamma)}{\partial k_\gamma^2} = \frac{27t}{11k_\gamma^4} - \frac{12t}{11k_\gamma^3} \leq 0$$

for $k \in [\frac{9}{4}, \infty[$

When selling information to one firms only, data broker has revenue

$$w(k_\gamma) = \frac{t}{7} - \frac{3t}{7k_\gamma} + \frac{9t}{28k_\gamma^2}$$

$$\frac{\partial w(k_\gamma)}{\partial k_\gamma} = \frac{3t}{7k_\gamma^2} - \frac{9t}{14k_\gamma^3}$$

Comparing both derivatives we have

$$2\frac{\partial w_{m_\gamma}(k_\gamma)}{\partial k_\gamma} > \frac{\partial w(k_\gamma)}{\partial k_\gamma}$$

for $k_\gamma \in [\frac{3}{2}, \infty[$

Costs are increasing in k.

Take $k^* s.t. \frac{\partial w(k_\gamma)}{\partial k_\gamma}|_{k_\gamma=k_\gamma^*} = \frac{\partial c(k_\gamma)}{\partial k_\gamma}|_{k_\gamma=k_\gamma^*}$, that is, the optimal k_γ when information is sold to one firm.

Plugging k^* into $\frac{\partial w_{m_\gamma}(k_\gamma)}{\partial k_\gamma}$, we have

$$2\frac{\partial w_{m_\gamma}(k_\gamma)}{\partial k_\gamma}|_{k_\gamma=k^*} > \frac{\partial c(k_\gamma)}{\partial k_\gamma}|_{k_\gamma=k^*}$$

Since costs are assumed to be convex and we have proved that revenues are concave, it is immediate that $\tilde{k}^*$ such that

$$2\frac{\partial w_{m_\gamma}(k)}{\partial k}|_{k_\gamma=\tilde{k}^*} = \frac{\partial c(k)}{\partial k}|_{k_\gamma=\tilde{k}^*}$$

verifies

$$\tilde{k}^* > k^*$$

Considering information selling to both firms, for any j:

$$2w_{m_\gamma}(k_\gamma) = \frac{4jt}{3k} - \frac{22j^2t}{9k^2} - \frac{2jt}{k^2}.$$

$$\frac{\partial w_{m_\gamma}(k_\gamma, j)}{\partial k_\gamma} = -\frac{4jt}{3k^2} + \frac{44j^2t}{9k^3} + \frac{4jt}{k^3}$$

$$\frac{\partial^2 w_{m_\gamma}(k_\gamma, j)}{\partial k_\gamma^2} = \frac{8jt}{3k_\gamma^3} - \frac{44j^2t}{3k_\gamma^4} - \frac{12jt}{k^4}$$

Take

$$k_\gamma^*(j) \quad s.t. \quad \frac{\partial w_{m_\gamma}(k_\gamma)}{\partial k_\gamma}|_{k_\gamma = k_\gamma^*} = \frac{\partial c(k_\gamma)}{\partial k_\gamma}|_{k_\gamma = k_\gamma^*}$$

that is, the optimal k_γ when information is sold until j.

Consider $j' \geq j$, then we have

$$\frac{\partial w_{m_\gamma}(k_\gamma, j')}{\partial k_\gamma}|_{k_\gamma = k_\gamma^*} \geq \frac{\partial w_{m_\gamma}(k_\gamma, j)}{\partial k_\gamma}|_{k_\gamma = k_\gamma^*}$$

Again, since costs are assumed to be weakly convex and we have proved that revenues are concave, it is immediate that $\tilde{k}^*$ such that

$$2\frac{\partial w_{m_\gamma}(k)}{\partial k}|_{k_\gamma = \tilde{k}^*} = \frac{\partial c(k)}{\partial k}|_{k_\gamma = \tilde{k}^*}$$

verifies

$$\tilde{k}^* > k^*$$

Information collection increases with competition

Costs are assumed to be such that $[2w_{m_1}(k_1) - c(k_1)]$ and $[2w_l(k_1) - c(k_1)]$ have interior solutions.

We prove now that the optimal collection of information k_1 is lower for $[2w_{m_1}(k_1) - c(k_1)]$ than for $[2w_l(k_1) - c(k_1)]$.

For $k_1 > k_2$, we have

$$w_l(k_1) = \frac{((12k_2 - 11)k_1^2 + (4k_2 - 12k_2^2)k_1 + 7k_2^2)t}{36k_1^2k_2^2}$$

$$2\frac{\partial w_l(k_1)}{\partial k_1} = \frac{((6k_1 - 2)k_2 - 7k_1)t}{9k_1^3k_2}$$

$$2w_{m_1}(k_1) = \frac{2t}{11} - \frac{6t}{11k_1} + \frac{9t}{22k_1^2}.$$

and

$$2\frac{\partial w_{m_1}(k_1)}{\partial k_1} = \frac{6t}{11k_1^2} - \frac{9t}{11k_1^3} \leq \frac{\partial w_l(k_1)}{\partial k_1} = \frac{((6k_1-2)k_2 - 7k_1)t}{9k_1^3 k_2}$$

Consider k^* such that

$$2\frac{\partial w_l(k_1)}{\partial k_1}|_{k_1=k^*} = \frac{((6k^*-2)k_2 - 7k^*)t}{9k^{*3}k_2} = \frac{\partial c(k_1)}{\partial k_1}|_{k_1=k^*}$$

Since

$$\frac{\partial^2 w_{m_1}(k_1)}{\partial k_1^2} = \frac{27t}{11k_1^4} - \frac{12t}{11k_1^3} \leq 0$$

for $k \in [2, \infty[$

revenues are concave, and necessarily, $\tilde{k}^*$ such that

$$\frac{\partial w_{m_1}(k_1)}{\partial k_1} = \frac{\partial c(k_1)}{\partial k_1}$$

verifies $\tilde{k}^* \leq k^*$.

Thus information collection is higher on markets where data brokers compete.

We show now that the higher the competition between data brokers, the higher the information collection by data broker 1 in equilibrium k_1^*. Costs are assumed to be such that $[2w_{m_1}(k_1) - c(k_1)]$ and $[2w_l(k_1) - c(k_1)]$ are respectively maximized for $\tilde{k_1} < \hat{k_1}$.

It is immediate to see that $\tilde{k_1} \geq k_1^* \geq \hat{k_1}$.

Thus, since $[2w_{m_1}(k_1) - c(k_1)]$ and $[2w_l(k_1) - c(k_1)]$ are strictly concave, we can conclude that $[2w_{m_1}(k_1^*) - c(k_1^*)]' < 0$ and $[2w_l(k_1^*) - c(k_1^*)]' > 0$.

In order to show that k_1^* increases with $\frac{l}{m_1}$, we consider l_1 and l_2 such that $\frac{l_1}{m_{1,1}} \geq \frac{l_2}{m_{1,2}}$, and $k_{1,1}^*$ and $k_{1,2}^*$ such that:

$$\begin{aligned} &\frac{\partial(m_{1,1}[2w_{m_1}(k_1) - c(k_1)] + l_1[2w_l(k_1) - c(k_1)])}{\partial k_1}(k_1 = k_{1,1}^*) = 0 \\ &\frac{\partial(m_{1,2}[2w_{m_1}(k_1) - c(k_1)] + l_2[2w_l(k_1) - c(k_1)])}{\partial k_1}(k_1 = k_{1,2}^*) = 0 \end{aligned} \tag{4.16}$$

Assume that $k_{1,1}^* \leq k_{1,2}^*$.

By concavity we have $[2w_l(k_{1,1}^*) - c(k_{1,1}^*)]' > [2w_l(k_{1,2}^*) - c(k_{1,2}^*)]'$, and since $\frac{l_1}{m_{1,1}} \geq \frac{l_2}{m_{1,2}}$:

$$\frac{l_1}{m_{1,1}}[2w_l(k_{1,1}^*) - c(k_{1,1}^*)]' > \frac{l_2}{m_{1,2}}[2w_l(k_{1,2}^*) - c(k_{1,2}^*)]'$$

A similar reasoning gives us:

$$[2w_{m_1}(k_{1,1}^*) - c(k_{1,1}^*)]' > [2w_{m_1}(k_{1,2}^*) - c(k_{1,2}^*)]'$$

which is contradictory with the fact that $k_{1,1}^*$ and $k_{1,2}^*$ maximize $\Pi_1 = m_1[2w_{m_1}(k_1) - c(k_1)] + l[2w_l(k_1) - c(k_1)]$.

Thus, $k_{1,1}^* \geq k_{1,2}^*$. Information collection by data broker 1 increases with the intensity of competition.

We show now that the higher the intensity of competition between data brokers, the less data broker 2 collects information.

The profits of data broker 2 are

$$\Pi_2 = m_2[2w_{m_2}(k_2) - c(k_2)] - lc(k_2)$$

In equilibrium, k_2^* verifies

$$2w'_{m_2}(k_2^*) = c'(k_2^*)[1 + \frac{l}{m_2}]$$

By strict concavity of w_{m_2} and convexity of c, it is immediate to see that k_2 decreases with $\frac{l}{m_2}$.

Proof of Lemma 1

In order to prove the convergence of the incremental share of identified consumers due to asymmetric competition to its value under symmetric competition, we explicit the numerical expression of both values:

$$\begin{aligned}\Delta s_0(k_1) &= s_0(k_1) - s(k_1)\\ &= \frac{1}{3} - \frac{1}{2k_1} - [\frac{3}{11} - \frac{9}{22k_1}]\\ &= \frac{2}{33} - \frac{1}{11k_1}\end{aligned}$$

$$\begin{aligned}\Delta s_1(k_1) &= s_1(k_1) - s(k_1)\\ &= \frac{1}{3} - \frac{1}{9k_2} - \frac{7}{18k_1} - [\frac{3}{11} - \frac{9}{22k_1}]\\ &= \frac{2}{33} - \frac{1}{9k_2} - \frac{7}{18k_1} + \frac{9}{22k_1}\end{aligned}$$

Subtracting both terms we have:

$$\begin{aligned}\Delta s_0(k_1) - \Delta s_1(k_1) &= -\frac{1}{11k_1} + \frac{1}{9k_2} + \frac{7}{18k_1} - \frac{9}{22k_1}\\ &= \frac{1}{9k_2} - \frac{1}{9k_1}\\ &\underset{k_1 \to^+ k_2}{\longrightarrow} 0\end{aligned}$$

Proof of Proposition 10

Th optimal information partition in monopoly and equilibrium are respectively $\frac{j^*_{m_1}}{k_1} = \frac{3}{11} - \frac{9}{22k_1}$ and $\frac{j^*_{l,1}}{k_1} = \frac{1}{3} - \frac{1}{9k_2} - \frac{7}{18k_1}$.

We compare the gain of consumer surplus in both cases, with information precision noted k_{m_γ} and $k_{l,\gamma}$ for respectively monopoly and competition:

$$\Delta CS(\frac{j_{m_1}}{k_{m_1}}) = \frac{36k_{m_1}^2 t + 24k_{m_1} t - 117t}{484k_{m_1}^2}$$

$$\Delta CS(\frac{j_{l,1}^*}{k_{l,1}}) = \frac{((36k_{l,2}^2 - 24k_{l,2} + 4)k_{l,1}^2 + (24k_{l,2}^2 - 8k_{l,2})k_{l,1} - 77k_{l,2}^2)t}{324k_{l,2}^2 k_{l,1}^2}$$

and

$$\begin{aligned}\Delta CS(\frac{j_{l,1}^*}{k_{l,1}}) - \Delta CS(\frac{j_{m_1}}{k_{m_1}}) = \\ \frac{((1440k_{l,2}^2 - 2904k_{l,2} + 484)k_{l,1}^2 + (2904k_{l,2}^2 - 968k_{l,2})k_{l,1} - 9317k_{l,2}^2)k_{m_1}^2 t}{39204k_{l,2}^2 k_{l,1}^2 k_{m_1}^2} \\ - \frac{(1944k_{l,2}^2 k_{l,1}^2 k_{m_1} + 9477k_{l,2}^2 k_{l,1}^2)t}{39204k_{l,2}^2 k_{l,1}^2 k_{m_1}^2}\end{aligned} \tag{4.17}$$

Which is always positive for

$$\begin{aligned}k_{l,1} \geq & \frac{11^{3/2} 9k_{l,2} k_{m_1} ((144k_{l,2}^2 - 264k_{l,2} + 44)k_{m_1}^2 - 168k_{l,2}^2 k_{m_1} + 819k_{l,2}^2)^{1/2}}{(1440k_{l,2}^2 - 2904k_{l,2} + 484)k_{m_1}^2 - 1944k_{l,2}^2 k_{m_1} + 9477k_{l,2}^2} \\ & + \frac{(484k_{l,2} - 1452k_{l,2}^2)k_{m_1}^2}{(1440k_{l,2}^2 - 2904k_{l,2} + 484)k_{m_1}^2 - 1944k_{l,2}^2 k_{m_1} + 9477k_{l,2}^2}\end{aligned} \tag{4.18}$$

Indeed, equation 4.17 is oriented upward in $k_{l,1}$. This can be proven since the factor of $k_{l,1}^2$ in this equation is $(1440k_{l,2}^2 - 2904k_{l,2} + 484)k_{m_1}^2 - 1944k_{l,2}^2 k_{m_1} + 9477k_{l,2}^2$. It is easy to show that this value is always positive for $k_{l,2} \geq 2$.

We show now that

$$\frac{11^{3/2}9k_{l,2}k_{m_1}((144k_{l,2}^2-264k_{l,2}+44)k_{m_1}^2-168k_{l,2}^2k_{m_1}+819k_{l,2}^2)^{1/2}}{(1440k_{l,2}^2-2904k_{l,2}+484)k_{m_1}^2-1944k_{l,2}^2k_{m_1}+9477k_{l,2}^2}$$
$$+\frac{(484k_{l,2}-1452k_{l,2}^2)k_{m_1}^2}{(1440k_{l,2}^2-2904k_{l,2}+484)k_{m_1}^2-1944k_{l,2}^2k_{m_1}+9477k_{l,2}^2} \quad (4.19)$$
$$\leq 4$$

and thus that for $k_{l,\gamma} \geq 4$, consumer surplus increases with competition between data brokers.

$$11^{3/2}9k_{l,2}k_{m_1}((144k_{l,2}^2-264k_{l,2}+44)k_{m_1}^2-168k_{l,2}^2k_{m_1}+819k_{l,2}^2)^{1/2}484k_{l,2}-1452k_{l,2}^2)k_{m_1}^2$$
$$\leq 4(1440k_{l,2}^2-2904k_{l,2}+484)k_{m_1}^2-1944k_{l,2}^2k_{m_1}+9477k_{l,2}^2 \quad (4.20)$$

for $k_{m_2} \geq 4$ and $k_{m_1} \geq 4$.

Thus, for $k_{l,1} \geq 4$, consumer surplus increases with competition between data brokers.

Conclusion

Consumer information has become a key strategic asset in most industries, shaping markets and societies. Yet, markets for information are characterized by a perplexing murkiness. Consumers and regulators thus face a spectacular lack of control over the collection and selling of consumer information. Data brokers stand out as the cornerstone of the emerging market for consumer information. They are now able to collect information on most consumers with a very high precision. Data brokers then use this information as information gatekeepers: they can shape competition on product markets by selling information on specific consumer segments to firms. We have shown that data brokers adopt competition-softening strategies that reduce consumer surplus. Thus, there is an urge for data protection agencies and competition authorities to adapt their regulatory practices to the new challenges raised by data brokers.

We have analyzed in this thesis how regulators can use the mechanisms through which information is sold to firms to limit consumer data collection and to increase consumer surplus. We argue that the choice of the mechanism through which information is sold to a firm is a simple yet powerful regulatory tool that should be further explored. Indeed, we show that selling information through a take it or leave it mechanism maximizes consumer welfare and minimizes the collection of consumer information. This mechanism fits with the objectives of competition authorities and of data protection agencies. However, data brokers never choose this selling mechanism as it minimizes their profits, and prefer to sell information by auction or sequential bargaining mechanisms.

Data brokers thus choose mechanisms that are harmful for consumers, as they lower their surplus, and intensify consumer data collection. Our results suggest that regulators should enforce that consumer information is sold through a take it or leave it selling, as it is optimal for consumers, but it is never adopted by data brokers.

Secondly, we focus on competition between data brokers. We show that the market structure influences the selling strategies of data brokers. Data brokers always sell more consumer segments when they compete than in a monopoly. The amount of consumer data that a data broker collects follows an inverted U-shape relation with the intensity of competition. On the one hand, when the data brokerage industry exhibits a firm dominant in terms of market share, we show that competition between data brokers is positively correlated with the amount of information that they collect. Data brokers collect and sell more consumer information in this case than in monopoly, which goes against the aims of a data protection authority. On the other hand, competition between data brokers increases competition on the product market, which benefits consumers, and is in line with the objectives of a competition authority. Thus, both agencies defend opposite views on the degree of competition needed in the data brokerage industry. Our results call for more integration between both data protection agencies and competition authorities, as competition and data are entangled and cannot be considered independently from each other.

Several questions should be further explored, in line with our conclusions. First, empirical tests of our results are welcome. How does the activity of data brokers affect competition on product markets? How does this relation changes when data broker compete more or less fiercely? Koski et al. (2018) gives interesting insights in this direction. Secondly, further research should explore practices of information sharing between data brokers. Sharing information allows data brokers to increase the precision of their data bases exponentially, which could strongly reduce consumer welfare. Thirdly, it is essential to understand the strategies of data brokers regarding on which consumer they collect information, which relates to their strategies of market expansion. Finally, the link between consumer information and innovation should be more than ever under the scope of scholars: do data brokers fuel or prevent innovation? Does consumer information represent a barrier to entry, and thus a threat for innovation? Both theoretical and empirical studies are welcome on this topic.

Bibliography

Anat R Admati and Paul Pfleiderer. A monopolistic market for information. *Journal of Economic Theory*, 39(2):400–438, 1986.

Anat R Admati and Paul Pfleiderer. Direct and indirect sale of information. *Econometrica: Journal of the Econometric Society*, pages 901–928, 1990.

George A Akerlof. The market for “lemons”: Quality uncertainty and the market mechanism. In *Uncertainty in Economics*, pages 235–251. Elsevier, 1978.

James J Anton and Dennis A Yao. The sale of ideas: Strategic disclosure, property rights, and contracting. *The Review of Economic Studies*, 69(3):513–531, 2002.

James J Anton and Dennis A Yao. Little patents and big secrets: managing intellectual property. *RAND Journal of Economics*, pages 1–22, 2004.

Ashish Arora and Andrea Fosfuri. Pricing diagnostic information. *Management Science*, 51(7):1092–1100, 2005.

Kenneth Joseph Arrow. Economic welfare and the allocation of resources for invention. In *Readings in industrial economics*, pages 219–236. Springer, 1972.

Matthew Backus, Thomas Blake, Bradley Larsen, and Steven Tadelis. Sequential bargaining in the field: Evidence from millions of online bargaining interactions. Technical report, National Bureau of Economic Research, 2018a.

Matthew Backus, Thomas Blake, and Steven Tadelis. Communication and bargaining breakdown: An empirical analysis. *NBER Working Paper No. xxxxxx*, 2018b.

Matthew Backus, Thomas Blake, and Steven Tadelis. On the empirical content of cheap-talk signaling: An application to bargaining. *Journal of Political Economy*, 127(4):000–000, 2019.

Patrick Bajari, Robert McMillan, and Steven Tadelis. Auctions versus negotiations in procurement: an empirical analysis. *The Journal of Law, Economics, & Organization*, 25(2):372–399, 2008.

Michael R Baye and John Morgan. Information gatekeepers on the internet and the competitiveness of homogeneous product markets. *American Economic Review*, 91 (3):454–474, 2001.

Juliane Begenau, Maryam Farboodi, and Laura Veldkamp. Big data in finance and the growth of large firms. *Journal of Monetary Economics*, 2018.

Paul Belleflamme, Wing Man Wynne Lam, and Wouter Vergote. Price discrimination and dispersion under asymmetric profiling of consumers. 2017.

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

Dirk Bergemann and Alessandro Bonatti. Markets for information: An introduction. *Annu. Rev. Econ*, 11:1–23, 2019.

Dirk Bergemann, Alessandro Bonatti, and Alex Smolin. The design and price of information. *American economic review*, 108(1):1–48, 2018.

Kostas Bimpikis, Davide Crapis, and Alireza Tahbaz-Salehi. Information sale and competition. *Management Science*, 2019.

Ken Binmore, Ariel Rubinstein, and Asher Wolinsky. The nash bargaining solution in economic modelling. *The RAND Journal of Economics*, pages 176–188, 1986.

Francis Bloch and Gabrielle Demange. Taxation and privacy protection on internet platforms. *Journal of Public Economic Theory*, 20(1):52–66, 2018.

David Bounie, Antoine Dubus, and Patrick Waelbroeck. Selling strategic information in digital competitive markets. 2018.

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

Jeremy M. Burke, Curtis R. Taylor, and Liad Wagman. Information acquisition in competitive markets: An application to the us mortgage market. *American Economic Journal: Microeconomics*, 4(4):65–106, May 2012. doi: 10.1257/mic.4.4.65. URL http://www.aeaweb.org/articles?id=10.1257/mic.4.4.65.

Andrea Canidio and Patrick Legros. Learning within or outside firms? labor market frictions and entrepreneurship. Technical report, Working Paper. Available at https://www. dropbox. com/s/i1meqmm4sr3wf0n . . . , 2017.

Ramon Casadesus-Masanell and Andres Hervas-Drane. Competing with privacy. *Management Science*, 61(1):229–246, 2015.

Giovanni Cespa. Information sales and insider trading with long-lived information. *The Journal of Finance*, 63(2):639–672, 2008.

Jay Pil Choi, Doh-Shin Jeon, and Byung-Cheol Kim. Privacy and personal data collection with information externalities. *Journal of Public Economics*, 173:113–124, 2019.

Markus Christen and Miklos Sarvary. Competitive pricing of information: A longitudinal experiment. *Journal of marketing research*, 44(1):42–56, 2007.

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

Jacques Crémer, Yves-Alexandre de Montjoye, and Heike Schweitzer. Competition policy for the digital era. *Report for the European Commission*, 2019.

Raymond J Deneckere and R Preston McAfee. Damaged goods. *Journal of Economics & Management Strategy*, 5(2):149–174, 1996.

Liran Einav and Jonathan Levin. Economics in the age of big data. *Science*, 346 (6210):1243089, 2014.

David Encaoua and Abraham Hollander. First-degree discrimination by a duopoly: pricing and quality choice. *The BE Journal of Theoretical Economics*, 7(1), 2007.

Rosa-Branca Esteves. Pricing with customer recognition. *International Journal of Industrial Organization*, 28(6):669–681, 2010.

Drew Fudenberg and J Miguel Villas-Boas. Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, 1:377–436, 2006.

Esther Gal-Or, Ronen Gal-Or, and Nabita Penmetsa. The role of user privacy concerns in shaping competition among platforms. *Information Systems Research*, 2018.

Diego García and Francesco Sangiorgi. Information sales and strategic trading. *The Review of Financial Studies*, 24(9):3069–3104, 2011.

Avi Goldfarb and Catherine Tucker. Privacy and innovation. *Innovation policy and the economy*, 12(1):65–90, 2012.

Yiquan Gu, Leonardo Madio, and Carlo Reggiani. Data brokers co-opetition. *Available at SSRN 3308384*, 2018.

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

Bengt R Holmstrom and Jean Tirole. The theory of the firm. *Handbook of industrial organization*, 1:61–133, 1989.

Johannes Hörner and Andrzej Skrzypacz. Selling information. *Journal of Political Economy*, 124(6):1515–1562, 2016.

Ganesh Iyer and David Soberman. Markets for product modification information. *Marketing science*, 19(3):203–225, 2000.

Ole Jann and Christoph Schottmüller. An informational theory of privacy. 2018.

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

Nicola Jentzsch, Geza Sapi, and Irina Suleymanova. Targeted pricing and customer data sharing among rivals. *International Journal of Industrial Organization*, 31(2): 131–144, 2013.

Pranav Jindal and Peter Newberry. To bargain or not to bargain: The role of fixed costs in price negotiations. *Journal of Marketing Research*, 55(6):832–851, 2018.

Stephen Kaisler, Frank Armour, J Alberto Espinosa, and William Money. Big data: Issues and challenges moving forward. In *2013 46th Hawaii International Conference on System Sciences*, pages 995–1004. IEEE, 2013.

Morton I Kamien. Patent licensing. *Handbook of game theory with economic applications*, 1:331–354, 1992.

Jakub Kastl, Marco Pagnozzi, and Salvatore Piccolo. Selling information to competitive firms. *The RAND Journal of Economics*, 49(1):254–282, 2018.

Michael L Katz and Carl Shapiro. How to license intangible property. *The Quarterly Journal of Economics*, 101(3):567–589, 1986.

Reinhold Kesler, Michael Kummer, and Patrick Schulte. User data, market power and innovation in online markets: Evidence from the mobile app industry. Technical report, Working Paper, 2017.

Jin-Hyuk Kim, Liad Wagman, and Abraham L Wickelgren. The impact of access to consumer data on the competitive effects of horizontal mergers and exclusive dealing. *Journal of Economics & Management Strategy*, 2018.

Heli Koski et al. How do competition policy and data brokers shape product market competition? Technical report, The Research Institute of the Finnish Economy, 2018.

Yassine Lefouili, Ying Lei Toh, et al. Privacy and quality. *TSE Working Paper 17 (795)*, 2017.

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

Qihong Liu and Konstantinos Serfes. Customer information sharing among rival firms. *European Economic Review*, 50(6):1571–1600, 2006.

Qihong Liu and Konstantinos Serfes. Price discrimination in two-sided markets. *Journal of Economics & Management Strategy*, 22(4):768–786, 2013.

Alessandro Lizzeri. Information revelation and certification intermediaries. *The RAND Journal of Economics*, pages 214–231, 1999.

Jacob Marschak. Economic measurements for policy and prediction. *Economic information, decision, and prediction*, pages 293–322, 1974.

Andrew McAfee, Erik Brynjolfsson, Thomas H Davenport, DJ Patil, and Dominic Barton. Big data: the management revolution. *Harvard business review*, 90(10): 60–68, 2012.

Paul R Milgrom and Steven Tadelis. How artificial intelligence and machine learning can impact market design. Technical report, National Bureau of Economic Research, 2018.

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

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

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

Patrick Rey and David Salant. Abuse of dominance and licensing of intellectual property. *International Journal of Industrial Organization*, 30(6):518–527, 2012.

John Riley and Richard Zeckhauser. Optimal selling strategies: When to haggle, when to hold firm. *The Quarterly Journal of Economics*, 98(2):267–289, 1983.

Michael Rothschild and Joseph Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. In *Uncertainty in economics*, pages 257–280. Elsevier, 1978.

Ariel Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, pages 97–109, 1982.

William Samuelson. A comment on the coase theorem. *game-theoretic Models of Bargaining*, 321:339, 1985.

Miklos Sarvary and Philip M Parker. Marketing information: A competitive analysis. *Marketing science*, 16(1):24–38, 1997.

Carl Shapiro. Consumer information, product quality, and seller reputation. *The Bell Journal of Economics*, pages 20–35, 1982.

Oz Shy and Rune Stenbacka. Investment in customer recognition and information exchange. *Information Economics and Policy*, 25(2):92–106, 2013.

Oz Shy and Rune Stenbacka. Customer privacy and competition. *Journal of Economics & Management Strategy*, 25(3):539–562, 2016.

Joel Sobel and Ichiro Takahashi. A multistage model of bargaining. *The Review of Economic Studies*, 50(3):411–426, 1983.

Michael Spence. Job market signaling. In *Uncertainty in Economics*, pages 281–306. Elsevier, 1978.

Lars A Stole. Price discrimination and competition. *Handbook of industrial organization*, 3:2221–2299, 2007.

Curtis Taylor and Liad Wagman. Consumer privacy in oligopolistic markets: Winners, losers, and welfare. *International Journal of Industrial Organization*, 34:80–84, 2014.

Jacques-Francois Thisse and Xavier Vives. On the strategic choice of spatial price policy. *The American Economic Review*, pages 122–137, 1988.

David Ulph and Nir Vulkan. *Electronic commerce and competitive first-degree price discrimination.* University of Bristol, Department of Economics, 2000.

William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961.

J Miguel Villas-Boas. Sleeping with the enemy: Should competitors share the same advertising agency? *Marketing Science*, 13(2):190–202, 1994.

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

Thomas A Weber and David C Croson. Selling less information for more: garbling with benefits. *Economics Letters*, 83(2):165–171, 2004.

Peter Weill. The relationship between investment in information technology and firm performance: A study of the valve manufacturing sector. *Information systems research*, 3(4):307–333, 1992.

Yi Xiang and Miklos Sarvary. Buying and selling information under competition. *Quantitative Marketing and Economics*, 11(3):321–351, 2013.

ÉCOLE DOCTORALE
Sciences de l'homme et de la société (SHS)

Titre : Vente stratégique d'informations et concurrence dans les marchés numériques

Mots clés : Données personnelles ; Concurrence ; Structure de marchés ; Régulation

Résumé : Cette thèse contribue à la littérature théorique sur le marché des données personnelles en trois points. Premièrement, elle analyse les stratégies des data brokers, ou courtiers en données, concernant la quantité de données personnelles qu'ils collectent et vendent. Deuxièmement, elle examine les effets de la concurrence entre data brokers sur leurs stratégies de monétisation des données personnelles. Troisièmement, elle propose des recommandations aux agences de protection des données personnelles et aux autorités de la concurrence afin de réglementer la concurrence entre courtiers en données. En fournissant des informations sur le fonctionnement de l'industrie de la vente d'information et en analysant son impact pour les consommateurs et les entreprises, cette thèse peut aider les décideurs à mieux répondre aux problèmes posés par l'émergence d'un marché des données personnelles.

Title : Strategic Information and Competition in Digital Markets

Keywords : Consumer information ; Competition ; Market design ; Regulation

Abstract : Consumer information has become a key strategic asset in most industries, shaping markets and societies. Yet, markets for information are characterized by a perplexing murkiness. Consumers and regulators face a spectacular lack of control over the collection and selling of consumer information. Data brokers stand out as the cornerstone of the emerging market for consumer information. They are now able to collect information on most consumers with a very high precision. Data brokers then use this information as information gatekeepers : they can shape competition on product markets by selling information on specific consumer segments to firms. We analyze in this thesis how brokers adopt competition-softening strategies that reduce consumer surplus. We argue that there is an urge for data protection agencies and competition authorities to adapt their regulatory practices to the new challenges raised by data brokers.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France