User data and endogenous entry in online markets

THE JOURNAL OF INDUSTRIAL ECONOMICS 0022-1821
Volume 0 January 2024 No. 0

USER DATA AND ENDOGENOUS ENTRY IN ONLINE MARKETS*

LAURA ABRARDI
CARLO CAMBINI
RAFFAELE CONGIU
FLAVIO PINO

We investigate how the presence of a Data Broker (DB), who sells consumer data to downstream firms, affects firm entry and competition in a horizontally differentiated oligopoly market, in which data allow firms to price discriminate. The DB chooses the price and amount of data sold to each firm. We show that the data sale by the DB reduces excessive market entry, as the competition induced by personalized prices exerts a downward pressure on prices and profits. The data-induced entry barrier and resulting weakened competition dominates the pro-competitive effect of personalized prices. Consequently, while the DB's presence might alleviate excessive market entry, it also diminishes consumer surplus.

I. INTRODUCTION

DATA BROKERS (DBs) TRACK CONSUMERS ONLINE, hoard massive amounts of information and sell that intelligence in the form of targeted market segments based on the customer's needs. Though consumers can benefit from firms' targeted commercial offers, DBs might also have the power to affect market entry and steer competition simply by choosing to which firms (and to what extent) data are sold. This paper analyzes a market where a DB sells consumer

*We thank Paul Belleflamme, Marc Bourreau, Federico Boffa, Francesco Clavorà Braulin, Joan Calzada, Alberto Cavaliere, Alexandre de Cornière, Antoine Dubus, Tomaso Duso, Marita Freimane, Axel Gautier, Lukasz Grzybowski, Johannes Johnen, Elisabetta Iossa, Alberto Iozzi, Leonardo Madio, Claudio Piga, Andrew Rhodes, Lorien Sabatino, Luca Sandrini, Adi Shany, Shiva Shekhar, Konstantinos Serfes, Robert Somogyi, Eliza Stenzhorn, Guillaume Thébaudin, Tommaso Valletti, Lorenzo Zirulia as well as participants to the Brown Bag Seminar (16th of February 2022, Louvain-la-Neuve), CESIFO 2022 doctoral workshop, 13th Conference on Digital Economics (Paris), NERI 2022 (Genova), SIEPI 2022 (L'Aquila), IIBEO (Alghero), CRESSE 2022 (Heraklion), EARIE 2022 (Vienna) JEI 2022 (Gran Canaria) conferences, for useful comments on previous versions of this manuscript. This work has been partially supported by Ministero dell'Istruzione, dell'Università e della Ricerca Award TESUN-83486178370409 finanziamento dipartimenti di eccellenza CAP. 1694 TIT. 232 ART. 6.

Authors' affiliation: Politecnico di Torino, Department of Management, Corso Duca degli Abruzzi, 24, Turin, 10129, Italy.
e-mail: carlo.cambini@polito.it

data to a number of horizontally differentiated downstream firms, which can use data for price discrimination. We highlight how the DB, by choosing the firms to which data are sold, and the price and quantity of data sold, can affect firm entry, firm profits and consumer surplus.

First-degree price discrimination, once extremely rare, has now become a reality through the use of big data (OECD [2018]). As early as 2000, Amazon delivered a proof of the feasibility of this form of price discrimination when it started to charge its consumers different prices for the same product, based on their purchase histories.[1] Consumers receive personalized prices based on geolocation, income level, browsing history and proximity to rival's stores, among others (Aparicio *et al.* [2023]; Valentino-DeVries *et al.* [2012]).[2]

Collecting and processing data at a scale that makes it valuable for personalized pricing requires unique resources and capabilities. The demand for such abilities has determined the growth of the DB sector, a highly concentrated industry whose revenue is estimated at USD 200 billion (Crain [2018]; FTC [2014]). DBs' business model compounds both online and offline sources, collecting data from commercial, governmental, and other publicly available sources—for example, blogs, social media. As they typically do not get their data directly from consumers, DBs are often away from the media's spotlight or people's awareness. Indeed, consumers are often unfamiliar with DBs like BlueKai, Experian or Teradata. Yet, DBs are building intricate profiles with thousands of records on almost every household (FTC [2014]). Working in the background, DBs mostly engage in business-to-business relations, selling the processed information to downstream firms who want to reach specific consumers with targeted offers.

Given the huge potential to influence downstream competition, policymakers have often expressed concerns regarding the reach and the lack of transparency of this highly concentrated, and yet virtually unregulated industry. Recent literature (see, e.g., Montes *et al.* [2019]) has pointed out how DBs have the incentive to increase some firms' market power by selectively selling data in downstream duopolistic markets. However, little is known on the strategies used by DBs when they serve markets populated by more than two competing firms, and how these strategies influence market entry and, in turn, competition, firms' profits and consumer surplus.

This paper aims to understand how a DB can influence firm entry and downstream competition in oligopolistic markets by deciding to whom and

[1] "On the Web, Price Tags Blur—What You Pay Could Depend on Who You Are", Washington Post, Sept. 27, 2000.

[2] Mikians *et al.* [2012] show that individual consumer data such as geolocalization are used by firms to price discriminate them, with price differences of up to 166%. In 2012, the New York Times also found evidence of personalized pricing in supermarket chains, with higher prices being set for loyal consumers (Clifford [2012]). More recently, Aparicio *et al.* [2023] show that the algorithms used by the leading online grocers in the U.S. personalize prices at the delivery zip code level.

how much data to sell. We consider a circular city model à la Salop [1979], where firms can enter the market and then acquire consumer data from a DB through Take-It-Or-Leave-It offers (as in Bergemann and Bonatti [2019]).[3] The DB sets the data price equal to the difference in firms' profits between buying or not buying data. Firms offer a basic price to unidentified consumers and, if they acquire data, location-specific tailored prices to the identified ones. The DB has information on all consumers, and he decides to which firms he sells data—making them informed—how much data he sells to each one (e.g., the full dataset or only a partition of it), and the price of data.

We find an *entry barrier effect* of data, whereby the equilibrium entails a reduction of firm entry, relative to the benchmark case in which data are not available or are exogenously provided to the firms (as in Taylor and Wagman [2014]). Intuitively, the DB charges a price of data such that he extracts most of the firms' profits, thus reducing entry in the downstream market. We also find that the DB affects competition through two additional channels besides the price of data, namely the number of firms to which data are sold, and the quantity of data. In particular, the DB's optimal strategy entails the sale of data to all firms. By doing so, the DB threatens firms should they choose not to buy data, thereby increasing their willingness to pay for them.

Interestingly, we find that, in an oligopolistic setting, differently from a duopolistic setup, the quantity of data sold in equilibrium depends on the degree of horizontal differentiation. When the horizontal differentiation is low (i.e., transportation costs are low, relative to the entry cost), the DB sells a sufficiently large partition (e.g., the whole dataset) to each firm, allowing them to identify all their consumers. Conversely, when horizontal differentiation is sufficiently high, the DB only sells to each firm data about a share of the consumers served in equilibrium. Intuitively, when horizontal differentiation is low, downstream competition is fierce. In this situation, the DB can effectively raise firms' willingness to pay for data by increasing the competitive threat firms face if they have to compete without data. This is achieved by selling the full dataset. Conversely, when horizontal differentiation is high, competition is milder. Therefore, the DB can extract higher profits from firms by further mitigating competition. This can be achieved by selling smaller quantities of data. Because of the entry barrier effect and the resulting higher market concentration, we also find that consumer surplus is always lower than in the benchmark case in which data are not available.

The literature studying the impact of data on competition is growing. Firms could use consumer data to identify naive consumers (Johnen [2020]) or to distinguish between consumer groups with different price sensitivities

[3] The use of direct sales when selling data is well documented in practice. For example, Acxiom, Experian, and Epsilon, among the top five Data Brokers companies, sell their datasets at a posted price on the Amazon Web Services marketplace.

(Colombo [2018]). de Cornière and Taylor [2020] provide a general framework in which data are a revenue-shifter, for a given level of consumers' utility. Although this framework finds a wide range of applications in which data increase the quality of the information, it is ill-suited for price discrimination in spatial competition settings where data provide information on the type of consumers (Armstrong and Vickers [2001]). When data are used for price discrimination and are exogenously available to firms, a pro-competitive effect may arise under competition, provided that the market is fully covered.[4] As informed firms compete more fiercely, consumers benefit from lower prices. However, recent contributions have highlighted conditions under which price discrimination can instead benefit firms. If the market is only partially covered under uniform pricing, personalized prices can raise firms' profits and harm consumers by expanding demand (Rhodes and Zhou [2022]). Nonexclusive contracts are another direction along which price discrimination can increase profits. If consumers can buy from multiple firms in the market, the milder competition allows personalized prices to benefit firms (Lu *et al.* [2022]). Consumers' ability to hide their identity from firms can also help to mitigate the pro-competitive effect of data, and raise firms' profits (Chen *et al.* [2020]). Finally, Jullien *et al.* [2023] show that personalized pricing can raise profits when a firm sells its product to consumers both directly and through a retailer.

Another stream of literature analyzes the effect of perfect price discrimination on entry. The seminal work by Spence [1976] shows that each firm's choice of quantity or product characteristic is socially optimal under perfect discrimination in a competitive market. Bhaskar and To [2004] extend the study by Spence [1976] and show that perfect discrimination leads to excessive entry from a social welfare point of view. The effect of price discrimination on excessive entry is also highlighted by Taylor and Wagman [2014] in a setting in which entry affects consumers' preferences for existing firms à la Salop [1979]. However, entry becomes socially efficient if the existing firms' choice of product characteristics is not affected by the entry of an additional firm, as also highlighted by Rhodes and Zhou [2022].

These studies focus on settings in which data are exogenously available to firms. A more recent strand of literature has endogenized the information acquisition process, either through firms' repeated interactions with consumers (Acquisti and Varian [2005]; Bergemann and Bonatti [2011]; Hagiu and Wright [2020]; Liu and Serfes [2004]; Villas-Boas [2004]) or by acquiring data from strategic actors (Bergemann and Bonatti [2015]; de Cornière [2016]; Choe *et al.* [2023]; Gu *et al.* [2019]). In particular, Braulin and Valletti [2016], Montes *et al.* [2019] and Bounie *et al.* [2021]

[4] See for instance (Bester and Petrakis [1996]; Liu and Serfes [2004, 2005]; Shaffer and Zhang [1995]; Shy and Stenbacka [2016]; Taylor [2003]; Taylor and Wagman [2014]; Thisse and Vives [1988]).

consider a monopolistic DB who sells data to a downstream duopoly through a series of auctions, as in Jehiel and Moldovanu [2000]. These studies highlight how a DB can limit competition between two existing firms by selling data exclusively to one of them, thus extracting higher industry profits at the expense of consumers and firms. However, when three firms are present, Delbono *et al.* [2021] find that the DB always sells data to two or more firms—depending on the selling mechanism—and thus exclusive sales are never part of the equilibrium. A parallel stream of literature studies the role of competition between DBs on data collection. In particular, Ichihashi [2021], by studying a market with many data intermediaries and one downstream firm, shows that the nonrivalrous nature of data can lead to significant concentration in data markets.

Our work is also related to the literature that analyzes the vertical relation between an upstream input monopolist and downstream firms (Cachon and Lariviere [2005]; DeGraba [1990]; Greenhut and Ohta [1976]; Tyagi [1999]). This literature has mostly focused on settings where the monopolist sells a good that is essential for the downstream firms' production. Our analysis builds on this literature by focusing on the sale of a nonessential input, as firms can enter the market even when not purchasing data. Moreover, the data acquisition only influences the firms' efficiency in extracting surplus from consumers.

We contribute to the existing literature in two ways. First, we extend the duopolistic setup to analyze how the number of competing firms in an oligopoly market influences the DB's strategy and the subsequent market outcomes. Second, we endogenize the number of firms present in the market by modeling their entry. We thus highlight a novel effect of data, which we denote the *entry barrier effect*, that emerges as a result of the DB's profit-maximizing strategy. Our analysis shows that the reduction in competition given by the DB's entry barrier effect outweighs the pro-competitive effect of data, so that consumer surplus is ultimately reduced. To our knowledge, this is the first work to highlight the entry barrier effect of the DB's strategy and its potential anti-competitive nature.

A key insight of previous literature on monopolistic DBs is that antitrust authorities should ban exclusive data deals to foster competition and protect consumers when the downstream market is a duopoly. However, our results suggest that, if firms' entry is taken into account, such a measure may be ineffective as the harm to competition stems from the entry barrier raised by a monopolistic DB.

The remainder of the paper is organized as follows. Section II presents the model, and Section III analyzes firms' equilibrium prices. Section IV studies the DB's profits and his optimal strategy and discusses the consequent market outcomes. Section V concludes. All proofs and technical details are contained in the Appendix.

II. THE MODEL

We consider a market in which horizontally differentiated firms sell a product to a mass of consumers. Each firm can observe consumers' preferences only if it purchases customer-specific data from a Data Broker (DB). For example, firms sell their products via e-commerce solutions, and the possibility of identifying the consumer through data acquired from a DB allows the firm to offer personalized prices. The following Section II illustrates the market and the specific role of data in downstream price competition. Section II provides details on the data selling mechanism, the DB's strategy, and the timing of the model.

II(i). *Consumers, Firms and the Data Broker*

We consider a free-entry game of a market represented by a circular city of length 1 (Salop [1979]; Vickrey [1964]), where firms sell competing products to consumers. Firms are indexed by $i \in \{0, 1, 2, \ldots, n-2, n-1\}$, where $n \geq 2$ is the number of symmetric firms that enter the market.[5] We assume that firms enter the market choosing equally spaced locations so that the position of a generic firm i is indexed by $\frac{i}{n}$. Their marginal cost of production is normalized to 0, whereas entry entails a fixed cost F.[6]

Consumers are uniformly distributed on the circumference and normalized to 1, and their locations are indexed by $x \in [0, 1)$ in counter-clockwise order. Each consumer buys at most one unit of the product. A consumer derives a gross utility v from consuming the product, with v sufficiently high to ensure full market coverage, and faces a linear transportation cost $t > 0$.

There is one Data Broker who has a dataset with the location of all consumers in the market. The DB can sell a partition of this dataset to each firm entering the market. We assume that the partition sold to each firm is centered on its location.[7] This assumption draws from evidence suggesting the importance of targeting consumers who have the strongest preference for a firm's product (Iyer *et al.* [2005]), who are located closest to a firm's location.

We denote with $d_i \in [0, 1]$ the partition of data offered to firm i. Thus, if firm i accepts the offer, it identifies consumers on the arch $\left[\frac{i}{n} - \frac{d_i}{2}, \frac{i}{n} + \frac{d_i}{2}\right]$, that

[5] As standard in the literature on markets with entry, we assume sequential entry to avoid coordination problems and ignore integer constraints on n in the baseline model. A similar approach has been recently adopted in Rhodes and Zhou [2022]. In Section IV, we discuss how our results change when the number of firms is an integer.

[6] We can think of F as the cost incurred in the process of digitization (see Anderson and Bedre-Defolie 2023), such as the creation of an online retail shop.

[7] In a duopolistic Hotelling model in which data partitions cannot contain disjoint intervals, Bounie *et al.* [2021] show that the partitions sold in equilibrium must contain the firm location. Although they focus on third-degree price discrimination, they also extend this result under perfect price discrimination.

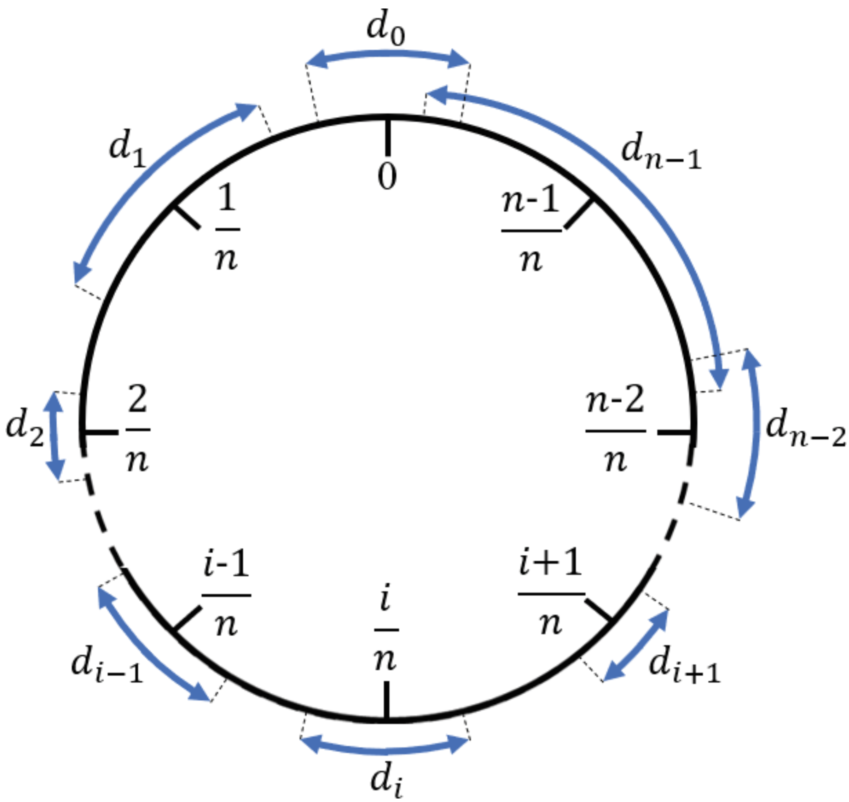

Figure 1
Data Partitions and Identified Consumers

Notes: [Colour figure can be viewed at wileyonlinelibrary.com]

is, it observes their location and performs first-degree price discrimination on them. The partition set containing the partitions offered by the DB to all firms is $\mathbf{P} = \left(d_0, d_1, d_2, \ldots, d_{n-1}\right)$. Figure 1 provides a graphical representation of a partition set, where each firm obtains a partition of (potentially different) size d_i.

A firm i offers location-specific tailored prices $p_i^{\mathrm{T}}(x) \geq 0$ to the consumer x in the identified segment, and a basic price $p_i^{\mathrm{B}} \geq 0$ to unidentified consumers.

A consumer in x buying from firm i maximizes his utility $U(x, i)$, defined as

$$U(x, i) = v - p_i^{\mathrm{T}}(x) - tD(x, i)$$

if firm i has data on consumer x, or

$$U(x, i) = v - p_i^{\mathrm{B}} - tD(x, i)$$

if it does not, where $D(x,i)$ is the shortest arch between the consumer and firm i. The location of an indifferent consumer between firms i and $i+1$ is $\hat{x}_{i,i+1}$, such that $U\left(\hat{x}_{i,i+1}, i\right) = U\left(\hat{x}_{i,i+1}, i+1\right)$.

A firm's profit before paying for data is given by the integral of its prices over its market segment. Let $\pi_i^{\mathrm{W}}(\mathbf{P})$ denote a firm's profits when it buys the partition (i.e., $d_i > 0$), and $\pi_i^{\mathrm{L}}(\mathbf{P})$ denote its profits when it does not (i.e., $d_i = 0$). If firm i buys a partition and identifies only part of its consumers, a portion d_i of its market share will receive tailored prices, while the remaining portion $\hat{x}_{i,i+1} - \hat{x}_{i-1,i} - d_i$ will receive its basic price. Then, its profits are equal to

$$(1) \qquad \pi_i^{\mathrm{W}}(\mathbf{P}) = \int_{\frac{i}{n}-\frac{d_i}{2}}^{\frac{i}{n}+\frac{d_i}{2}} p_i^{\mathrm{T}}(x)dx + p_i^{\mathrm{B}}\left(\hat{x}_{i,i+1} - \hat{x}_{i-1,i} - d_i\right) - F,$$

where the first term on the right-hand side represents firm profits over the identified consumers, and the second term represents its profits over unidentified consumers.

If firm i's partition is large enough that it identifies all consumers it serves, its profits become

$$(2) \qquad \pi_i^{\mathrm{W}}(\mathbf{P}) = \int_{\hat{x}_{i-1,i}}^{\hat{x}_{i,i+1}} p_i^{\mathrm{T}}(x)dx - F,$$

as it serves all consumers through tailored prices. Conversely, if a firm does not obtain data, it only serves unidentified consumers through basic prices, so that its profits are

$$(3) \qquad \pi_i^{\mathrm{L}}(\mathbf{P}) = p_i^{\mathrm{B}}\left(\hat{x}_{i,i+1} - \hat{x}_{i-1,i}\right) - F.$$

II(ii). *The Data Sale and Timing*

The DB sells data through simultaneous Take-It-Or-Leave-It (TIOLI) offers (Bergemann *et al.* [2018]; Bounie *et al.* [2022]). In particular, the DB chooses the partition set $\mathbf{P}$ and the price of each partition w_i, offering them to the downstream firms. Each firm observes $\mathbf{P}$ and simultaneously accepts or refuses to buy its respective partition $d_i \in \mathbf{P}$ at the price w_i.[8]

The DB sets the partitions' prices equal to the firm's willingness to pay for data, given by the difference between its profits when buying or not buying its partition:

$$(4) \qquad w_i = \pi_i^{\mathrm{W}}(\mathbf{P}) - \pi_i^{\mathrm{L}}(\mathbf{P}).$$

[8] By assuming public DB's offers, we rule out situations like secret contracting games as in Hart and Tirole [1988]. See also (Bounie *et al.* [2021]; Montes *et al.* [2019]; Braulin and Valletti [2016]).

Note that the DB is allowed to sell partitions of size zero, implying that he can exclude some firms from the data sale.

The DB's profits can be expressed as the sum of firms' willingness to pay for data:

$$\pi_{\mathrm{DB}}(\mathbf{P}) = \sum_{i=0}^{n-1} w_i. \tag{5}$$

The timing of the model is as follows:[9]

Stage 1. Firms enter the market and pay the fixed cost F.
Stage 2. The DB chooses a partition set $\mathbf{P}$ and the vector of partition prices $\mathbf{w} = (w_0, w_1, \dots, w_{n-1})$. $\mathbf{P}$ is common knowledge.
Stage 3. Firms that entered the market individually and simultaneously accept or refuse the DB's offers.
Stage 4. Firms simultaneously set basic prices p_i^{B} for the unidentified consumers.
Stage 5. Firms observe all basic prices and set tailored prices $p_i^{\mathrm{T}}(x)$ for the identified consumers. Consumers purchase the product and profits are achieved.

As a benchmark, we refer to the Salop [1979] model with marginal costs normalized to 0. In this setting, each firm sets a price $\widetilde{p}_i = \frac{t}{n}$ and obtains a market share of $\frac{1}{n}$, resulting in profits $\widetilde{\pi}_i = \frac{t}{n^2} - F$. The number of entering firms is $\widetilde{n} = \sqrt{\frac{t}{F}}$, resulting in firms' prices $\widetilde{p}_i = \sqrt{tF}$ and profits $\widetilde{\pi}_i = 0$.

III. EQUILIBRIUM PRICES

We rely on the equilibrium concept of Perfect Nash Equilibrium (PNE) and proceed by backward induction, starting from the firms' pricing stage. We distinguish between the two subgames, depending on whether a generic firm i either accepts or rejects the DB's offer in Stage 3.

III(i). *Firm i Accepts the DB's Offer*

When firm i accepts the DB's offer of the data partition d_i, its pricing strategy depends on the size of the data partitions offered to its direct competitors $i-1$ and $i+1$ in the earlier Stage 2. In particular, the data partitions d_i and d_{i-1}, or analogously d_i and d_{i+1}, might overlap or not.

[9] Stage 5 follows Stage 4 to ensure the existence of an equilibrium in pure strategies and is supported by managerial practices (Fudenberg and Villas-Boas [2006]). See also Montes *et al.* [2019] and Bounie *et al.* [2021] for an analogous approach.

Nonoverlapping partitions. If $d_i + d_k < \frac{2}{n}$, where $k \in \{i-1, i+1\}$ denotes either one of i's direct competitors, the data partitions offered to the two adjoining firms do not overlap, implying that the farthest consumer of firm i is served through a basic price. Thus, basic prices define the total demand, and the indifferent consumers between firms $i-1$ and i, and between i and $i+1$, are, respectively:[10]

$$\hat{x}_{i-1,i} = \frac{2i-1}{2n} + \frac{p_i^{\mathrm{B}} - p_{i-1}^{\mathrm{B}}}{2t} \qquad \text{and} \qquad \hat{x}_{i,i+1} = \frac{2i+1}{2n} + \frac{p_{i+1}^{\mathrm{B}} - p_i^{\mathrm{B}}}{2t}. \tag{6}$$

Nonoverlapping partitions imply that each consumer can be identified by at most one firm. Then, if a consumer located in x is identified by firm i through the data partition d_i, such a consumer is offered the tailored price from firm i, and the basic price from the direct competitor. In this case, firm i observes the competitors' basic prices and offers to that consumer a tailored price $p_i^{\mathrm{T}}(x)$ that matches the consumer's utility when buying from the competitor at the basic price:

$$p_i^{\mathrm{T}}(x) = \begin{cases} p_{i-1}^{\mathrm{B}} + 2tx - \frac{t}{n}(2i-1) & \text{for } x \in \left[\frac{i}{n} - \frac{d_i}{2}, \frac{i}{n}\right], \\ p_{i+1}^{\mathrm{B}} - 2tx + \frac{t}{n}(2i+1) & \text{for } x \in \left[\frac{i}{n}, \frac{i}{n} + \frac{d_i}{2}\right]. \end{cases} \tag{7}$$

Note that tailored prices decrease as the direct rival's basic price decreases due to downstream competition.

The profits of firm i when it buys data, before paying for them, are given by (1). Using (6) and (7), they can be expressed as

$$\begin{aligned} \pi_i^{\mathrm{W}}(\mathbf{P}) = {} & \frac{d_i}{2n}\left(2t + np_{i-1}^{\mathrm{B}} + np_{i+1}^{\mathrm{B}} - ntd_i\right) \\ & + p_i^{\mathrm{B}}\left(\frac{n\left(p_{i+1}^{\mathrm{B}} + p_{i-1}^{\mathrm{B}} - 2p_i^{\mathrm{B}}\right) + 2t}{2nt} - d_i\right) - F, \end{aligned} \tag{8}$$

where the first component represents the profit on the identified segment and the second component represents the profit on the unidentified segment. The first-order condition of (8) with respect to $p_i^{\mathrm{B}}(\mathbf{P})$ is:

$$p_i^{\mathrm{B}} = \frac{t}{2n} - \frac{td_i}{2} + \frac{p_{i+1}^{\mathrm{B}} + p_{i-1}^{\mathrm{B}}}{4}, \tag{9}$$

which differs from the reaction function of the standard Salop (1979) model for the negative term $-\frac{td_i}{2}$.

[10] All basic prices are a function of $\mathbf{P}$. Whenever possible, we omit the argument to streamline the exposition.

By jointly considering equations (8) and (9), we find that data have two opposite effects on firms' profits. On the one hand, more data increase the share of identified consumers, who are charged with a tailored price that exactly matches their willingness to pay for the product. This is the *surplus extraction effect* of data (Thisse and Vives [1988]), which increases firm profits through the first term of equation (8). On the other hand, as firm i acquires more data, its unidentified consumers are on average farther from its location, requiring the firm to lower its basic price (equation (9)). A lower basic price reduces firm profits by the second term in equation (8), which constitutes the *competition effect* of data (Thisse and Vives [1988]).

The system of reaction functions in equation (9) for all firms allows us to obtain the subgame equilibrium basic prices and profits, the properties of which are illustrated in the following lemma.

Lemma 1. In the subgame where all firms buy their respective partition, if partitions do not overlap, we have that:

(i) $\frac{\partial p_i^{B*}}{\partial d_j} \leq 0$, $\forall i, j$ (firm i's subgame equilibrium basic price is decreasing in d_j),

(ii) $\frac{\partial \pi_i^{W*}(\mathbf{P})}{\partial d_j} \leq 0$, $\forall j \neq i$ (firm i's subgame equilibrium profit is decreasing in d_j),

(iii) There exists a threshold $\overline{d}_i$ such that $\frac{\partial \pi_i^{W*}(\mathbf{P})}{\partial d_i} > 0$ if $d_i < \overline{d}_i$, and $\frac{\partial \pi_i^{W*}(\mathbf{P})}{\partial d_i} \leq 0$ otherwise, $\forall i$.

Proof. See the Appendix. ■

As firm i acquires more data, it offers its basic price to consumers who are on average farther from its location and consequently lowers its basic price, leading also other firms to lower their basic prices as a strategic reaction (point (i) of Lemma 1). Moreover (point (ii) of Lemma 1), other firms' partitions always lower firm i's profits, as they drive firms to price more aggressively. Instead (point (iii) of Lemma 1), the effect of d_i on firm i's profits is ambiguous. On the one hand, a bigger partition allows firm i to identify more consumers, increasing the *surplus extraction effect*. On the other hand, a bigger partition also entails fiercer competition, which erodes firm i's profits through the *competition effect* of data. Whether the former or the latter of these two effects dominates depends on the size of d_i. When $d_i < \overline{d}_i$, the partition increases firm i's profits. The reason is that a small partition allows firm i to identify the most valuable consumers (i.e., those near the firm's location). However, as d_i increases, the marginal gain of identifying consumers farther from the firm's location decreases, as the firm can extract less surplus from them. Instead, the profit erosion caused by the *competition effect* of data remains constant, and can more than offset the *surplus extraction effect* when $d_i \geq \overline{d}_i$. The previous literature (Bounie *et al.* [2021]; Chen *et al.* [2020]) has highlighted how the size of the partitions affects the relative weights of the *competition* and *surplus extraction effects* in a duopolistic setting, possibly leading to the dominance of the latter over the former.[11] Point (iii) of Lemma 1 shows that such property extends to an oligopolistic setting.

Overlapping partitions. If $d_i + d_k \geq \frac{2}{n}$, with $k \in \{i-1, i+1\}$ denoting i's direct competitor, data partitions sold to firms d_i and d_k overlap, implying that all consumers between firms i nd k are identified by at least one firm. If a consumer located in x is identified by only one firm, the informed firm offers a tailored price that matches the consumer's utility when buying at the competitor's basic price, as in (7). Conversely, if a consumer located in x is identified by both adjacent firms, firm i offers to that consumer a tailored price $p_i^{\mathrm{T}}(x)$ that matches the consumer's utility when buying from the competitor at the tailored price $p_k^{\mathrm{T}}(x)$:

$$(10)\quad p_i^{\mathrm{T}}(x) = \begin{cases} p_{i-1}^{\mathrm{T}}(x) + 2tx - \frac{t}{n}(2i-1) & \text{for } x \in \left[\frac{i}{n} - \frac{d_i}{2}, \frac{i-1}{n} + \frac{d_{i-1}}{2}\right], \\ p_{i+1}^{\mathrm{T}}(x) - 2tx + \frac{t}{n}(2i+1) & \text{for } x \in \left[\frac{i+1}{n} - \frac{d_{i+1}}{2}, \frac{i}{n} + \frac{d_i}{2}\right]. \end{cases}$$

In this case, it is as if both adjacent firms were competing *à la* Bertrand on the consumer located in x. Competition drives the firm with a location disadvantage (e.g., $i+1$) to offer the lowest possible tailored price, that is, $p_{i+1}^{\mathrm{T}}(x) = 0$. Such an offer is always dominated by the tailored price offered by the firm with the location advantage (e.g., firm i), which sets $p_i^{\mathrm{T}}(x) = -2tx + \frac{t}{n}(2i+1)$. This implies that all consumers identified by both firms always buy from the closest competitor.

To study the location of the indifferent consumer, consider the cases illustrated in Figure 2.

In particular, in panel a) d_i is sufficiently small $\left(\text{i.e., } d_i < \frac{1}{n}\right)$, but partitions still overlap (i.e., $d_i + d_{i+1} \geq \frac{2}{n}$). In this case, firm i has a positional advantage over all consumers it identifies $\left(\text{i.e., the segment } \left(\frac{i}{n}, \frac{i}{n} + \frac{d_i}{2}\right]\right)$, and serves them through tailored prices. In addition, firm i also has a positional advantage over some consumers only identified by its competitor $\Big($i.e., the segment $\left(\frac{i}{n} + \frac{d_i}{2}, \frac{i}{n} + \frac{1}{2n}\right]\Big)$, as the overlapping segment is sufficiently close to firm i's location. Due to this advantage, some of these consumers prefer firm i's basic price to the competitor's (zero) tailored price. The indifferent consumer (denoted by $\hat{x}_{i,i+1}$ in the figure) obtains the same utility by buying from firm i at p_i^B or buying from firm $i+1$ at $p_{i+1}^T(x) = 0$, implying that the demand of firm i includes some consumers served through its basic price.

[11] In particular, Chen *et al.* [2020] focus on the effects of consumer hiding, and Bounie *et al.* [2021] analyze a data sale through first-price auctions.

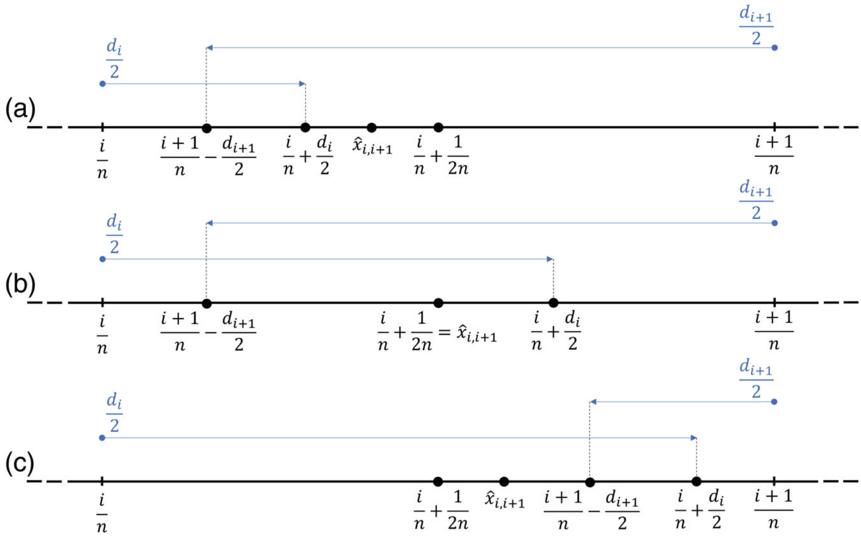

Figure 2

Location of the Indifferent Consumers when Partitions Overlap

Notes: [Colour figure can be viewed at wileyonlinelibrary.com]

A different situation emerges when the overlap between data partitions includes the middle consumer, as illustrated in panel (b) of Figure 2, where $d_i \geq \frac{1}{n}$ and $d_{i+1} \geq \frac{1}{n}$. In this case, neither firm has a positional advantage over consumers only identified by the competitor, hence the indifferent consumer is equidistant from the two firms. Basic prices are only offered to consumers who are identified by the rival and are closer to its location. Competition with the rival's tailored price brings the firm's basic price to zero.[12] Panel (c) of Figure 2 mirrors the situation described in panel a), and has the indifferent consumer closer to firm $i+1$.

The following equation summarizes the discussion above and formally expresses the position of the indifferent consumer:

$$(11) \qquad \hat{x}_{i,i+1} = \begin{cases} \frac{2i+1}{2n} - \frac{p_i^B}{2t} & \text{for } d_i < \frac{1}{n} \quad \textit{and} \quad d_i + d_{i+1} \geq \frac{2}{n}, \\ \frac{i}{n} + \frac{1}{2n} & \text{for } d_i \geq \frac{1}{n} \quad \textit{and} \quad d_{i+1} \geq \frac{1}{n}, \\ \frac{2i+1}{2n} + \frac{p_{i+1}^B}{2t} & \text{for } d_{i+1} < \frac{1}{n} \quad \textit{and} \quad d_i + d_{i+1} \geq \frac{2}{n}. \end{cases}$$

[12] Note that firm i sets its basic price only if it does identify some, but not all, consumers in the market, that is, $d_i < 1$. If $d_i = 1$, firm i identifies all consumers and only competes on tailored prices. Nonetheless, the sale of the whole dataset still entails an indifferent consumer who is equidistant from the two adjacent firms.

Firm i's profits are given by (1) if $d_i < \frac{1}{n}$ and $d_i + d_{i+1} \geq \frac{2}{n}$, and by (2) otherwise. Using (10) and (11), we obtain the effects of the data partitions on firm i's profits in this subgame, which are summarized in the following lemma.

Lemma 2. In the subgame where all firms buy their respective partition, if partitions overlap, we have that:
(i) $\frac{\partial \pi_i^{W*}(\mathbf{P})}{\partial d_k} \leq 0, k \in \{i-1, i+1\}$ (firm i's subgame equilibrium profit is weakly decreasing in d_k),
(ii) $\frac{\partial \pi_i^{W*}(\mathbf{P})}{\partial d_i} \geq 0$ (firm i's subgame equilibrium profit is weakly increasing in d_i).

Proof. See the Appendix. ■

An increase in the size of the rivals' partitions reduces firm i's profit, as it expands the set of consumers identified by both firms, on which competition is more intense (point (i) of Lemma 2). Moreover, when partitions overlap, an increase of d_i strictly increases firm i's profits when it allows firm i to identify additional consumers on which it has a location advantage (namely, when $d_i < \frac{1}{n}$), whereas it has no effect on firm i's profits otherwise (point (ii) of Lemma 2).

III(ii). *Firm i Rejects the DB's Offer*

If firm i does not buy data, it becomes uninformed and competes having $d_i = 0$. In this case, its profit is given by (3) that, using (6), can be expressed as

$$\pi_i^{L}(\mathbf{P}) = p_i^{B}\left(\frac{n\left(p_{i+1}^{B} + p_{i-1}^{B} - 2p_i^{B}\right) + 2t}{2nt}\right) - F. \tag{12}$$

In this case, data have no direct effect on (12), which is affected by data only indirectly through the competitors' basic prices.

The first-order condition of (12) with respect to $p_i^{B}(\mathbf{P})$ is:

$$p_i^{B} = \frac{t}{2n} + \frac{p_{i+1}^{B} + p_{i-1}^{B}}{4}. \tag{13}$$

Equation (13) is the reaction function of the standard Salop [1979] model. Note that the other firms, by accepting their respective partitions, exhibit a reaction function as described by (9).

The main properties of firms' equilibrium prices and profits in this subgame in which firm i does not buy data are described in the following lemma.

Lemma 3. In the subgame where all firms except firm i buy data, the equilibrium basic prices of all firms are higher, and firm i's profits are lower than in the subgame where also firm i buys data.

Proof. See the Appendix. ■

As firm i does not buy data, the competitive pressure in the market decreases, and all firms set higher basic prices. Moreover, competing without data puts firm i at a disadvantage vis-à-vis its rivals, leading to lower profits. This result also implies that firms' willingness to pay for data w_i is always positive.

IV. DB'S EQUILIBRIUM STRATEGY, ENTRY, AND WELFARE

We now analyze the DB's profits and identify its optimal strategy in terms of the partition set **P**. We then find the level of firms' entry in the Perfect Nash equilibrium of the game and the implications of the equilibrium on consumer surplus and welfare.

IV(i). *DB's Optimal Strategy*

As data are a key strategic input to compete in the downstream market, the DB aims to extract most of the surplus from firms. To do so, the DB sets the data prices w_i equal to firms' willingness to pay for data, namely the difference in profits between buying or not their respective partitions. As a tie-breaker rule, we assume that if a firm is indifferent between purchasing or not purchasing data, it prefers buying them. It follows that, after paying for data, firms are left with profits equal to $\pi_i^{L*}(\mathbf{P})$. Thus, the DB solves the following problem:

$$\max_{d_0, d_1, \dots, d_{n-1}} \pi_{DB}(\mathbf{P}) = \sum_{i=0}^{n-1} \left(\pi_i^{W*}(\mathbf{P}) - \pi_i^{L*}(\mathbf{P}) \right). \tag{14}$$

The following lemma illustrates the properties of the optimal partition chosen by the DB, and its effects on firms' profits.

Lemma 4. In equilibrium, the DB offers same-sized partitions to all firms, that is, $d_i = d$, for all i. The DB's profit $\pi_{DB}(\mathbf{P})$ has two local maxima at $d < \frac{1}{n}$, and for any $d \in \left[\frac{3}{2n}, 1\right]$.

Proof. See the Appendix. ■

Intuitively, due to firms' symmetry, DB's profits are influenced by all the partitions he sells in the same way, and thus he offers symmetric partitions. To gain an intuition on the second part of the Lemma, it is convenient to refer to Figure 3, which represents a firm's profit with respect to d in the two subgames in which all firms buy data (the solid line in the figure) or the firm rejects the data offer and competes against informed rivals (the dashed line).

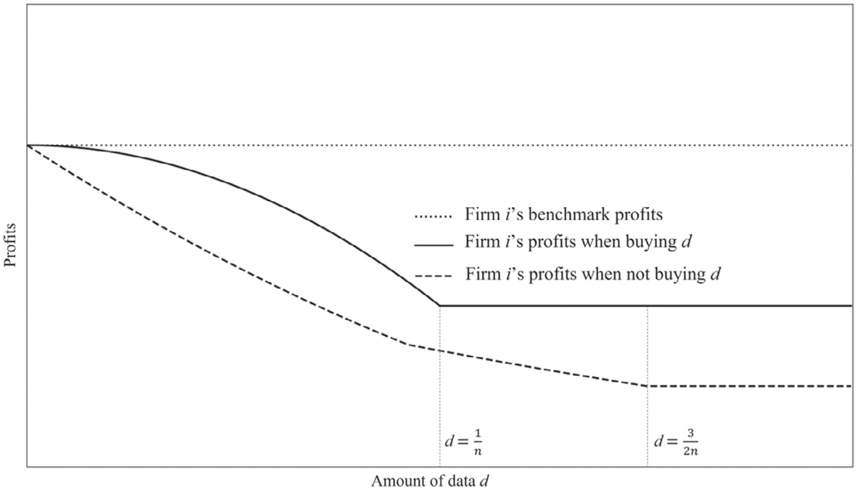

Figure 3
Firms' Profits with Respect to d ($n = 4$ and $t = 20$)

The DB's profit is given by the difference between the two curves and is always positive due to the competitive advantage provided by data. The figure highlights how the amount of data sold affects the two curves, determining a trade-off for the DB. On the one hand, more data reduce a firm's profit if it rejects the DB's offer and competes without data against informed rivals. This increases firms' willingness to pay for data, hence the DB's profit. On the other hand, more data intensify competition between informed firms, which reduces their basic prices, an effect magnified by the fact that all firms have the same amount of data. This reduces firms' willingness to pay for data and the DB's profit. When d is small enough, both these effects are present, implying a local maximum in the DB's profit and positive basic prices. Conversely, when $d \geq \frac{1}{n}$ and all firms purchase data, partitions overlap. Although all consumers are served through tailored prices, basic prices are still offered to consumers who are closer to a firm's rivals. Competition brings basic prices to the marginal cost, that is, to zero. The second effect thus disappears, as additional data would not influence the firms' basic prices, that is, firms' profits become constant for $d \geq \frac{1}{n}$ when all firms purchase data. Moreover, when $d \in \left[\frac{3}{2n}, 1\right]$, the first effect also disappears as any additional data would identify consumers who are too close to the uninformed firm to be poached due to horizontal differentiation. This implies that firms' profits when not buying data are constant for $d \in \left[\frac{3}{2n}, 1\right]$.[13] Thus, any $d \in \left[\frac{3}{2n}, 1\right]$ represents a local maximum for the DB's profit.

[13] Since the uninformed firm's market share is lower due to its competitive disadvantage, its rivals need a larger partition $\left(d \geq \frac{3}{2n}\right)$ to identify the indifferent consumer.

IV(ii). *Equilibrium Data Partitions*

Given that the DB offers same-sized partitions in equilibrium from Lemma (4), his profit-maximization problem in (14) can be expressed as

$$\max_{d} \pi_{DB}(\mathbf{P}) = n \left(\pi_i^{W*}(\mathbf{P}) - \pi_i^{L*}(\mathbf{P}) \right). \tag{15}$$

The solution of problem (15) is the optimal amount of data d^*, and it depends on the number of entering firms. We thus find the following proposition.

Proposition 1. There exists a number $\hat{n}$ of firms such that in equilibrium the DB offers $d^* \in \left[\frac{3}{2n}, 1\right]$ if $n < \hat{n}$, and nonoverlapping partitions otherwise.

Proof. See the Appendix. ■

Proposition 1 highlights a discontinuity in the DB's strategy, whereby he sells overlapping partitions when the market is concentrated, and nonoverlapping partitions otherwise. The intuition behind the result of Proposition 1 relies on the role of indirect competition by nonadjoining rivals, who emerge when $n \geq 4$.[14] When $n \leq 3$, firms only have direct rivals. If a firm declines to purchase data, both its competitors compete against an uninformed rival. The resulting milder competition leads to a stark increase in basic prices if partitions are nonoverlapping. In turn, higher basic prices reduce the firm's willingness to pay for data. To avoid the reduction in firms' willingness to pay for data stemming from the increase of basic prices if a firm decides not to buy data, the DB offers overlapping partitions, so that all consumers are served through tailored prices. When $n \geq 4$, firms face both the direct competition of adjoining rivals and the indirect competition of more distant ones via the complementarity of the pricing strategies. Even if a firm declines to purchase data, its indirect rivals still directly compete against informed firms. The strong competitive pressure faced by indirect rivals exercises a downward pressure on basic prices and on the profits of the uninformed firm. In this situation, selling overlapping partitions, relative to nonoverlapping partitions, mildly raises the (already high) threat faced by uninformed firms, but sharply decreases the equilibrium profits and, in turn, the value of data for firms. Thus, the DB opts for the sale of nonoverlapping partitions of data when indirect rivals are present in the market, thereby tempering competition and increasing the value of data.

The previous literature (Bounie *et al.* [2022]), has analyzed the role of a DB selling data to a downstream duopoly assuming nonoverlapping partitions, showing that a DB sells data to both firms under TIOLI offers. Our results

[14] In the Appendix, we show that $\hat{n}$ is a constant such that $\hat{n} \in (3, 4)$. Moreover, it does not depend on the degree of horizontal differentiation t.

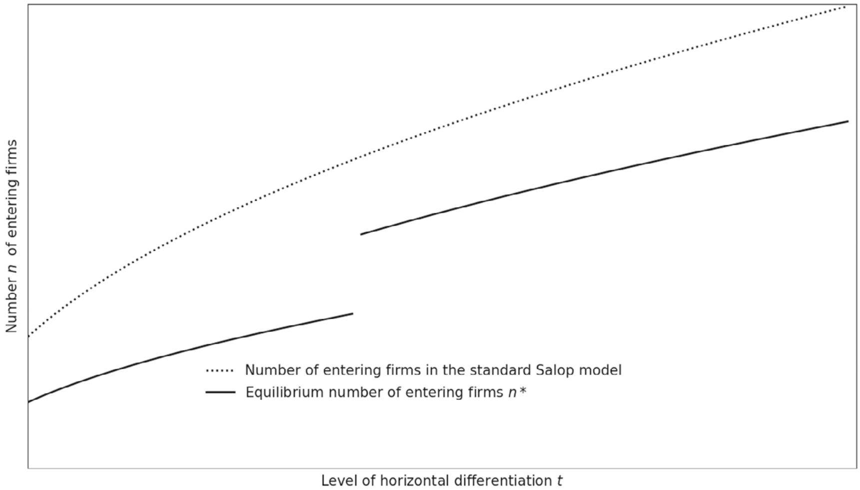

Figure 4
Number of Entering Firms with Respect to the Level of Horizontal Differentiation t ($F = 1$)

show that the sale to all firms is still optimal when entry is accounted for. However, selling nonoverlapping partitions is suboptimal when the market is sufficiently concentrated, as it entails a too mild threat for firms that choose not to buy data.

IV(iii). *Entry, Consumer Surplus, and Welfare*

We now study the implications of the equilibrium characterized in the previous section on the number of entering firms and, consequently, on consumer surplus and welfare.

As a benchmark, recall that the number of entering firms in the standard Salop [1979] model (see Section II), absent the DB, is $\tilde{n} = \sqrt{\frac{t}{F}}$. In our model, the number of entering firms n^* is obtained through the free-entry condition, by imposing that firms' profits after paying for entry and data are equal to zero. We find the following result.

Proposition 2. There exists a value of $\hat{t}$ such that the number of entering firms in equilibrium is $n^* = \frac{1}{2}\sqrt{\frac{t}{F}}$ if $t < \hat{t}$, and $n^* \approx \frac{3}{4}\sqrt{\frac{t}{F}}$ otherwise.

Proof. See the Appendix. ■

Figure 4 illustrates the number of entering firms with respect of the level of horizontal differentiation in the downstream market, highlighting the three main properties of n^* established in Proposition 2. First, the equilibrium number of entering firms is always lower than in the benchmark case where data are absent. Intuitively, such *entry barrier effect* arises because firms' profits are lower. Indeed, data increase the intensity of competition, leading firms to price more aggressively. In addition, the need to pay for data further erodes firms' profits and leaves less room for entry. Interestingly, the number of entering firms in equilibrium, n^*, cannot be lower than $\frac{\tilde{n}}{2}$, that is, the entry deterrence caused by data is limited by firms' horizontal differentiation. This result complements the entry barrier effect of data identified by de Cornière and Taylor [2020] in a setting in which data affect the quality of the information held by firms. Our analysis shows that the entry barrier effect also emerges when data can be used for price discrimination as they carry information on consumer preferences. Second, Proposition 2 highlights that n^* is increasing in t. As is typical of the localized competition setup à la Salop, a higher horizontal differentiation softens competition, increasing profits and the scope for entry. Third, Proposition 2 shows that n^* has a discontinuity in t, given by the DB changing his selling strategy from overlapping to nonoverlapping partitions. When horizontal differentiation is low, the reduced profitability decreases the scope for entry, which makes the sale of overlapping partitions optimal (from Proposition 1). Conversely, the higher scope for entry when horizontal differentiation is high leads to the sale of nonoverlapping partitions. The discontinuity of n^* is due to the rise in firms' profits when the DB suddenly shifts from selling sufficiently large partitions $d^* \in \left[\frac{3}{2n}, 1\right]$ to selling nonoverlapping partitions $d^* < \frac{1}{n}$.

The key message of Proposition 2 is that the DB reduces the firms' incentives to enter the market via the data sale. The degree of horizontal differentiation influences how much such incentives are reduced, and the channel through which they are reduced. If the market is mildly competitive due to a high level of horizontal differentiation, the DB sells small quantities of data, thus limiting the entry barrier effect of data. This tempers the erosion of firms' profits caused by data, and increases their willingness to pay for them. When instead the market is highly competitive due to low horizontal differentiation, the DB sells sufficiently large partitions, thus reinforcing the entry barrier and the firms' disincentive to enter. By doing so, the DB reduces firms' profit should they decide to forgo data, which raises their willingness to pay for them.

Let us now study the implications of our equilibrium on consumer surplus and welfare. The surplus of consumers buying from firm i is defined as the integral of consumers' utility:

$$CS_i = \int_{\hat{x}_{i-1,i}}^{\hat{x}_{i,i+1}} U(x, i)dx. \tag{16}$$

Total consumer surplus is $CS = \sum_i CS_i$. In particular, consumer surplus can be expressed as[15]

$$CS = \begin{cases} v - \frac{5t}{4n} + \frac{ntd^2}{2} & \text{for } d \in \left[0, \frac{1}{n}\right), \\ v - \frac{5t}{4n} + \frac{t}{2n} & \text{for } d \in \left[\frac{1}{n}, 1\right]. \end{cases} \tag{17}$$

Let us define welfare TW as the weighted sum of consumer surplus CS, firm profits and the DB' profits:

$$TW = CS + \alpha\left(\sum_{i=0}^{n-1}(\pi_i - w_i) + \pi_{\text{DB}}\right), \tag{18}$$

where $\alpha \in [0, 1]$ is the weight of industry profits in the welfare function and represents the potentially lower weight that the policymaker could attribute to industry profits because of, for instance, their international reach (Baron and Myerson [1982]). Note that, in equilibrium, $\pi_i^{W*} - w_i^* = \pi_i^{L*} = 0$ by the free-entry condition. By denoting with $\widetilde{CS} = \widetilde{TW} = v - \frac{5}{4}\sqrt{tF}$ the consumer surplus and welfare in the standard Salop [1979] model, respectively, the following proposition summarizes the impact of the DB's equilibrium strategy on consumers surplus and welfare.

Proposition 3. In equilibrium, $CS^* < \widetilde{CS}$. Moreover, $TW^* > \widetilde{TW}$ if and only if α is sufficiently high.

Proof. See the Appendix. ■

Proposition 3 highlights that the entry barrier effect that arises from the DB's presence lowers consumer surplus, relative to a situation in which data are not available. Proposition 3 is in contrast with the findings of the previous literature highlighting the positive effect of data on consumer surplus due to the increase in competition (Braulin and Valletti [2016]; Bounie *et al.* [2021]; Montes *et al.* [2019]). Our results instead point out that, when firm entry is endogenous, the limited entry hurts consumers, and more than offsets the decrease in basic prices caused by the competition effect of data.[16]

Proposition 3 also shows that, if the weight α of the industry profits in the welfare function is high enough, total welfare is higher than in the

[15] As already shown in Lemma 4, all firms buy data in equilibrium and additional data have no effect when $d \geq \frac{1}{n}$. Thus, consumer surplus is constant for $d \geq \frac{1}{n}$.

[16] Notably, the reduction of consumer surplus stems from the presence of a DB and its effect on entry. In the absence of a DB, the possibility of firms to price discriminate through data increases consumer surplus by weakening the entry barrier effect (see, e.g., Taylor and Wagman [2014]).

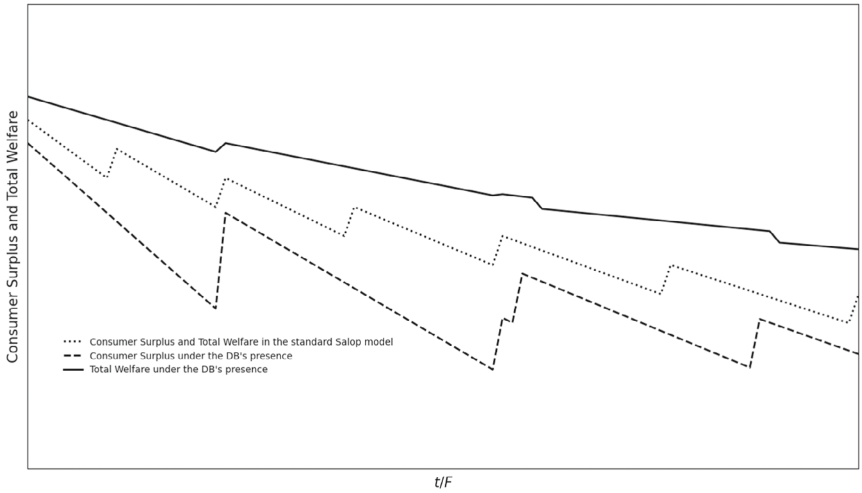

Figure 5

Consumer Surplus and Total Welfare for $v = 100$, $\frac{t}{F} \in [16, 100]$, $\alpha = 1$. Consumer surplus and total welfare under the DB's presence are reported as confidence intervals, as there is no explicit solution for n^*.

benchmark case.[17] The DB's equilibrium strategy either partially (when he sells nonoverlapping partitions) or totally (when he sells sufficiently large partitions) solves the excessive entry problem identified by Salop (1979). While being higher, total welfare is mostly appropriated by the DB, with redistributive implications from a policymaking point of view.

The results of Proposition 3, while obtained for a continuous number of firms, can easily be extended to the case in which n is integer. In this case, n^* is obtained by rounding down the result of Proposition 2 to the closest integer.[18] Figure 5 illustrates the levels of consumer surplus and total welfare when $n^* \in \mathbb{N}$ and $\alpha = 1$. Note that the two functions follow a step-like curve, corresponding to the steps of the n^* function at the integer levels.

When n is rounded down, market concentration is (weakly) higher than in the case where n is continuous. This in turn implies lower consumer surplus and higher industry profits than in the case of a continuous n^*. Then, $CS^* < \widetilde{CS}$ when $n^* \in \mathbb{N}$, extending the first part of Proposition 3 to an integer number of firms.

[17] In the Appendix, we show that total welfare increases iff $\alpha \geq \frac{1}{2}$ when $n < \hat{n}$, and iff $\alpha \geq \frac{3}{7}$ when $n \geq \hat{n}$.

[18] Indeed, $n^* \in \mathbb{R}$ is the value of n that is consistent with the free-entry condition. Any value higher than $n^* \in \mathbb{R}$ entails negative profits once the fixed entry cost is taken into account.

Moreover, when $n^* \in \mathbb{N}$, total welfare is higher than in the benchmark provided that α is sufficiently high, confirming also the second part of Proposition 3. This is due to the fact that the presence of the DB, by reducing entry by at least 25%, brings entry significantly closer to the efficient, welfare maximizing level.[19] Thus, the efficiency gain generated by the DB's presence always outweighs any inefficiency that could originate by rounding down n^*.

V. CONCLUDING REMARKS

With the steady growth of online services, DBs have become central players in the digital economy. Their ability to extract valuable information from consumers' data allows them to influence competition in retail markets, with important welfare implications. Our work contributes to the growing literature on the competitive effects of DBs by modeling an oligopoly market where the number of firms is endogenous and data sold by a DB are used for price discrimination.

We show that the presence of a DB reduces the entry of firms in the downstream market. The DB benefits from the increased concentration, as he can extract firms' profits through the price of data. Previous literature on price discrimination in spatial competition settings has often highlighted a pro-competitive effect of data, as firms engage in price wars over the identified consumers. We show that, when entry endogenously depends on the DB's strategy, the entry barrier effect dominates the competition effect, leading to an overall decrease in competition in the market. Due to the increased market concentration stemming from the entry barrier effect, the data sale by a monopolistic DB reduces consumer surplus. Moreover, welfare is mostly appropriated by the DB. As a consequence, if the weight of the industry profits in the welfare function is sufficiently low, the presence of a DB is welfare decreasing.

Our analysis also highlights that the DB influences downstream competition not only by limiting entry but also by affecting how the entered firms compete. When horizontal differentiation is low, the DB floods the market with data by selling sufficiently large partitions. Firms thus engage in Bertrand-like competition over each consumer through tailored prices. Conversely, when horizontal differentiation is high, the DB tempers downstream competition by selling nonoverlapping partitions of data, which leaves some consumers unidentified by all firms. Firms imperfectly compete through uniform prices over the segments of unidentified consumers and offer tailored prices only to the closest consumers.

[19] In Salop [1979], welfare is maximized when the number of entering firms is $\frac{1}{2}\sqrt{\frac{t}{F}}$. Proposition 2 shows that the DB's presence reduces entry by $\frac{\tilde{n}-n^*}{\tilde{n}} = \frac{1}{4}$ when $t \geq \hat{t}$, and by $\frac{\tilde{n}-n^*}{\tilde{n}} = \frac{1}{2}$ otherwise.

From a policymaking point of view, our results suggest that the presence of a DB that can steer the competitive dynamics by raising entry barriers is detrimental for consumers, despite the fact that the use of data intensifies competition between firms.

An important issue that remains to be addressed deals with the presence of competition at the DB's level. Indeed, competition between DBs is likely to limit the DB's bargaining power, possibly tempering the entry barrier effect. A careful analysis is needed to fully assess the implications of competition between DBs on entry in the downstream market and for consumers.

APPENDIX

Proof of Lemma 1. The expression (9) of firms' basic prices can be rewritten as

$$4p_i^{\mathrm{B}} - p_{i-1}^{\mathrm{B}} - p_{i+1}^{\mathrm{B}} = \frac{2t}{n} - 2td_i,$$

$\forall i \in \{0, \ldots, n-1\}$. The firms' equilibrium basic prices are obtained by solving the system composed by the above n equations. The system in matricial form is expressed by $\mathbf{A} * \mathbf{p} = \mathbf{b}$, where $\mathbf{p}$ is the vector containing basic prices, and $\mathbf{b}$ is the vector containing the known terms:

$$\begin{bmatrix} 4 & -1 & \ldots & 0 & 0 & 0 & \ldots & -1 \\ -1 & 4 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 4 & -1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & -1 & 4 & -1 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & -1 & 4 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ -1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 4 \end{bmatrix} * \begin{bmatrix} p_0^{\mathrm{B}} \\ p_1^{\mathrm{B}} \\ \ldots \\ p_{i-1}^{\mathrm{B}} \\ p_i^{\mathrm{B}} \\ p_{i+1}^{\mathrm{B}} \\ \ldots \\ p_{n-1}^{\mathrm{B}} \end{bmatrix} = \begin{bmatrix} \frac{2t}{n} - 2td_0 \\ \frac{2t}{n} - 2td_1 \\ \ldots \\ \frac{2t}{n} - 2td_{i-1} \\ \frac{2t}{n} - 2td_i \\ \frac{2t}{n} - 2td_{i+1} \\ \ldots \\ \frac{2t}{n} - 2td_{n-1} \end{bmatrix}.$$

Matrix $\mathbf{A}$ is circulant, tridiagonal and symmetric. Given a general circulant tridiagonal matrix of form $\mathbf{M} = (a, b, 0, 0, \ldots, 0, c)$, where a, b, c express the non-null elements of the first line, the general expression of its inverse is provided by Searle [1979], and it is given by

$$\mathbf{A}^{-1} = \begin{bmatrix} a_0 & a_1 & \ldots & a_{n-1} \\ a_{n-1} & a_0 & \ldots & a_{n-2} \\ \ldots & \ldots & \ldots & \ldots \\ a_1 & a_2 & \ldots & a_0 \end{bmatrix},$$

where $a_j = \frac{z_1 z_2}{b(z_1 - z_2)} \left(\frac{z_1^j}{1 - z_1^n} - \frac{z_2^j}{1 - z_2^n} \right)$ and $z_1, z_2 = \frac{\sqrt{-a \pm (a^2 - 4bc)}}{2c}$. In our case, as $a = 4, b = -1$ and $c = -1$, we obtain that $a_j = -\frac{1}{2\sqrt{3}} \left(\frac{\left(2+\sqrt{3}\right)^j}{1-\left(2+\sqrt{3}\right)^n} - \frac{\left(2-\sqrt{3}\right)^j}{1-\left(2-\sqrt{3}\right)^n} \right)$.

Using $\mathbf{A}^{-1}$, we obtain the equilibrium basic prices through $\mathbf{p} = \mathbf{A}^{-1} * \mathbf{b}$:

$$\begin{bmatrix} p_0^{B*} \\ p_1^{B*} \\ \dots \\ p_{n-1}^{B*} \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & \dots & a_{n-1} \\ a_{n-1} & a_0 & \dots & a_{n-2} \\ \dots & \dots & \dots & \dots \\ a_1 & a_2 & \dots & a_0 \end{bmatrix} * \begin{bmatrix} \frac{2t}{n} - 2td_0 \\ \frac{2t}{n} - 2td_1 \\ \dots \\ \frac{2t}{n} - 2td_{n-1} \end{bmatrix}. \tag{A.1}$$

Thus, equilibrium basic prices are

$$p_i^{B*} = \left(\frac{2t}{n} * \sum_{j=0}^{n-1} a_j \right) - 2t \sum_{j=0}^{n-1} d_{i+j} a_j. \tag{A.2}$$

A useful property of the matrix $\mathbf{A}^{-1}$ in our framework is that $\sum_{j=0}^{n-1} a_j = \frac{1}{2}$, so that (A.2) can be simplified as

$$p_i^{B*} = \frac{t}{n} - 2t \sum_{j=0}^{n-1} d_{i+j} a_j. \tag{A.3}$$

From (A.3), all equilibrium basic prices are decreasing in all firms' partitions (point *(i)* of the Lemma).

To study the effects of the data partitions on firm i's profits in (8), it is useful to rewrite (A.3) as:

$$p_i^{B*} = \frac{t}{n} - 2td_i a_0 - 2t \sum_{j=1}^{n-1} d_{i+j} a_j \tag{A.4}$$

and

$$p_{i+1}^{B*} = \frac{t}{n} - 2td_i a_1 - 2t \sum_{j=0, j\neq 1}^{n-1} d_{i+1+j} a_j. \tag{A.5}$$

Moreover, by exploiting the property of circulant matrices that $a_j = a_{n-j}\ \forall j \neq \{0, \frac{n}{2}\}$, we also have $p_{i-1}^{B*} = p_{i+1}^{B*}$. By substituting (A.4) and (A.5) in (8) and simplifying, we obtain

$$\begin{aligned} \pi_i^{W*}(\mathbf{P}) = & \underbrace{\frac{d_i}{2n} \left(4t - ntd_i \left(1 + 4a_1\right) - 4nt \sum_{j=0, j\neq 1}^{n-1} d_{i+1+j} a_j \right)}_{\text{Firm } i'\text{s profits through tailored prices}} \\ & + \underbrace{\left(\frac{t}{n} - 2td_i a_0 - 2t \sum_{j=1}^{n-1} d_{i+j} a_j \right)}_{\text{Firm } i\text{'s basic price}} \\ & \times \underbrace{\left(\frac{1}{n} + 2d_i \left(a_0 - a_1 - \frac{1}{2}\right) + 2 \sum_{j=1}^{n-1} d_{i+j} \left(a_j - a_{j-1}\right) \right)}_{\text{Consumers served by firm } i \text{ through the basic price}} - F. \end{aligned} \tag{A.6}$$

From (A.6), an increase of $d_{k\neq i}$ decreases the profits firm i makes through the tailored price, as well as its basic price and its market share. In fact, the coefficient a_j is decreasing in j $\forall j \in \{0, \frac{n}{2}\}$, implying $a_j - a_{j-1} < 0$. We thus conclude that the rivals' partitions always decrease firm i's profits (point *(ii)* of the Lemma).

From the FOC of (A.6) with respect to d_i, we obtain

$$\overline{d_i} = \frac{1-2a_1}{n\left(1+4a_1-4a_0+8a_0^2-8a_0a_1\right)} - 4\frac{\sum_{j=0,j\neq 1}^{n-1} d_{i+1+j}a_j\left(a_0-a_1\right)+a_0\sum_{j=0,j\neq 1}^{n-1} d_{i+1+j}\left(a_{j-1}-a_j\right)}{1+4a_1-4a_0+8a_0^2-8a_0a_1}. \tag{A.7}$$

For any $d_i > \overline{d_i}$, $\frac{d\pi_i^W}{dd_i} < 0$, while $\frac{d\pi_i^W}{dd_i} \geq 0$ otherwise (point *(iii)* of the Lemma). We are only left to show that $\overline{d_i}$ exists and may be positive. To this aim, note that the first term on the right side of (A.7) only depends on n, is always positive and lower than $\frac{1}{n}$. Conversely, the second term is negative or null. In particular, when $d_i \geq 0$ and $d_{k\neq i} = 0$, the second term on the right side of (A.7) is equal to zero, implying that a $\overline{d_i} > 0$ exis■

Proof of Lemma 2. The case of overlapping partitions emerges if $d_i + d_k \geq \frac{2}{n}$, with $k \in \{i-1, i+1\}$ denoting i's direct competitor. Without loss of generality, we focus on firm i's profits on the arch between firm i and $i+1$, denoted as $\pi_{i,i+1}^W$. We distinguish between five cases, depending on the sizes of the two data partitions.

Case 1: $d_i < \frac{1}{n}$ and $\frac{1}{n} < d_{i+1} < \frac{2}{n}$. In this case, the two partitions overlap on an arch that is closer to firm i's location. Then, firm i serves all consumers identified only by firm i, plus all consumers identified by both firms, on which firm i has a positional advantage. In particular, consumers identified only by firm i are those located between firm i' position—in $\frac{i}{n}$—and the last consumer identified by its rival—in $\frac{i+1}{n} - \frac{d_{i+1}}{2}$. Firm i serves such consumers through the tailored price in equation (7). Consumers identified by both firms are located between $\frac{i+1}{n} - \frac{d_{i+1}}{2}$ and $\frac{i}{n} + \frac{d_i}{2}$, and firm i serves them through the tailored price in equation (10).

Moreover, firm i also has a positional advantage over some unidentified consumers, namely those located between $\frac{i}{n} + \frac{d_i}{2}$ and $\frac{i}{n} + \frac{1}{2n}$. Then, firm i can beat firm $i+1$'s tailored offer to some of these consumers if it sets p_i^B low enough, and thus has an incentive to set a positive basic price. The last consumer that buys from firm i through its basic price is the one that is indifferent between buying from firm i at p_i^B and buying from firm $i+1$ at $p_{i+1}^T(x) = 0$. By equating the consumer utilities under the two options, we find $\hat{x}_{i,i+1} = \frac{2i+1}{2n} - \frac{p_i^B}{2t}$.

Equilibrium profits of firm i on the arch between firm i and $i+1$ are:

$$\pi_{i,i+1}^{\mathrm{W}*}(\mathbf{P}) = \int_{\frac{i}{n}}^{\frac{i+1}{n}-\frac{d_{i+1}}{2}} p_{i+1}^{\mathrm{B}} - 2tx + \frac{t}{n}(2i+1)dx + \int_{\frac{i+1}{n}-\frac{d_{i+1}}{2}}^{\frac{i}{n}+\frac{d_i}{2}} -2tx + \frac{t}{n}(2i+1)dx + p_i^B\left(\hat{x}_{i,i+1} - \frac{i}{n} - \frac{d_i}{2}\right) - \frac{F}{2}. \tag{A.8}$$

The first component of (A.8) represents the profits firm i makes on the consumers identified only by itself, whereas the second component represents the profits it makes on consumers that are also identified by firm $i+1$. The third component are firm i's profits obtained through its basic price. Note that fixed costs are divided by two, so that when we consider total firm i's profits the fixed costs are equal to F. By computing the FOC of (A.8) with respect to p_i^B we find $p_i^{B*} = \frac{t}{2n} - \frac{td_i}{2}$ and $\hat{x}^*_{i,i+1} = \frac{i}{n} + \frac{1}{4n} + \frac{d_i}{4}$. Replacing them in (A.8) we obtain

$$\text{(A.9)} \quad \pi^{W*}_{i,i+1}(\mathbf{P}) = \int_{\frac{i}{n}}^{\frac{i+1}{n}-\frac{d_{i+1}}{2}} p^{B}_{i+1} - 2tx + \frac{t}{n}(2i+1)dx + \int_{\frac{i+1}{n}-\frac{d_{i+1}}{2}}^{\frac{i}{n}+\frac{d_i}{2}} -2tx + \frac{t}{n}(2i+1)dx + \frac{t}{8n^2}\left(d_i n - 1\right)^2 - \frac{F}{2}.$$

By deriving equation (A.9) with respect to d_i and d_{i+1}, we find that $\frac{d\pi^{W*}_{i,i+1}}{dd_i} > 0$ if $d_i < \frac{1}{n}$ and $\frac{d\pi^{W*}_{i,i+1}}{dd_{i+1}} < 0$.

Case 2: $d_i < \frac{1}{n}$ and $d_{i+1} \geq \frac{2}{n}$. In this case, all consumers identified by firm i are also identified by firm $i+1$, as the latter identifies all consumers on the arch. Then, firm i's equilibrium profits obtained through tailored prices only depend on the consumers that it serves through the tailored price in equation (10), which are those located between its location—that is, $\frac{i}{n}$—and the last consumer that it identifies—that is, $\frac{i}{n} + \frac{d_i}{2}$. Equilibrium profits of firm i on the arch between firm i and $i+1$ become:

$$\text{(A.10)} \quad \pi^{W*}_{i,i+1}(\mathbf{P}) = \int_{\frac{i}{n}}^{\frac{i}{n}+\frac{d_i}{2}} -2tx + \frac{t}{n}(2i+1)dx + \frac{t}{8n^2}\left(d_i n - 1\right)^2 - \frac{F}{2}.$$

By deriving equation (A.10) with respect to d_i and d_{i+1}, we find that $\frac{d\pi^{W*}_{i,i+1}}{dd_i} > 0$ if $d_i < \frac{1}{n}$ and $\frac{d\pi^{W*}_{i,i+1}}{dd_{i+1}} = 0$.

Case 3: $d_i \geq \frac{1}{n}$ and $\frac{2}{n} > d_{i+1} \geq \frac{1}{n}$. In this case, the two partitions overlap over the consumer who is equidistant from firm i and $i+1$, that is, the consumer located in $x = \frac{i}{n} + \frac{1}{2n}$. Then, firm i serves all consumers between its location $\frac{i}{n}$ and the middle consumer in $\frac{i}{n} + \frac{1}{2n}$, as firm i cannot poach consumers who are located after $x = \frac{i}{n} + \frac{1}{2n}$, due to firm $i+1$'s positional advantage over those consumers. In particular, consumers located between $\frac{i}{n}$ and $\frac{i+1}{n} - \frac{d_{i+1}}{2}$ are identified only by firm i, which serves them through the tailored price in (7). Conversely, the remaining share of consumers located between $\frac{i+1}{n} - \frac{d_{i+1}}{2}$ and $\frac{i}{n} + \frac{1}{2n}$ are identified by both firms and served by firm i through the tailored price (10). Firm i's equilibrium profits on the arch between firm i and $i+1$ thus become

$$\text{(A.11)} \quad \pi^{W*}_{i,i+1}(\mathbf{P}) = \int_{\frac{i}{n}}^{\frac{i+1}{n}-\frac{d_{i+1}}{2}} p^{B}_{i+1} - 2tx + \frac{t}{n}(2i+1)dx + \int_{\frac{i+1}{n}-\frac{d_{i+1}}{2}}^{\frac{i}{n}+\frac{1}{2n}} -2tx + \frac{t}{n}(2i+1)dx - \frac{F}{2}.$$

By deriving equation (A.11) with respect to d_i and d_{i+1}, we find that $\frac{d\pi^{W*}_{i,i+1}}{dd_i} = 0$ and $\frac{d\pi^{W*}_{i,i+1}}{dd_{i+1}} < 0$.

Case 4: $d_i \geq \frac{1}{n}$ and $d_{i+1} \geq \frac{2}{n}$. As in Case 2, firm $i+1$ identifies all consumers on the arch. Then, firm i's equilibrium profits only depend on the consumers that it serves through the tailored price in equation (10), which are those located between its location—$\frac{i}{n}$—and the last consumer on which it has a positional advantage—$\frac{i}{n} + \frac{1}{2n}$. Equilibrium profits of firm i on the arch between firm i and $i+1$ become:

$$\pi^{\mathrm{W}*}_{i,i+1}(\mathbf{P}) = \int_{\frac{i}{n}}^{\frac{i}{n}+\frac{1}{2n}} -2tx + \frac{t}{n}(2i+1)dx - \frac{F}{2}. \tag{A.12}$$

By deriving equation (A.12) with respect to d_i and d_{i+1}, we find that $\frac{d\pi^{W*}_{i,i+1}}{dd_i} = 0$ and $\frac{d\pi^{W*}_{i,i+1}}{dd_{i+1}} = 0$.

Case 5: $d_i > \frac{1}{n}$ and $d_{i+1} < \frac{1}{n}$. In this case, specularly to Case 1, the overlap is closer to firm $i+1$ and, as such, firm $i+1$ follows the same strategy that firm i follows in Case 1. Firm $i+1$ is able to serve some consumers through its basic price. The last consumer that buys from firm $i+1$ is the one indifferent between buying from firm $i+1$ at a price p^B_{i+1} and buying from firm i at a price $p^T_i(x) = 0$. Such consumer is located in $\hat{x}_{i,i+1} = \frac{2i+1}{2n} + \frac{p^B_{i+1}}{2t}$. Like in case 1, in equilibrium firm $i+1$ sets $p^{B*}_{i+1} = \frac{t}{2n} - \frac{td_{i+1}}{2}$, leading to $\hat{x}^*_{i,i+1} = \frac{i}{n} + \frac{3}{4n} - \frac{d_{i+1}}{4}$. Then, firm i's equilibrium profits are equal to:

$$\pi^{\mathrm{W}*}_{i,i+1}(\mathbf{P}) = \int_{\frac{i}{n}}^{\frac{i}{n}+\frac{3}{4n}-\frac{d_{i+1}}{4}} \frac{t}{2n} - \frac{td_{i+1}}{2} - 2tx + \frac{t}{n}(2i+1)dx - \frac{F}{2}. \tag{A.13}$$

By deriving equation (A.13) with respect to d_i and d_{i+1}, we find that $\frac{d\pi^{W*}_{i,i+1}}{dd_i} = 0$ and $\frac{d\pi^{W*}_{i,i+1}}{dd_{i+1}} < 0$.

By jointly considering the results of all cases, we conclude that $\frac{d\pi^{W*}_{i,i+1}}{dd_{i+1}} \leq 0$ and $\frac{d\pi^{W*}_{i,i+1}}{dd_i} \geq 0$. Moreover, as the same results would hold when analyzing the segment $[i-1, i]$, we conclude that $\frac{d\pi^{W*}_i}{dd_k} \leq 0, k \in \{i-1, i+1\}$ (point *(i)* of Lemma 2) and $\frac{d\pi^{W*}_i}{dd_i} \geq 0$ (point *(ii)* of Lemma 2). ■

Proof of Lemma 3. When firm i is the only one not buying data, basic prices are obtained from the system of equations (A.1) in the proof of Lemma 1 by imposing $d_i = 0$. Hence, the i-th component of **b** in (A.1) is $\frac{2t}{n}$. By solving the system, the equilibrium basic prices in this subgame—denoted with the superscript L—are

$$p^{\mathrm{B\,L}*}_i = \frac{t}{n} - 2t\sum_{j=1}^{n-1} d_{i+j}a_j, \tag{A.14}$$

$$p^{\mathrm{B\,L}*}_{i+k} = p^{\mathrm{B\,L}*}_{i-k} = \frac{t}{n} - 2t\sum_{j=0, j\neq k}^{n-1} d_{i+k+j}a_j, \tag{A.15}$$

By directly comparing (A.14) and (A.15) with (A.4) and (A.5), respectively, we immediately observe that $p^{B*}_i > p^{BL*}_i$ and $p^{B*}_{i+k} > p^{BL*}_{i+k}$.

By substituting (A.14) and (A.15) in (12), we obtain

$$\pi_i^{L*}(\mathbf{P}) = \left(\frac{t}{n} - 2t\sum_{j=1}^{n-1} d_{i+j}a_j\right)\left(\frac{1}{n} + 2\sum_{j=1}^{n-1} d_{i+j}\left(a_j - a_{j-1}\right)\right) - F. \tag{A.16}$$

By comparing (A.16) with (A.6), we find that $\pi_i^{W*}(\mathbf{P}) > \pi_i^{L*}(\mathbf{P})$. ■

Proof of Lemma 4. The proof proceeds in two steps. In Step 1, we show that any DB's equilibrium strategy implies offering same-sized partitions to all entering firms, that is, $d_i = d\ \forall i \in \{0, \dots, n-1\}$. In Step 2, we compute the DB's profits when he offers same-sized partitions.

Step 1. We first show that competition in our model is localized, that is, a firm is never able to serve consumers who are located after its direct rivals' locations, for any **P**. To this aim, let us focus on the case where a generic firm i is at the highest competitive disadvantage, and show that it is still able to obtain a positive market share in equilibrium.

From Lemmas 1 and 2, firm i's profits are (weakly) decreasing with the size of all the rivals' partitions, and from Lemma 3 we know that firm i's profits are lower when it does not obtain data, all else equal. Thus, firm i's profits are minimized under the partition set $\hat{\mathbf{P}} = (1, 1, \dots, 1, 0, 1, \dots, 1)$, where $d_i = 0$ and $d_{k \neq i} = 1$, that is, $\hat{\mathbf{P}}$ places firm i at the highest competitive disadvantage.

We now show that, under $\hat{\mathbf{P}}$, firm i's market share is still strictly positive.

First note that as $i+1$ and $i-1$ have the same information on all consumers, the last consumer they serve is located no farther than firm i's location, else it would be poached by the rival by means of its locational advantage. This implies that the last consumer firm $i+1$ serves, denoted by $\hat{x}_{i,i+1}$, is located between its location and firm i's location, that is, $\hat{x}_{i,i+1} \in \left[\frac{i}{n}, \frac{i+1}{n}\right]$, and the same hold for firm $i-1$.

We now show that the last consumers served by $i+1$ and $i-1$ are not located exactly in $\frac{i}{n}$, but rather allow to firm i a strictly positive market share. As firms $i+1$ and $i-1$ identify all consumers, they only offer tailored prices, whereas firm i, which is uninformed, uses the basic price. Let us now focus on a generic consumer located between firms i and $i+1$. Its net utility from consuming from either firm is:

$$\begin{aligned} U(x, i) &= v - t\left(x - \frac{i}{n}\right) - p_i^B, \\ U(x, i+1) &= v - t\left(\frac{i+1}{n} - x\right) - p_{i+1}^T(x). \end{aligned} \tag{A.17}$$

The indifferent consumer location is obtained by equating the consumer utilities in (A.17) and setting $p_{i+1}^T(x) = 0$, which is the tailored price offered by $i+1$ to the last consumer it serves. The position of the indifferent consumer is thus:

$$\hat{x}_{i,i+1} = \frac{2i+1}{2n} - \frac{p_i^B}{2t}. \tag{A.18}$$

Analogously, the indifferent consumer between firms $i-1$ and i is

$$\hat{x}_{i-1,i} = \frac{2i-1}{2n} + \frac{p_i^B}{2t}. \tag{A.19}$$

Using (A.18) and (A.19) in (3), we obtain firm i's profit

$$\pi_i^{\mathrm{L}}(\hat{\mathbf{P}}) = p_i^{\mathrm{B}}\left(\frac{2t - 2np_i^{\mathrm{B}}}{2nt}\right) - F. \tag{A.20}$$

By computing the FOC of (A.20) with respect to p_i^{B}, we obtain $p_i^{\mathrm{B}*} = \frac{t}{2n}$ and

$$\hat{x}^*_{i,i+1} = \frac{i}{n} + \frac{1}{4n} \quad \textit{and} \quad \hat{x}^*_{i-1,i} = \frac{i}{n} - \frac{1}{4n}, \tag{A.21}$$

that is, firm i obtains strictly positive market shares also when competing without data in a market of fully informed rivals, allowing us to conclude that in our model competition is localized.

The DB solves the following maximization problem

$$\max_{d_0,d_1,\ldots,d_{n-1}} \pi_{DB}(\mathbf{P}) = \sum_{i=0}^{n-1}\left(\pi_i^{\mathrm{W}*}(\mathbf{P}) - \pi_i^{\mathrm{L}*}(\mathbf{P})\right). \tag{A.22}$$

As segments are ex-ante identical, in equilibrium the DB replicates the strategy on one segment on all segments, that is, $d_0 = d_2 = d_4 = \cdots = d_{n-2}$ and $d_1 = d_3 = d_5 = \cdots = d_{n-1}$. Moreover, as competition is localized, $d_i + d_{i+1} > 0\ \forall i$, otherwise the DB fails to extract the value of data from one segment and could profitably deviate by selling a positive amount of data to one of the two firms. Depending on the partition sizes, we have two possible cases.

Case 1: $d_i + d_{i+1} < \frac{2}{n}$. In this case, any pair of adjacent partitions does not overlap. Then, $\pi_i^{\mathrm{W}*}$ in (A.22) is expressed by (A.6), whereas $\pi_i^{\mathrm{L}*}$ in (A.22) is expressed by (A.16), for all i. By the symmetry of the DB's profits with respect to any partition d_i we thus have that, in equilibrium, the DB sets $d_i = d\ \forall i \in \{0, \ldots, n-1\}$.

Case 2: $d_i + d_{i+1} \geq \frac{2}{n}$. In this case, any pair of adjacent partitions does overlap. Without loss of generality, we focus on firms' profits on the arch between firm i and $i+1$. It is useful to distinguish between two subcases: when either $d_i < \frac{1}{n}$ or $d_{i+1} < \frac{1}{n}$, and when $d_i, d_{i+1} \geq \frac{1}{n}$.

Case 2a: $d_i < \frac{1}{n}$. Firms' profits functions are those expressed in the proof of Lemma 2. In particular, as $d_i + d_{i+1} \geq \frac{2}{n}$, firm $i+1$ identifies all the consumers it serves on the arch $\left[\frac{i}{n}, \frac{i+1}{n}\right]$. Moreover, as $d_i = d_{i+2}$, the same is also true on the arch $\left[\frac{i+1}{n}, \frac{i+2}{n}\right]$. Then, firm $i+1$ does not serve any consumer through its basic price p^B_{i+1}, which is only offered to consumers who are located closer to firm i and to which firm i offers a tailored price $p_i^T(x)$. Bertrand-like competition on that segment of consumers leads to $p^{B*}_{i+1} = 0$. Then, firm i's profits are expressed by either (A.9)—if $\frac{1}{n} < d_{i+1} < \frac{2}{n}$—or (A.10)—if $d_{i+1} < \frac{2}{n}$. However, as $p^{B*}_{i+1} = 0$, we have that (A.9) coincides with (A.10). Firm $i+1$'s profits are instead expressed by (A.13), once adjusted with the $i+1$ subscript. By simplifying those expressions, we obtain:

$$\begin{aligned} \pi^{W*}_{i,i+1} &= \frac{td_i}{4n}\left(2 - d_i n\right) + \frac{t}{8n^2}\left(d_i n - 1\right)^2 - \frac{F}{2}, \\ \pi^{W*}_{i+1,i} &= \frac{t}{16n^2}\left(d_i n - 3\right)^2 - \frac{F}{2}, \end{aligned} \tag{A.23}$$

where $\pi^{W*}_{i,i+1}$ and $\pi^{W*}_{i+1,i}$ are respectively firm i and firm $i+1$'s profits on the arch $\left[\frac{i+1}{n}, \frac{i+2}{n}\right]$. It is useful to rewrite (A.22) as

$$\text{(A.24)} \qquad \max_{d_0,d_1,\dots,d_{n-1}} \pi_{DB}(\mathbf{P}) = \sum_{i=0}^{n-1} \left(\pi^{\mathrm{W}*}_{i,i+1}(\mathbf{P}) + \pi^{\mathrm{W}*}_{i+1,i}(\mathbf{P})\right) - \sum_{i=0}^{n-1} \left(\pi^{\mathrm{L}*}_{i}(\mathbf{P})\right).$$

By the symmetry of the FOCs of the DB's profits with respect to any partition d_i, the solution of problem (A.24) implies $d_i = d \; \forall i \in \{0, \dots, n-1\}$, which contradicts our initial hypothesis of $d_i < d_{i+1}$. Thus, we conclude that there exists no equilibrium with $d_i < \frac{i}{n}$ and $d_i + d_{i+1} \geq \frac{2}{n}$. By symmetry, the same result holds when $d_{i+1} < \frac{i}{n}$ and $d_i + d_{i+1} \geq \frac{2}{n}$.

Case 2b: $d_i, d_{i+1} \geq \frac{1}{n}$. Then, firms identify all consumers they have a positional advantage on and serve them through tailored prices. As in Case 2a, firm i's basic price is only offered to consumers who are unidentified by firm i, who are located closer to its rivals and to which the rivals offer a tailored price. Bertrand-like competition on that segment of consumers leads to $p^{B*}_i = 0 \; \forall i \in \{0, \dots, n-1\}$. Firms' profits are thus expressed as in (A.11)—if $\frac{1}{n} < d_{i+1} < \frac{2}{n}$—or as in (A.12)—if $d_{i+1} \geq \frac{2}{n}$. However, as $p^{B*}_i = 0$, we have that (A.11) coincides with (A.12). As firms' equilibrium winning profits do not depend on any partition, and firms' equilibrium losing profits $\pi^{\mathrm{L}*}_i$ are expressed in (A.16), we conclude by symmetry that, if $d_i, d_{i+1} \geq \frac{1}{n}$, then in equilibrium the DB sets $d_i = d \; \forall i \in \{0, \dots, n-1\}$.

Jointly taking the results from Cases 1, 2a and 2b, we conclude that, in equilibrium, the DB always sets $d_i = d \; \forall i \in \{0, \dots, n-1\}$.

Step 2. Let us focus on firms' prices and profits when firm i buys data. First, suppose that partitions do not overlap (i.e., $d < \frac{1}{n}$). By setting $d_i = d \; \forall i \in \{0, \dots, n-1\}$ in (A.4) and (A.5), we obtain $p^{B*}_i = \frac{t}{n} - td \; \forall i \in \{0, \dots, n-1\}$. As all basic prices are equal, indifferent consumers are located in the middle between firms' locations, that is, $\hat{x}_{i,i+1} = \frac{2i+1}{2n}$. By substituting the equilibrium prices in (8), we obtain firms' profits under nonoverlapping partitions:

$$\text{(A.25)} \qquad \pi^{W*}_i = \frac{t}{n^2} - \frac{td^2}{2} - F.$$

Firms identify all consumers they serve when $\frac{i}{n} + \frac{d}{2} \geq \hat{x}_{i,i+1}$, which we can rewrite as $d \geq \frac{1}{n}$. In this case, partitions overlap and firms set their basic prices equal to zero, as shown in Step 1 of this proof. By substituting $p^{B*}_i = 0 \; \forall i \in \{0, \dots, n-1\}$ in (A.11), we obtain firms' profits under overlapping partitions

$$\text{(A.26)} \qquad \pi^{W*}_i = \frac{t}{2n^2} - F.$$

We now focus on the subgame in which firm i does not buy data. We have to consider three separate cases: (i) all informed firms serve both identified and unidentified consumers, (ii) all informed firms except firm i's direct rivals only serve identified consumers and (iii) all informed firms only serve identified consumers.

(i) When firm i does not buy data, it becomes the only uninformed firm in the market, and its profits are given by (12). By setting $d_i = d \; \forall i \in \{0, \dots, n-1\}$ in (A.14) and (A.15), we obtain

(A.27) $$p_i^{\mathrm{B\,L}*}(\mathbf{P}) = \frac{t}{n} - td + 2tda_0 \quad \text{and} \quad p_{i-j}^{\mathrm{B\,L}*}(\mathbf{P}) = p_{i+j}^{\mathrm{B\,L}*}(\mathbf{P}) = \frac{t}{n} - td + 2tda_j.$$

By substituting (A.27) in (12), we find

(A.28) $$\pi_i^{\mathrm{L}*}(\mathbf{P}) = \left(\frac{t}{n} - td + 2tda_0\right)\left(2d\left(a_1 - a_0\right) + \frac{1}{n}\right) - F.$$

From (A.27), we know that informed firms set different equilibrium basic prices, depending on their distance from firm i. In particular, basic prices are higher, the closer a firm is to firm i. Let us focus on the indifferent consumer between firms $i-2$ and $i-1$. Using (A.27), we obtain

$$\widehat{x}_{i-2,i-1} = \frac{2i-3}{2n} + d(a_1 - a_2).$$

Firm $i-2$ can identify consumers up to $\frac{i-2}{n} + \frac{d}{2}$. Then, if

$$d \geq d_1 \equiv \frac{1}{2n\left(\frac{1}{2} + a_1 - a_2\right)},$$

firm $i-2$ only serves identified consumers and sets its basic price equal to 0. As $(a_j - a_{j+1})$ decreases with j, all other informed firms except $i+1$ and $i-1$ also set their basic prices equal to 0. Thus, this case only holds as long as $d < d_1$.

(ii) Without loss of generality, we focus on firms $i-1$ and i. If $d \geq d_1$, firm $i-1$ identifies all consumers on the arch it shares with firm $i-2$, whereas it still serves some unidentified consumers on the arch it shares with firm i. We can write firm $i-1$'s profits as

(A.29) $$\pi_{i-1}^{\mathrm{W}}(\mathbf{P}) = \int_{\widehat{x}_{i-2,i-1}}^{\frac{i-1}{n}} p_{i-1,i-2}^{\mathrm{T}}(x)dx + \int_{\frac{i-1}{n}}^{\frac{i-1}{n}+\frac{d}{2}} p_{i-1,i}^{\mathrm{T}}(x)dx + p_{i-1}^{\mathrm{B}}(\mathbf{P})\left(\widehat{x}_{i-1,i} - \frac{i-1}{n} - \frac{d}{2}\right) - F.$$

Firm i's profits are given by (12). The FOCs of (A.29) and (12) give us the equilibrium basic prices:

(A.30) $$p_{i-1}^{\mathrm{B}*}(\mathbf{P}) = \frac{t(3-2nd)}{5n} \quad \text{and} \quad p_i^{\mathrm{B\,L}*}(\mathbf{P}) = \frac{t(4-nd)}{5n}.$$

Substituting (A.30) into (12), we obtain

(A.31) $$\pi_i^{\mathrm{L}*}(\mathbf{P}) = \frac{t(nd-4)^2}{25n^2} - F.$$

From (A.30), informed firms set positive basic prices as long as $d < \frac{3}{2n}$. After this threshold, firms $i-1$ and $i+1$ identify all the consumers they serve, and thus set their equilibrium basic prices equal to zero.

(iii) Suppose $d \geq \frac{3}{2n}$. Then, firms $i+1$ and $i-1$ identify all the consumers they serve. In turn, firm i's profits are given by (A.20), and only depend on its basic price. The FOC of (A.20) with respect to p_i^B leads to

(A.32) $$p_i^{B*}(\mathbf{P}) = \frac{t}{2n} \quad \text{and} \quad \pi_i^{L*}(\mathbf{P}) = \frac{t}{4n^2} - F.$$

Comparing (A.28), (A.31) and (A.32) we conclude that firm i's profits when not buying data are strictly decreasing in d if $d < \frac{3}{2n}$, and constant otherwise.

By combining the expressions for $\pi_i^{W*}(\mathbf{P})$ (in (A.25) and (A.26)) and $\pi_i^{L*}(\mathbf{P})$ (in (A.28), (A.31) and (A.32)) above, we obtain the function of DB's profits:

$$\max_d \pi_{DB} = \begin{cases} n\left(\frac{t}{n^2} - \frac{td^2}{2} - \left(\frac{t}{n} - td + 2tda_0\right)\left(2d\left(a_1 - a_0\right) + \frac{1}{n}\right)\right) & \text{for } d < d_1, \\ n\left(\frac{t}{n^2} - \frac{td^2}{2} - \frac{t(nd-4)^2}{25n^2}\right) & \text{for } d_1 \leq d < \frac{1}{n}, \\ n\left(\frac{t}{2n^2} - \frac{t(nd-4)^2}{25n^2}\right) & \text{for } \frac{1}{n} \leq d < \frac{3}{2n}, \\ n\left(\frac{t}{4n^2}\right) & \text{for } d \geq \frac{3}{2n}, \end{cases}$$

which is continuous in d. We prove that the strategy of setting $d \in \left[d_1, \frac{3}{2n}\right)$ (corresponding to the second and third part of DB's profits) is always suboptimal. By computing the FOC of the first part with respect to d, we find that it is first increasing and then decreasing in d, with a local maximum for some $d \in (0, d_1)$. By computing the FOC of the second part with respect to d, we find that it is monotonically decreasing in d over its domain. Thus, the DB's profits in the first part are higher than those in the second part for some d. By computing the FOC of the third part with respect to d, we find that it is monotonically increasing in d over its domain. Thus, as the DB's profits are continuous, the DB profits in the fourth part are constant and higher than those in the third one. We conclude that in equilibrium the DB either sells $d < d_1$ or $d \geq \frac{3}{2n}$. ∎

Proof of Proposition 1. Let us consider the two strategies $d < d_1$ and $d \geq \frac{3}{2n}$. If the DB sets $d < d_1$, FOC with respect to d gives us

(A.33) $$d^* = \frac{1 - 2a_1}{n\left(-8a_0^2 + a_0\left(8a_1 + 4\right) - 4a_1 + 1\right)}.$$

By substituting (A.33) in DB's profits, we obtain

(A.34) $$\pi_{DB}^* = \frac{t}{2}(1 - 2a_1)d^*.$$

Instead, if the DB sets $d \geq \frac{3}{2n}$, his profits are equal to

(A.35) $$\pi_{DB}^* = \frac{t}{4n}.$$

By comparing (A.34) and (A.35), the DB sets $d = d^*$ iff

(A.36) $$\frac{(1 - 2a_1)^2}{(-8a_0^2 + a_0(8a_1 + 4) - 4a_1 + 1)} \geq \frac{1}{2}.$$

Numerical analysis allows us to find that inequality (A.36) is satisfied for $n \geq \hat{n} \approx 3.34$. Therefore, when $n < \hat{n}$, the DB sets $d \geq \frac{3}{2n}$, whereas when $n \geq \hat{n}$, the DB sets $d = d^*$. ■

Proof of Proposition 2. The DB chooses his equilibrium strategy given the number of entering firms. Suppose that $n < \hat{n}$, where $\hat{n}$ is the implicit solution of condition (A.36) holding with the equality. Then, firms' profits after paying for data and entry coincide with the profits if they do not buy data in (A.32). The number of entering firms is given by the corresponding free entry condition

$$\pi_i^{L*}(\mathbf{P}) = \frac{t}{4n^2} - F = 0,$$

leading to $n^* = \frac{1}{2}\sqrt{\frac{t}{F}}$. We are only left to check under which parameter values we have that $n^* < \hat{n}$. It is immediate to see that $n^* < \hat{n}$ iff $t < \hat{t} = F\hat{n}^2$.

Suppose instead that $t \geq \hat{t}$. Then, as $n \geq \hat{n}$, the DB sets $d = d^*$. Firms' profits after paying for data and entry coincide with the profits if they do not buy data in (A.27). The number of entering firms is given by the corresponding free entry condition

$$\text{(A.37)} \qquad \pi_i^{L*}(\mathbf{P}) = \left(\frac{t}{n} - td^* + 2td^*a_0\right)\left(2d^*(a_1 - a_0) + \frac{1}{n}\right) - F = 0,$$

which has no explicit solution, as the coefficients a_j exponentially depend on n. To solve the condition, it is useful to rewrite d^* as

$$\text{(A.38)} \qquad d^* = \frac{\alpha(n)}{n},$$

where

$$\alpha(n) = \frac{1 - 2a_1}{-8a_0^2 + a_0(8a_1 + 4) - 4a_1 + 1}.$$

By substituting (A.38) in (A.37), we obtain

$$\text{(A.39)} \qquad \left(1 - \alpha(n) + 2a_0\alpha(n)\right)\left(2(a_1 - a_0\alpha(n) + 1\right) = \frac{F}{t}n^2.$$

Let us denote with $A(n)$ the left hand-side of (A.39). $A(n)$ is monotonically decreasing in n and has an asymptote for $n \to \infty$:

$$\lim_{n\to\infty} A(n) = \frac{36\sqrt{3} - 99}{1644\sqrt{3} - 2915}.$$

We can approximate $A(n)$ with

$$\text{(A.40)} \qquad \overline{A}(n) \approx \frac{1}{n^3} + \frac{36\sqrt{3} - 99}{1644\sqrt{3} - 2915} \quad \text{or} \quad \underline{A}(n) \approx \frac{1}{n^3} + \frac{53}{100}.$$

The first (second) approximation overestimates (underestimates) the true value of $A(n)$ for any $n > 2$. By substituting (A.40) in (A.39), we obtain

(A.41) $$\frac{1}{n^3} + \frac{36\sqrt{3}-99}{1644\sqrt{3}-2915} - \frac{F}{t}n^2 = 0 \quad or \quad \frac{1}{n^3} + \frac{53}{100} - \frac{F}{t}n^2 = 0.$$

To find an explicit solution to (A.41), we use the Newton-Raphson approximation method. We obtain

(A.42) $$\overline{n^*} \approx \sqrt{\frac{t}{F}} \frac{4096\left(1644\sqrt{3}-2915\right)\frac{F}{t} + 243\left(1708\sqrt{3}-3091\right)\sqrt{\frac{t}{F}}}{8\left(1644\sqrt{3}-2915\right)\left(512\frac{F}{t}+81\sqrt{\frac{t}{F}}\right)}$$

$$or \quad \underline{n^*} \approx \frac{102400\frac{F}{t} + 11799\sqrt{\frac{t}{F}}}{200\left(\frac{512}{\frac{t}{F}^{\frac{3}{2}}}+81\right)},$$

which are both slightly above $\frac{3}{4}\sqrt{\frac{t}{F}}$. ■

Proof of Proposition 3. Suppose $t < \hat{t}$. Then, $n < \hat{n}$, and the DB sells overlapping partitions. As all firms obtain the same quantity of data, indifferent consumers are located at the center of each arch. Without loss of generality, let us focus on the arch between firms i and $i+1$. The indifferent consumer in the middle of the arch is located in $\frac{2i+1}{2n^*}$. Firm i serves all its consumers in $\left[\frac{i}{n^*}, \frac{2i+1}{2n^*}\right]$ through its tailored price. Consumer surplus on this semi-arch is given by integrating consumer net utility between $\frac{i}{n^*}$ and $\frac{2i+1}{2n^*}$. The total consumer surplus on all $2n^*$ semi-arches of the market is:

(A.43) $$CS^*_{t<\hat{t}} = 2n^*\left(\int_{\frac{i}{n^*}}^{\frac{2i+1}{2n^*}} v - p_i^{\mathrm{T}*}(x) - t\left(x-\frac{i}{n^*}\right)dx\right),$$

where

(A.44) $$p_i^{\mathrm{T}}(x) = -2tx + \frac{t}{n^*}(2i+1).$$

By replacing (A.44) in (A.43) we obtain

$$CS^*_{t<\hat{t}} = v - \frac{3t}{4n^*} = v - \frac{3}{2}\sqrt{tF},$$

that is, $CS^*_{t<\hat{t}} < \widetilde{CS} = v - \frac{5}{4}\sqrt{tF}$.

When $t \geq \hat{t}$, all firms offer equal basic prices and have equal market shares. Thus, indifferent consumers are located in the middle points between firms. To compute total consumer surplus, we evaluate the consumer surplus of consumers located in $\left[0, \frac{1}{2n^*}\right]$, and multiply it by $2n^*$. We obtain

(A.45)
$$CS^*_{t\geq\hat{t}} = 2n^*\left(\int_0^{\frac{d^*}{2}} v - tx - p_0^T(x)dx + \int_{\frac{d^*}{2}}^{\frac{1}{2n^*}} v - tx - p_0^{B*}dx\right) = v - \frac{5t}{4n^*} + \frac{1}{2}n^*td^{*2}$$

Using (A.42) and (A.45), we find

$$\overline{CS}^*_{t\geq\hat{t}} = v - \frac{5t}{4\overline{n^*}} + \frac{1}{2}\overline{n^*}td^{*2} \quad or \quad \underline{CS}^*_{t\geq\hat{t}} = v - \frac{5t}{4\underline{n^*}} + \frac{1}{2}\underline{n^*}td^{*2}.$$

By comparing both expressions of $CS^*_{t\geq\hat{t}}$ with $\widetilde{CS} = v - \frac{5}{4}\sqrt{tF}$, we find that consumer surplus is lower than in the benchmark case, both when underestimating or overestimating n^*.

We now look at total welfare and recall that in equilibrium firms' profits are equal to zero due to the free entry condition. Then, the only components of total welfare are consumer surplus and the DB's profits. When $t < \hat{t}$, we have $CS^*_{t<\hat{t}} = v - \frac{3}{2}\sqrt{tF}$ and $\pi^*_{DB} = \frac{1}{2}\sqrt{tF}$. Using (18), total welfare is thus equal to

$$TW^*_{t<\hat{t}} = v - \frac{3}{2}\sqrt{tF} + \frac{\alpha}{2}\sqrt{tF}.$$

We find that $TW^*_{t<\hat{t}} \geq \widetilde{TW} = v - \frac{5}{4}\sqrt{tF}$ iff $\alpha \geq \frac{1}{2}$.

If instead $t \geq \hat{t}$, we have $CS^*_{t\geq\hat{t}} = v - \frac{5t}{4n^*} + \frac{1}{2}n^*td^{*2}$ and $\pi^*_{DB} = \frac{t}{2}(1-2a_1)d^*$. By substituting the equilibrium values of n^* and d^*, we obtain

$$CS^*_{t\geq\hat{t}} \approx v - \frac{7}{5}\sqrt{tF}$$

and

$$\pi^*_{DB} \approx \frac{7}{20}\sqrt{tF}.$$

Total welfare is thus equal to

$$TW^*_{t\geq\hat{t}} = u - \frac{7}{5}\sqrt{tF} + \frac{7\alpha}{20}\sqrt{tF}.$$

We find that $TW^*_{t\geq\hat{t}} \geq \widetilde{TW} = v - \frac{5}{4}\sqrt{tF}$ iff $\alpha \geq \frac{3}{7}$. ■

REFERENCES

Acquisti, A. and Varian, H. R., 2005, 'Conditioning Prices on Purchase History,' *Marketing Science*, 24(3), pp. 367–381.

Anderson, S. and Bedre-Defolie, O., 2023, 'Hybrid Platform Model: Monopolistic Competition and a Dominant Firm,' *The RAND Journal of Economics, forthcoming*. Available at: *https://www.eui.eu/people?id=ozlem-bedre-defolie*

Aparicio, D.; Metzman, Z. and Rigobon, R., 2023, 'The Pricing Strategies of Online Grocery Retailers,' *Quantitative Marketing and Economics*, forthcoming. Available at: *https://link.springer.com/article/10.1007/s11129-023-09273-w#citeas*

Armstrong, M. and Vickers, J., 2001, 'Competitive Price Discrimination,' *The RAND Journal of Economics*, 32(4), pp. 579–605.

Baron, D. P. and Myerson, R. B., 1982, 'Regulating a Monopolist with Unknown Costs,' *Econometrica: Journal of the Econometric Society*, 50(4), pp. 911–930.

Bergemann, D. and Bonatti, A., 2011, 'Targeting in Advertising Markets: Implications for Offline versus Online Media,' *The RAND Journal of Economics*, 42(3), pp. 417–443.

Bergemann, D. and Bonatti, A., 2015, 'Selling Cookies,' *American Economic Journal: Microeconomics*, 7(3), pp. 259–294.

Bergemann, D. and Bonatti, A., 2019, 'Markets for Information: An Introduction,' *Annual Review of Economics*, 11, pp. 85–107.

Bergemann, D.; Bonatti, A. and Smolin, A., 2018, 'The Design and Price of Information,' *American Economic Review*, 108(1), pp. 1–48.

Bester, H. and Petrakis, E., 1996, 'Coupons and Oligopolistic Price Discrimination,' *International Journal of Industrial Organization*, 14(2), pp. 227–242.

Bhaskar, V. and To, T., 2004, 'Is Perfect Price Discrimination Really Efficient? An Analysis of Free Entry,' *RAND Journal of Economics*, 35, pp. 762–776.

Bounie, D.; Dubus, A. and Waelbroeck, P., 2021, 'Selling Strategic Information in Digital Competitive Markets,' *The RAND Journal of Economics*, 52(2), pp. 283–313.

Bounie, D.; Dubus, A. and Waelbroeck, P., 2022, 'Market for Information and Selling Mechanisms,' *Economics Working Paper Series*, 22/367, pp. 1–67.

Braulin, F. C. and Valletti, T., 2016, 'Selling Customer Information to Competing Firms,' *Economics Letters*, 149, pp. 10–14.

Cachon, G. P. and Lariviere, M. A., 2005, 'Supply Chain Coordination with Revenue-Sharing Contracts: Strengths and Limitations,' *Management Science*, 51(1), pp. 30–44.

Chen, Z.; Choe, C. and Matsushima, N., 2020, 'Competitive Personalized Pricing,' *Management Science*, 66(9), pp. 4003–4023.

Choe, C.; Cong, J. and Wang, C., 2023, 'Softening Competition Through Unilateral Sharing of Customer Data,' *Management Science, forthcoming*. Available at: *https://pubsonline.informs.org/doi/10.1287/mnsc.2023.4689*

Clifford, S., 2012, 'Shopper Alert: Price May Drop for You Alone,' *The New York Times http://www.nytimes.com/2012/08/10/business/supermarkets-try-customizing-prices-for-shoppers.html?pagewanted=all*.

Colombo, S., 2018, 'Behavior- and Characteristic-Based Price Discrimination,' *Journal of Economics and Management Strategy*, 27(2), pp. 237–250.

Crain, M., 2018, 'The Limits of Transparency: Data Brokers and Commodification,' *New Media & Society*, 20(1), pp. 88–104.

de Cornière, A., 2016, 'Search Advertising,' *American Economic Journal: Microeconomics*, 8(3), pp. 156–188.

de Cornière, A. and Taylor, G., 2020, *Data and Competition: A General Framework with Applications to Mergers, Market Structure, and Privacy Policy*. SSRN Scholarly Paper ID 3547379. (Social Science Research Network, Rochester, NY).

DeGraba, P., 1990, 'Input Market Price Discrimination and the Choice of Technology,' *The American Economic Review*, 80(5), pp. 1246–1253.

Delbono, F.; Reggiani, C. and Sandrini, L., 2021, *Strategic Data Sales to Competing Firms*. Technical Report JRC126568 (JRC Digital Economy Working Paper, Seville, Spain).

FTC, 2014, *Data Brokers: A Call for Transparency and Accountability*. Tech. rept. (Federal Trade Commission, Washington, DC).

Fudenberg, D. and Villas-Boas, J. M., 2006, 'Behavior-Based Price Discrimination and Customer Recognition,' *Handbook on Economics and Information Systems*, 1, pp. 377–436.

Greenhut, M. L. and Ohta, H., 1976, 'Related Market Conditions and Interindustrial Mergers,' *The American Economic Review*, 66(3), pp. 267–277.

Gu, Y.; Madio, L. and Reggiani, C., 2019, 'Exclusive Data, Price Manipulation and Market Leadership,' *Cesifo Working Paper No. 7853*, pp. 1–42.

Hagiu, A. and Wright, J., 2020, 'Data-Enabled Learning, Network Effects and Competitive Advantage,' *The RAND Journal of Economics, forthcoming*.

Hart, O. D. and Tirole, J., 1988, 'Contract Renegotiation and Coasian Dynamics,' *The Review of Economic Studies*, 55(4), pp. 509–540.

Ichihashi, S., 2021, 'Competing Data Intermediaries,' *The RAND Journal of Economics*, 52(3), pp. 515–537.

Iyer, G.; Soberman, D. and Villas-Boas, J. M., 2005, 'The Targeting of Advertising,' *Marketing Science*, 24(3), pp. 461–476.

Jehiel, P. and Moldovanu, B., 2000, 'Auctions with Downstream Interaction among Buyers,' *The RAND Journal of Economics*, 31(4), pp. 768–791.

Johnen, J., 2020, 'Dynamic Competition in Deceptive Markets,' *The RAND Journal of Economics*, 51(2), pp. 375–401.

Jullien, B.; Reisinger, M. and Rey, P., 2023, 'Personalized Pricing and Distribution Strategies,' *Management Science*, 69(3), pp. 1687–1702.

Liu, Q. and Serfes, K., 2004, 'Quality of Information and Oligopolistic Price Discrimination,' *Journal of Economics and Management Strategy*, 13(4), pp. 671–702.

Liu, Q. and Serfes, K., 2005, 'Imperfect Price Discrimination, Market Structure, and Efficiency,' *Canadian Journal of Economics/Revue canadienne d'économique*, 38(4), pp. 1191–1203.

Lu, Q. and Matsushima, N., 2022, *Personalized Pricing when Consumers can Purchase Multiple Items* (Osaka University, Institute of Social and Economic Research).

Mikians, J.; Gyarmati, L.; Erramilli, V. and Laoutaris, N., 2012, 'Detecting Price and Search Discrimination on the Internet,' *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pp. 79–84. HotNets-XI (Association for Computing Machinery, New York, NY, USA).

Montes, R.; Sand-Zantman, W. and Valletti, T., 2019, 'The Value of Personal Information in Online Markets with Endogenous Privacy,' *Management Science*, 65(3), pp. 955–1453.

OECD, 2018, *Personalised Pricing in the Digital Era*. Technical Report DAF/COMP (2018)13 (OECD, Directorate for Financial and Enterprise Affairs Competition Committee, Paris).

Rhodes, A. and Zhou, J., 2022, 'Personalized Pricing and Competition,' *Working Paper*.

Salop, S. C., 1979, 'Monopolistic Competition with Outside Goods,' *The Bell Journal of Economics*, 10(1), pp. 141–156.

Searle, S. R., 1979, 'On Inverting Circulant Matrices,' *Linear Algebra and Its Applications*, 25, pp. 77–89.

Shaffer, G. and Zhang, Z. J., 1995, 'Competitive Coupon Targeting,' *Marketing Science*, 14(4), pp. 395–416.

Shy, O. and Stenbacka, R., 2016, 'Customer Privacy and Competition,' *Journal of Economics and Management Strategy*, 25(3), pp. 539–562.

Spence, M., 1976, 'Product selection, fixed costs, and monopolistic competition,' *The Review of Economic Studies*, 43(2), pp. 217–235.

Taylor, C. and Wagman, L., 2014, 'Consumer Privacy in Oligopolistic Markets: Winners, Losers, and Welfare,' *International Journal of Industrial Organization*, 34, pp. 80–84.

Taylor, C. R., 2003, 'Supplier Surfing: Competition and Consumer Behavior in Subscription Markets,' *The RAND Journal of Economics*, 34(2), pp. 223–246.

Thisse, J.-F. and Vives, X., 1988, 'On The Strategic Choice of Spatial Price Policy,' *The American Economic Review*, 78(1), pp. 122–137.

Tyagi, R. K., 1999, 'On the Effects of Downstream Entry,' *Management Science*, 45(1), pp. 59–73.

Valentino-DeVries, J.; Singer-Vine, J. and Soltani, A., 2012, 'Websites Vary Prices, Deals Based on Users' Information,' *The Wall Street Journal*. Available at: *https://www.wsj.com/articles/SB10001424127887323777204578189391813881534*

Vickrey, W. S., 1964, *Microstatics* (Harcourt, Brace & World, Inc., New York-Burlingame, USA).

Villas-Boas, J. M., 2004, 'Price Cycles in Markets with Customer Recognition,' *The RAND Journal of Economics*, 35(3), pp. 486–501.