CER-ETH – Center of Economic Research at ETH Zurich

Competition Between Strategic Data Intermediaries with Implications for Merger Policy

D. Bounie, A. Dubus, P. Waelbroeck

Working Paper 23/383
April 2023

Economics Working Paper Series

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Competition Between Strategic Data Intermediaries with Implications for Merger Policy*

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

April 12, 2023

Abstract

We build a model of competition between strategic data intermediaries collecting consumer information that they sell to firms competing in a product market. Each intermediary has access to exclusive information on a group of consumers and competes with other intermediaries on a common group of consumers. Information allows firms to distinguish different segments of the consumer demand, and an equilibrium has the following properties. (*i.*) The largest intermediary collects the highest number of segments and sells information in the competitive market. (*ii.*) The incentives of the largest intermediary to collect data increase with the competitive pressure exerted by smaller intermediaries through an escape-competition effect. (*iii.*) Intermediaries sell information on a larger group of consumers in the competitive market than in the monopoly markets, increasing the intensity of competition among firms. (*iv.*) Competition reduces the incentives of intermediaries to collect data, thus increasing consumer surplus. These results have important implications for merger policy. Indeed, mergers increase the amount of data collected by intermediaries, which reduces consumer surplus due to enhanced price discrimination. This effect takes place in the market where the merging intermediaries operate, and also in other related markets through a ripple effect.

Keywords: Competition; Data intermediaries; Data collection; Selling strategies.

JEL: L13; L4; L51.

*Patrick Waelbroeck thanks for insightful discussions the members of the Chair Values and Policies of Personal Information of Institut Mines Télécom, Paris. Portions of the work on this article were done while as Antoine Dubus was receiving financial support from the FNRS Grant PDR T.01.47.19. This research has also been conducted within the Chair "Digital Finance" under the aegis of the Risk Foundation, a joint initiative by Cartes Bancaires CB, La Banque Postale, la Caisse des Dépôts, Telecom Paris and University of Paris 2 Panthéon - Assas.

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

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

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

1 Introduction

1.1 Motivation

In recent years, Big Tech companies such as Alphabet (Google), Amazon, Meta (Facebook), and Microsoft have acquired large digital companies including WhatsApp, LinkedIn, FitBit, Skype, or Nest (De Loecker et al., 2020).[1] These acquisitions have been controversial and contested by scholars and practitioners (Kwoka Jr and Valletti, 2020; CMA, 2020).

Central to these debates is the growing importance of data as a competitive asset. Besides providing consumers with online services such as social networking, online retailing, or web searches, major Big Tech companies such as Alphabet, Amazon, and Meta also act as data intermediaries in a new market for information (Bergemann and Bonatti, 2019). They extensively collect information on consumers, thus threatening privacy, while providing a competitive advantage to firms by selling them targeted information. By controlling the provision of data, data intermediaries impact the intensity of competition between firms. It is therefore crucial for scholars and policymakers to understand how mergers between data intermediaries change their ability to change the intensity of competition in product markets, and therefore consumer surplus.

Data intermediaries have recently been the topic of intense research (Bergemann and Bonatti, 2015; Bergemann et al., 2018; Montes et al., 2019; Elliott et al., 2021; Bounie et al., 2021; Ali et al., 2022; Bergemann et al., 2022), but a critical question remains: how does competition between data intermediaries change their incentives to collect personal data and their selling strategies? By answering this question, we will also be able to analyze and understand the effect of mergers in the market for information on consumer surplus. Interestingly, the literature generally ignores competition between data intermediaries, focusing instead on information design and on consumer privacy in a monopoly setting.

[1] In the year 2017 alone, Alphabet, Amazon, Apple, Meta, and Microsoft spent a total of \$31.6bn on start-ups acquisitions. (American tech giants are making life tough for startups; The Economist, June 2 2018.) The UK Competition and Market Authority reported, "Over the last 10 years the five largest firms have made over 400 acquisitions globally. None has been blocked and very few have had conditions attached to approval, in the UK or elsewhere, or even been scrutinised by competition authorities (CMA, 2020)."

Moreover, information is assumed to be exogenous, and data intermediaries are not strategic when they acquire consumer data. Yet, data intermediaries such as Alphabet and Meta determine how many data points they collect on consumers. The finer the data the higher the cost, but this allows information buyers to better target consumers and extract their surplus. Therefore, firms are ready to pay a high price for fine-grained consumer information, and considering intermediaries that endogenously collect data might change the equilibrium in the market for information, especially when intermediaries compete.

In this article, we develop a new $m-l$ framework of competition between data intermediaries. Each intermediary $i = 1, .., n$ has access to exclusive information on a group of consumers of size m_i, but faces competition from other intermediaries on a common group of consumers in what we will refer to as the competitive market of size l.[2] Our approach builds on the models of competition developed by Rosenthal (1980) and Varian (1980), where firms can reach potentially different groups of consumers and have access to locked-in consumers who do not know the existence of other firms. This framework solves a central challenge of the literature on markets for information, namely that a perfectly competitive market does not allow intermediaries to invest in costly data acquisition. This question dates back to Diamond (1971) and Grossman and Stiglitz (1976, 1980) for financial markets. Our $m-l$ approach solves this issue by considering imperfect competition between intermediaries that have access to different groups of consumers, in the spirit of Armstrong and Vickers (2022).

In our model, intermediary i collects consumer information in markets m_i and l. Information partitions consumer demand into segments, and collecting more information allows an intermediary to better identify consumers by reducing the size of the segments. Firms can use information to price-discriminate consumers, thereby extracting more surplus (we refer to this effect as rent extraction). Firms are willing to pay a higher price for finer consumer segments. While previous literature considers an exogenous amount of data available to firms, we endogenize the decision of intermediaries to collect more or less consumer data. When choosing

[2] When there is no confusion, m_i will represent the monopoly market of intermediary i or the size of market m_i. Similarly, l will represent the competitive market or the size of this market.

the number of segments to collect, each intermediary balances the increased willingness to pay of firms and a cost to collect data. Additionally, we assume that an intermediary cannot distinguish consumers who belong to market m_i from those belonging to market l before collecting data, therefore, intermediary i collects data to maximize the sum of profits in markets m_i and l.

Data intermediaries also sell information strategically by determining which combination of consumer segments they sell to each firm. After having collected information, intermediaries can distinguish markets m_i and l, and sell information differently in both markets.[3] An intermediary i sells an optimal number of segments that maximizes the price of information in its monopoly market m_i and in the competitive market l. In both markets, the price of information is given by the profits of an informed firm minus its outside option if it does not purchase information and either remains uninformed, or acquires information from another intermediary. Hence, the price naturally depends on whether the intermediary sells information in its monopoly market or in the competitive market, and the selling strategies may differ in m_i and in l. In a given market, selling more information allows firms to target more consumers, but also increases the competitive pressure between firms (Armstrong and Zhou, 2022), and this competitive effect of information reduces their willingness to pay.

1.2 Results

Our key findings are the following. First, competition between data intermediaries changes the outside option of information buyers since they can purchase data from other intermediaries. It is therefore profitable for competing intermediaries to sell information to all firms, while a monopolist intermediary only sells information to one firm. This has important implications for merger policies that we discuss later. Secondly, when data intermediaries compete, they sell more consumer segments than under monopoly, thus intensifying competition between firms. Thirdly, when intermediaries compete, they collect fewer consumer segments, and firms extract less surplus from consumers. Finally, competition at the

[3] For instance, cookies used for analytics can provide information and allow an intermediary to identify consumer visits to competitors' websites.

top increases the incentives of the leading intermediary to collect consumer data. This escape-competition effect creates an externality that can increase the amount of data collected by intermediaries in other markets where they operate, reducing, in turn, the surplus of consumers in these markets.

Competition between intermediaries and the number of informed firms

Data intermediaries sell information to only one firm in their monopoly markets, but to both firms in the competitive market. On the one hand, a monopolist intermediary can increase the willingness to pay of a prospective buyer by threatening it to sell information to its competitor in case it does not purchase information. Therefore, in equilibrium, a monopolist intermediary sells information to only one firm. On the other hand, this threat cannot be exerted when intermediaries compete, as firms can purchase information from other intermediaries. In the competitive market all firms acquire similar information, they compete on a level playing field, and competition in the product market becomes fiercer.

Competition between intermediaries and the number of consumers identified

A monopolist intermediary sells to Firm 1 (w.l.o.g.) information on fewer consumers than in the competitive market. As discussed previously, selling more consumer segments increases the rent-extraction effect of information, but also increases the competitive pressure in the product market. We have already shown in Bounie et al. (2021) that a monopolist intermediary balances these two opposing effects of information by selling only a subset of available consumer segments. We generalize this result to a competitive setting and establish that all intermediaries propose to firms a partition that identifies only some segments of consumers, leaving others unidentified. At first glance, this new result is surprising; competition could force intermediaries to sell all their available information in the spirit of Bertrand competition. However, the competitive effect of information is so strong that no intermediary makes positive profits from selling all the available information. Therefore, all intermediaries restrict the total number of segments proposed to firms.

Overall, competition between intermediaries reduces the equilibrium amount of information sold to individual firms. However, the aggregate number of segments sold to both firms is higher than in the monopoly markets. Consumer surplus is therefore higher in the competitive market than in monopoly markets.

Competition between intermediaries and the amount of data collected

Competition between intermediaries also impacts their incentives to collect data, thus changing the ability of information buyers to personalize their prices and to extract consumer surplus. The incentives of intermediaries to collect data are determined by two effects. On the one hand, more identified consumers increase the marginal value of an additional segment collected. This effect is stronger when intermediaries compete as they sell information on more consumers than in a monopoly market. On the other hand, a monopolist intermediary sells information to only one firm, as it can threaten a prospective buyer to sell data to its competitor. The value of this threat increases with the precision of information, yielding additional incentives to collect data in a monopoly market through an outside-option effect.

Taking these two effects into account, we show that an equilibrium exists under general conditions, in which the intermediary with the largest monopoly market collects more data than other intermediaries and sells information in the competitive market as a constrained monopolist. Other intermediaries collect an amount of data that increases with the size of their monopoly markets, and do not sell information in market l. Other equilibria may exist in which another intermediary collects the largest number of segments, but in which the equilibrium strategies of other intermediaries do not change. We provide the necessary and sufficient conditions for all equilibria to exist. Overall, intermediaries have higher incentives to collect data in their monopoly markets than in the competitive market, which is a novel result in the literature.

Escape-competition and ripple effects

Finally, the incentives of the largest intermediary to collect consumer information increase with the intensity of the competitive pressure. This effect will be referred to as an escape-competition effect (see Aghion et al. (2005)), which reduces consumer surplus in the competitive market and in the local monopoly market of the largest intermediary. We can generalize this escape-competition effect to a setting with several competitive markets to show that mergers can generate effects outside the relevant market through what we will refer to as ripple effects. The escape-competition effect in a given market can change the incentives to collect data of top-sellers intermediaries in other competitive markets.

To illustrate these two effects, we consider three intermediaries with monopoly markets m_1, m_2, m_3, and two competitive markets l_1 and l_2. We focus on the case where intermediaries 1 and 2 compete in market l_1 and intermediaries 2 and 3 compete in market l_2, intermediary 1 sells information in market l_1 and intermediary 2 sells information in market l_2. After a merger between intermediaries 2 and 3, the merged intermediary collects more consumer segments and exerts a stronger competitive pressure in market l_1 than intermediary 2 before the merger. This increases the incentives of intermediary 1 to collect data in markets l_1 and m_1 through the escape-competition effect. In general, this effect also takes place in any competitive market where intermediary 1 has the second highest information precision and so on.

1.3 Contributions

This article contributes to the academic literature on markets for information and to the policy debates on Big Tech mergers on two points.

First, we develop a novel framework of competition between data intermediaries who can reach different groups of consumers. The incentives to collect data depend on the intensity of competition between data intermediaries. This result contributes to the literature on the market for information that has focused on monopolist intermediaries and on exogenous data acquisition, as well as to the general literature on the effects of information on markets.

Secondly, our work also contributes to the literature on data-driven mergers and acquisitions (De Corniere and Taylor, 2020; Chen et al., 2022; Dubus and Legros, 2022; Prat and Valletti, 2022). By endogenizing the decision of intermediaries to collect data, we show that consumer surplus decreases after a merger due to a stronger rent-extraction effect. This reduction of surplus is not accounted for in models in which the amount of data available to firms is exogenous. This new theoretical result is supported by recent empirical evidence on the impact of mergers on the amount of data collected by digital companies (Kesler et al., 2020; Affeldt and Kesler, 2021).

The remainder of this article is organized as follows. In Section 2 we describe the model. We characterize the data strategies of intermediaries in the competitive market in Section 3, and in their monopoly markets in Section 4. We analyze consumer surplus in Section 5 and mergers between intermediaries in Section 6 in order to study how they impact competition and consumer surplus in product markets. Section 7 concludes the paper.

2 Description of the Model

We build a model of competition between data intermediaries collecting customer data that they sell to firms competing in a product market, where the data sold allows firms to price discriminate consumers. We first characterize the nature of competition between data intermediaries. We then define consumer utility, the strategies of competing data intermediaries collecting and selling consumer data, the decision of firms to purchase consumer data, and finally, the timing of the game. We discuss in detail our main assumptions in Appendix A.1.

2.1 Nature of Competition Between Data Intermediaries

We consider n competing data intermediaries that collect and sell consumer information to firms (with $n \geq 2$). Each data intermediary can collect information on a mass $m_i \in \mathbb{R}_+$ (with $i = 1, .., n$) of consumers who belong to its monopoly market, and on a market of mass $l \in \mathbb{R}_+$ of consumers where all intermediaries compete à la Bertrand for the sale of information.[4] Consumers, therefore, either belong to a monopoly market or a competitive market, so that the total mass of consumers is $\mu = m_1 + ... + m_n + l$.[5] By convention, intermediary DI_1 has a larger

[4] We analyze in Section 4 a situation in which data intermediaries only collect and sell information on their monopoly market.

[5] We assume that $m_i > 0 \ \ \forall i$ and $l > 0$ in the remainder of the article. This framework has as special cases $l = 0$ and $m_i = 0$, that are analyzed in Sections 3 and 4.

monopoly market than data intermediary DI_2 and so on: $m_1 \geq m_2 ... \geq m_n$.[6]

This $m-l$ approach is relevant to describe the market for information, where an intermediary (DI_i) has access to exclusive information on a group of consumers (corresponding to monopoly markets m_i), but faces competition from other intermediaries on another group of consumers (corresponding to market l). For instance, Meta collects data on its users, which other data intermediaries cannot sell in the product market. Meta also collects information on users who visit other platforms or online services such as the ones offered by Alphabet; Meta and Alphabet have therefore similar information on these consumers.

To the best of our knowledge, the literature has not yet analyzed the strategies of intermediaries collecting and selling consumer data in a competitive framework. Indeed, analyzing competition between data intermediaries raises a major conceptual challenge. Consider intermediaries that can only collect data in the competitive market l, where they compete head-to-head to sell information. If collecting data is costly, and intermediaries make zero profits, it is not profitable for them to collect data. This is a well-known issue in financial markets (Grossman and Stiglitz, 1976, 1980), which has been recently formalized by Ichihashi (2021) in the case of intermediaries collecting consumer data. By introducing markets on which each intermediary has monopoly access to information, our $m-l$ approach addresses this conceptual issue and enables the existence of a competitive market for information.

2.2 Consumers

Consumers are divided into $n+1$ mutually exclusive markets of different sizes such that $\mu = m_1 + .. + m_n + l$. Each market is characterized by a Hotelling line $[0, 1]$ of mass $m_1, .., m_n, l$ on which consumers are uniformly distributed. On each line, a consumer can buy one product at a price $p_1 \in \mathbb{R}_+$ from Firm 1 located at 0, or $p_2 \in \mathbb{R}_+$ from Firm 2 located at 1.[7]

[6] As a special case, we will also allow for symmetric data intermediaries in terms of size of their monopoly markets: $m_1 = m_2 ... = m_n$; we will show that they collect a different amount of information in the only equilibrium of the game.

[7] We assume that all markets are covered. This assumption is common in the literature. See for instance Thisse and Vives (1988), Liu and Serfes (2004), Stole (2007), Ulph and Vulkan (2000), Montes et al. (2019), and Bounie et al. (2021).

Consumers located at $x \in [0, 1]$ derive a utility V from purchasing the product. They incur 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. Hence on each unit line, consumers located at x have a utility function defined by:[8]

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

2.3 Data Intermediaries

At the beginning of the game, data intermediaries collect information that divides each market into consumer segments.[9] More data is costly to collect but allows an intermediary to have a finer partition of a unit line. In the second stage of the game, intermediaries sell partitions that allow firms to identify and price-discriminate consumers.

2.3.1 Collecting Data

A data intermediary DI_i collects data on $m_i + l$ consumers. We assume the intermediary cannot distinguish consumers who belong to market m_i and market l before collecting data, and the intermediary collects the same amount of information on consumers in markets m_i and l.[10] A data intermediary collects data points (such as gender, age, or zip code) that partition consumer demand into $k \in \mathbb{N}_+^*$ segments of size $\frac{1}{k}$.[11] Several technologies allow firms to collect data on consumers such as cookies and pixels (Bergemann and Bonatti, 2015; Choe et al., 2018).

[8] Prices p_1 and p_2 can be potentially different on different unit lines, and for distinct locations of a given unit line.

[9] Other modeling choices are possible, such as the sale of a signal analyzed by Admati and Pfleiderer (1988) and Bergemann et al. (2018) among others, in the case of a monopolist information seller. Our focus on information that partitions a Hotelling line is motivated by the tractability of this competition model, yet it provides us with a rich set of strategy space for data intermediaries.

[10] We discuss this specification in Appendix A.1.

[11] We drop index i when there is no confusion.

We illustrate the partition collected by a data intermediary in Figure 1. In a given market in $\{m_1, .., m_n, l\}$, the k segments of size $\frac{1}{k}$ form a partition $\mathcal{P}^k$ that we refer to as the reference partition. These elementary segments generate a sigma-field $\mathbb{P}_k$ containing the 2^{k-1} possible partitions of the unit line, among which the intermediary can select the partition to sell.[12]

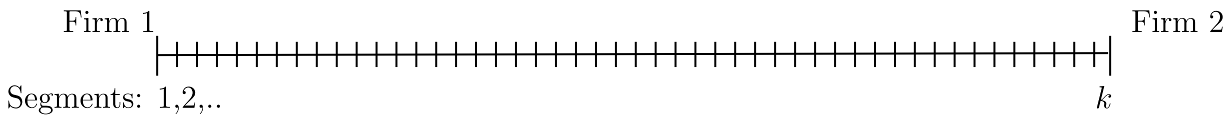

Figure 1: Reference partition $\mathcal{P}^k$

The number of consumer segments k corresponds to the precision of information, and a firm that has information can third-degree price-discriminate consumers by charging different prices on different segments. For instance, when $k = 2$, the partition is coarse, and firms can only distinguish whether consumers belong to $[0, \frac{1}{2}]$ or to $[\frac{1}{2}, 1]$.

This approach allows us to analyze varying levels of information precision and characterize the amount of data collected by intermediaries. We will show how competition between data intermediaries impacts consumer surplus in product markets by changing the amount of data collected, changing in turn the ability of firms to price discriminate consumers.

The cost of collecting data is given by $c(k) : \mathbb{N}^*_+ \to \mathbb{R}_+$ for a mass one of consumers, and satisfies standard convexity conditions.[13] As we assume that a data intermediary collects the same amount of information in markets m_i and l, the total cost to collect consumer data is $(l + m_i)c(k)$. This cost encompasses various dimensions of the activity of data intermediaries, such as installing trackers or storing and handling data. Collecting more information by increasing the number of segments allows a firm to extract more surplus on consumers, increasing in turn its willingness to pay for information and the price of information.

[12]The choice of the partition corresponds to the selling strategy of an intermediary, which we analyze in detail in Sections 3 and 4.

[13]We assume that $c(0) = c'(0) = 0$ and $c(.), c''(.) > 0$.

2.3.2 Selling Information

Once they have collected data, intermediaries can distinguish markets m_i and l, and sell information differently in both markets. For instance, cookies used for analytics can provide information on the websites visited by consumers, and allow a firm to know whether consumers have visited the websites of its competitors.

In each market m_i and l, data intermediary i can sell any combination of segments of the consumer demand. We denote by $\mathcal{P}_1^m(DI_i) \in \mathbb{P}_k$ and $\mathcal{P}_2^m(DI_i) \in \mathbb{P}_k$ the partitions offered to Firm 1 and Firm 2 respectively by intermediary DI_i on its monopoly market. These partitions can be potentially different for each firm. Similarly, intermediary DI_i proposes to firms partitions $\mathcal{P}_1^l(DI_i) \in \mathbb{P}_k$ and $\mathcal{P}_2^l(DI_i) \in \mathbb{P}_k$ in the competitive market l.

We have shown in Bounie et al. (2021) that a monopolist data intermediary can weaken or strengthen the intensity of competition in the product market by determining the quantity of information available to firms, which has two opposing effects on consumer surplus. On the one hand, an informed firm can price discriminate consumers, thus increasing its profits through this rent extraction effect. On the other hand, information also increases competition in the product market, which reduces the profits of both firms. An optimal partition thus maximizes consumer rent extraction while softening the competitive effect of information. We will see how a merger between data intermediaries increases their ability to soften the competitive effect of information, impacting in turn competition in the product market and consumer surplus.

To illustrate how information impacts the intensity of competition in the product market, consider a situation in which $k = 4$ segments are available (see Figure 2). By allowing Firm 1 to distinguish consumers located close to Firm 2 and to charge them prices p_{13} and p_{14}, the data intermediary also increases the competitive pressure on Firm 2 that lowers price p_2. Now suppose that the data intermediary only sells the first segment to Firm 1, which charges price p'_{11} to these consumers: the competitive pressure will be much lower, and Firm 2 will increase its price $p'_2 > p_2$. By keeping a share of consumers unidentified, the data intermediary will keep a low level of competition between firms, while still allowing Firm 1 to extract more surplus from identified consumers close to its location.

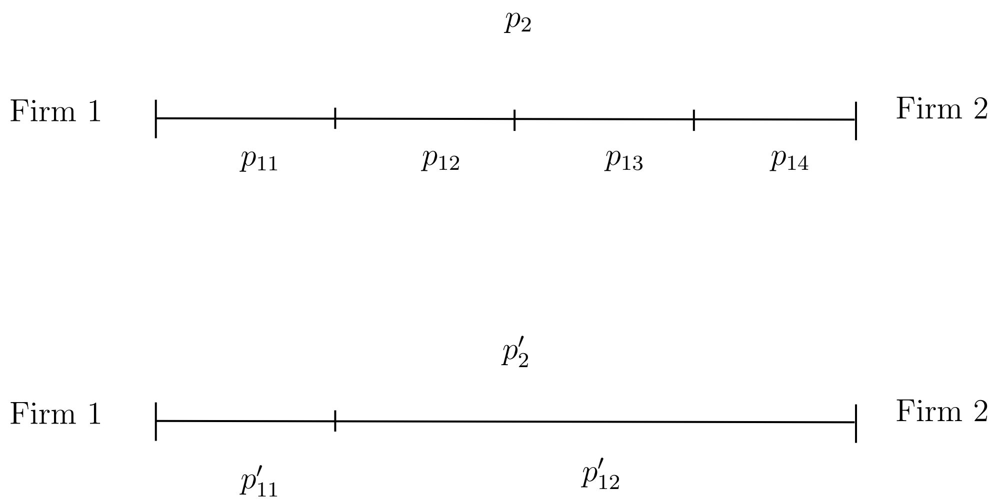

Figure 2: Example of partitions, $k = 4$

In the competitive market, data intermediaries compete à la Bertrand in the sale of information. We assume that firms cannot combine information acquired from different data intermediaries, and a firm only purchases information from one data intermediary.[14] Firms choose intermediary $\overline{DI}$ with the highest information precision, so that competition in market l leads to a winner takes all situation. Other data intermediaries collect data in market l, which is therefore contestable. Hence, intermediary $\overline{DI}$ sells information under the competitive pressure exerted by data intermediary $\underline{DI}$ with the second highest number of segments collected.

In monopoly market m_i, data intermediary DI_i sells information to one or two firms. We assume that data intermediaries sell information through first-price auctions.[15] Intermediaries know the willingness to pay of firms for information, and an auction is used in our case to maximize the price of information by increasing surplus extraction from firms. This selling mechanism is frequently used by major data intermediaries such as Alphabet[16] and in data marketplaces (Sheehan and Yalif, 2001; O'kelley and Pritchard, 2009), and is also commonly used in the literature on markets for information (Montes et al., 2019; Bounie et al., 2021).

[14] We discuss this assumption in Appendix A.1.

[15] Note that intermediaries can sell information to both firms using a first-price auction by setting the appropriate reserve price. See Appendix B.3 for a discussion.

[16] First-price Auction, Second-price, and the Header-Bidding, Smartyads, February 2018.

2.4 Firms

In the third stage of the game, firms compete in prices using the information that they have purchased from intermediaries. Firms can acquire information from the monopolist data intermediary DI_i in each monopoly market m_i and from one of the competing data intermediaries in the competitive market l. Without information, firms only know that consumers are uniformly distributed on the unit line. A firm that has purchased information knows to which segment of the partition a consumer belongs.

In order to compute the profits of a firm that has acquired partition $\mathcal{P}_\theta$, we need to compute demand and prices on each consumer 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 sells to all consumers and sets a price under the constraint exerted by its competitor. Firm θ sets prices in two steps.[17] First, it sets prices on all segments. In a second step, Firm θ updates its prices in segments where it is a constrained monopolist. On a given Hotelling line, each firm knows whether its competitor is informed, and the partition $\mathcal{P}_{-\theta}$.[18] We will show that the partitions sold by intermediaries have the following features. An intermediary sells information that allows firms to identify segments of close-by consumers on which firms charge constrained-monopolist prices. Consumers in the middle of the line are left unidentified, and firms charge a homogeneous price to these consumers.

For any partition $\mathcal{P}_1$ composed of n segments, we denote by $d_{\theta i}$ the demand of Firm $\theta = \{1, 2\}$ on the ith segment. Firm θ maximizes the following profit function with respect to the vector of prices $\mathbf{p}_\theta := (p_{\theta 1}, .., p_{\theta i}, .., p_{\theta n})$:

$$\pi_\theta(\mathbf{p}_\theta) = \sum_{i=1}^{n} d_{\theta i} p_{\theta i}$$

[17] We discuss this sequential pricing assumption in Appendix A.1.

[18] This assumption is also standard in the literature (Thisse and Vives, 1988; Braulin and Valletti, 2016; Montes et al., 2019; Elliott et al., 2021).

2.5 Timing

A data intermediary DI_i first collects data and sells potentially different partitions to firms in market m_i and in market l. Then, in each market $m_1, .., m_n, l$, firms set prices on the different segments. The timing of the game is the following:

- Stage 1: data intermediary DI_i collects data on k_i consumer segments in markets m_i and l.
- Stage 2: data intermediary DI_i offers information partitions for sale in its monopoly market and in the competitive market.
- Stage 3: firms set prices $p_{\theta i}$ on the different consumer segments.
- Stage 4: consumers observe prices and make their purchase decisions.

We analyze in the next sections the strategies of intermediaries collecting and selling consumer data in the competitive and in monopoly markets.

3 Competitive Market

We first study the selling strategies of intermediaries in the competitive market, as well as their incentives to collect consumer data. We show that (*i.*) the data intermediary with the highest precision sells information to both firms, but does not sell all consumer segments in order to reduce the competitive effect of information; (*ii.*) other intermediaries do not sell information; (*iii.*) the amount of data collected depends on the marginal value of data, which is determined by two opposing forces: a rent-extraction effect, and on an outside-option effect. We explain these mechanisms in Section 3.2.

3.1 Selling Information in the competitive market

We first characterize the selling strategies of data intermediaries. We solve the game by backward induction, and in this stage the amount of data collected is given. We first determine the profits of the firms, then the price of information, and finally the optimal selling strategy and the information structures offered by intermediaries. To simplify the exposition, we consider a market of size 1, and we will show that the selling strategy does not depend on the size of the market.

Profits of the firms in the competitive market

We denote by $\pi_1(\mathcal{P}_1^l(DI_i), \mathcal{P}_2^l(DI_j))$ and $\pi_2(\mathcal{P}_2^l(DI_j), \mathcal{P}_1^l(DI_i))$ the profits of Firm 1 and Firm 2 when they respectively acquire partitions $\mathcal{P}_1^l(DI_i)$ from data intermediary DI_i and $\mathcal{P}_2^l(DI_j)$ from DI_j. Let $\overline{DI}$ be the intermediary with the highest information precision $\overline{k} \in \mathbb{N}_+$, and $\underline{DI}$ be the intermediary with the second highest number of segments collected $\underline{k} \in \mathbb{N}_+$.

Proposition 1
$\overline{DI}$ sells information to both firms; other intermediaries do not sell information.

Intermediaries compete à la Bertrand in prices and information structures in the competitive market. Hence, each intermediary i maximizes the price of information, given for each firm by the difference between profits with information purchased from i and with information purchased from $j \neq i$ that maximizes their profits. As the profits of a firm increase with the number of segments available on a given group of consumers, only intermediary $\overline{DI}$ with the highest number of segments $\overline{k}$ sells information in market l. Other data intermediaries make zero profits in this market.

Prices of information in the competitive market

Data intermediary $\overline{DI}$ sells information to Firm 1 and Firm 2 at prices corresponding to the willingness to pay of each firm for information. Consider the incentive of Firm 1 to purchase information. Firm 1 can acquire $\mathcal{P}_1(\overline{DI}) \in \mathbb{P}_{\overline{k}}$ at price $\overline{p}_1$ from $\overline{DI}$, or $\mathcal{P}_1(\underline{DI}) \in \mathbb{P}_{\underline{k}}$ at price $\underline{p}_1$ from $\underline{DI}$ and make profit $\pi_1(\mathcal{P}_1(\underline{DI}), \mathcal{P}_2(\overline{DI}))$.[19] The willingness to pay of Firm 1 for information is thus $\pi_1(\mathcal{P}_1(\overline{DI}), \mathcal{P}_2(\overline{DI})) - \pi_1(\mathcal{P}_1(\underline{DI}), \mathcal{P}_2(\overline{DI}))$. The price that Firm 2 is ready to pay for information is defined in a similar way: $\pi_2(\mathcal{P}_2(\overline{DI}), \mathcal{P}_1(\overline{DI})) - \pi_2(\mathcal{P}_2(\underline{DI}), \mathcal{P}_1(\overline{DI}))$. Lemma 1 summarizes this discussion.

[19] We drop subscripts l from the notations of the partitions $\mathcal{P}_1^l(.)$ and $\mathcal{P}_2^l(.)$ throughout this section.

Lemma 1

The prices of information charged by data intermediary $\overline{DI}$ to Firm 1 and Firm 2 in the competitive market are:

$$\begin{cases} \overline{p}_1(\mathcal{P}_1(\overline{DI}), \mathcal{P}_2(\overline{DI})) = \pi_1(\mathcal{P}_1(\overline{DI}), \mathcal{P}_2(\overline{DI})) - \pi_1(\mathcal{P}_1(\underline{DI}), \mathcal{P}_2(\overline{DI})), \\ \qquad \textit{and} \\ \overline{p}_2(\mathcal{P}_2(\overline{DI}), \mathcal{P}_1(\overline{DI})) = \pi_2(\mathcal{P}_2(\overline{DI}), \mathcal{P}_1(\overline{DI})) - \pi_2(\mathcal{P}_2(\underline{DI}), \mathcal{P}_1(\overline{DI})). \end{cases}$$

Optimal information structure

Data intermediary $\overline{DI}$ chooses the partitions that maximize the profits of the firms by combining segments of the reference partition $\mathcal{P}^{\overline{k}}$. For instance, the data intermediary can combine segments 2, 3 and 4 of the partition on top of Figure 2 to form a new partition that consists of just two segments, given at the bottom of Figure 2. Even though we allow for any partition of the unit line, some partitions can be easily ruled out. Selling consumer segments far away from a firm will only increase the competitive effect of information, while selling coarse segments close to a firm's location is not optimal since more precise information would increase its willingness to pay for information. In Proposition 2, we characterize the optimal partitions $\mathcal{P}_1^*(\overline{DI})$ and $\mathcal{P}_2^*(\overline{DI})$ represented in Figure 3.

Proposition 2

An optimal partition $\mathcal{P}_1^*(\overline{DI})$ for Firm 1 (and $\mathcal{P}_2^*(\overline{DI})$ for Firm 2) divides the unit line into two intervals:

- *The first interval consists of $\overline{j_1}$ ($\overline{j_2}$) segments of size $\frac{1}{\overline{k}}$ on $[0, \frac{\overline{j_1}}{\overline{k}}]$ ($[1 - \frac{\overline{j_2}}{\overline{k}}, 1]$) where consumers are identified.*
- *Consumers in the second interval of size $1 - \frac{\overline{j_1}}{\overline{k}}$ ($1 - \frac{\overline{j_2}}{\overline{k}}$) are unidentified.*
- *The optimal numbers of segments sold are:*

$$\overline{j_1}^*(\overline{k}) = \overline{j_2}^*(\overline{k}) = \frac{\overline{k}}{3} - \frac{\overline{k}}{9\underline{k}} - \frac{7}{18}.$$

Proof: see Appendix B.1.

The optimal partitions divide the unit line into two intervals. Firms can price discriminate identified consumers, and charge a uniform price on the second interval of unidentified consumers. Data intermediary $\overline{DI}$ does not sell all consumer segments to reduce the competitive pressure of information. It is easy to understand that selling all consumer segments is not optimal for a data intermediary: selling more segments increases competition and reduces the willingness to pay of firms for information. Partitions $\mathcal{P}_1^*(\overline{DI})$ and $\mathcal{P}_2^*(\overline{DI})$ balance the competition and rent-extraction effects of information. Similarly, $\underline{DI}$ will offer to Firm 1 and Firm 2 partitions composed of $\underline{j_1}$ and $\underline{j_2}$ segments closest to their locations. Using these partitions, the optimization problem for data intermediary $\overline{DI}$ in the competitive market boils down to choosing $\overline{j_1}$ and $\overline{j_2}$ under the constraint exerted by intermediary $\underline{DI}$.

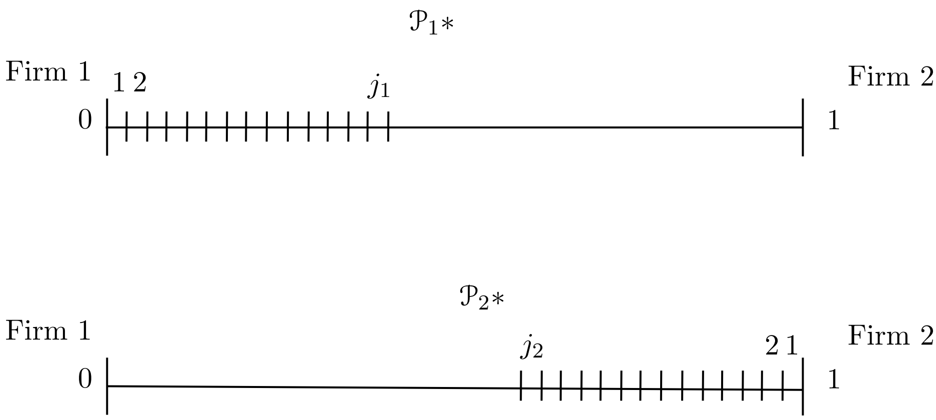

Figure 3: Selling partitions $\mathcal{P}_1^*$ to Firm 1 and $\mathcal{P}_2^*$ to Firm 2

Hence, $\overline{DI}$ internalizes the competitive effect of information by selling to each firm information partitions on their high-valuation consumers, keeping low-valuation consumers unidentified. This selling strategy softens the competitive effect of information, and reduces consumer surplus.

3.2 Collecting Consumer Data in the Competitive Market

We analyze in this section the incentives of intermediary $\overline{DI}$ to collect consumer data in the competitive market.[20] Remember that data intermediaries cannot distinguish to which market consumers belong before collecting data, and the number of consumer segments collected by each data intermediary DI_i is identical on m_i and l.

In the competitive market l, $\overline{DI}$ sells information and makes profits equal to the sum of the prices paid by each firm (net of the cost to collect consumer data), times the size of the competitive market l:[21]

$$\Pi_l(\overline{k}) = l[\overline{p}_1(\overline{k}) + \overline{p}_2(\overline{k}) - c(\overline{k})].$$

The incentives to collect data depend on the marginal effect of collecting more data on the prices $\overline{p}_1(\overline{j_1}^*(\overline{k}), \overline{j_2}^*(\overline{k}))$ and $\overline{p}_2(\overline{j_2}^*(\overline{k}), \overline{j_1}^*(\overline{k}))$. The price of information is the difference between the profits of a firm with information purchased from $\overline{DI}$ (equal to $\pi_\theta(\overline{j_\theta}^*(\overline{k}), \overline{j}^*_{-\theta}(\overline{k})))$ and with information purchased from intermediary $\underline{DI}$ (equal to $\pi_\theta(\underline{j_\theta}^*(\overline{k}), \overline{j}^*_{-\theta}(\overline{k})))$. We analyze these two terms separately.

The profits of Firm 1 with information (and similarly for Firm 2) can be decomposed as the sum of two terms: a term corresponding to profits with perfect information for a given share of identified consumers; and a term that represents the surplus of identified consumers when information is imperfect, which decreases with $\overline{k}$.[22]

The effects of more data collected on the outside options of the firms must account for the strategic response of intermediary $\underline{DI}$ with the second highest number of segments to an increase in $\overline{k}$, which limits the ability of $\overline{DI}$ to exert a threat on a prospective buyer. Indeed, intermediary $\underline{DI}$ exerts a competitive pressure on $\overline{DI}$ by proposing a best-response partition that accounts for the precision of information of $\overline{DI}$. In other words, the values of $\underline{j_1}^*$ and $\underline{j_2}^*$ decrease with $\overline{k}$.

[20] Other intermediaries make zero profit in this market, and overall, they have incentives to collect data only through the profits they make on their respective monopoly markets, and which we analyze in Section 4.2.

[21] The convexity of the cost function ensures that the profit functions are strictly concave with a unique maximum at $\overline{k}^*$.

[22] The expression of profits is given in Appendix B.3.

Overall, given this strategic response by $\underline{DI}$, the marginal effect of collecting more data for intermediary $\overline{DI}$ is a pure rent-extraction effect, which we can identify by considering the FOC w.r.t. $\overline{k}$:

$$\begin{aligned}\overline{p}_1(\overline{j_1}^*(\overline{k}),\overline{j_2}^*(\overline{k})) + \overline{p}_2(\overline{j_2}^*(\overline{k}),\overline{j_1}^*(\overline{k})) &= \pi_1(\overline{j_1}^*(\overline{k}),\overline{j_2}^*(\overline{k})) - \pi_1(\underline{j_1}^*(\underline{k}),\overline{j_2}^*(\overline{k})) \\ &+ \pi_2(\overline{j_2}^*(\overline{k}),\overline{j_1}^*(\overline{k})) - \pi_2(\underline{j_2}^*(\underline{k}),\overline{j_1}^*(\overline{k}))\end{aligned}$$

$$\implies \frac{\partial \overline{p}_1(\overline{j_1}^*(k),\overline{j_2}^*(k)) + \overline{p}_2(\overline{j_2}^*(k),\overline{j_1}^*(k))}{\partial k}\Big|_{k=\overline{k}} = \underbrace{\frac{\overline{j_1}^*(\overline{k}) + \overline{j_2}^*(\overline{k})}{\overline{k}}\frac{t}{\overline{k}^2}}_{\substack{\text{Rent-extraction}\\ \text{effect}}}.$$

Proposition 3

The marginal gain from collecting data depends on a rent-extraction effect, which is proportional to the transportation cost t, as well as to the number of consumers identified by the firms: $\frac{\overline{j_1}^*(\overline{k})+\overline{j_2}^*(\overline{k})}{\overline{k}}$.

The proof follows a direct application of the envelope theorem to the optimal price of information in the competitive market given in Appendix B.1. Collecting more data intensifies the rent-extraction effect as Firm 1 can better price-discriminate consumers on thinner segments of the demand. This effect is stronger when the intermediary sells information to firms on a larger share of consumers: the marginal gain from having thinner segments is stronger when more consumers are identified. Moreover, this effect increases proportionally with t: as t increases, the intensity of competition between firms decreases and the potential for rent extraction increases.

4 Monopoly Markets

We now characterize the selling strategies of data intermediaries in their monopoly markets, as well as their incentives to collect consumer data (we drop subscript i, m_i and DI_i from the notations in this section).

4.1 Selling Information in Monopoly Markets

We determine the price of information, the optimal selling strategy and the information structure sold by an intermediary in a monopoly market when selling information to Firm 1 only. We show that the profits of an intermediary in the monopoly market are always higher when selling information to one firm than when selling information to both firms.

Price of information in a monopoly market

We characterize the equilibrium when the intermediary sells information to Firm 1 only. We prove in Proposition 5 that in this case, the price of information is always higher than when the intermediary sells information to both firms.

Let $\pi_1(\mathcal{P}^k, \emptyset)$ and $\pi_2(\mathcal{P}^k, \emptyset)$ be the respective profits of Firm 1 and Firm 2 when they acquire the reference partition $\mathcal{P}^k$ and their competitor has no information. Similarly, let $\pi_1(\emptyset, \mathcal{P}^k)$ and $\pi_2(\emptyset, \mathcal{P}^k)$ be their profits when they are uninformed but face a competitor that has acquired partition $\mathcal{P}^k$. The profits of an uninformed firm are minimized when its competitor has information $\mathcal{P}^k$. Thus, this partition represents the maximal level of threat for a firm that does not purchase information. The resulting price of information is given by the difference between the profits of Firm 1 with information and this maximal threat, and is given in Equation 2.

Lemma 2

The monopoly price of information when selling partition $\mathcal{P}_1$ to Firm 1 is:

$$p_1^{m*} = max_{\mathcal{P}_1}\{\pi_1(\mathcal{P}_1, \emptyset) - \pi_1(\emptyset, \mathcal{P}^k)\}. \tag{2}$$

Optimal information structure

The partition that maximizes the price of information given by Lemma 2 is similar to the partitions sold in the competitive market: consumers located close to Firm 1 are identified, and far-away consumers are kept unidentified to soften the competitive effect of information. In Proposition 4, we characterize the features of this optimal partition $\mathcal{P}_1^*$.

Proposition 4

In a monopoly market, the optimal partition $\mathcal{P}_1^*$ divides the unit line into two intervals:

- *The first interval consists of $j_1 \in [\![0, k]\!]$ segments of size $\frac{1}{k}$ on $[0, \frac{j_1}{k}]$ where consumers are identified.*
- *Consumers in the second interval of size $1 - \frac{j_1}{k}$ are unidentified.*
- *The optimal number of segments sold is:*

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

Proof: see Appendix B.2.

Partition $\mathcal{P}_1^*$ divides the unit line into two intervals. Firm 1 can price discriminate identified consumers, and firms charge a uniform price on the second interval of unidentified consumers. The data intermediary does not sell all consumer segments to Firm 1 to reduce the competitive effect of information.[23] Fewer consumers are identified in the monopoly market than in the competitive market: $\frac{j_1^*(k)}{k} < \frac{\overline{j_1}^*(\overline{k}) + \overline{j_2}^*(\overline{k})}{\overline{k}}$.

Selling information to one or to both firms

The data intermediary compares its profits when selling information to one or to both firms. Proposition 5 shows that the data intermediary will only sell information to Firm 1 on its monopoly market (Firm 2 remains uninformed).

Proposition 5

In monopoly markets m_i, intermediary DI_i sells information to Firm 1 only.

Proof: see Appendix B.3.

Proposition 5 implies that Firm 2 does not acquire information and stays uninformed, which allows a monopolist data intermediary to maximize the profits of Firm 1, equal to $\pi_1(j_1, \emptyset)$. As the competitive effect of information is stronger

[23]The integer value of j_1^* that maximizes the profits of the data intermediary is chosen by comparing $\pi(\lfloor j_1^* \rfloor)$ and $\pi(\lfloor j_1^* \rfloor + 1)$: $max(\pi(\lfloor j_1^* \rfloor), \pi(\lfloor j_1^* \rfloor + 1))$.

when both firms are informed, their resulting profits are lower, which decreases their willingness to pay for information. On the contrary, a monopolist intermediary can charge a high price when selling information to Firm 1 only, by threatening to sell information to Firm 2 in case Firm 1 declines the offer. This threat increases the willingness to pay of Firm 1 for information, and as there is no competing intermediary from which Firm 1 could acquire information, selling information to Firm 1 only is optimal.

A monopolist data intermediary can fully internalize the competitive effect of information by selling information to only one firm, keeping its competitor uninformed, and by selling fewer segments than in the competitive market. This selling strategy softens the competitive effect of information, and reduces consumer surplus compared to a case where both firms have information.

4.2 Collecting Consumer Data in Monopoly Markets

We analyze in this section the incentives of intermediaries to collect data in monopoly markets, which we compare to those in the competitive market. We find that the incentives of data intermediaries to collect consumer data in monopoly markets are driven by (*i.*) a rent extraction effect that is proportional to the number of consumers that Firm 1 can identify, and by (*ii.*) an outside-option effect that reduces the profits of Firm 1 if it remains uninformed and faces Firm 2 that has acquired information. We show that the incentives to collect data in monopoly markets are higher than those in the competitive market.

Incentives to collect data in monopoly markets

The profits of a data intermediary on its monopoly market are given by the price of information net of the cost to collect data, times the size of the monopoly market m:

$$\Pi_m(k) = m[p_m(k) - c(k)].$$

Consider Firm 1 (w.l.o.g.). The marginal effects of increasing the amount of data collected on the price of information can be decomposed into two effects: a rent-extraction effect and an outside-option effect:

$$
\begin{aligned}
p_m(k) &= \pi_1(j_1^*(k), \emptyset) - \pi_1(\emptyset, \mathcal{P}^k) \\
\implies \frac{\partial p_m(k)}{\partial k} &= \underbrace{\frac{j_1^*(k)}{k}\frac{t}{k^2}}_{\substack{\text{Rent-extraction} \\ \text{effect}}} - \underbrace{\frac{\partial \pi_1(\emptyset, \mathcal{P}^k)}{\partial k}}_{\substack{\text{Outside-option} \\ \text{effect}}}
\end{aligned}
$$

We identify these two effects by applying the envelope theorem to the optimal price of information $p_m(k)$. On top of the rent-extraction effect that we have described in the analysis of the competitive market, the incentives to collect data in monopoly markets are also determined by an outside-option effect. In a monopoly market, more precise information lowers the profits of an uninformed firm facing an informed competitor. A monopolist data intermediary threatens a prospective buyer to sell the reference partition (that includes all available consumer segments) to its competitor, and more segments collected increase the value of this threat.

Proposition 6
The incentives of an intermediary to collect data in its monopoly market depend on:

- *A rent-extraction effect, which is proportional to the transportation cost t, and to the equilibrium number of consumers identified by Firm 1: $\frac{j_1^*(k)}{k}$.*
- *An outside-option effect determined by the profits of Firm 1 when it remains uninformed and Firm 2 acquires information: $\pi_1(\emptyset, \mathcal{P}^k)$*

Comparing the incentives to collect data in the competitive and in monopoly markets

Competition has two impacts on the incentives of $\overline{DI}$ to collect data. First, we have shown that $\overline{DI}$ sells consumer segments on a larger share of consumers in the competitive market than in its monopoly market, and the marginal gains of an increase in $\overline{k}$ are greater in market l according to this stronger rent-extraction effect. Secondly, $\overline{DI}$ cannot exert a threat on firms in market l, because of the competitive pressure exerted by $\underline{DI}$. Hence, an increase in $\overline{k}$ will have no impact on the outside option of a firm in market l, and this second effect lowers the incentives of $\overline{DI}$ to collect data in the competitive market.

Overall, we show in Proposition 7 that the outside-option effect dominates the rent-extraction effect, and $\overline{DI}$ has higher incentives to collect data in its monopoly market than in the competitive market.

Proposition 7

Data intermediaries have higher incentives to collect data in their monopoly markets than in the competitive market.

Proof: See Appendix B.4.

Proposition 7 states that the marginal gains from collecting data are higher in monopoly markets than in the competitive market. This is a central result of this article that has important implications for merger policy, which we discuss in Section 6.

While Proposition 7 may sound intuitive, we have shown that the incentives of $\overline{DI}$ to collect data in its monopoly market and in the competitive market result from two opposite effects: a rent-extraction effect and an outside-option effect. Understanding how these effects may change with different market structures is therefore crucial to better understand the incentives of intermediaries to collect consumer data, and Proposition 7 opens the door to this new research direction.

Amount of data collected in equilibrium

We have established that intermediaries have greater incentives to collect data in their monopoly markets than in the competitive market l. In this section, we use this result to characterize the number of consumer segments k_i collected by each data intermediary, and we analyze which intermediary sells data in market l. We show that, either an equilibrium exists in which a single intermediary i collects more data than its competitors and sells data in market l, or that there is no equilibrium. We provide conditions for an equilibrium to exist where DI_1 collects more data than its competitors.[24]

We introduce further notations. Let $k_i^* = \text{argmax}_{k_i}\{m_i p_{m_i}(k_i) - (m_i + l)c(k_i)\}$ and $\tilde{k}_1^* = \text{argmax}_{k_1}\{m_1 p_{m_1}(k_1) + 2l p_l(k_1, k_2^*) - (m_1 + l)c(k_1)\}$, so that k_i^* and $\tilde{k}_1^*$

[24] In Appendix B.5 we provide conditions for an equilibrium to exist where another intermediary $i > 1$ collects more data than other intermediaries.

denote the optimal number of segments collected respectively by DI_i when selling information only in its monopoly market, and by DI_1 selling information on both its monopoly market and the competitive market l.

We provide in Appendix B.5 a set of conditions $\mathcal{C}$ over the cost function for an equilibrium to exist where DI_1 collects the largest number of segments $\tilde{k}_1^*$ and sells information in the competitive market. Conditions $\mathcal{C}$ require that the cost to collect consumer data increases more than a threshold value after $\tilde{k}_1^*$. This increase ensures that intermediary DI_2 has no interest to collect more data than $\tilde{k}_1^*$. This condition holds for general values of $m_1, .., m_n$, and in particular, in the limit case where all intermediaries have equal market sizes: $m_1 = .. = m_n$. Proposition 8 summarizes this discussion:

Proposition 8

- *(a) Under condition $\mathcal{C}$ an equilibrium exists in which DI_1 collects the highest number of segments $\tilde{k}_1^*$ and sells data in market l.*

  In this equilibrium other intermediaries collect a number of segments that increases with the size of their monopoly markets:

$$\tilde{k}_1^* > k_2^* \geq ... \geq k_n^*.$$

- *(b) When data intermediaries have identical market sizes $m_1 = m_2 = ... = m_n$, an equilibrium has the following property. One data intermediary (1, w.l.o.g.) collects strictly more information than the others who all collect the same number of segments with:*

$$\tilde{k}_1^* > k_2^* = ... = k_n^*.$$

Proof: see Appendix B.5.

Proposition 8 (a) highlights a positive relation between market power, captured by the size of the monopoly market, and the amount of data collected by intermediaries. A data intermediary that is dominant in terms of size of its monopoly market collects more consumer segments than other intermediaries, and DI_1 is the only intermediary that sells information in the competitive market l. Other data intermediaries only sell information in their monopoly markets and collect data proportionally to the sizes of their monopoly markets $\left(\frac{\partial k_i^*}{\partial m_i} > 0\right)$.

Proposition 8 (b) shows that the only possible equilibrium when intermediaries have monopoly markets of identical sizes $m_1 = m_2 = .. = m_n$ is such that one of the intermediaries collects more information than the others. The existence of equilibria depends on the degree of convexity of $c(.)$, and an equilibrium does not necessarily exist. If an equilibrium exists, it must have the features described in Proposition 8 (b).

In the remaining of the article, and for clarity of the analysis we focus on the case characterized in Proposition 8, where DI_1 – with the largest monopoly market – collects more data than its competitors. Under conditions $\mathcal{C}$, another type of equilibrium may exist in which another intermediary i collects more data than DI_1, and sells on the competitive market l. For this equilibrium to exist, there must be an interior solution for DI_i when it sells in market $m_i + l$. We refer to this situation as a leapfrogging strategy that sustains the following equilibrium strategies: DI_i sells in market m_i and l and collects $\tilde{k}_i^*$; DI_1 collects k_1^*. To prove that these strategies constitute an equilibrium, we discuss the incentives of DI_i and DI_1 to deviate from the proposed equilibrium. Consider first DI_1. For cost functions that increase more than a certain threshold after $\tilde{k}_i^*$, DI_1 will not deviate from its strategy. Note that since other DI_j $(j \geq 2 \neq i)$ collect fewer data than DI_1, they will not deviate from their equilibrium strategy if DI_1 does not deviate. Consider now DI_i, by definition, if there is an interior solution, it is deviation proof. We have thus established that there can exist a leapfrogging equilibrium.[25]

5 Consumer Surplus

We compare in this section consumer surplus in monopoly markets m_i and in the competitive market l. Information has two effects on the prices paid by consumers. On the one hand, consumer surplus depends on the number of segments sold to firms, which is larger in the competitive market than in the monopoly markets.

[25] We characterize all equilibria and their existence conditions in Appendix B.5.

Fewer segments sold reduce the intensity of competition in the product market and harm consumers. On the other hand, DI_1 collects more data than other intermediaries, and consumer rent-extraction is greater in the competitive market l than in markets m_i ($i \geq 2$). We then analyze how consumer surplus changes first with the number of consumer segments sold to firms (holding k constant), and second with the amount of data collected k.

Selling information and consumer surplus

As we have seen in Section 3, more consumer segments are sold in the competitive market than in monopoly markets. Suppose that Firm 1 has information on j_1 consumer segments, and Firm 2 has information on j_2 consumer segments. If Firm 1 obtains additional information on segment $[\frac{j_1}{k}, \frac{j_1+1}{k}]$, there are two effects on consumer surplus:

1. A rent extraction effect: Firm 1 price discriminates consumers on $[\frac{j_1}{k}, \frac{j_1+1}{k}]$, which reduces their surplus.

2. A competitive effect: Firm 1 lowers its price on $[\frac{j_1+1}{k}, 1]$, which increases the competitive pressure on Firm 2. In turn, Firm 2 also lowers its price, which has a positive effect on the surplus of consumers over the whole line.

Overall, the second effect always dominates the first, and consumer surplus increases when more consumer segments are sold. Indeed, the rent extraction effect only increases profits on one additional segment, while the competitive effect operates on the whole Hotelling line. Lemma 3 shows that consumer surplus, denoted $CS(j_1, j_2, k)$, increases with the number of consumer segments j_1 and j_2 sold to Firm 1 and to Firm 2.

Lemma 3

For a given j_2, consumer surplus always increases with the number of consumer segments j_1 sold to Firm 1:

$$\forall \;\; j_2, k : \quad \frac{\partial CS(j_1, j_2, k)}{\partial j_1} > 0.$$

Proof: See Appendix B.6

Amount of data collected and consumer surplus

We now discuss the effect of a change in the amount of data collected k on consumer surplus. We perform comparative statics with respect to k holding $\frac{j_1}{k}$ and $\frac{j_2}{k}$ constant. Increasing k reduces the size of the segments and allows firms to better extract consumer surplus. Lemma 4 shows that in equilibrium consumer surplus decreases with k.

Lemma 4

Consumer surplus in equilibrium decreases with k:

$$\frac{\partial CS(j_1, j_2, k)}{\partial k} < 0.$$

Proof: see Appendix B.7.

Surplus Comparison

We can now compare consumer surplus in the competitive market and in the monopoly markets. Let CS_l and CS_i ($i = 1, .., n$) denote aggregate consumer surplus in the competitive market l and in monopoly market i respectively.

Proposition 9

Aggregate consumer surplus is ranked as follows:

$$CS_l > CS_n > CS_{n-1} > .. > CS_1.$$

In the competitive market, information is sold on more consumers than in monopoly markets, and the resulting competitive effect of information yields a higher aggregate consumer surplus.[26] In monopoly markets, surplus decreases with the number of consumer segments collected, leading to the ranking in Proposition 9.

[26] The proof is available upon request and simply compares surplus in equilibrium in the competitive market and in monopoly markets.

6 Implications for Merger Analysis

We now discuss the implications of our results for mergers analysis. We analyze how mergers between data intermediaries change their data strategies as well as the strategies of competing intermediaries, and the impact of these changes on consumer surplus. We show the existence of a new ripple effect: a merger in a given market can have repercussions on other markets where the merging intermediaries do not operate.

In this last section, we study two types of mergers: a merger between a duopoly of data intermediaries and the acquisition of a small intermediary by DI_2. We analyze in Appendix A.2 other important market configurations. Namely, we consider the acquisition of a small intermediary by DI_1, the acquisition of a medium-size intermediary by DI_2, and a merger between DI_1 and DI_2 when the remaining intermediaries are small. Finally, we generalize in Appendix A.4 our $m-l$ framework to multiple competitive markets. We show that a merger between intermediaries can also impact consumer surplus in competitive markets where none of the merging intermediaries sell information.

6.1 Merger Between Duopolist Intermediaries

We consider a situation with two duopolist data intermediaries DI_1 and DI_2. A merger between DI_1 and DI_2 eliminates competition in market l, where the merged entity collects and sells information as a monopolist. This merger is detrimental to consumers and market competition for two main reasons. On the one hand, we have shown in Propositions 4 and 5 that fewer segments are sold in market l after the merger, and to only one firm, which reduces the intensity of competition in the product market. On the other hand, as stated in Proposition 7, more segments are collected after the merger which increases rent extraction from consumers. Overall, consumers are worse off after the merger of two duopolist data intermediaries in the market for information.

6.2 Mergers in an Oligopolistic Market for Information

We now consider a merger in an oligopolistic market for information, and we analyze the acquisition of a small intermediary DI_i ($i > 2$) by DI_2.[27] We assume that the merged entity DI_{2i} has a market share $m_{2i} < m_1$, and we leave the case $m_{2i} > m_1$ to Appendix A.2. DI_{2i} has the following additional incentives to collect data. First, DI_{2i} enjoys a natural cost efficiency. Indeed, DI_2 and DI_i both collect data in market l, while DI_{2i} only collects information once in market l. This cost efficiency can be interpreted as an ability for the merged firm to avoid redundancy when collecting data, increasing the amount of data that it collects. Secondly, DI_{2i} can leverage on a larger market size than DI_2 and DI_i respectively and has higher incentives to collect data.

Data synergies are an important dimension of the market for information. While policymakers have in general interpreted cost efficiencies in a positive way, our results point to a negative effect of such mergers on consumer surplus through a greater amount of data collected.

Contrary to a merger between duopolists, there are still $n - 2 > 0$ remaining competitors in market l after the merger, so that the merged entity still sells information to both firms. However, we will show in this section, that a merger has an impact on consumer surplus through changes in the amount of data collected by DI_{2i} and DI_1.

Indeed, an important determinant of the decision of DI_1 to collect consumer data is the amount of data collected by its direct competitor DI_2 (k_2). As k_2 becomes closer to k_1 – for instance after DI_2 merges with a smaller intermediary – the incentives of DI_1 to differentiate by collecting more segments increase as well. We refer to this incentive as the escape-competition effect.[28] Proposition 10 summarizes this discussion.

Proposition 10 (Escape-competition effect)

[27] For instance, Kesler et al. (2020) highlight a high intensity of competition in the market for data-driven mobile applications, to which our analysis applies. Such market would remain therefore highly competitive, even after a merger between several of these applications.

[28] This effect is reminiscent of models of innovation in a competitive environment (Aghion et al., 2005).

The incentives of DI_1 to collect consumer segments in the competitive market increase with k_2: $\frac{\partial k_1^*}{\partial k_2} > 0$.

Proof: See Appendix A.3.
Increasing k_2 has a positive effect on the marginal value of information for DI_1, given that the equilibrium number of consumers that firms can identify in the competitive market increases with k_2 (see Section 3.2.) The resulting rent-extraction effect is stronger for DI_1, which collects more data when k_2 increases.

Hence, after the merger, DI_{2i} collects more segments than DI_2 due to efficiency gains, and DI_1 collects more segments due to the stronger escape-competition effect. According to Lemma 4, this reduces surplus in markets m_1, m_2, m_i and l. Hence, a merger reduces consumer surplus through changes in the amount of data collected by intermediaries.

We can generalize this escape-competition effect to a setting with several competitive markets. There are potentially $2^n - n$ markets where the n intermediaries may be competing, with variable sizes $l_j \geq 0$ ($j = 1, .., 2^n - n$). We denote by $\mathcal{M}_i$ the set of competitive markets where intermediary DI_i operates. Its total profits Π_i can be written as the sum of profits on its monopoly market and on all markets in $\mathcal{M}_i$:[29]

$$\Pi_i(k_i) = m_i[p_{m_i}(k_i) - c(k_i)] + \sum_{l_h \in \mathcal{M}_i} l_h[p_{1l_h}(k_i) + p_{2l_h}(k_i) - c(k_i)]. \tag{3}$$

In this setting, when an intermediary DI_i collects more consumer segments – for instance because competing intermediaries merge – the escape-competition effect takes place in any competitive market where DI_i is the second largest intermediary. Hence, even markets where none of the merging intermediaries and none of their direct competitors operate can be impacted by the merger. We refer to this mechanism as a *ripple effect*. In these markets, more data is collected after the merger due to this ripple effect, and consumer surplus is reduced.

[29] See Appendix A.4 for an illustration of this ripple effect.

7 Conclusion

Data intermediaries can strategically collect and sell consumer data to maximize the price of information by softening competition in product markets. Ensuring a competitive market for information protects consumer surplus through two main channels. On the one hand, competition between data intermediaries lowers the amount of data collected on consumers, which reduces the ability of information buyers to extract rent from consumers. On the other hand, a competitive market for information is also essential to promote the competitiveness of product markets by limiting the competition-softening strategies of data intermediaries. Hence, with the growing importance of consumer data in most industries, we argue that any merger should be analyzed more carefully in the market for information than in regular markets because the data sold by intermediaries also impact competition in other markets. Our flexible $m-l$ competitive framework provides a new theoretical background to study competition and mergers in the market for information.

Our results support policies that reduce the amount of data on which an intermediary has monopoly access. Recent calls for open-data regulations and data-sharing mandates could therefore limit the ability of intermediaries to soften competition in product markets (Crémer et al., 2019). Similarly, the right to data portability enacted in the General Data Protection Regulation[30] and in the Digital Market Act[31] in Europe, which allows consumers to access and move their data from one intermediary to the other, is likely to increase competition in the market for information and increase consumer surplus.

Our findings also bear significant implications for personal data protection and advocate closer collaboration between data protection agencies and competition authorities to protect consumer privacy and consumer surplus. New regulations limit the amount of personal data collected by intermediaries. For example, the California Consumer Privacy Act provides a detailed list of safeguards to protect personal data. Similarly, a data minimization principle has been enacted in the Health Insurance Portability and Accountability Act in the US, and in the General

[30] General Data Protection Regulation, last accessed, March 15, 2023.

[31] The Digital Markets Act: ensuring fair and open digital markets, last accessed, March 15, 2023.

Data Protection Regulation in Europe. Moreover, recent actions from the Federal Trade Commission call for regulation of the data brokerage industry,[32] and, in the US, states such as Vermont and California have recently passed laws to gain control over the practices of data intermediaries.[33] We show that promoting competition between intermediaries is an efficient way to reduce the amount of data that they collect, and in turn to protect consumer privacy. It remains to be seen how recent regulations, such as the General Data Protection Regulation in the European Union, which has introduced new ways to protect consumers through opt-in, right to be forgotten, data minimization, and privacy by design, will impact the data strategies of large intermediaries and competition in the product market.

References

Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar. Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, 14(4):218–56, 2022.

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

Anat R Admati and Paul Pfleiderer. Selling and trading on information in financial markets. *The American Economic Review*, 78(2):96–103, 1988.

Pauline Affeldt and Reinhold Kesler. Competitors' reactions to big tech acquisitions: Evidence from mobile apps. 2021.

Philippe Aghion, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. Competition and innovation: An inverted-u relationship. *The quarterly journal of economics*, 120(2):701–728, 2005.

S Nageeb Ali, Nima Haghpanah, Xiao Lin, and Ron Siegel. How to sell hard information. *The Quarterly Journal of Economics*, 137(1):619–678, 2022.

Mark Armstrong and John Vickers. Patterns of competitive interaction. *Econometrica*, 90(1):153–191, 2022.

Mark Armstrong and Jidong Zhou. Consumer information and the limits to competition. *American Economic Review*, 112(2):534–77, 2022.

Paul Belleflamme, Wing Man Wynne Lam, and Wouter Vergote. Competitive imperfect price discrimination and market power. *Marketing Science*, 39(5):996–1015, 2020.

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

[33] Column: Shadowy data intermediaries make the most of their invisibility cloak; Los Angeles Times, November 5, 2019.

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. *Annual Review of Economics*, 11, 2019.

Dirk Bergemann, Alessandro Bonatti, and Alex Smolin. The design and price of information. *American economic review*, 108(1):1–48, 2018.

Dirk Bergemann, Alessandro Bonatti, and Tan Gan. The economics of social data. *The RAND Journal of Economics*, 53(2):263–296, 2022.

David Bounie, Antoine Dubus, and Patrick Waelbroeck. Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, 52(2):283–313, 2021. doi: https://doi.org/10.1111/1756-2171.12369. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12369.

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

Zhijun Chen, Chongwoo Choe, and Noriaki Matsushima. Competitive personalized pricing. *Management Science*, 66(9):4003–4023, 2020.

Zhijun Chen, Chongwoo Choe, Jiajia Cong, and Noriaki Matsushima. Data-driven mergers and personalization. *The RAND Journal of Economics*, 53(1):3–31, 2022.

Chongwoo Choe, Stephen King, and Noriaki Matsushima. Pricing with cookies: Behavior-based price discrimination and spatial competition. *Management Science*, 64(12):5669–5687, 2018.

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.

Vidyanand Choudhary, Anindya Ghose, Tridas Mukhopadhyay, and Uday Rajan. Personalized pricing and quality differentiation. *Management Science*, 51(7):1120–1130, 2005.

CMA. Online platforms and digital advertising. 2020.

Dominic Coey, Bradley J Larsen, Kane Sweeney, and Caio Waisman. Scalable optimal online auctions. *Marketing Science*, 2021.

Jacques Crémer, Yves-Alexandre de Montjoye, and Heike Schweitzer. Competition policy for the digital era. *Report for the European Commission*, 2019.

Alexandre De Corniere and Greg Taylor. Data and competition: a general framework with applications to mergers, market structure, and privacy policy. 2020.

Jan De Loecker, Jan Eeckhout, and Gabriel Unger. The Rise of Market Power and the Macroeconomic Implications*. *The Quarterly Journal of Economics*, 135(2):561–644, 01 2020. ISSN 0033-5533. doi: 10.1093/qje/qjz041. URL https://doi.org/10.1093/qje/qjz041.

Peter A Diamond. A model of price adjustment. *Journal of economic theory*, 3(2): 156–168, 1971.

Antoine Dubus and Patrick Legros. The sale of data: Learning synergies before m&as. 2022.

Matthew Elliott, Andrea Galeotti, Andrew Koh, and Wenhao Li. Market segmentation through information. *Available at SSRN 3432315*, 2021.

Drew Fudenberg and J Miguel Villas-Boas. Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, 1:377–436, 2006.

Sanford J Grossman and Joseph E Stiglitz. Information and competitive price systems. *The American Economic Review*, 66(2):246–253, 1976.

Sanford J Grossman and Joseph E Stiglitz. On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408, 1980.

Shota Ichihashi. Competing data intermediaries. *The RAND Journal of Economics*, 52 (3):515–537, 2021.

Ganesh Iyer, David Soberman, and J Miguel Villas-Boas. The targeting of advertising. *Marketing Science*, 24(3):461–476, 2005.

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.

Reinhold Kesler, Michael Kummer, and Patrick Schulte. Competition and privacy in online markets: Evidence from the mobile app industry. In *Academy of Management Proceedings*, volume 2020, page 20978. Academy of Management Briarcliff Manor, NY 10510, 2020.

John E Kwoka Jr and Tommaso M Valletti. Scrambled eggs and paralyzed policy: Breaking up consummated mergers and dominant firms. *Industrial and Corporate Change*, 2020.

Josh Lerner and Jean Tirole. Efficient patent pools. *American Economic Review*, 94(3): 691–711, 2004.

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

Toshihiro Matsumura and Noriaki Matsushima. Should firms employ personalized pricing? *Journal of Economics & Management Strategy*, 24(4):887–903, 2015.

Rodrigo Montes, Wilfried Sand-Zantman, and Tommaso Valletti. The value of personal information in online markets with endogenous privacy. *Management Science*, 65(3): 1342–1362, 2019.

Charles Brian O'kelley and Adam Roger Pritchard. Data marketplace and broker fees, January 8 2009. US Patent App. 11/772,965.

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

Andrea Prat and Tommaso Valletti. Attention oligopoly. *American Economic Journal: Microeconomics*, 14(3):530–57, 2022.

Robert W Rosenthal. A model in which an increase in the number of sellers leads to a higher price. *Econometrica: Journal of the Econometric Society*, pages 1575–1579, 1980.

Andrew Sheehan and Guy Yalif. Multi-round auction and internet marketplace, December 6 2001. US Patent App. 09/746,022.

Lars A Stole. Price discrimination and competition. *Handbook of industrial organization*, 3:2221–2299, 2007.

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.

Hal R Varian. A model of sales. *The American economic review*, 70(4):651–659, 1980.

A Appendix

A.1 Discussion of the Assumptions

Covered markets. We focus our analysis on covered markets, in which firms compete for consumers. When a market is not covered, a firm can charge a monopoly price to close-by consumers, and has no chance of reaching the consumers of its competitor. The optimal selling strategy of an intermediary in this case is simply to sell all available information. As information does not impact competition between firms, there is only a rent extraction effect, which enables the firm to extract more surplus from consumers. Our focus on covered markets allows to enrich the analysis by considering the competitive impact of information, and how data intermediaries can strategically collect and sell consumer information accounting for this impact. As we are interested in the impact of information on competition between firms we focus our analysis on product markets that are covered.

Collecting consumer data. We assume that an intermediary cannot distinguish to which market a consumer belongs before collecting its data. In other words, an intermediary with no information about consumers can simply not discriminate between them before collecting at least some information.

To understand the implications of this assumption, consider the following alternative scenario. Suppose that intermediaries could distinguish consumers in their monopoly markets m_i from those in market l. As the decision of intermediaries to collect data is taken before the information selling stage, the unique equilibrium of this game is characterized as follows: one intermediary collects data and sells it as a monopolist in the competitive market; other intermediaries anticipate that they would achieve zero profits in this market, and therefore, they do not collect data on these consumers. In this case, there is no competitive pressure in market l, where an intermediary sells information as a monopolist. This issue is emphasized by Ichihashi (2021). As we are interested in studying competition between intermediaries, we rule out this scenario from our analysis. Moreover, besides the potential theoretical interest of such equilibrium, finding that intermediaries never compete contrasts with the actual structure of the data brokerage industry, where for instance, Equifax competes with Transunion and Experian in the consumer financial data sector (Pasquale, 2015).

Finally, we adopt for simplicity the specification that an intermediary i collects the same amount of data on consumers in markets m_i and l. However, our results would hold without necessarily assuming that the same data is collected, but by assuming that marginal costs to collect data are positively correlated in both markets – hence the amount of data collected in these markets are strategic complements. This specification captures potential data externalities across consumers, in line with recent literature (Choi et al., 2019; Acemoglu et al., 2022; Bergemann et al., 2022).

Sequential pricing decision. Firm θ sets prices in two steps. First, it sets prices on competitive segments where it shares consumer demand with its competitors. Then, on segments where it is a monopolist, it sets a monopoly price. On a given Hotelling line, each firm knows whether its competitor is informed, and the partition $\mathcal{P}_{-\theta}$.[34]

Sequential pricing decision avoids the nonexistence of Nash equilibrium in pure strategies. Consider indeed the case where Firm 1 sets prices p_{11} and p_{12} simultaneously on a monopoly segment and on a competitive segment. For simplicity consider the case where Firm 2 charges a homogeneous price p_2. The equilibrium prices of Firm 1 are chosen as best responses p_2. Yet in general, p_{11} is not optimal as Firm 1 is a monopolist in this segment. Hence, it can increase p_{11} up to the limit point where Firm 2 would reach a positive demand on the monopoly segment. This situation is not an equilibrium either as Firm 2 has now incentives to increase p_2 and reach a positive demand on both the monopoly and the competitive segments. Setting prices sequentially on monopoly and competitive segments allows firms to charge the highest possible price on their monopoly segments while not having their competitor increase their competitive price as a result. Moreover, sequential pricing allows us to focus our analysis on pure strategy Nash equilibria.

This specification is also common in the literature and is supported by managerial practices. For instance, Acquisti and Varian (2005) use sequential pricing to analyze intertemporal price-discrimination with incomplete information on consumer demand. Choudhary et al. (2005), Jentzsch et al. (2013), Matsumura and Matsushima (2015), Chen et al. (2020) and Belleflamme et al. (2020) also focus on sequential pricing where a higher personalized price is charged to identified consumers after a firm sets a uniform price. Sequential pricing is also common in business practices (see also, Fudenberg and Villas-Boas (2006)). Recently, Amazon has been accused to show higher prices for Amazon Prime subscribers, who pay an annual fee for unlimited shipping services, than for non-subscribers (Lawsuit alleges Amazon charges Prime members for ”free” shipping, Consumer affairs, August 29 2017). Thus Amazon first sets a uniform price and then increases prices for high-consumers who are better identified when they join the Prime program.

Firms purchase information from only one intermediary. We have

[34] This assumption is also standard in Thisse and Vives (1988).

assumed that a firm cannot combine information purchased from different intermediaries There are two justifications for this assumption. First, compared with data intermediaries, firms do not have the same technical tools to combine data from different sources and in different formats in a cost-efficient way. Secondly, most of the time data intermediaries do not sell individual data but rather give information on whether a particular consumer belongs to a specific segment of the demand such as age group, gender, profession, buying frequency, purchase interests, or geographical location.

Overall, allowing the combination of data from different intermediaries would soften the intensity of competition in the market for information, increasing the incentives of intermediaries to collect consumer data and reducing consumer surplus. It may also give intermediaries incentives to organize the market in data pools to exploit the complementarity of different data sets, in the spirit of the market for innovations (Lerner and Tirole, 2004).

Non-overlapping partitions. We assume that in the competitive market, a data intermediary proposes non-overlapping partitions to Firm 1 and Firm 2. Formally, let $[s_{1n}, 1]$ denote the segment located the furthest away from Firm 1 in $\mathcal{P}_1$, and $[0, s_{2n}]$ the segment located furthest away from Firm 2 in $\mathcal{P}_2$: we assume that $s_{1n} \leq s_{2n}$.

With this assumption we do not need to compute profits with all possible partitions, and we can focus only on segments that allow firms to extract surplus from high-valuation consumers. In particular, we can show that selling all available segments to firms yields lower profits than the optimal partition that we characterize under this assumption. Moreover, focusing on non-overlapping partitions simplifies the analysis. A general proof without this assumption is not tractable since there is a high cardinality of the possible combinations of consumer segments, and of the resulting prices and demands.

A.2 Merger Cases

We analyze how alternative merger configurations impact consumer surplus. There are two important elements to consider: the respective sizes of the merged firms before and after the merger, and whether the merger changes the competitive pressure between the two largest intermediaries. We will analyze two typical cases in the following sections.

A.2.1 Start-up acquisition by DI_1

When the largest intermediary acquires a smaller intermediary of market size m_i, two opposite effects change its incentives to collect consumer data. On the one hand, the data intermediary has incentives to collect more data following the cost efficiency described in the previous section. Data intermediary DI_2 does not change the amount of data it collects, and the competitive pressure in market l remains the same before and after the acquisition. On the other hand, the escape-competition effect is now weaker as the merged entity is larger than before the acquisition: $m_1 + m_i > m_1$.

Overall we can show that the first effect always dominates the second, and DI_1 collects more consumer segments after the acquisition than DI_1 before the acquisition.[35] Hence such an acquisition reduces consumer surplus in markets m_1, m_i, and l through an increased ability of firms to extract rent. Consumers in market m_i incur the highest increase in the amount of data collected and the greatest loss of surplus from the merger.

A.2.2 Start up acquisition by DI_2

When DI_2 acquires a smaller intermediary so that $m_2 + m_i > m_1$, the merger changes the identity of the seller in market l: after the merger, DI_{2i} sells information in market l, under the competitive pressure exerted by DI_1. Hence, the incentives of DI_1 to collect data decrease after the merger as it achieves zero profits in market l. Consumers in market m_1 are better identified before the acquisition, and their surplus increases after the acquisition.

This change in the competitive balance in market l also impacts the incentives of DI_{2i} to collect consumer information. Indeed, DI_2 collects more segments after the acquisition than DI_1 before the acquisition following two effects. On the one hand, the relative size of the merged entity is greater than DI_1: $m_2+m_i > m_1$. On the other hand, the competitive pressure after the acquisition is exerted by DI_1,

[35] The proof by contradiction is available upon request.

which collects more segments after the acquisition than DI_2 collected before the acquisition. There are thus winners and losers of such merger, as consumer surplus decreases in markets m_2, m_i, and l where more segments are collected and sold after the merger, but increases in market m_1. Overall, we can show that aggregate consumer surplus decreases after the acquisition. This result is intuitive, and the proof is available upon request.

A.3 Proof of Proposition 10

We prove that the number of consumer segments $\overline{k}$ collected by $\overline{DI}$ increases with the number of consumer segments $\underline{k}$ collected by $\underline{DI}$. We first write the price of information in the competitive market. We substitute the values of $\overline{j_1}^*$, $\overline{j_2}^*$ and $\underline{j_1}^*$ in $\pi_1(\overline{j_1}, \overline{j_2}) - \pi_1(\underline{j_1}, \overline{j_2})$ in the profit functions of Firm 1 and Firm 2. The price of information is identical for both firms and can be written as

$$\begin{aligned} p_l(\overline{k}, \underline{k}) &= [\pi_1(\overline{j_1}, \overline{j_2}) - \pi_1(\underline{j_1}, \overline{j_2})] \\ &= \frac{((12\underline{k} - 11)\overline{k}^2 + (4\underline{k} - 12\underline{k}^2)\overline{k} + 7\underline{k}^2)t}{36\underline{k}^2\overline{k}^2} \end{aligned} \tag{4}$$

Consider the second degree derivative of p_l with respect to $\overline{k}$ and $\underline{k}$:

$$\frac{\partial^2 p_l(\overline{k}, \underline{k})}{\partial \overline{k} \partial \underline{k}} = \frac{1}{9\overline{k}^2\underline{k}^2} \geq 0$$

Thus, the larger the value of $\underline{k}$, the larger the value of the first degree derivative of p_l with respect to $\overline{k}$, and the higher the marginal gain from collecting data.

A.4 Mergers with Several Competitive Markets

We generalize our $m - l$ competitive framework in this section to multiple competitive markets. We show that a merger between intermediaries can also impact consumer surplus in competitive markets where none of the merging intermediaries sell information, thus generating ripple effects beyond the market where the merger takes place.

Consider three intermediaries with monopoly markets $m_1 > m_2 > m_3$, and two competitive markets l_1 and l_2. We focus on the case where DI_1 and DI_2 compete in market l_1 and DI_2 and DI_3 compete in market l_2. Market sizes are chosen so that DI_1 sells information at a positive price in market l_1 and DI_2 sells information at a positive price in market l_2. These different markets correspond for instance to different geographical areas where different subsets of all intermediaries are active. Alternatively, these markets could correspond to different sectors of activities: health data, financial data, and credit ratings.

Using this set-up, we consider a merger between DI_2 and DI_3, and we show how it impacts DI_1 through the escape-competition effect in market l_1. We assume $m_1 > m_2 + m_3 + l_2$, so that DI_1 collects more data than the merged intermediary after the merger.

The profits of each intermediary can be written:

$$\begin{cases} \Pi_1(k_1) = m_1 p_{m_1}(k_1) + l_1(p_1(k_1) + p_2(k_1)) - (m_1 + l_1)c(k_1), \\ \Pi_2(k_2) = m_2 p_{m_2}(k_2) + l_2(p_1(k_2) + p_2(k_2)) - (m_2 + l_1 + l_2)c(k_2), \\ \Pi_3(k_3) = m_3 p_{m_3}(k_3) - (m_3 + l_2)c(k_3). \end{cases}$$

After a merger between DI_2 and DI_3, the merged intermediary collects k_{23} consumer segments, with a monopoly market of size $m_2 + m_3 + l_2$, and competes with DI_1 in market l_1. The efficiency gain allows the merged intermediary to collect more segments than the separate entities, which reduces consumer surplus in markets l_2, m_2 and m_3. Moreover, even though the merger does not directly impact competition in market l_1 – the number of intermediaries in this market remains the same before and after the merger –, the merged intermediary exerts a stronger competitive pressure than DI_2 before the merger. This increases in turn the incentives of DI_1 to collect data in markets l_1 and m_1 through the escape-competition effect.

As DI_1 collects more data after the merger, the escape-competition effect also takes place in any competitive market where DI_1 has the second highest information precision and so on.

B Online Appendix

B.1 Proof of Proposition 2

Optimal information structures in the competitive market

We characterize the partitions that maximize the profits of the firms in the competitive market. We show that a data intermediary 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 characterize the optimal information structure for one of the firms (say Firm 1), and the result directly applies to its competitor.

Suppose that the intermediary sells information to Firm 1 (without loss of generality). The data intermediary can choose any partitions in the sigma-field $\mathbb{P}$ generated by the elementary segments of size $\frac{1}{k}$. There are three types of segments to consider:

- Segments A, where Firm θ serves all consumers but Firm 2 exerts a competitive pressure.
- Segments B, where Firms 1 and 2 compete; both have a positive demand.
- Segments C, where Firm θ has no demand and makes zero profits.

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 a firm. 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 firms, and of size $1 - \frac{j}{k}$ (with j an integer, $j \leq k$). Finally, we can discard segments of type C because information on consumers on these segments does not increase profits.

Throughout the analysis we consider the sale of information to Firm 1, and the result directly generalizes to Firm 2.

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 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. In Figure 4 and in all similar figures, the blue curves represent the demand for Firm 1 (demand for Firm 2 is not represented and corresponds to the complementary demand on the segments). To illustrate, for segments of type A, the blue curve covers the whole segment. For segments of type B, the blue curve only covers part of the segment. 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 $\mathcal{P}'$.

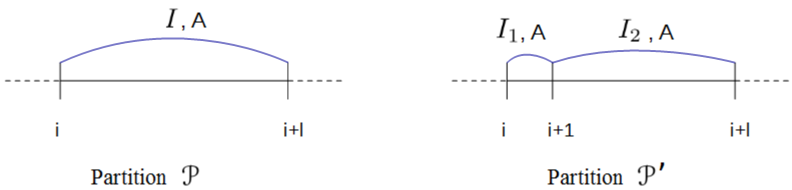

Figure 4: Step 1: segments of type A

To prove this claim, we establish that the profits of Firm 1 is lower with a coarser sub-partition $\mathcal{P}$ with $I = [\frac{i}{k}, \frac{i+l}{k}]$, than with a finer sub-partition $\mathcal{P}'$ obtained by replacing I with two segments: $I_1 = [\frac{i}{k}, \frac{i+1}{k}]$ and $I_2 = [\frac{i+1}{k}, \frac{i+l}{k}]$ (other segments are unchanged).

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 serves all consumers; 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 by their equilibrium values obtained in the previous equations:

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

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 intermediary 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).

We consider the prices set by Firm 1 with different partitions, on the segment where Firm 2 sets a homogeneous price. Going from left to right on the Hotelling line, we look for the first time 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 (right panel of Figure 5). We now show that profits are higher when the data intermediary switches segments I_1 and J. The resulting sub-partition is now $I_1' = [\frac{i}{k}; \frac{i+1}{k}]$ followed by $J' = [\frac{i+1}{k}, \frac{i+l+1}{k}]$ (right panel of Figure 5).

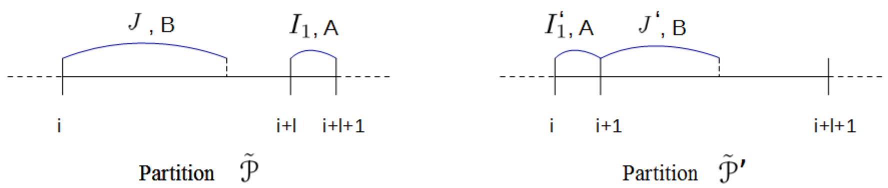

Figure 5: Step 2: relative position of type A and type B segments

The two cases are shown in Figure 5 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 $J \cup I_1$ under partition $\tilde{\mathcal{P}}$ and with $I_1' \cup J'$ 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 partitions, we first characterize type B segments. On segments of type B, both firms must have a positive demand. Eq. 5 gives the conditions for the demands addressed to Firm 1 and to Firm 2 to be positive on such segments.

$$\begin{gathered}\forall \;\; i, l \in \mathbb{N} \;\; s.t. \;\; 0 \leq i \leq k-1 \;\; \text{and} \;\; 1 \leq l \leq k-i-1, \\ \frac{i}{k} \leq \frac{\tilde{p}_2 + t}{2t} \qquad \text{and} \qquad \frac{\tilde{p}_2 + t}{2t} - \frac{l}{k} \leq \frac{i+l}{k}.\end{gathered} \tag{5}$$

Condition $\frac{i}{k} \leq \frac{\tilde{p}_2+t}{2t}$ guarantees that Firm 1 serves consumers on segment J, and $\frac{\tilde{p}_2+t}{2t} - \frac{l}{k} \leq \frac{i+l}{k}$ guarantees that Firm 2 serves positive demand on segment J.

In particular, we use the relation that Eq. 5 characterizes 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 $\tilde{\mathcal{P}}$.

To facilitate the computation of demands on segments of type A, we introduce intermediary notations that characterize the location of these segments (u_i). 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 $\frac{l_i}{k}$.[36] 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 $\frac{1}{k}$. On the segment where Firm 2 sets a homogeneous price and Firm 1 has information, 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:

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

[36] With u_i and s_i integers below k.

Focusing on the side of the line where Firm 2 charges a homogeneous price, 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. Note that the overall profits of Firm 1 also account for the right side of the line, where Firm 1 charges a homogeneous price and Firm 2 might charge personalized prices according to the partition that it acquires. We will show that the changes of partitions owned by Firm 1 will induce an upward shifting of prices by Firm 2, increasing the total profits of Firm 1.

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

The profits of Firm 2 on the left of the line where it charges a homogeneous price are generated on segments of type B only, where the demand for Firm 2 is:

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

The profits of Firm 2 on its left side segment can be written therefore as:

$$\pi_2(\tilde{\mathcal{P}}) = \sum_{i=1}^{n} \tilde{p}_2[\frac{\tilde{p}_{1i}^{B} - \tilde{p}_2 - t}{2t} + \frac{s_i + l_i}{k}]. \tag{6}$$

Firm 1 sets $\tilde{p}_{1i}^{A}$ and $\tilde{p}_{1i}^{B}$ to maximize its profits on the left of the line $\pi_1(\tilde{\mathcal{P}})$. The price set on the segment located furthest right is set to maximize profits also on segments where Firm 2 personalizes its prices, and we will analyze the equilibrium prices and demands on this price in the last stage of the proof. Firm 2 sets its homogeneous price $\tilde{p}_2$ to maximize profits on the left side of the line $\pi_2(\tilde{\mathcal{P}})$, both profits are strictly concave.

Equilibrium prices (except the homogeneous price set by Firm 1 the segment located furthest right) are:

$$\begin{aligned}
\tilde{p}_{1i}^{A} &= t + \tilde{p}_2 - 2\frac{u_i t}{k} \\
\tilde{p}_{1i}^{B} &= \frac{\tilde{p}_2 + t}{2} - \frac{s_i t}{k} = \frac{t}{3} + \frac{2t}{3n}[\sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}]] - \frac{s_i t}{k} \\
\tilde{p}_2 &= -\frac{t}{3} + \frac{4t}{3n}\sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}].
\end{aligned} \tag{7}$$

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. 5: $\frac{\tilde{p}_2+t}{2t} - \frac{l_i}{k} \leq \frac{s_i+l_i}{k}$ for all segments of type B located at $[\frac{s_i}{k}, \frac{s_i+l_i}{k}]$. 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. 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. 5 do not hold:

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

We now compare the profits of Firm 1 with sub-partition $\tilde{\mathcal{P}}$ ($J \cup I_1$) and with sub-partition $\tilde{\mathcal{P}}'$ ($I_1' \cup J'$). 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 $\tilde{\mathcal{P}}$. 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. 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. 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 function of Firm 1 to an increase in the equilibrium price of Firm 2 ($\tilde{p}_2' \geq \tilde{p}_2$).

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. 5.

Thus for any segment, the profits of Firm 1 increase with $\tilde{\mathcal{P}}'$ compared to $\tilde{\mathcal{P}}$. The resulting upward shifting in the homogeneous price set by Firm 2 increases also the profits of Firm 1 on the right side of the line.

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 softens competition between firms and yields a higher profit for Firm 1. 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 $\hat{\mathcal{P}}'$.

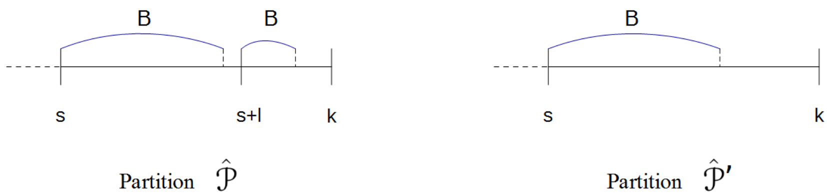

Figure 6: Step 3: demands of Firm 1 on segments of type B

Figure 6 depicts partition $\hat{\mathcal{P}}$ on the left panel, and partition $\hat{\mathcal{P}}'$ on the right panel. Partition $\hat{\mathcal{P}}$ divides the interval in two segments $[\frac{i}{k}, \frac{i+l}{k}]$ and $[\frac{i+l}{k}, \frac{i+l+j}{k}]$, whereas $\hat{\mathcal{P}}'$ only includes segment $[\frac{i}{k}, \frac{i+l+j}{k}]$. 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 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 $0 \leq m \leq n$ segments of type B in partition $\hat{\mathcal{P}}$ that are no longer 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 the homogeneous price set by Firm 2 on the left side of the line and the personalized prices set by Firm 1 (except the homogeneous price on the right side of the line) can be written as:

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

We consider a change of partition under which the two type B segments located furthest on the left are merged. Let $\hat{p}_{1s}^B$ and $\hat{p}_{1s+l}^B$ be the prices on these two segments when the partition is $\hat{\mathcal{P}}$.

$$\hat{p}_{1s}^B = \frac{\hat{p}_2 + t}{2} - \frac{st}{k}, \quad \hat{p}_{1s+l}^B = \frac{\hat{p}_2 + t}{2} - \frac{s+l}{k}t.$$

$\hat{p}_2'$ 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. 5 might not hold as price $\hat{p}_2$ varies depending on the partition acquired by Firm 1. As $\hat{p}_2$ is greater with coarser partitions, some segments that are of type B with partition $\hat{\mathcal{P}}$ are then of type A with partition $\hat{\mathcal{P}}'$. We note $\tilde{s}_i$ the m segments for which it is the case. We then have:

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

The profits of Firm 1 on all segments located on the left except the homogeneous price located on the right are written as follows, with partitions $\hat{\mathcal{P}}$ and $\hat{\mathcal{P}}'$.

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

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

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

We consider the terms separately. First,

$$\begin{aligned}
&\frac{t}{2}\sum_{i=1,s_i\neq\tilde{s}_i}^{n}[\frac{\hat{p}'_2+t}{2t} - \frac{s_i}{k}]^2 - \frac{t}{2}\sum_{i=1,s_i\neq\tilde{s}_i}^{n}[\frac{\hat{p}_2+t}{2t} - \frac{s_i}{k}]^2 \\
=&\frac{t}{2}\sum_{i=1,s_i\neq\tilde{s}_i}^{n}[[\frac{2}{3(n-m)}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}]]^2 \\
&+ [\frac{\hat{p}_2+t}{2t} - \frac{s_i}{k}][\frac{4}{3(n-m)}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}]]] \\
\geq&\frac{t}{2}[\frac{\hat{p}_2+t}{2t} - \frac{s+l}{k}]\frac{4}{3}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}].
\end{aligned} \tag{8}$$

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. 5 hold for price $\hat{p}'_2$ and do not hold for price $\hat{p}_2$. 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:

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

We replace $\tilde{s}_i$ by its upper bound value and then $\tilde{l}_i$ by its lower bound value. We can rewrite Eq. 8 for all permissible values of $\hat{p}'_2$:

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

Getting back to the difference in profits, we obtain:

$$\begin{aligned}\Delta\pi_1 &\geq \frac{t}{2}[\frac{\hat{p}_2+t}{2t} - \frac{s+l}{k}]\frac{4}{3}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}] - \frac{t}{2}[\frac{\hat{p}_2+t}{2t} - \frac{s+l}{k}]^2 \\ &\geq \frac{t}{2}[\frac{\hat{p}_2+t}{2t} - \frac{s+l}{k}][\frac{\hat{p}_2}{2t} + \frac{s+l}{3k} - \frac{1}{6}].\end{aligned} \tag{9}$$

The first bracket of Equation 9 is positive given Eq. 5. The second bracket is positive if $\frac{\hat{p}_2}{2t} + \frac{s+l}{3k} \geq \frac{1}{6}$. A sufficient condition for this result to hold is $\hat{p}_2 \geq \frac{t}{3}$. We prove that this inequality is always satisfied by showing that the reference partition minimizes the price and profit of Firm 2, and that in this case, $\hat{p}_2 \geq \frac{t}{2}$.[37] And as this price is greater than $\frac{t}{3}$, the second bracket of Equation 9 is positive. This proves that $\Delta\pi_1 \geq 0$.

The upward shifting of the homogeneous price set by Firm 2 implies that the total profits of Firm 1 increase.

Conclusion

Applying these three steps to Firm 1 and Firm 2 increases their profits, and yields a situation in which:

- Each firm charges prices as a constrained monopolist on segments located closest to their locations.
- Firm 1 has at most two segments s_1 and s_2 on the rest of the line where it sets competitive prices.

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

- Firm 2 has at most two segments s_3 and s_4 on the rest of the line where it sets competitive prices.

Considering this situation, it is easy to compare profits under different values of s_1, s_2, s_3, s_4 to show that the optimal partition for Firm 1 (and similarly for Firm 2) includes only two intervals. 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 $[\frac{j}{k}, 1]$. ■

Equilibrium numbers of segments sold

We characterize the equilibrium prices and numbers of segments sold in the competitive market. We compute the number of segments proposed by data intermediaries $\overline{DI}$ and $\underline{DI}$ to Firm 1 and Firm 2 in equilibrium.

We compute prices and profits in equilibrium when both firms are informed with the optimal partitions found above. Firm 1 is a monopolist on the $\overline{j_1}$ segments of size $\frac{1}{k}$ in $[0, \frac{\overline{j_1}}{k}]$ and Firm 2 has information on $[1-\frac{\overline{j_2}}{k}, 1]$. On $[\frac{\overline{j_1}}{k}, 1]$ Firm 1 sets a unique price p_1 and gets demand d_1. Similarly, on $[0, 1-\frac{\overline{j_2}}{k}]$ 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 $[0, \frac{\overline{j_2}}{k}]$ where it cannot identify consumers.

Prices set by Firm 2 on segments in interval $[\frac{\overline{j_2}}{k}, 1]$ 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: $d_1 = \frac{p_2 - p_1 + t}{2t} - \frac{\overline{j_1}}{k}$ (resp. $d_2 = 1 - \frac{\overline{j_2}}{k} - \frac{p_2 - p_1 + t}{2t}$).

Step 2: profits of the firms.

The profits of the firms are:

$$\pi_1 = \sum_{i=1}^{\overline{j_1}} d_{1i}p_{1i} + d_1 p_1 = \sum_{i=1}^{\overline{j_1}} \frac{1}{k}(p_2 + t - 2t\frac{i}{k}) + (\frac{p_2 - p_1 + t}{2t} - \frac{\overline{j_1}}{k})p_1,$$
$$\pi_2 = \sum_{i=1}^{\overline{j_2}} d_{2i}p_{2i} + d_2 p_2 = \sum_{i=1}^{\overline{j_1}} \frac{1}{k}(p_1 + t - 2t\frac{i}{k}) + (\frac{p_1 - p_2 + t}{2t} - \frac{\overline{j_2}}{k})p_2.$$

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:

$$p_1 = t[1 - \frac{2}{3}\frac{\overline{j_2}}{k} - \frac{4}{3}\frac{\overline{j_1}}{k}], \quad p_2 = t[1 - \frac{2}{3}\frac{\overline{j_1}}{k} - \frac{4}{3}\frac{\overline{j_2}}{k}].$$

These best responses yield the following demands, prices, and profits:

$$p_{1i} = 2t - \frac{4}{3}\frac{\overline{j_2}t}{k} - \frac{2}{3}\frac{\overline{j_1}t}{k} - 2\frac{it}{k}, \quad p_{2i} = 2t - \frac{4}{3}\frac{\overline{j_1}t}{k} - \frac{2}{3}\frac{\overline{j_2}t}{k} - 2\frac{it}{k}.$$

$$d_1 = \frac{1}{2} - \frac{2}{3}\frac{\overline{j_1}}{k} - \frac{1}{3}\frac{\overline{j_2}}{k}, \quad d_2 = \frac{4}{3}\frac{\overline{j_2}}{k} - \frac{1}{2} - \frac{1}{3}\frac{\overline{j_1}}{k}.$$

$$\begin{aligned}\pi_1^* &= \sum_{i=1}^{\overline{j_1}} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{\overline{j_1}}{k} - \frac{2}{3}\frac{\overline{j_2}}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{\overline{j_1}}{k} - \frac{1}{3}\frac{\overline{j_2}}{k})t[1 - \frac{2}{3}\frac{\overline{j_2}}{k} - \frac{4}{3}\frac{\overline{j_1}}{k}] \\ &= \frac{t}{2} - \frac{7}{9}\frac{\overline{j_1}^2 t}{k^2} + \frac{2}{9}\frac{\overline{j_2}^2 t}{k^2} - \frac{4}{9}\frac{\overline{j_1 j_2}t}{k^2} + \frac{2}{3}\frac{\overline{j_1}t}{k} - \frac{2}{3}\frac{\overline{j_2}t}{k} - \frac{\overline{j_1}t}{k^2}.\end{aligned} \tag{10}$$

$$\begin{aligned}\pi_2^* &= \sum_{i=1}^{\overline{j_2}} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{\overline{j_2}}{k} - \frac{2}{3}\frac{\overline{j_1}}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{\overline{j_2}}{k} - \frac{1}{3}\frac{\overline{j_1}}{k})t[1 - \frac{2}{3}\frac{\overline{j_1}}{k} - \frac{4}{3}\frac{\overline{j_2}}{k}] \\ &= \frac{t}{2} - \frac{7}{9}\frac{\overline{j_2}^2 t}{k^2} + \frac{2}{9}\frac{\overline{j_1}^2 t}{k^2} - \frac{4}{9}\frac{\overline{j_1 j_2}t}{k^2} + \frac{2}{3}\frac{\overline{j_2}t}{k} - \frac{2}{3}\frac{\overline{j_1}t}{k} - \frac{\overline{j_2}t}{k^2}.\end{aligned} \tag{11}$$

Data intermediary $\underline{DI}$ with the second highest number of segments collected competes à la Bertrand with $\overline{DI}$. It exerts the maximal competitive pressure by proposing respectively to Firm 1 and Firm 2 information partitions $\underline{j_1}$ and $\underline{j_2}$ that maximize their profits $\pi_1(\underline{j_1}, \overline{j_2})$ and $\pi_2(\underline{j_2}, \overline{j_1})$.

By replacing variables $\underline{j_1}$ and $\overline{j_2}$ into π_1 (and respectively for π_2), we obtain the following expressions:

$$\pi_1(\underline{j_1}, \overline{j_2}) = \frac{t}{2} - \frac{7}{9}\frac{(\underline{j_1})^2 t}{\underline{k}^2} + \frac{2}{9}\frac{(\overline{j_2})^2 t}{\overline{k}^2} - \frac{4}{9}\frac{\underline{j_1}\overline{j_2} t}{\underline{k}\overline{k}} + \frac{2}{3}\frac{\underline{j_1} t}{\underline{k}} - \frac{2}{3}\frac{\overline{j_2} t}{\overline{k}} - \frac{\underline{j_1} t}{\underline{k}^2}$$

$$\pi_2(\underline{j_2}, \overline{j_1}) = \frac{t}{2} - \frac{7}{9}\frac{(\underline{j_1})^2 t}{\underline{k}^2} + \frac{2}{9}\frac{(\overline{j_1})^2 t}{\overline{k}^2} - \frac{4}{9}\frac{\underline{j_2}\overline{j_1} t}{\underline{k}\overline{k}} + \frac{2}{3}\frac{\underline{j_2} t}{\underline{k}} - \frac{2}{3}\frac{\overline{j_1} t}{\overline{k}} - \frac{\underline{j_2} t}{\underline{k}^2}$$

Data intermediary $\underline{DI}$ maximizes simultaneously these two profit functions with respect to $\underline{j_1}$ and $\underline{j_2}$. Simultaneously, data intermediary $\overline{DI}$ with the highest information precision $\overline{k}$ maximizes its profits by maximizing with respect to $\overline{j_1}$ and $\overline{j_2}$ the sum:

$$p_l(\overline{j_1}, \overline{j_2}) + p_l(\overline{j_2}, \overline{j_1}) = \pi_1(\overline{j_1}, \overline{j_2}) - \pi_1(\underline{j_1}, \overline{j_2}) + \pi_2(\overline{j_2}, \overline{j_1}) - \pi_2(\underline{j_2}, \overline{j_1})$$

$$= \frac{(7\underline{k}\overline{k}(\underline{j_2})^2 + (4\underline{k}\overline{j_1} - 6\underline{k} + 9)\overline{k}\underline{j_2} + 7\underline{k}\overline{k}(\underline{j_1})^2 + (4\underline{k}\overline{j_2} - 6\underline{k} + 9)\overline{k}\underline{j_1})t}{9\underline{k}\overline{k}} + \frac{((-7\underline{k}(\overline{j_2})^2 + (6\underline{k} - 8\underline{k}\overline{j_1})\overline{j_2} - 7\underline{k}(\overline{j_1})^2 + 6\underline{k}\overline{j_1})\overline{k} - 9\underline{k}\overline{j_2} - 9\underline{k}\overline{j_1})t}{9\underline{k}\overline{k}} \tag{12}$$

Thus equilibrium variables $\overline{j_1}, \overline{j_2}, \underline{j_1}, \underline{j_2}$ are chosen as simultaneous best responses. FOCs on $\overline{j_1}$, $\overline{j_2}$, $\underline{j_1}$ and $\underline{j_2}$ give respectively in equilibrium:

$$\overline{j_1}^* = \overline{j_2}^* = \frac{\overline{k}}{3} - \frac{\overline{k}}{9\underline{k}} - \frac{7}{18}$$

$$\underline{j_1}^* = \underline{j_2}^* = \frac{\underline{k}}{3} - \frac{11}{18} + \frac{\underline{k}}{9\overline{k}}$$

In order to find the optimal integer value of j_1 and j_2, we can consider the best outcome among the two integers closest to the optimum values of $\overline{j_1}$, $\overline{j_2}$, $\underline{j_1}$ and $\overline{j_2}$.

B.2 Proof of Proposition 4: Optimal information structure in a monopoly market when the data intermediary sells information to one firm

We begin with describing the mechanism used by an intermediary to sell information in its monopoly market.

Selling mechanism

In order to maximize the price of information, the data intermediary designs two simultaneous auctions, and only the partition with the highest bid will be sold. We focus on pure strategy Nash equilibria. Consider a given partition $\mathcal{P}_1$. We first characterize the price of information and then obtain the optimal partition.

Firm 1 with the highest willingness to pay knows the bid of Firm 2 and has interest to underbid from its true valuation. Thus, a firm can bid just above the willingness to pay of its competitor and win the auction, which reduces the price of information. To avoid underbidding by Firm 1, in auction 1 $\mathcal{P}_1$ is auctioned with a reserve price p_1^m.[38] The reference partition $\mathcal{P}^k$ that includes all k information segments is auctioned in auction 2, in order to exert a maximal threat on Firm 1 and to maximize its willingness to pay for $\mathcal{P}_1$. Participation of both firms is guaranteed as the data intermediary sets no reserve price in auction 2.

Consider the optimal strategies of Firm 1 and Firm 2. Firm 2 will bid $\pi_2(\mathcal{P}^k, \emptyset) - \pi_2(\emptyset, \mathcal{P}^k)$ in auction 2 that corresponds to its willingness to pay for partition $\mathcal{P}^k$, as its worst outside option is to face Firm 1 informed with k. However, Firm 2 will never bid above the reserve price for $\mathcal{P}_1$. Consider now the optimal strategy of Firm 1. Firm 1 can bid for partition $\mathcal{P}^k$, pay a price $\pi_1(\mathcal{P}^k, \emptyset) - \pi_1(\emptyset, \mathcal{P}^k)$, and make profits $\pi_1(\emptyset, \mathcal{P}^k)$. On the other hand, Firm 1 can also participate to the auction with $\mathcal{P}_1$, win the auction by bidding the reserve price p_1^m, and make profits $\pi_1(\mathcal{P}_1, \emptyset) - p_1^m$. The data intermediary will set a reserve price $p_1^{m*} = \pi_1(\mathcal{P}_1, \emptyset) - \pi_1(\emptyset, \mathcal{P}^k) - \epsilon$, where ϵ is an arbitrary small positive number. Thus, $\pi_1(\mathcal{P}_1, \emptyset) - p_1^{m*} > \pi_1(\emptyset, \mathcal{P}^k)$, and since only one partition is sold, it will be $\mathcal{P}_1$. In equilibrium, Firm 1 bids p_1^{m*} for $\mathcal{P}_1$, and Firm 2 bids $\pi_2(\mathcal{P}^k, \emptyset) - \pi_2(\emptyset, \mathcal{P}^k)$. The partitions are therefore $(\mathcal{P}_1, \mathcal{P}^k)$.

The proof of the optimal partition is identical to the proof of Proposition 2, in

[38] For instance, Coey et al. (2021) analyze the role of reserve prices in repeated online auctions.

the case where one of the firms is not informed. Hence we only need to prove that the profits of an uninformed firm are minimized when its competitor acquires $\mathcal{P}^k$.

The profits of an uninformed firm are minimized when its competitor acquires $\mathcal{P}^k$

To prove this claim we consider Firm 1 informed and Firm 2 uninformed. We consider prices and demand on a segment of length $\frac{l}{k}$, $[\frac{s}{k}, \frac{s+l}{k}]$, and we show that partitioning this segment into two subsegments $[\frac{s}{k}, \frac{s+1}{k}]$ and $[\frac{s+1}{k}, \frac{s+l}{k}]$ reduces the price set by Firm 2 as well as it demand on $[\frac{s}{k}, \frac{s+l}{k}]$, which overall lowers its profits. By iterating this argument, we can conclude that the reference partition $\mathcal{P}^k$ minimizes the profit of the uninformed firm.

We can write the equilibrium price set by Firm 2 with partition $\mathcal{P}$:

$$p_2 = -\frac{t}{3} + \frac{4t}{3n} \sum_{i=1}^{n} [\frac{s_i}{2k} + \frac{l_i}{k}]$$

This term is proportional to the average of $\frac{s_i}{2k} + \frac{l_i}{k}$'s. We show that this value is smaller with finer partitions.

We rule out the case where Firm 1 is a monopolist on $[\frac{s}{k}, \frac{s+l}{k}]$, as prices and profit of Firm 2 do not change with finer subsegments in this case.

Consider the case where Firm 1 and Firm 2 compete on $[\frac{s}{k}, \frac{s+l}{k}]$. There are two cases to consider when partitioning this segment into $[\frac{s}{k}, \frac{s+1}{k}]$ and $[\frac{s+1}{k}, \frac{s+l}{k}]$.

First, Firm 1 is a monopolist on $[\frac{s}{k}, \frac{s+1}{k}]$, and firms compete on $[\frac{s+1}{k}, \frac{s+l}{k}]$. The price set by Firm 2 with this second partition decreases as on segment $[\frac{s+1}{k}, \frac{s+l}{k}]$ we have $\frac{s}{2k} + \frac{l}{k} > \frac{s+1}{2k} + \frac{l-1}{k}$. It is clear that demand for Firm 2 also decreases as Firm 1 sets a price on $[\frac{s+1}{k}, \frac{s+l}{k}]$ instead of $[\frac{s}{k}, \frac{s+l}{k}]$. In reaction the aggregate profit of Firm 2 over the unit line decreases.

Secondly, Firm 1 and Firm 2 compete on $[\frac{s}{k}, \frac{s+1}{k}]$ and on $[\frac{s+1}{k}, \frac{s+l}{k}]$.

To show that the price set by Firm 2 decreases with this new partition, we compare the right side of the expression of price p_2: $\frac{4t}{3n} \sum_{i=1}^{n} [\frac{s_i}{2k} + \frac{l_i}{k}]$. This term is the average of $\frac{s_i}{2k} + \frac{l_i}{k}$. To prove that the price set by Firm 2 decreases, we show that this average is lower with the second than with the first partition.

Consider a typical element $\frac{s}{2k} + \frac{l}{k}$ of $[\frac{s}{k}, \frac{s+l}{k}]$. Similarly, consider a typical element $\frac{1}{2}[\frac{s}{2k} + \frac{s+1}{2k} + \frac{l-1}{k} + \frac{1}{k}]$ of the finer partition $[\frac{s}{k}, \frac{s+1}{k}] \cup [\frac{s+1}{k}, \frac{s+l}{k}]$.

The first term is larger than the second as

$$\frac{s}{2k} + \frac{l}{k} > \frac{1}{2}[\frac{s}{2k} + \frac{s+1}{2k} + \frac{l-1}{k} + \frac{1}{k}].$$

It is clear that demand for Firm 2 also decreases as Firm 1 can better target consumers and compete more fiercely with finer segments. In reaction the aggregate profits of Firm 2 over the unit line are smaller with the finer partition than with the coarser one. This establishes the result.

Equilibrium number of segments in the monopoly markets

We compute prices and profits in equilibrium when information is sold to one firm. Without loss of generality we consider the case where Firm 1 is informed only. Firm 1 has the partition on $[0, \frac{j_1}{k}]$ that includes j_1 segments of size $\frac{1}{k}$, and has no information on consumers on $[\frac{j_1}{k}, 1]$. We write in step 1 prices and demands, in step 2 we give the profits, and solve prices and profits in equilibrium in step 3.

Step 1: prices and demands.

On each segment of size $\frac{1}{k}$, Firm 1 sets a price p_{1i}, $i = 1, .., j_1$, and consumer demand is: $d_{1i} = \frac{1}{k}$. Let's p_2 denote the unique price set by Firm 2. Prices on each segment are determined by the indifferent consumer of each segment located at its right extremity, $\frac{i}{k}$: $V - t\frac{i}{k} - p_{1i} = V - t(1 - \frac{i}{k}) - p_2 \implies \frac{i}{k} = \frac{p_2 - p_{1i} + t}{2t}$ $\implies p_{1i} = p_2 + t - 2t\frac{i}{k}$. On the rest of the unit line, 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, 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 $d_1 = x - \frac{j_1}{k} = \frac{p_2 - p_1 + t}{2t} - \frac{j_1}{k}$. 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:

$$\pi_1 = \sum_{i=1}^{j_1} d_{1i} p_{1i} + d_1 p_1 = \sum_{i=1}^{j_1} \frac{1}{k}(p_2 + t - 2t\frac{i}{k}) + (\frac{p_2 - p_1 + t}{2t} - \frac{j_1}{k})p_1,$$

$$\pi_2 = d_2 p_2 = \frac{p_1 - p_2 + t}{2t} p_2.$$

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 $p_1 = t[1 - \frac{4}{3}\frac{j_1}{k}]$ and $p_2 = t[1 - \frac{2}{3}\frac{j_1}{k}]$. By replacing these values in profits and demands we find that: $p_{1i} = 2t[1 - \frac{i}{k} - \frac{1}{3}\frac{j_1}{k}]$, $d_1 = \frac{1}{2} - \frac{2}{3}\frac{j_1}{k}$ and $d_2 = \frac{1}{2} - \frac{1}{3}\frac{j_1}{k}$. Profits are:[39]

$$\begin{aligned}\pi_1^* &= \sum_{i=1}^{j_1} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j_1}{k}] + \frac{t}{2}(1 - \frac{4}{3}\frac{j_1}{k})^2 \\ &= \frac{t}{2} + \frac{2j_1 t}{3k} - \frac{7t}{9}\frac{j_1^2}{k^2} - \frac{tj_1}{k^2} \\ \pi_2^* &= \frac{t}{2} + \frac{2t}{9}\frac{j_1^2}{k^2} - \frac{2}{3}\frac{j_1 t}{k}.\end{aligned} \tag{13}$$

We can now determine the optimal size j_1^* when the data intermediary only sells information to Firm 1, by maximizing profits with respect to j_1. The profits of the data intermediary when it sells to one firm are:[40]

$$\begin{aligned}\Pi_1(j_1) &= \pi_1(j_1, \emptyset) - \pi_1(\emptyset, \mathcal{P}^k) \\ &= \frac{3t}{8} + \frac{2j_1 t}{3k} - \frac{t}{4k} - \frac{7j_1^2 t}{9k^2} - \frac{j_1 t}{k^2} - \frac{t}{8k^2}.\end{aligned}$$

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

We show that more consumers are identified in the competitive market than in monopoly markets by comparing $\frac{\overline{j_1}(k)+\overline{j_2}(k)}{k} = 2[\frac{1}{3} - \frac{1}{9\underline{k}} - \frac{7}{18\overline{k}}]$ with $\frac{j_1^m(k')}{k'} = \frac{6k'-9}{14k'}$:

$$2[\frac{1}{3} - \frac{1}{9\underline{k}} - \frac{7}{18\overline{k}}] - \frac{6k'-9}{14k'} = \frac{((30\underline{k} - 28)k' + 81\underline{k})\overline{k} - 98\underline{k}k'}{126\underline{k}k'\overline{k}}$$

which is clearly positive for $k', \underline{k}, \overline{k} \geq 2$. ■

B.3 Proof of Proposition 5

Suppose now that the data intermediary sells information to both firms. We first describe the selling mechanism. Then we characterize the optimal information structure in a lemma that we prove below, and that we will use to derive the profits of the intermediary in equilibrium, and to show that it is more profitable to sell information to only one firm.

By abuse of notation, let $\mathcal{P}_1$ and $\mathcal{P}_2$ denote now the optimal partitions sold

[39] For $p_{1i} \geq 0 \implies \frac{j_1}{k} \leq \frac{3}{4}$. Profits are equal whatever $\frac{j_1}{k} \geq \frac{3}{4}$.

[40] The expression of $\pi(\emptyset, \mathcal{P}^k)$ is provided in Liu and Serfes (2004).

to Firm 1 and Firm 2 respectively. $\pi_1(\mathcal{P}_1, \mathcal{P}_2)$ and $\pi_2(\mathcal{P}_2, \mathcal{P}_1)$ are the respective profits of Firm 1 and Firm 2 with partitions $\mathcal{P}_1$ and $\mathcal{P}_2$. Partitions are potentially different from those when the data intermediary sells information to Firm 1 only.

Simultaneous auctions, selling to both firms in a monopoly market:

The data intermediary simultaneously auctions partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ in two separate auctions: Firm 1 (Firm 2) can bid in the two auctions but is only interested in partition $\mathcal{P}_1$ ($\mathcal{P}_2$). Since both firms are guaranteed to obtain their preferred partitions, they will underbid in both auctions from their true valuation. To avoid underbidding, a data intermediary respectively sets reserve prices p_{21}^m and p_{22}^m that correspond to the willingness to pay of Firm 1 for $\mathcal{P}_1$ and of Firm 2 for $\mathcal{P}_2$. Since partition $\mathcal{P}_2$ is optimal for Firm 2, and since Firm 1 has a lower valuation for this partition, Firm 1 will not bid above p_{22}^m in the auction for $\mathcal{P}_2$, and similarly Firm 2 will not bid above p_{21}^m in the auction for $\mathcal{P}_1$. In equilibrium, the data intermediary maximizes the sum of the willingness to pay of each firm for information: $p_{21}^m + p_{22}^m = \pi_1(j_{21}, j_{22}) - \pi_1(\emptyset, j_{22}) + \pi_2(j_{22}, j_{21}) - \pi_2(\emptyset, j_{21})$. We now derive the optimal information structure.

Optimal information structure: selling information to both firms in a monopoly market

We show that the optimal partitions include all available consumer segments:

$$j_{21}^*(k) = j_{22}^*(k) = k.$$

There are two classes of partitions to consider in order to find the optimal partitions sold to firms. A first class $\mathcal{C}$ is composed of segments closest to a firm's location. The other class is composed of the remaining partitions. There are two local maximum, one in each class, which we will characterize and then compare profits in both cases.

We first characterize the optimal information structure in class $\mathcal{C}$.

Part a: optimal information structure when the data intermediary sells information to both firms in class $\mathcal{C}$

For each firm, the partition divides the unit line into two intervals. The first interval is partitioned in j segments of size $\frac{1}{k}$ identifying the closest consumers to a firm. The second interval of size $1 - \frac{j}{k}$ 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 assume that the unit line is composed of one interval where firms compete, located at the middle of the line. Information structures that are ruled out by this assumption are those that allow firms to poach consumers located far away from their locations. Selling consumer segments far away from the location of a firm has two conflicting effects on the profits of the data intermediary. On the one hand, partitions ruled out by this assumption lower the valuation of the firms for information because they intensify competition in the market. On the other hand, these partitions also worsen the outside options of firms by lowering their profits if they remain uninformed, which increases their valuation for information. Hence the two effects go in opposite directions. Showing that the first effect dominates the second is not tractable without this assumption, given the high cardinality of the possible combinations of consumer segments. Additionally, there is no evidence of firm strategies targeting consumers who do not belong to their core market. On the contrary, the marketing literature has emphasized the benefits of targeting ads to consumer segments with the strongest preferences (Iyer et al., 2005). As we will show, the optimal partition under this assumption is similar to the optimal partition when the data intermediary sells information to one firm.

Inequalities in Eq. 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. 5 guarantees that there is positive demand for Firm 1, whereas the second part guarantees positive demand for Firm 2. Inequalities in Eq. 5 are expressed as a function of p_2 without loss of generality. We use Eq. 5 to characterize type A and type B segments, in order to compute the profits of the firms.

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

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

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 the profits of a firm when both of them are informed and 2) such a partition does not change the profits of an uninformed firm facing an informed competitor.

1) *A partition that maximizes the profits of a firm when both of them are informed 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 in Appendix B.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 the profits of an uninformed firm facing an informed competitor.*

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_1 segments of size $\frac{1}{k}$, and is located at $[0, \frac{j_1}{k}]$. Consumers are unidentified on the second interval of size $1 - \frac{j_1}{k}$ located at $[\frac{j_1}{k}, 1]$. For Firm 2, the first interval is partitioned in j_2 segments of size $\frac{1}{k}$, and is located at $[1 - \frac{j_2}{k}, 1]$. Consumers are unidentified on the second interval of size $1 - \frac{j_2}{k}$ located at $[0, 1 - \frac{j_2}{k}]$.

In the following section we will compute interior solutions with $j_1, j_2 \in [0, \frac{k}{2}]$, and we will compare profits with the interior solution with profits with the corner solution where all information is sold to both firms. ■

Part b: the intermediary sells symmetric information to both firms in class $\mathcal{C}$

We show now that selling symmetric information is optimal for the data intermediary, that is, in equilibrium $j_1 = j_2$.

Prices and profits in equilibrium are provided in Appendix B.1:

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

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

The data intermediary maximizes the following profit function:

$$\Pi_2(j_1, j_2) = (\pi_1(j_1, j_2) - \pi_1(\emptyset, j_2)) + (\pi_2(j_1, j_2) - \pi_2(\emptyset, j_1))$$
$$= -\frac{7}{9}\frac{j_2^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_2 t}{k} - \frac{j_2 t}{k^2} - \frac{7}{9}\frac{j_1^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_1 t}{k} - \frac{j_1 t}{k^2}.$$

At this stage, straightforward FOCs with respect to j_1 and j_2 confirm that, in equilibrium, $j_1 = j_2$. The fact that the solution is a maximum is directly found using the determinant of the Hessian matrix.

The profit of the data intermediary when both firms are informed with partitions $j_1 = j_2 = j \in [0, \frac{k}{2}]$ is: $\Pi_2(j) = 2w_2 = 2[\frac{2jt}{3k} - \frac{11j^2t}{9k^2} - \frac{jt}{k^2}]$.

FOC with respect to j leads to $j_2^* = \frac{6k-9}{22}$ and: $\Pi_2^* = \frac{2t}{11} - \frac{6t}{11k} + \frac{9t}{22k^2}$.

The remaining partitions out of class $\mathcal{C}$ necessarily include all segments closest to a firm's location and nothing after $\frac{1}{2}$. We show that the price of information is maximized in this class when all information is sold to both firms.

Different partitions in this class include different numbers of segments closest to a firm's competitor. Because selling only part of the segments does not yield the highest feasible threat, and that the profits of the firms when both of them are informed are identical for all partitions in this class, we can conclude that the profits are maximized when all information is sold to both firms.

We can write the profit of the data intermediary when all information is sold by replacing j_1, j_2 by $\frac{k}{2}$ to obtain firms' profits when both firms are informed ($\pi_\theta(k, k) = \frac{t}{4} - \frac{t}{2k}$), and by considering the profits of an uninformed firm facing a competitor informed with all segments, given in Liu and Serfes (2004) ($\pi_\theta(\emptyset, \mathcal{P}^k) = \frac{t}{8} + \frac{t}{4k} + \frac{t}{8k^2}$): $\Pi_2^{all} = \frac{t}{4} - \frac{3t}{2k} - \frac{t}{4k^2}$.

Profits are higher in corner solution where all information is sold than with partitions belonging to $\mathcal{C}$, and the intermediary sells all information to both firms. ■

Profit comparison

We compare profits when the data intermediary sells information to both firms and to Firm 1 only in the monopoly market, and we prove that the data intermediary sells information to Firm 1 only in equilibrium.

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

The profits of the data intermediary when it sells to one firm are:[41]

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

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

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

2) Profits when the intermediary sells information to both firms.

The profits of the data intermediary when both firms are informed are:

$$\Pi_2^{all} = \frac{t}{4} - \frac{3t}{2k} - \frac{t}{4k^2}.$$

3) DI's selling strategy in equilibrium.

We compare the profits of the intermediary when it sells information to one firm or to both firms. The difference between the profits is: $\Pi_1^* - \Pi_2^* \geq \frac{15t}{56k^2} > 0$. ■

B.4 Proof of Proposition 7

We show that the incentives for an intermediary to collect segments are greater in its monopoly market than in the competitive market. This result is straightfor-

[41] The expression of $\pi_\theta(\emptyset, \mathcal{P}^k)$ is provided in Liu and Serfes (2004).

ward for intermediaries that do not sell information and make zero profits in the competitive market, and we focus our proof on intermediary $\overline{DI}$.

We first write the price of information in the competitive market. We substitute the values of $\overline{j_1}^*$, $\overline{j_2}^*$ and $\underline{j_1}^*$ in $\pi_1(\overline{j_1}, \overline{j_2}) - \pi_1(\underline{j_1}, \overline{j_2})$ in the profit functions of Firm 1 and Firm 2. The price of information increases in $\overline{k}$:

$$\begin{aligned} p_l(\overline{k}, \underline{k}) &= [\pi_1(\overline{j_1}, \overline{j_2}) - \pi_1(\underline{j_1}, \overline{j_2})] \\ &= \frac{((12\underline{k} - 11)\overline{k}^2 + (4\underline{k} - 12\underline{k}^2)\overline{k} + 7\underline{k}^2)t}{36\underline{k}^2\overline{k}^2} \end{aligned} \tag{14}$$

When selling information to Firm 1 on the monopoly market, a monopolist data intermediary DI_i has revenue: $p_m(k_i) = \frac{t}{7} - \frac{3t}{7k_i} + \frac{9t}{28k_i^2}$.

with marginal revenue equal to: $\frac{\partial p_m(k_i)}{\partial k_i} = \frac{3t}{7k_i^2} - \frac{9t}{14k_i^3}$.

We prove that the optimal number of segments collected is larger for $p_m(k_i) = \frac{t}{7} - \frac{3t}{7k_i} + \frac{9t}{28k_i^2}$ than in the competitive market where the revenue of $\overline{DI}$ is $2p_l(\overline{k})$.

For $\overline{k} > \underline{k}$, we have

$$p_l(\overline{k}) = \frac{((12\underline{k} - 11)\overline{k}^2 + (4\underline{k} - 12\underline{k}^2)\overline{k} + 7\underline{k}^2)t}{36\overline{k}^2\underline{k}^2}$$

$$2\frac{\partial p_l(\overline{k})}{\partial \overline{k}} = \frac{((6\overline{k} - 2)\underline{k} - 7\overline{k})t}{9\overline{k}^3\underline{k}}$$

$$p_m(k_i) = \frac{t}{7} - \frac{3t}{7k_i} + \frac{9t}{28k_i^2}.$$

$$\frac{\partial p_m(\overline{k})}{\partial \overline{k}} = \frac{3t}{7\overline{k}^2} - \frac{9t}{14\overline{k}^3} \geq \frac{\partial p_l(\overline{k})}{\partial \overline{k}} = \frac{((6\overline{k} - 2)\underline{k} - 7\overline{k})t}{9\overline{k}^3\underline{k}}$$

Consider $\overline{k}^*$ such that $2\frac{\partial p_l(\overline{k})}{\partial \overline{k}}|_{\overline{k}=\overline{k}^*} = \frac{((6\overline{k}^* - 2)\underline{k} - 7\overline{k}^*)t}{9\overline{k}^{*3}\underline{k}} = \frac{\partial c(\overline{k})}{\partial \overline{k}}|_{\overline{k}=\overline{k}^*}$. Since $\frac{\partial^2 p_m(\overline{k})}{\partial \overline{k}^2} = \frac{27t}{14\overline{k}^4} - \frac{6t}{7\overline{k}^3} \leq 0 \;\; \forall \, \overline{k} \in [2, \infty[$, revenues are concave, and necessarily we have for k: $\frac{\partial p_m(k)}{\partial k} = \frac{\partial c(k)}{\partial k} \iff k \leq \overline{k}^*$. The incentives to collect consumer data are higher in monopoly markets than in the competitive market l. ■

B.5 Proof of Proposition 8

We first state that the number of consumer segments k_i^* collected by the $n-1$ intermediaries which do not sell information in market l increases with m_i. $k_i^* = argmax\{m_i p_{m_i}(k_i) - (m_i + l)c(k_i)\} \implies p'_{m_i}(k_i^*) = (1 + \frac{l}{m_i})c'(k_i^*)$. An increase in

m_i reduces the value of the right-hand-side term and, under the concavity of p_{m_i} and the convexity of $c(.)$, increases the equilibrium value of k_i^*.

B.5.1 Condition $\mathcal{C}$: degree of convexity of prices

$$\exists\ \underline{c_1'}, \underline{c_1} \in \mathbb{R}_+\ s.t.:$$

$$\begin{aligned} \mathcal{C}_1 :\ & c'(\tilde{k}_1^*) > \underline{c_1'} \text{ or } c'(\tilde{k}_1^*) < \underline{c_1'}\ \&\ c(\tilde{k}_1^*) - c(k_1^*) > \underline{c_1}. \\ \mathcal{C}_2 :\ & c'(\tilde{k}_1^*) < \underline{c_1'}\ \&\ c(\tilde{k}_1^*) - c(k_1^*) < \underline{c_1} \\ & \&\ m_2 < \underline{m_2} \text{ for a given } 0 \leq \underline{m_2} < m_1. \end{aligned} \tag{15}$$

B.5.2 Existence of equilibria

Suppose that intermediary i sells in the competitive market l, and collects $\tilde{k}_i^* = \text{argmax}\{m_i p_{m_i}(k_i) + 2l p_l(k_i, \underline{k}_j^*) - (m_i + l)c(k_i)\}$ segments. Following the discussion at the beginning of this section, if $i = 1$, the second highest number of segments collected ($\underline{k}_j^*$ in the above equation) is equal to k_2^*, and if $i > 1$ it is equal to k_1^*.

There are two types of equilibria in this game.

- DI_1 sells in market l. Collecting $\tilde{k}_1^* > k_j^*,\ j = 2, .., n$ is an equilibrium for DI_1 if deviation is not profitable for DI_2, that is if

$$m_2 p_m(k_2^*) - (m_2 + l)c(k_2^*) > m_2 p_m(\tilde{k}_2^*) + 2l p_l(\tilde{k}_2^*, \tilde{k}_1^*) - (m_2 + l)c(\tilde{k}_2^*).$$

- DI_i $i > 1$ sells in market l. Collecting $\tilde{k}_i^* > k_1^*$ is an equilibrium for intermediary DI_i if deviation is not profitable for DI_1:

$$m_1 p_m(k_1^*) - (m_1 + l)c(k_1^*) > m_1 p_m(\tilde{k}_1^*) + 2l p_l(\tilde{k}_1^*, \tilde{k}_i^*) - (m_1 + l)c(\tilde{k}_1^*).$$

The intermediary with the second highest number of segments collected has the highest incentives to deviate, hence if the above conditions hold for respectively DI_2 and DI_1, they hold for the other intermediaries.

The existence of an equilibrium depends on two structural factors that impact the decisions of intermediaries to deviate: the size of the monopoly markets of the two intermediaries collecting the highest numbers of segments m_1 and m_2, and the degree of convexity of the cost to collect data in the neighborhood of the highest number of segments collected $\tilde{k}_i^*$. The conditions for an equilibrium to exist are stated below:

- Consider $\tilde{k}_1^* > k_j^*,\ j = 1, .., n$, there exists $\underline{c_1'}, \underline{c_1} \in \mathbb{R}_+$ such that:
  - $c'(\tilde{k}_1^*) > \underline{c_1'}$ or $c'(\tilde{k}_1^*) < \underline{c_1'}$ & $c(\tilde{k}_1^*) - c(k_1^*) > \underline{c_1} \implies \forall\, m_2, \tilde{k}_1^*$ is an equilibrium.
  - $c'(\tilde{k}_1^*) < \underline{c_1'}$ & $c(\tilde{k}_1^*) - c(k_1^*) < \underline{c_1} \implies \exists!\ m_1 > \underline{m_2} \geq 0\ s.t.\ \underline{m_2} > m_2 \iff \tilde{k}_1^*$ is an equilibrium.

  In the equilibrium where DI_1 collects the highest number of segments, the amount of data collected by other intermediaries is positively related to the size of their monopoly markets: $\tilde{k}_1^* > k_2^* \geq ... \geq k_n^*$.

- Consider $\tilde{k}_i^* > k_j^*,\ i > 1,\ j = 1, .., n,\ j \neq i$, there exists $\underline{c_i'}, \underline{c_i} \in \mathbb{R}_+$ such that:
  - $c'(\tilde{k}_i^*) > \underline{c_i'} \implies \forall\, m_1, \tilde{k}_i^*$ is an equilibrium.
  - $c'(\tilde{k}_i^*) < \underline{c_i'}$ & $\exists\ 0 \leq k \leq \tilde{k}_i^*\ s.t.\ c(\tilde{k}_i^*) - c(k) > \underline{c_i} \implies \exists!\ \underline{m_i} > m_i\ s.t.\ \underline{m_i} > m_1 \iff \tilde{k}_i^*$ is an equilibrium.

  In the equilibrium where intermediary DI_i collects the highest number of segments, the amount of data collected by the other intermediaries is positively related to the size of their monopoly markets: $\tilde{k}_i^* > k_1^* > ... > k_n^*$.

- When data intermediaries have identical market sizes $m_1 = m_2 = ... = m_n$, an equilibrium has the following property. One data intermediary (1, w.l.o.g.) collects strictly more information than the others who all collect the same number of segments with: $\tilde{k}_1^* > k_2^* = ... = k_n^*$.

We prove these three points in the following sections.

B.5.3 DI_1 collects the highest number of segments in equilibrium

The incentives to deviate of DI_2 are the strongest when $m_2 = m_1$. Consider the limit case where $m_2 = m_1$. We want to show that there exists an equilibrium where $k_2^* < \tilde{k}_1^*$. We know that DI_2 has interest to collect more data than DI_1 if and only if: $m_2 p_m(k_2^*) - (m_2+l)c(k_2^*) < m_2 p_m(\tilde{k}_2^*) - (m_2+l)c(\tilde{k}_2^*) + 2lp_l(\tilde{k}_2^*, \tilde{k}_1^*)$. Consider the first degree derivative of the profits of DI_2 at $k_2 = \tilde{k}_1^*$:

$$\frac{\partial\left(m_2 p_m(k_2) - (m_2+l)c(k_2) + 2lp_l(k_2, \tilde{k}_1^*)\right)}{\partial k_2}|_{k_2=\tilde{k}_1^*}.$$

On the one hand, $\frac{\partial(m_2 p_m(k_2)-(m_2+l)c(k_2))}{\partial k_2}|_{k_2=\tilde{k}_1^*}$ is necessarily negative as $k_2^* < \tilde{k}_1^*$, and its value depends on the degree of convexity of $c(.)$. In particular, by continuity there exists $c(.)$ such that $\frac{\partial(m_2 p_m(k_2)-(m_2+l)c(k_2))}{\partial k_2}|_{k_2=\tilde{k}_1^*}$ tends to zero. On the other hand, $\frac{\partial\left(2lp_l(k_2,\tilde{k}_1^*)\right)}{\partial k_2}|_{k_2=\tilde{k}_1^*}$ is strictly positive.

Hence, there exists a threshold value $\underline{c_1'}$ such that the profits of DI_2 always decrease at $\tilde{k}_1^*$ if $c'(\tilde{k}_1^*) > \underline{c_1'}$. In this case deviation is not profitable and the strategies proposed in this section constitute an equilibrium.

Suppose now that the profits of DI_2 increase at $\tilde{k}_1^*$. Deviation by DI_2 is profitable only if $m_2 p_m(k_2^*) - (m_2+l)c(k_2^*) < m_2 p_m(\tilde{k}_2^*) - (m_2+l)c(\tilde{k}_2^*) + 2lp_l(\tilde{k}_2^*, \tilde{k}_1^*)$.

If $c(\tilde{k}_1^*) - c(k_1^*)$ is greater than a threshold $\underline{c_1}$, this inequality never holds and the strategies described in this section constitute an equilibrium.

If $c(\tilde{k}_1^*) - c(k_1^*) < \underline{c_1}$, the above inequality holds and deviation is profitable for $m_2 = m_1$. We now show that an equilibrium can exist if m_2 is small compared to m_1. We know that the incentives of DI_2 to deviate decrease as m_2 decreases. Hence, there are two cases to consider.

First, if when $m_2 = 0$, $2lp_l(\tilde{k}_2^*, \tilde{k}_1^*) - lc(\tilde{k}_2^*) > 0$, then deviation is always profitable and no equilibrium exists to the game.

Secondly, if DI_2 has no incentive to deviate when $m_2 = 0$, then by continuity of the profit function of DI_2 w.r.t. m_2, there exists a cut-off point $\underline{m_2}$ such that deviation is profitable for $m_2 > \underline{m_2}$ and is not profitable otherwise. Therefore, for values of m_2 such that $0 < m_2 < \underline{m_2}$, these strategies constitute an equilibrium.

This establishes the proof of existence of an equilibrium focusing on intermediaries 1 and 2. As DI_2 is the one with the greatest incentives to deviate (the loss from deviating in the monopoly market is the lowest), these results imply that smaller intermediaries do not have interest to deviate either.

B.5.4 DI_i $i > 1$ collects the highest number of segments in equilibrium

We use the same reasoning as in the previous section to show that there can exist an equilibrium in which a smaller intermediary collects more data than DI_1.

We analyze the incentives to DI_1 to deviate from this strategy. There are two cases to consider. On the one hand, if the first degree derivative of the cost function at $c(\tilde{k}_2^*)$ is greater than a threshold $\underline{c_i'}$, deviation is never profitable.

On the other hand, if the first degree derivative of the cost function at $\tilde{k}_2^*$ is lower than $\underline{c_i'}$, the profits of DI_1 increase at $\tilde{k}_2^*$ and a deviation is profitable under the condition that there exists k for which the loss of profits in the monopoly market resulting from the deviation is more than covered by the positive gains achieved in the competitive market: $c(\tilde{k}_2^*) - c(k) < \underline{c_i}$. If there exists a value of $\underline{c_i}$ such that this equality holds, then $\tilde{k}_2^*, k_1^*$ is an equilibrium.

B.5.5 Equilibrium with symmetric intermediaries

Consider data intermediaries that are symmetric in terms of market size $m_1 = m_2$ (the reasoning generalizes easily to any number of intermediaries). In this case there is no symmetric pure strategy Nash equilibrium in the amount of data collected by intermediaries. Indeed, consider a symmetric equilibrium under which both intermediaries collect the same amount of consumer data k^*. Necessarily we have $mp_m'(k^*) - (m+l)c'(k^*) = 0$. However, a data intermediary has interest to deviate from this situation by increasing the amount of consumer data collected and make $mp_m(\tilde{k}^*) + 2lp_l(\tilde{k}^*, k^*) - (m+l)c(\tilde{k}^*)$, given that the derivative is positive at $k^* + \epsilon$ for an arbitrary small ϵ.

However, there can exist asymmetric equilibria in pure strategies. In such equilibrium, one intermediary collects $\tilde{k}^*$ that maximizes $mp_m(\tilde{k}^*) + 2lp_l(\tilde{k}^*, k^*) - (m+l)c(\tilde{k}^*)$ and the other collects k^* that maximizes profits on their monopoly markets $mp_m(k^*) - (m+l)c(k^*)$, with $\tilde{k}^* > k^*$.

To prove that these strategies can sustain an equilibrium, we analyze the incentives of two intermediaries to deviate.

Consider first the intermediary with the highest precision. This intermediary has no interest to deviate since its profits are maximized at $\tilde{k}^*$. The other intermediary has interest to deviate and to collect $\hat{k}^* > \tilde{k}^*$, that is, to collect more data than the intermediary with the highest precision, if the profits of doing so are greater than the monopoly profits: $mp_m(\hat{k}^*) + 2lp_l(\hat{k}^*, \tilde{k}^*) - (m+l)c(\hat{k}^*) \geq mp_m(k^*) - (m+l)c(k^*)$.

If this equation is satisfied, deviation to $\hat{k}^*$ is profitable for intermediary j, and intermediary DI_i has interest to collect k^* data. In this case, $\hat{k}^*$ is not an equilibrium, and no equilibrium exists.

There exists an asymmetric equilibrium if this equation is not satisfied. In this case, there exists an equilibrium to the game in which one data intermediary collects k^*, and the other collects $\tilde{k}^*$.

B.6 Proof of Lemma 3

We show that consumer surplus always increases with the number of consumer segments sold. Consumer surplus when Firm 1 has j_1 consumer segments and

Firm 2 has j_2 consumer segments is defined as follows:

$$
\begin{aligned}
CS(j_1, j_2, k) = & \sum_{i=1}^{j_1}[\int_0^{\frac{1}{k}} V - 2t[1 - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k} - \frac{i}{k}] - tx\mathrm{d}x] \\
& + \int_{\frac{j_1}{k}}^{\frac{1}{2}+\frac{j_1}{3k}-\frac{j_2}{3k}} V - t[1 - \frac{4}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] - tx\mathrm{d}x + \int_{\frac{1}{2}+\frac{j_1}{3k}-\frac{j_2}{3k}}^{1-\frac{j_2}{k}} V - t[1 - \frac{2}{3}\frac{j_1}{k} - \frac{4}{3}\frac{j_2}{k}] - tx\mathrm{d}x \\
& + \sum_{i=1}^{j_2}[\int_0^{\frac{1}{k}} V - 2t[1 - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k} - \frac{i}{k}] - tx\mathrm{d}x] \\
& = \sum_{i=1}^{j_1} \frac{1}{k}(V - 2t[1 - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k} - \frac{i}{k}]) - \frac{j_1 t}{2k^2} \\
& + \sum_{i=1}^{j_2} \frac{1}{k}(V - 2t[1 - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k} - \frac{i}{k}]) - \frac{j_2 t}{2k^2} \\
& + V[1 - \frac{j_2}{k} - \frac{j_1}{k}] - [\frac{1}{2} - \frac{2j_1}{3k} - \frac{j_2}{3k}]t[1 - \frac{4}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] \\
& - [\frac{1}{2} - \frac{2j_2}{3k} - \frac{j_1}{3k}]t[1 - \frac{4}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k}] - t[\frac{1}{4} - \frac{1}{9}\frac{j_1 j_2}{k^2} - \frac{7}{18}\frac{j_2^2}{k^2} - \frac{7}{18}\frac{j_1^2}{k^2}] \\
& = \frac{j_1}{k}[V - 2t[1 - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] + \frac{j_1(j_1+1)t}{k^2} - \frac{j_1 t}{2k^2} \\
& + \frac{j_2}{k}[V - 2t[1 - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k}] + \frac{j_2(j_2+1)t}{k^2} - \frac{j_2 t}{2k^2} \\
& + V[1 - \frac{j_2}{k} - \frac{j_1}{k}] + t[-\frac{5}{4} + \frac{1}{3}\frac{j_1}{k} + \frac{1}{3}\frac{j_2}{k} + \frac{5}{6}\frac{j_1^2}{k^2} + \frac{5}{6}\frac{j_2^2}{k^2} - 2\frac{j_1 j_2}{k^2}] \\
& = V + t[-\frac{5}{4} + \frac{17}{18}\frac{j_1^2}{k^2} + \frac{17}{18}\frac{j_2^2}{k^2} + \frac{j_1 j_2}{k^2}] + \frac{1}{2}\frac{j_1 t}{k^2} + \frac{1}{2}\frac{j_2 t}{k^2}
\end{aligned}
\tag{16}
$$

The first degree derivative with respect to j_1 is $\frac{\partial CS}{\partial j_1} = \frac{17j_1}{9k} + \frac{j_2}{k} + \frac{1}{2k}$, which is larger than zero for $\frac{j_1}{k} \geq -\frac{18j_2+9}{34k}$ (i.e. always above zero). ■

B.7 Proof of Lemma 4

The first degree derivative of CS with respect to k, for given $\frac{j_1}{k}$, $\frac{j_2}{k}$ is equal to $\frac{\partial CS}{\partial k} = -\frac{j_1 t}{k^3} - \frac{j_2 t}{k^3}$, which is clearly always negative, and consumer surplus always decreases with information precision. ■

Working Papers of the Center of Economic Research at ETH Zurich

(PDF-files of the Working Papers can be downloaded at www.cer.ethz.ch/research/working-papers.html).

23/383 D. Bounie, A. Dubus, P. Waelbroeck
Competition Between Strategic Data Intermediaries with Implications for Merger Policy

23/382 J. Lehtomaa, C. Renoir
The Economic Impact of Tropical Cyclones: Case Studies in General Equilibrium

23/381 S. Srinivasan
Social Policies and Adaptation to Extreme Weather: Evidence from South Africa

23/380 L. Barrage
Fiscal Costs of Climate Change in the United States

23/379 S. Bansal, M. Filippini, S. Srinivasan
How Regulation Might Fail to Reduce Energy Consumption While Still Stimulating Total Factor Productivity Growth

22/378 A. Jo, C. Karydas
Firm Heterogeneity, Industry Dynamics and Climate Policy

22/377 V. da Cruz
Cap-and-Innovate: Evidence of regulation-induced innovation in California

22/376 G. Loumeau
Land Consolidation Reforms: A Natural Experiment on the Economic and Political Effects of Agricultural Mechanization

22/375 L. Bretschger
Green Road is Open: Economic Pathway with a Carbon Price Escalator

22/374 A. Goussebaïle
Democratic Climate Policies with Overlapping Generations

22/373 H. Gersbach, O. Tejada, J. Wagner
Policy Reforms and the Amount of Checks & Balances

22/372 S. Houde, W. Wang
The Incidence of the U.S.-China Solar Trade War

22/371 J. A. Bingler
Expect the worst, hope for the best: The valuation of climate risks and opportunities in sovereign bonds

22/370 A. Bommier, A. Fabre, A. GoussebaÃ¯le, and D. Heyen
Disagreement Aversion

22/369 A. Jo, A. Miftakhova
How Constant is Constant Elasticity of Substitution? Endogenous Substitution between Clean and Dirty Energy

22/368 N. Boogen, M. Filippini, A. L. Martinez-Cruz
Value of co-benefits from energy saving ventilation systems–Contingent valuations on Swiss home owners

22/367 D. Bounie, A. Dubus, P. Waelbroeck
Market for Information and Selling Mechanisms

22/366 N. Kumar, N. Kumar Raut, S. Srinivasan
Herd behavior in the choice of motorcycles: Evidence from Nepal

21/365 E. Komarov
Capital Flows and Endogenous Growth

21/364 L. Bretschger, A. Jo
Complementarity between labor and energy: A firm-level analysis

21/363 J. A. Bingler, C. Colesanti Senni, P. Monnin
Climate Transition Risk Metrics: Understanding Convergence and Divergence across Firms and Providers

21/362 S. Rausch, H. Yonezawa
Green Technology Policies versus Carbon Pricing: An Intergenerational Perspective

21/361 F. Landis, G. Fredriksson, S. Rausch
Between- and Within-Country Distributional Impacts from Harmonizing Carbon Prices in the EU

21/360 O. Kalsbach, S. Rausch
Pricing Carbon in a Multi-Sector Economy with Social Discounting

21/359 S. Houde, T. Wekhof
The Narrative of the Energy Efficiency Gap

21/358 F. Böser, H. Gersbach
Leverage Constraints and Bank Monitoring: Bank Regulation versus Monetary Policy

21/357 F. Böser
Monetary Policy under Subjective Beliefs of Banks: Optimal Central Bank Collateral Requirements

21/356 D. Cerruti, M. Filippini
Speed limits and vehicle accidents in built-up areas: The impact of 30 km/h zones

21/355 A. Miftakhova, C. Renoir
Economic Growth and Equity in Anticipation of Climate Policy

21/354 F. Böser, C. Colesanti Senni
CAROs: Climate Risk-Adjusted Refinancing Operations

21/353 M. Filippini, N. Kumar, S. Srinivasan
Behavioral Anomalies and Fuel Efficiency: Evidence from Motorcycles in Nepal

21/352 V. Angst, C. Colesanti Senni, M. Maibach, M. Peter, N. Reidt, R. van Nieuwkoop
Economic impacts of decarbonizing the Swiss passenger transport sector

21/351 N. Reidt
Climate Policies and Labor Markets in Developing Countries

21/350 V. Britz, H. Gersbach
Pendular Voting

21/349 E. Grieg
Public opinion and special interests in American environmental politics

21/348 N. Ritter, J. A. Bingler
Do homo sapiens know their prices? Insights on dysfunctional price mechanisms from a large field experiment

20/347 C. Daminato, M. Filippini, F. Haufler
Personalized Digital Information and Tax-favoured Retirement Savings: Quasi-experimental Evidence from Administrative Data

20/346 V. Britz, H. Gersbach
Open Rule Legislative Bargaining

20/345 A. Brausmann, E. Grieg
Resource Discoveries and the Political Survival of Dictators

20/344 A. Jo
The Elasticity of Substitution between Clean and Dirty Energy with Technological Bias

20/343 I. van den Bijgaart, D. Cerruti
The effect of information on market activity; evidence from vehicle recalls