Data Broker competition and downstream market entry

Laura Abrardi* Carlo Cambini* Flavio Pino*†

April 2023

Abstract

We investigate how the level of competition in a Data Broker (DB) market, and the level of information precision, affect downstream entry and competition. Two vertically differentiated DBs, with different levels of information precision, compete to sell consumer data to a horizontally differentiated oligopoly market. We show that only the DB with higher precision sells data in equilibrium, while the other exerts competitive pressure on him. The data sale always reduces firm entry, which results in an increase in total welfare. The magnitude of the effect of the data sale on consumer surplus is mainly determined by the accuracy of the information provided. However, whether the effect is positive or negative depends on the level of competition in the upstream market. Maximum consumer surplus is reached when information and competition in the DB market are perfect, whereas the minimum is reached when information is perfect and the DB market is monopolistic. Instead, if both DBs have enough exclusive data on some consumer groups, the resulting market power leads to a decrease in firm entry, which ultimately harms consumers. We thus argue that policymakers should focus on granting a level playing field in the DB market rather than enforcing limits on information accuracy.

Keywords: *Data Broker, competition, data sale, price discrimination*

JEL codes: *L12, L41, L86.*

*Politecnico di Torino, Department of Management, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy.
†Corresponding author. Email: flavio.pino@polito.it

1 Introduction

Information plays a critical role in shaping competition in digital markets. In these markets, firms rely on data to make strategic decisions, and the quality and availability of information can determine the success or failure of a business. In particular, the use of data for price discrimination has been documented in digital markets (Mikians et al., 2012). This practice has particularly caught the attention of policymakers, as it could prove detrimental to consumers (Bergemann and Bonatti, 2019). However, to collect and process data on a large enough scale to make it valuable for personalized pricing, a company must have unique resources and capabilities.

Data Brokers (DBs) are thus an important part of this ecosystem, as they collect and combine multiple data sources to sell them to firms. The DB market is a highly concentrated multi-billion dollar industry(Pasquale, 2015), which thus has the potential to influence downstream competition. The previous literature has highlighted how monopolistic DBs can have incentives to underserve the downstream markets (Montes et al., 2019; Bounie et al., 2021b), which in turn can lead to a reduction in firm entry (Abrardi et al., 2022).

In this chapter, we investigate how the level of competition and the precision of information in the DB market impact downstream competition and consumer welfare. The recent literature on DBs and downstream markets has mostly focused on spatial competition models *la Hotelling* (Montes et al., 2019; Bounie et al., 2021b; Bounie et al., 2021a). To stay in line with previous literature and at the same time allow us to endogenize the number of firms in the downstream market, we implement a Salop (1979) model, where an endogenous number of symmetric firm enters and can then acquire information regarding consumers from DBs. Firms can operate first-degree price discrimination on the identified consumers. We model the DB market as a vertically differentiated duopoly, where DB_1 sells information with precision $\alpha \in [0, 1]$ and DB_2 sells information with precision $\beta\alpha, \beta \in [0, 1]$. We assume that when DB_1 (DB_2) sells data with regards to a consumer segment, only a uniformly distributed share of size α ($\beta\alpha$) is actually identified. Thus, α represents the level of data accuracy, while β represents the level of competition in the DB market: if $\beta = 0$, the DB market is monopolistic, whereas if $\beta = 1$, it exhibits perfect competition.

We find that in equilibrium DB_1 sells data to all entering firms, and the data price he sets depends on the level of competition β. Intuitively, DB_1 anticipates that firms' outside option is buying data from DB_2, and thus sets the data price equal to the difference in firms' profits between buying data from him or from DB_2. Our analysis shows that the data sale always reduces firm entry, as firms engage in price wars and pay for the acquisition of data.

Instead, the result with regard to consumer surplus is more nuanced. The level of information precision α can be either surplus-increasing or surplus-decreasing, depending on the intensity of DB competition β. In particular, for any α, there exists a cutoff value β^* such that, if $\beta \geq \beta^*$, consumer surplus increases with respect to the standard Salop model. Consumer surplus is maximized when both competition and information are perfect ($\beta = 1, \alpha = 1$).

Our analysis also highlights how this result critically depends on the level of overlap between the DB's datasets. In the baseline model, DB_2's dataset completely overlaps with DB_1's, and firms would not get any value from DB_2's dataset if they have already purchased data from DB_1. Instead, if both DBs have enough proprietary data on some consumers, they are both able to charge high prices for their datasets. In turn, firms are left with lower profits, which results in reduced entry and, ultimately, consumer harm. The reduction in consumer surplus always takes place when datasets are *super-additive* (i.e., the accuracy of the combined datasets is higher than the sum of the individual datasets' accuracies) and can take place if datasets are sub-additive (i.e., the accuracy of the combined datasets is lower than the sum of the individual datasets accuracies) and overlaps between the datasets are small enough.

The remainder of the chapter is organized as follows: Section 2 describes the relevant previous literature and discusses the chapter's contribution to it. Section 3 describes the model, while, in Section 4, we find firms' equilibrium prices. In Section 5, we compute the DB's profits and find his optimal strategy, and in Section 6, we conduct a welfare analysis. In Section 7, we explore the scenario where both DBs have some proprietary data regarding different groups of consumers. Finally, Section 8 concludes.

2 Literature review

This chapter focuses on how competition between DBs and data accuracy affects market outcomes, with a particular emphasis on firm entry and consumer surplus. The closest papers in the literature are Belleflamme et al. (2020), Bounie et al. (2021a) and Abrardi et al. (2022).

Belleflamme et al. (2020) focus on the effect of data accuracy when data can be used by two firms to price discriminate in a homogeneous goods market. They show that data only results in market power when both firms can price discriminate but with different accuracy levels. The intuition is that if the two firms identify the same consumer set, they will engage in price wars, leading to marginal cost pricing.

Bounie et al. (2021a) analyze competition between DBs who sell data to a series of duopolistic downstream markets. In their setting, data allows third-degree price discrimination, and each DB is a monopolist in a specific downstream market and competes with all other DBs in a competitive market. They show that, in equilibrium, the DB with the biggest monopolistic market has an incentive to collect the most accurate data, leading him to also serve the competitive market. Instead, the DB with the second-highest accuracy exerts competitive pressure on him, leading to lower data prices. Furthermore, they analyze how mergers between DBs with different sizes of monopolistic markets affect welfare.

Abrardi et al. (2022) focus on a monopolistic DB that sells consumer data to a downstream oligopolistic market with free entry. Data allows firms to operate first-degree price discrimination, and the DB can choose to which firms he wants to sell data partitions, as well as the size of the partitions. Irrespective of the selling mechanism adopted by the DB, in equilibrium, the data sale always results in an *entry barrier effect*, which reduces firm entry and, in turn, consumer surplus.

Other recent contributions to the literature have addressed the topic of DB competition. Ichihashi (2021) focuses on the non-rivalrous nature of data and how consumers sharing data with multiple competing DBs decreases the value of data. Anticipating this, DBs offer a low compensation for data and can even sustain a monopoly outcome if consumer data is then used to extract surplus from them. Gu et al. (2022) focus instead on the complementarity

of different datasets and show under which conditions competing DBs would be better off by merging their datasets to sell them as a single unit.

Other studies have focused on related issues regarding the use of data for price discrimination. Montes et al. (2019) analyze a setting where a DB sells data to a duopolistic downstream market and show that, when consumers can hide at a cost, consumer surplus is directly proportional to said cost. In a similar setting, Bounie et al. (2021b) finds that a DB maximizes his profits by only selling some consumer data instead of all of them. The intuition is that firms that obtain information on all consumers price too fiercely; thus, limiting the amount of data sold relaxes competition and allows the DB to extract higher rents. Other works have instead focused on settings where data are exogenously available to firms (Thisse and Vives, 1988; Shaffer and Zhang, 1995; Liu and Serfes, 2004; Taylor and Wagman, 2014; Chen et al., 2020), or where firms directly obtain data from consumers (Villas-Boas, 2004; Bergemann and Bonatti, 2011; Hagiu and Wright, 2020)[1] For recent surveys regarding data markets, refer to Bergemann and Bonatti (2019), Goldfarb and Tucker (2019) and Pino (2022).

Our study contributes to the literature by combining imperfect price discrimination, DB competition, and endogenous entry. In particular, we show how the data sale can always result in an increase in consumer surplus if the DB market is competitive enough. The level of competition needed to benefit consumers is directly proportional to the level of data accuracy. However, this result critically depends on the absence of synergies between the datasets. If both DBs have enough proprietary data regarding some groups of consumers, or the datasets show strong synergies, the DBs' increase in market power instead results in consumer harm.

3 The model

We study two interconnected markets. In the upstream market, two DBs (DB_1 and DB_2) exogenously have data regarding some consumers. In the downstream market, horizontally

[1]This literature strand also includes behavior-based price discrimination. Surveys on the subject can be found in Fudenberg and Villas-Boas (2006) and Esteves et al. (2009).

differentiated firms can purchase such data to observe individual consumers' preferences and, in turn, make them personalized offers.

3.1 Consumers, firms and Data Brokers

In the downstream market, we consider a circular city with free entry (Vickrey, 1964; Salop, 1979). Firms (he), indexed by $i \in \{0, 1, 2, ..., n-1\}$ where n is the number of firms that enter the market, sell competing products to consumers. Following the previous literature (Rhodes and Zhou, 2021), we assume sequential entry to avoid coordination problems and ignore integer constraints on n. Furthermore, we assume that firms enter the market choosing equally spaced locations, such that a generic firm i is located in $\frac{i}{n}$. Firms' marginal costs are normalized to zero, while the entry fixed cost is F. This cost can be interpreted as the cost of digitization, such as the investment needed to open an online retail shop.

Consumers (she) are uniformly distributed over the circle, and their mass is normalized to 1. Their location is indexed by $x \in [0, 1)$ in counter-clockwise order, and each of them buys at most one unit of the product. Gross utility derived from consumption is v, and consumers face a linear transportation cost t.

In the upstream market, two DBs (it) have datasets containing customers' information that can allow firms to identify consumers with a certain probability. Following Belleflamme et al. (2020), DB_1's dataset contains information that grant firms a probability $\alpha \in [0, 1]$ of identifying consumers. Instead, DB_2's dataset contains information that grant firms a probability $\beta\alpha, \beta \in [0, 1]$ of identifying consumers. We interchangeably refer to α as the data accuracy or precision and to β as the level of competition between DBs. We assume that the DB_2's dataset is contained by DB_1's dataset. In other words, a firm has no advantage in buying both datasets, as it would still result in a probability α of identifying consumers.

DBs can sell partitions of their datasets to downstream firms. In particular, to maximize the value of data and, in turn, firms' willingness to pay, the partition offered to each firm contains his location as in Bounie et al. (2021b). Moreover, due to the symmetry of the market, partitions are centered on a given firm's location. The size of the partition offered to

firm i by DB_k is labelled as $d_{i,k} \in [0, \frac{1}{n})$.[2] Thus, partitions sold by DB_1 allow firms to identify a share $d_{i,1}$ of consumers with probability α, while partitions sold by DB_2 allow to identify $d_{i,2}$ consumers with probability $\beta\alpha$.[3] Firms can perform first-degree price discrimination on the identified consumers.

A given firm i thus offers a basic price $p^B_{i,k} \geq 0$ to all unidentifed consumers, and location-specific tailored prices $p^T_{i,k}(x) \geq 0$ to identified ones, where k indicates whether the firms has purchased data from DB_1 or DB_2. We assume that each consumer only observes one price from a given firm, and, as a tie-breaking rule, we assume that consumers prefer tailored prices over basic prices when they are indifferent.[4] A consumer utility is thus defined as

$$U(x, i) = v - p_i(x) - t * D(x, i),$$

where $p_i(x) \in \{p^B_{i,k}, p^T_{i,k}(x)\}$ and $D(x, i)$ is the shortest arch between the consumer's location x and firm's location $\frac{i}{n}$. We denote the location of the indifferent consumer between firms i and $i+1$ as $\hat{x}_{i,i+1}$. Figure 1 shows the scenario where firm i buys data from DB_1.

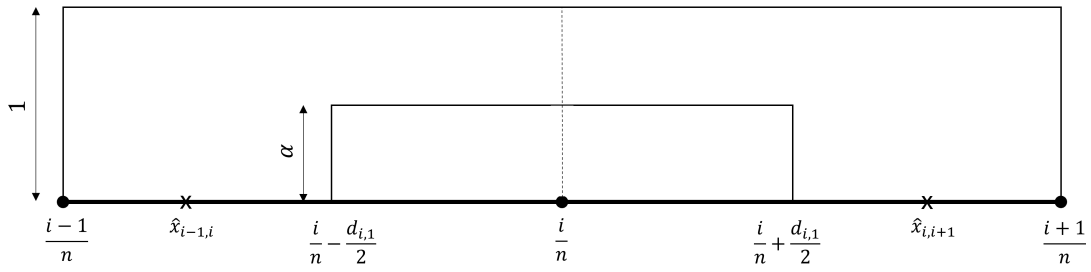

Figure 1: Firm i's market share when buying from DB_1. Firm i has a probability α of identifying consumers on the arch $[\frac{i}{n} - \frac{d_{i,1}}{2}, \frac{i}{n} + \frac{d_{i,1}}{2}]$ and offer them tailored prices, while it always offers his basic price on the consumers located on the arches $[\hat{x}_{i-1,i}, \frac{i}{n} - \frac{d_{i,1}}{2}]$ and $[\frac{i}{n} + \frac{d_{i,1}}{2}, \hat{x}_{i,i+1}]$. If instead firm i buys from DB_2, the probability would be $\beta\alpha$.

[2]Following Bounie et al. (2021b), we assume that DBs do not sell overlapping partitions, i.e., each consumer is at most identified by one firm. While this assumption allows the model to be tractable, it is also supported by previous literature in marketing that has stressed how targeting consumers with strong preferences is more beneficial to firms (Iyer et al., 2005).

[3]The level of information accuracy α could also be interpreted as the share of consumers, uniformly distributed, that a firm can identify over the arch $[\frac{i}{n} - \frac{d_{i,k}}{2}, \frac{i}{n} + \frac{d_{i,k}}{2}]$ once he buys a partition from DB_k.

[4]An example of consumers only observing one price would be a consumer accessing an online retail shop: if the firm can identify her, he can directly show her a tailored price instead of the basic one. In a setting similar to ours, Baik, Larson, et al. (2022) have shown that allowing consumers to see both prices, which then forces firms to only offer targeted discounts, has no effect on market outcomes when transportation costs are linear, like in our case.

A firm's profits when purchasing data from DB_1 can be thus written as

$$\pi_{i,1} = \alpha \int_{\frac{i}{n}-\frac{d_{i,1}}{2}}^{\frac{i}{n}+\frac{d_{i,1}}{2}} p_{i,1}^{\mathrm{T}}(x)\, dx + (1-\alpha) d_{i,1} p_{i,1}^{\mathrm{B}} + p_{i,1}^{\mathrm{B}} \left(\widehat{x}_{i,i+1} - \widehat{x}_{i-1,i} - d_{i,1}\right) - F, \tag{1}$$

where the first term on the right-hand side represents his profits when he is able to identify consumers, the second represents his profits when he is not able to identify consumers, and the third one represents his profits over unidentified consumers.

3.2 Data sale and timing

We assume that DBs simultaneously sell data partitions through non-renegotiable Take It Or Leave It (TIOLI) offers. In other words, DBs cannot change the offer they made to one firm based on other firms' behavior. Intuitively, DBs will set the data price w_{ix} equal to a firm's difference in profits between obtaining or not obtaining the partition they are proposing.

The timing of the model is as follows:[5]

Stage 1. Firms enter the market and pay the fixed cost F.

Stage 2. Each DB $k \in \{1, 2\}$ chooses a partition $d_{i,k}$ for each firm and offers it to that firm at a price w_{ix}.

Stage 3. Each firm that entered the market chooses whether to accept or decline the DBs' offers.

Stage 4. Firms set basic prices $p_{i,k}^{B}$ for unidentified consumers.

Stage 5. Firms that obtained a partition set tailored prices $p_{i,k}^{T}(x)$ for the identified consumers.

We solve the model through backward induction. As a useful benchmark, we refer to the standard Salop (1979) model, where entering firms make zero profits in equilibrium, resulting in $\tilde{n} = \sqrt{\frac{t}{F}}$ and $\tilde{CS} = \tilde{TW} = v - \frac{5}{4}\sqrt{tF}$.

[5]The sequentially of Stages 4 and 5 is common in the literature (see Montes et al. (2019), Bounie et al. (2021b), and Bounie et al. (2021a) among others), as it grants the existence of Pure Strategy Nash Equilibria. Moreover, it is supported by observed managerial practices (Fudenberg and Villas-Boas, 2006).

4 Equilibrium prices

Given our framework, where DB_2's dataset is contained within DB_1's dataset, only the latter will sell data partitions in the downstream market as these partitions are more valuable to firms. However, DB_2 will exert competitive pressure on DB_1 and limit its ability to extract surplus from firms. First, we analyze the equilibrium case where all firms acquire data from DB_1, and then we move to the subgame where a generic firm i instead buys data from DB_2.

Without loss of generality, we focus on firm i located in $\frac{i}{n}$, who buys data from DB_1. Indifferent consumers' locations are as in the standard Salop model, resulting in

$$\hat{x}_{i-1,i} = \frac{2i-1}{2n} + \frac{p^B_{i,1} - p^B_{i-1,1}}{2t} \quad and \quad \hat{x}_{i,i+1} = \frac{2i+1}{2n} + \frac{p^B_{i+1,1} - p^B_{i,1}}{2t}. \tag{2}$$

If firm i obtains a data partition, he can offer tailored prices $p^T_{i,1}(x)$ to the identified consumers. The tailored prices match the direct rivals' basic prices in utility levels, resulting in

$$p^{\mathrm{T}}_{i,1}(x) = \begin{cases} p^{\mathrm{B}}_{i-1,1} + 2tx - \frac{t}{n}(2i-1) & \text{for } \; x \in [\frac{i}{n} - \frac{d_i}{2}, \frac{i}{n}] \\ p^{\mathrm{B}}_{i+1,1} - 2tx + \frac{t}{n}(2i+1) & \text{for } \; x \in [\frac{i}{n}, \frac{i}{n} + \frac{d_i}{2}] \end{cases} \tag{3}$$

Using the expressions from (2) and (3), we can derive firm i's FOC of Equation (1) with respect to $p^B_{i,1}$, obtaining

$$p^{\mathrm{B}}_{i,1} = \frac{t}{2n} - \frac{t\alpha d_{i,1}}{2} + \frac{p^{\mathrm{B}}_{i+1,1} + p^{\mathrm{B}}_{i-1,1}}{4} \tag{4}$$

As highlighted in the previous literature (Thisse and Vives, 1988; Bounie et al., 2021b), we also find that data-enabled price discrimination has an ambiguous effect on firms' profits. On the one hand, the ability to offer tailored prices allows firms to extract more surplus from consumers, which is profit-increasing: this is referred to as *surplus extraction effect* (Thisse and Vives, 1988). On the other hand, as highlighted in (4), an increase in the acquired data leads to a reduction of firms' basic prices, as, on average, they serve consumers farther from their locations. The price reduction leads to fiercer competition, referred to as *competition effect* (Thisse and Vives, 1988).

The system of reaction functions of all firms allows us to obtain firms' equilibrium prices,

the properties of which are described in the following lemma.

Lemma 1 *Firms' equilibrium prices when DBs sell data through TIOLI offers are decreasing in α and in $d_{i,1} \forall i \in \{0, ..., n-1\}$. Partitions that are sold to firms closer to firm i have a stronger effect on his prices.*

Proof. See Appendix A. ■

The intuition of the above result is the following. Without loss of generality, we focus on firm i. First, as either α or $d_{i,1}$ increase, firm i identifies a larger share of consumers close to his location. Therefore, as the basic price is offered to consumers who are, on average, farther from the firm's location, the basic price decreases with α and $d_{i,1}$. Second, all the other firms that obtain data also decrease their basic price due to the same effect described above. As shown in Equation 3, tailored prices are based on the rivals' basic prices, and thus they also decrease.

As shown above, both basic and tailored prices decrease in the presence of data. However, at this stage of the game, we cannot draw conclusions with regard to firm profits. Indeed, obtaining more data allows a firm to identify more consumers, from which he extracts more surplus (i.e., tailored prices are higher than basic prices). Thus the effect of information precision and partition size on firm profits is ambiguous.

We now focus on the subgame where firm i buys data from DB_2 instead. His profit function will thus be

$$\pi_{i,2} = \beta\alpha \int_{\frac{i}{n}-\frac{d_{i,2}}{2}}^{\frac{i}{n}+\frac{d_{i,2}}{2}} p_{i,2}^{\mathrm{T}}(x)\, dx + (1-\beta\alpha) d_{i,2} p_{i,2}^{\mathrm{B}} + p_{i,2}^{\mathrm{B}} \left(\widehat{x}_{i,i+1} - \widehat{x}_{i-1,i} - d_{i,2}\right) - F, \qquad (5)$$

while all other firms' profits remain as in Equation (1). The properties of firms' prices are described in the following lemma.

Lemma 2 *In the subgame where firm i buys data from DB_2, all firms' prices are higher than the equilibrium case, and they are decreasing in β.*

Proof. See Appendix A. ■

The intuition of this result is the following. As firm i obtains less accurate data, he identifies fewer consumers and must thus offer his basic price to consumers who are, on average,

closer to his location, leading to higher basic prices. Predicting this behavior, all other firms will also charge higher basic prices with respect to the equilibrium case. Intuitively, as β increases, firm i identifies more consumers and lowers his basic price accordingly.

5 DBs equilibrium profits

Having analyzed firms' profits, we now focus on the upstream market for data. As stated before, only DB_1 will sell data in equilibrium, as its partitions contain those of DB_2 and are thus more valuable for firms. However, DB_1's data price also depends on DB_2's strategy, as firms can acquire data from DB_2 as an alternative. DB_1 will set the price for data equal to firms' willingness to pay, which is the difference in firms' profits between buying data from DB_1 or DB_2. Thus, DB_1 solves the following problem:

$$\max_{d_{0,1},d_{1,1},...,d_{n-1,1}} \pi_{DB_1} = \sum_{i=0}^{n-1} \pi_{i,1} - \pi_{i,2}. \tag{6}$$

Instead, DB_2 competes *la Bertrand* with DB_1, aiming to set its partitions to maximize firms' profits when they buy from him. DB_2 solves the following problem:

$$\max_{d_{0,2},d_{1,2},...,d_{n-1,2}} \sum_{i=0}^{n-1} \pi_{i,2}. \tag{7}$$

By simultaneously solving the two problems, we obtain the results described in the following proposition.

Proposition 1 *In equilibrium, both DB_1 and DB_2 offer equally sized partitions to all entering firms, i.e. $d_{i,1} = d_{i,2} = d^* \quad \forall \quad i \in \{0, ..., n-1\}$. The size of the equilibrium partitions d^* is decreasing in the information accuracy α and increasing in the level of competition between DBs β. In equilibrium, all firms buy from DB_1, and DB_1's profits are increasing in α and decreasing in β.*

Proof. See Appendix A. ■

Since firms are symmetric, DB_1's profits are influenced in the same way by any partition it sells and thus offers same-sized partitions. The same also holds for DB_2. Confirming

the results from Bounie et al. (2021b), we find that in equilibrium, both DBs offer non-overlapping partitions to temper the *competition effect* of data.

To better describe the intuition behind Proposition 1, Figure 2 shows firms' equilibrium profits as a function of the (symmetric) partitions offered by DBs. Having concluded that both DBs offer same-sized partitions to firms, we refer to the partition's size offered by DB_1 and DB_2 as d_1 and d_2, respectively. Firms' equilibrium profits are not influenced by d_2 and are decreasing with respect to d_1. The trend is given by the interplay between the *surplus extraction* and the *competition effects*. Since the partition's size is equal for all firms, an increase in it exacerbates the competitive pressure as all firms reduce their prices. The combined effect of all firms' pricing strategies makes the *competition effect* outweigh the *surplus extraction effect*, leading to lower profits. Instead, firms' profits when buying from DB_2 exhibit an inverse U-shaped curve with respect to d_2, while they are also decreasing in d_1. At first, an increase in d_2 allows firms to compete better against their more informed rivals, resulting in higher profits. However, as d_2 increases, the *competition effect* of data erodes the firms' profits, resulting in a concave function. By observing the functions, it is clear that DB_1 chooses an intermediate partition size to balance the decrease in firms' equilibrium profits and the decrease in firms' profits when buying from DB_2. Instead, DB_2 chooses an intermediate partition size to temper the *competition effect* on firms' profits when buying from him.

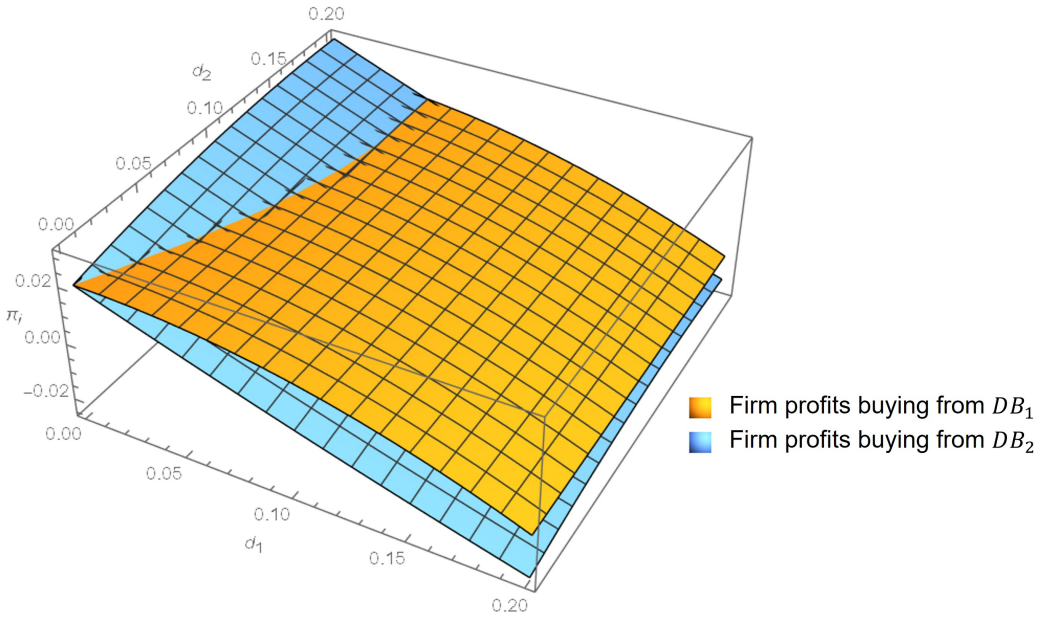

Figure 2: Firms' profits when buying and not buying data as a function of the partitions sold by DB_1 (d_1) and by DB_2 (d_2). $\alpha = 0.5, \beta = 0.5, t = 3, n = 5, F = 0.1$.

The influence of α and β on d^* is also guided by the *surplus extraction* and *competition effects*. An increase in α allows firms to extract more surplus from closer consumers, and DB_1 thus limits the partition's size to temper the *competition effect*. Instead, an increase in β favors firms that buy from DB_2, and DB_1 opts to increase the partition size to decrease those firms' profits and, in turn, increase their willingness to pay for data. Intuitively, DB_1's profits increase with information accuracy as data become more valuable to firms and decrease as the competitive pressure from DB_2 gets stronger.

6 Number of entering firms and welfare analysis

Having found the DBs' equilibrium strategies, we solve the game's first stage regarding firm entry. As in Salop (1979), firms enter as long as their profits, after paying for data and entry, are greater than 0. We obtain the results described in the following Proposition by binding this constraint.

Proposition 2 *The number of entering firms in equilibrium is always lower than in the benchmark, i.e.,* $n^* < \tilde{n}$. n^* *is decreasing in the information accuracy* α *and increasing in the level of competition between DBs* β.

Proof. See Appendix A. ■

As discussed in Section 5, in equilibrium, all firms buy data from DB_1. In this scenario, firms' profits prior to paying for data are lower than in the benchmark (as visible in Figure 2 when $d_1 = 0$) due to the profit-decreasing *competition effect* of data more than offsetting the profit-increasing *surplus extraction effect*. As the data price is positive, further decreasing firms' profits with respect to the benchmark, firm entry is always reduced. The level of information accuracy α further exacerbates the *competition effect*, leading to lower entry. In contrast, the level of competition β increases competitive pressure and reduces the data price, leading to higher entry. However, as we can see from Figure 3, the level of competition can never overcome the reduction in entry caused by the information accuracy.[6]

[6]Note that the size of the *entry barrier effect* is a function of both the transportation and the entry cost. As an example, the number of entering firms for $\alpha = 1, \beta = 0$ is $n^* \approx \frac{3}{4}\sqrt{\frac{t}{F}}$. Thus, the entry reduction could be higher or lower than unity depending on these two variables. To keep the analysis straightforward, we abstract from this problem by treating n as a continuous variable.

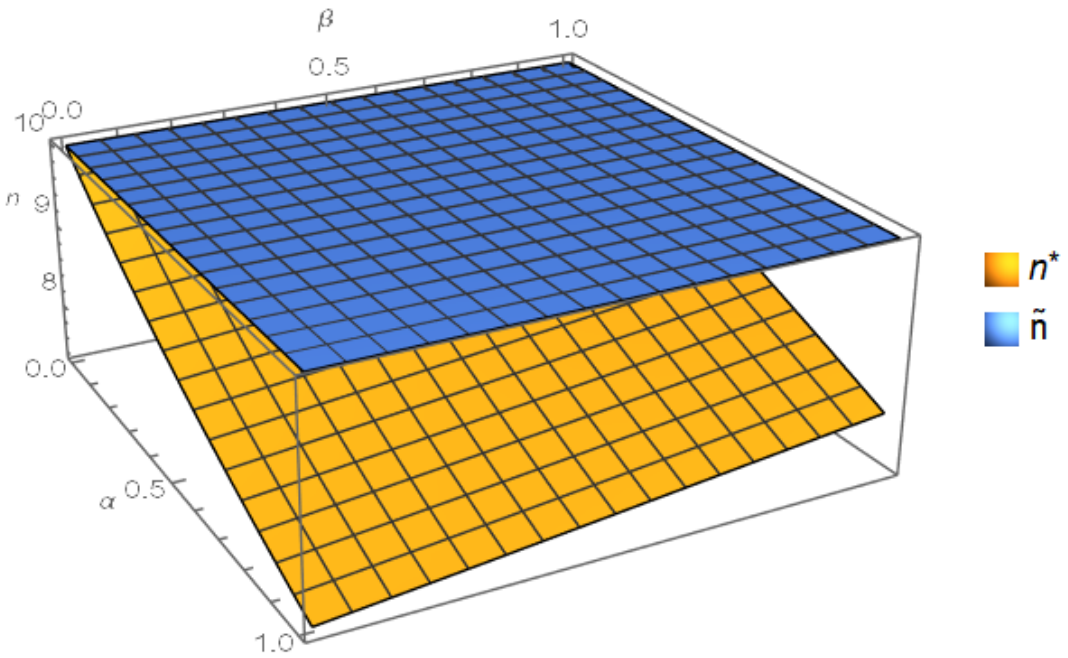

Figure 3: Number of entering firms in equilibrium and in the standard Salop model as a function of α and β. $u = 10, t = 10, F = 0.1$.

The data sale from DBs also affects the welfare analysis. On the one hand, the *competition effect* induced by data leads to fiercer competition and lower prices, which overall benefit consumers. However, the reduction in firm entry due to lower firms' profits increases the firm concentration in the downstream market, which in turn harms consumers. The following Proposition describes the results with regard to consumer surplus and total welfare.

Proposition 3 *For any level of information accuracy α, there exists a threshold level of DB competition β^* such that, if $\beta \geq \beta^*$, then $CS^* \geq \tilde{CS}$. For any level of α and β, $TW^* \geq \tilde{TW}$.*

Proof. See Appendix A. ■

To better understand the Proposition above, comparing our results with those of the relevant existing literature is useful. Bounie et al. (2021b) study a duopoly downstream market, where a DB can sell data to operate third-degree price discrimination. They show that consumer surplus increases under the presence of the DB, but a higher information accuracy lowers consumer surplus as firms improve their ability to extract surplus from consumers. Abrardi et al. (2022) move to an oligopolistic downstream market and find that

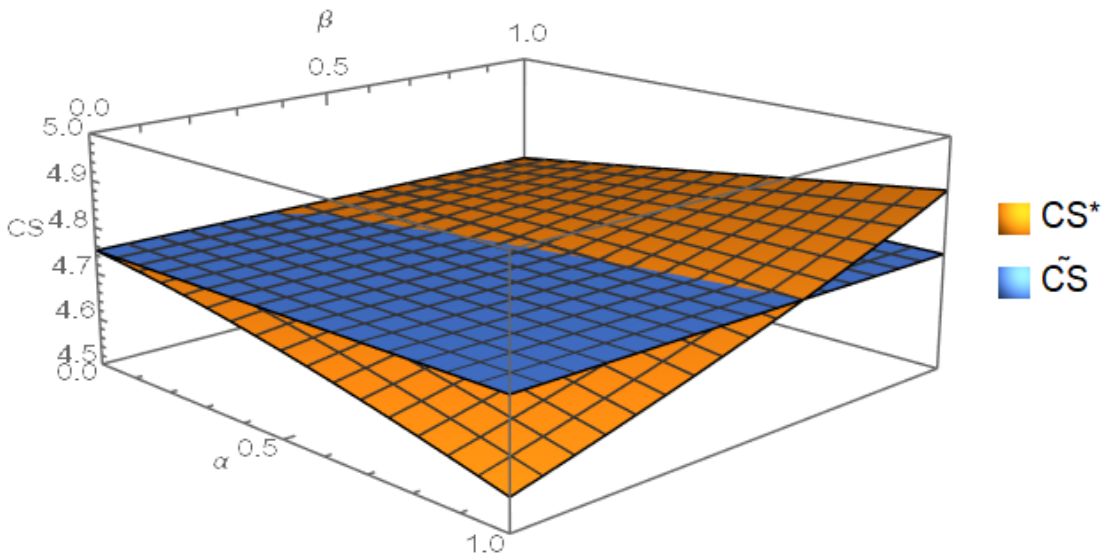

Figure 4: Consumer surplus in equilibrium and in the standard Salop model as a function of α and β. $u = 10, t = 10, F = 0.1$.

the data sale results in a reduction of consumer surplus, as the reduction in entry, referred to as *entry barrier* effect, more than offsets the price reduction derived from the *competition effect of data*. Our analysis thus highlights two novel results.

First, we find that information accuracy magnifies the data sale's effect on consumer surplus, as shown in Figure 4.

As information become more accurate, firms compete more fiercely, lowering prices and dissipating profits. However, this increase in competition has ambiguous effects on consumer surplus. While consumers would benefit from lower prices, the reduction in firm entry increases the concentration in the downstream market, which in turn harms consumers.

Second, we find that the level of competition in the DB market determines whether the data sale's effect on consumer surplus is positive or negative. As β increases, firms pay a lower price to acquire data, and the *entry barrier effect* of data is reduced, benefiting consumers. In particular, consumer surplus is maximized when $\alpha = 1, \beta = 1$: in this scenario, perfect information leads firms to fiercely compete in prices, while perfect competition in the DB market drives the data price to zero, leading to high firm entry. We thus argue that knowing the level of information accuracy is not enough to predict the effects of the data sale on consumer surplus, as it acts as a mere amplifier of the welfare effects of data. Instead, the

level of competition in the DB market determines whether these effects will benefit or harm consumers. From a policy perspective, ensuring a level playing field in the DB market is thus more effective than intervening in information accuracy when aiming to improve consumer surplus.

Finally, we find that total welfare always increases with respect to the benchmark case, confirming the results from Abrardi et al. (2022). The data sale always reduces firms' profits, resulting in lower entry, which lowers the amount of profits dissipated in paying the entry cost F. In other words, the data sale partially solves the excessive entry problem typical of the standard Salop model, leading to higher total welfare.

7 Synergic datasets

In the baseline model, we have analyzed a scenario where two DBs compete in selling datasets to an oligopolistic downstream market. However, DB_2's dataset was contained in DB_1's one, and thus, in equilibrium, firms only buy from DB_1. DB_2 only exerted competitive pressure on DB_1, influencing his pricing strategies, but could not sell his dataset in equilibrium. In this section, we expand the baseline model by dropping the complete overlap assumption. To remain consistent with the previous analysis, both DBs can still sell data partitions that grant accuracies of α and $\beta\alpha$, respectively. However, if a firm obtains data from both DBs regarding the same consumer location, it will then have an accuracy γ over those consumers. Figure 5 gives a visual representation of the new setup.

The accuracy of the combined datasets γ can be seen as a proxy of the level of synergy between the two datasets. On the one hand, the two datasets could contain some overlapping information: in such a scenario, we would have $\gamma \leq \alpha + \beta\alpha$. Following the previous literature (Gu et al., 2022), we will refer to this scenario as *sub-additive*. On the other hand, the combination of both datasets could also result in information that are more valuable than the same of the individual datasets' values, i.e., $\gamma > \alpha + \beta\alpha$. We refer to this scenario as *super-additive*.

To simplify the exposition, we introduce additional notation. We define $\pi_{i,0}$ as firm i's profits when not buying any dataset. Instead, $\pi_{i,k}$ define firm i's profits when buying the

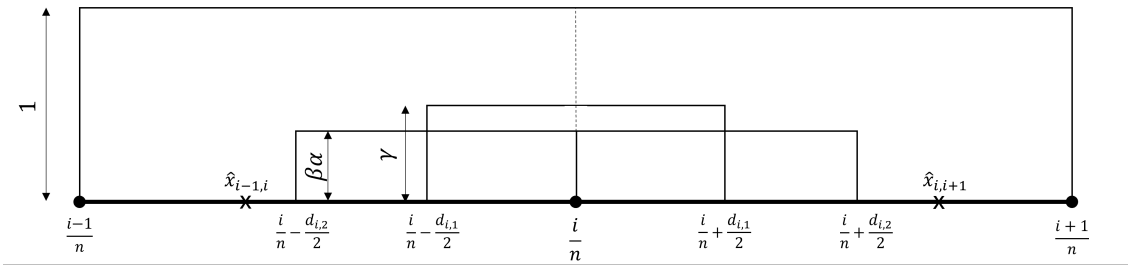

Figure 5: Firm i's market share when buying from both DB_1 and DB_2, assuming $d_2 > d_1$. Firm i has a probability γ of identifying consumers on the arch $[\frac{i}{n} - \frac{d_{i,1}}{2}, \frac{i}{n} + \frac{d_{i,1}}{2}]$ and offer them tailored prices, while it has a probability $\beta\alpha$ on consumers on the arches $[\frac{i}{n} - \frac{d_{i,2}}{2}, \frac{i}{n} - \frac{d_{i,1}}{2}]$ and $[\frac{i}{n} + \frac{d_{i,1}}{2}, \frac{i}{n} + \frac{d_{i,2}}{2}]$. Finally, it always offers its basic price on the consumers located on arches $[\hat{x}_{i-1,i}, \frac{i}{n} - \frac{d_{i,2}}{2}]$ and $[\frac{i}{n} + \frac{d_{i,2}}{2}, \hat{x}_{i,i+1}]$.

dataset from DB_k. Finally, $\pi_{i,12}$ are firm i's profits when buying both datasets. All these profits are computed prior to paying the datasets. Moreover, we define

$$ES_k = \pi_{i,k} - \pi_{i,0}, \qquad k \in \{1, 2, 12\}$$

as the Extra Surplus firm i firm obtains when purchasing the dataset(s) k. Moreover, $w_{i,1}$ and $w_{i,2}$ are the datasets' prices that DB_1 and DB_2 respectively offer to firm i. The superscript *sup* refers to the *super-additive* scenario, whereas the superscript *sub* refers to the *sub-additive one*.

In the updated setting, both datasets are valuable for downstream firms. This, in turn, influences the DBs' pricing strategies, which we summarize in the following Lemma.

Lemma 3 *If datasets are super-additive, any pair* (w_1^{*sup}, w_2^{*sup}) *such that* $w_1^{*sup} + w_2^{*sup} = ES_{12}$ *is a Nash equilibrium in the DBs' pricing game.*

If datasets are sub-additive, there exists a unique Nash equilibrium in the DB's pricing game, where $w_k^{*sub} = ES_{12} - ES_{-k}, k \in 1, 2$.

Proof. See Gu et al. (2022). ■

The results presented in the Lemma above are those described in Gu et al. (2022). Indeed, their analysis is also valid in our setting with regard to the DB's pricing strategies. When datasets are *super-additive*, DBs prefer that firms buy both datasets so that they can try to appropriate some of the positive synergies created by the datasets. In particular, any

pair of prices that fully extracts the Extra Surplus generated by the combined datasets is an equilibrium. Instead, if datasets are *sub-additive*, the DBs prefer trying to undercut their rival and, in equilibrium, set prices equal to the marginal value of their dataset.

However, the main difference between our model and that of Gu et al. (2022) is that in our setting, the Extra Surpluses ES_k are endogenously determined by the DBs choices of $d_{i,1}$ and $d_{i,2}$ respectively. Thus, even if the DBs' equilibrium pricing strategies have been defined in Lemma 3, we must solve the game to find the equilibrium partition sizes offered by both DBs. The following Proposition describes the market outcomes in the *super-additive* scenario.

Proposition 4 *If datasets are super-additive, both DB_1 and DB_2 offer equally sized partitions to all entering firms, i.e. $d_{i,1}^{sup} = d_{i,2}^{sup} = d^{*sup}$. Moreover, the equilibrium partitions are smaller than in the benchmark model.*

Consumer Surplus is always lower, and Total Welfare is always higher than in the benchmark model.

Proof. See Appendix A. ■

The intuition behind the results in the Proposition above is straightforward. When datasets are *super-additive*, both DBs simultaneously try to maximize the Extra Surplus generated by their combined datasets. By doing so, the DBs effectively act as a monopolistic DB that offers a dataset with accuracy γ by offering same-sized partitions. In turn, the market outcomes are the same as the baseline model in the scenario where $\alpha = \gamma, \beta = 0$.

In the baseline model, the level of competition β between DBs induced the sale of larger partitions as a way to exert competitive pressure. As the datasets super-additivity effectively allows DBs to avoid competition, in equilibrium DBs offer smaller partitions with respect to the baseline model to temper the *competition effect* of data.

The DBs' ability to extract surplus from entering firms by effectively avoiding competition among themselves leads to a higher entry barrier, which in turn increases downstream market concentration and harms consumers.

Instead, when datasets are *sub-additive*, DBs' equilibrium pricing strategy entails undercutting each other. Thus, each DB tries to maximize his own dataset's value. The following

Proposition summarizes the market outcomes stemming from this scenario.

Proposition 5 *If datasets are sub-additive, in equilibrium* DB_1 *sets* $d_{i,1}^{sub} = d_1^{*sub} \quad \forall \quad i$ *and* DB_2 *sets* $d_{i,2}^{sub} = \frac{d_1^{*sub}}{\beta} \quad \forall \quad i$.

For any level of β*, there exists a threshold* $\bar{\gamma} \in [\alpha, \alpha + \beta\alpha]$ *such that, if* $\gamma > \bar{\gamma}$, $CS^{*sub} < \tilde{CS}$.

Proof. See Appendix A. ■

In the *sub-additive scenario*, both DBs aim to maximize the marginal value of their respective datasets. Recall that, under the model's assumptions, DB_2's dataset is less accurate than DB_1's, as $\beta \leq 1$. Then, in equilibrium, DB_2 opts to sell larger partitions than DB_1. The intuition is that DB_2's partitions entail a lower *competition effect* for any identified location, as the share of identified consumers is lower. As the intensive margin of data is lower, DB_2 maximizes the partitions' values by increasing their size.

Interestingly, this result departs from that of the baseline model, where both DBs offer same-sized partitions. Indeed, in the baseline model, DB_2 has no market power, as his dataset has no value once a firm obtains a partition from DB_1. Then, DB_2 aims to maximize a firm's profits when it obtains its dataset, as it would be better off selling it for any price above zero. Conversely, when datasets do not completely overlap, DB_2's partitions are valuable for firms even if they already purchased a partition from DB_1. Then, DB_2 aims to maximize the firms' willingness to pay for the dataset, which in turn entails selling larger partitions in equilibrium.

From a welfare perspective, we find that a high enough value of γ results in consumer harm with respect to the standard Salop model. Indeed, as γ increases, so does the DBs' market power, as the value of the combined datasets is higher. As firms pay more to obtain both datasets, they are left with lower profits, and entry is reduced, ultimately harming consumers.

8 Conclusions

With the growing centrality of consumer data in the digital economy, DBs have become key enablers of data-driven technologies. Their ability to transform data into valuable information allows them to influence downstream competition with relevant welfare implications. This work contributes to the expanding literature regarding the effect on DBs, by analyzing how DB competition and information accuracy affect DBs' strategies and, in turn, economic outputs.

We show that in equilibrium if the DB market is vertically differentiated, only the DB with the highest information accuracy sells its partitions. However, the rival DB exerts competitive pressure and influences its strategy, leading to a lower data price. This effect is relevant with regard to welfare implications. Previous literature (Abrardi et al., 2022) has highlighted how a monopolist DB in a similar setting causes a fierce reduction in firm entry, resulting in consumer harm with respect to the standard Salop model. We expand on the previous literature by introducing imperfect information and competition in the DB market and find that both features influence consumer surplus. In particular, the intensity of DB competition can subvert the effect on consumer surplus, leading to consumer benefit with respect to the standard Salop model. Instead, information accuracy acts more as an amplifier of the effect that the data sale has on consumer surplus: the highest consumer surplus is reached when both information and DB competition is perfect, while the lowest is reached when information is perfect and there is no competition in the DB market.

However, if competing DBs have information on different sets of consumers, the competitive pressure rapidly reduces, as firms are better off buying both datasets. In particular, if data are *super-additive*, meaning that the combined dataset is more valuable than the sum of the individual datasets' values, the DBs set prices to extract all available surplus from firms. The rise in datasets' prices, in turn, reduces downstream firms' entry, leading to higher market concentration and consumer harm.

From a policy perspective, we thus argue that ensuring a level playing field in the DB market is a stronger lever than information accuracy to ensure a positive outcome for consumers. In particular, mergers in the DB sector could lead to lower competition and, in

turn, consumer harm. However, such a level playing field must not only be limited to the dataset's size. Indeed, DBs' market power stems from having proprietary data on specific consumers, which in turn allows them to raise the datasets' prices. This can be particularly harmful when data are *super-additive*, as DBs can then extract all available surplus from firms, in turn increasing the downstream market concentration. Further analyses should then be required to better understand how policymakers could invert such an outcome so that the positive effects of data stemming from price discrimination are not overwhelmed by the reduction in entry given by the datasets' prices.

References

Abrardi, Laura et al. (2022). "User data and endogenous entry in online markets". In: *Available at SSRN 4256544*.

Baik, Simon Anderson, Nathan Larson, et al. (2022). "Price Discrimination in the Information Age: Prices, Poaching, and Privacy with Personalized Targeted Discounts". In: *The Review of Economic Studies*.

Belleflamme, Paul, Wing Man Wynne Lam, and Wouter Vergote (2020). "Competitive imperfect price discrimination and market power". In: *Marketing Science* 39.5, pp. 996–1015.

Bergemann, Dirk and Alessandro Bonatti (2011). "Targeting in advertising markets: implications for offline versus online media". In: *The RAND Journal of Economics* 42.3, pp. 417–443.

— (2019). "Markets for information: An introduction". In: *Annual Review of Economics* 11, pp. 85–107.

Bounie, David, Antoine Dubus, and Patrick Waelbroeck (2021a). "Competition and Mergers with Strategic Data Intermediaries". In: *Available at SSRN*.

— (2021b). "Selling strategic information in digital competitive markets". In: *The RAND Journal of Economics* 52.2, pp. 283–313.

Chen, Zhijun, Chongwoo Choe, and Noriaki Matsushima (2020). "Competitive personalized pricing". In: *Management Science* 66.9, pp. 4003–4023.

Esteves, Rosa Branca et al. (2009). *A survey on the economics of behaviour-based price discrimination*. Tech. rep. NIPE-Universidade do Minho.

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

Goldfarb, Avi and Catherine Tucker (2019). "Digital economics". In: *Journal of Economic Literature* 57.1, pp. 3–43.

Gu, Yiquan, Leonardo Madio, and Carlo Reggiani (2022). "Data brokers co-opetition". In: *Oxford Economic Papers* 74.3, pp. 820–839.

Hagiu, Andrei and Julian Wright (2020). "Data-enabled learning, network effects and competitive advantage". In: *Unpublished manuscript*.

Ichihashi, Shota (2021). "Competing data intermediaries". In: *The RAND Journal of Economics* 52.3, pp. 515–537.

Iyer, Ganesh, David Soberman, and J Miguel Villas-Boas (2005). "The targeting of advertising". In: *Marketing Science* 24.3, pp. 461–476.

Liu, Qihong and Konstantinos Serfes (2004). "Quality of information and oligopolistic price discrimination". In: *Journal of Economics & Management Strategy* 13.4, pp. 671–702.

Mikians, Jakub et al. (2012). "Detecting price and search discrimination on the internet". In: *Proceedings of the 11th ACM workshop on hot topics in networks*, pp. 79–84.

Montes, Rodrigo, Wilfried Sand-Zantman, and Tommaso Valletti (2019). "The value of personal information in online markets with endogenous privacy". In: *Management Science* 65.3, pp. 1342–1362.

Pasquale, Frank (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Pino, Flavio (2022). "The microeconomics of data–a survey". In: *Journal of Industrial and Business Economics* 49.3, pp. 635–665.

Rhodes, Andrew and Jidong Zhou (2021). *Personalized pricing and privacy choice*. Tech. rep. Working paper.

Salop, Steven C (1979). "Monopolistic competition with outside goods". In: *The Bell Journal of Economics*, pp. 141–156.

Searle, SR (1979). "On inverting circulant matrices". In: *Linear algebra and its applications* 25, pp. 77–89.

Shaffer, Greg and Z John Zhang (1995). "Competitive coupon targeting". In: *Marketing Science* 14.4, pp. 395–416.

Taylor, Curtis and Liad Wagman (2014). "Consumer privacy in oligopolistic markets: Winners, losers, and welfare". In: *International Journal of Industrial Organization* 34, pp. 80–84.

Thisse, Jacques-Francois and Xavier Vives (1988). "On the strategic choice of spatial price policy". In: *The American Economic Review*, pp. 122–137.

Vickrey, William Spencer (1964). *Microstatics*. Harcourt, Brace & World.

Villas-Boas, J Miguel (2004). "Consumer learning, brand loyalty, and competition". In: *Marketing Science* 23.1, pp. 134–145.

Appendix for online publication only

Appendix A Proofs

Proof of Lemma 1. To obtain firms' equilibrium prices, we need to solve a system of equations composed by (4) $\forall i \in \{0, 0..., n-1\}$. In matrix form we have $\mathbf{A} * \mathbf{p} = \mathbf{b}$, where $\mathbf{p}$ is the price vector, and $\mathbf{b}$ is the known terms vector:

$$\begin{bmatrix} 4 & -1 & \dots & 0 & 0 & 0 & \dots & -1 \\ -1 & 4 & \dots & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 4 & -1 & 0 & \dots & 0 \\ 0 & 0 & \dots & -1 & 4 & -1 & \dots & 0 \\ 0 & 0 & \dots & 0 & -1 & 4 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -1 & 0 & \dots & 0 & 0 & 0 & \dots & 4 \end{bmatrix} * \begin{bmatrix} p^{\mathrm{B}}_{0,1} \\ p^{\mathrm{B}}_{1,1} \\ \dots \\ p^{\mathrm{B}}_{i-1,1} \\ p^{\mathrm{B}}_{i,1} \\ p^{\mathrm{B}}_{i+1,1} \\ \dots \\ p^{\mathrm{B}}_{n-1,1} \end{bmatrix} = \begin{bmatrix} \frac{2t}{n} - 2t\alpha d_{0,1} \\ \frac{2t}{n} - 2t\alpha d_{1,1} \\ \dots \\ \frac{2t}{n} - 2t\alpha d_{i-1,1} \\ \frac{2t}{n} - 2t\alpha d_{i,1} \\ \frac{2t}{n} - 2t\alpha d_{i+1,1} \\ \dots \\ \frac{2t}{n} - 2t\alpha d_{n-1,1} \end{bmatrix}$$

Matrix $\mathbf{A}$ is circulant, tridiagonal and symmetric. Exploiting the solution provided by Searle (1979) for the inverse of this type of matrix, we have that

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

where, $a_j = -\frac{1}{2\sqrt{3}} * \left(\frac{\left(2+\sqrt{3}\right)^j}{1-\left(2+\sqrt{3}\right)^n} - \frac{\left(2-\sqrt{3}\right)^j}{1-\left(2-\sqrt{3}\right)^n} \right)$. A property of this type of matrices is that $a_j = a_{n-j}$ $\forall j \neq 0, \frac{n}{2}$. In our particular case, coefficient a_j is decreasing in j $\forall j \in \{0, \frac{n}{2}\}$, and

$\sum_{j=0}^{n-1} a_j = \frac{1}{2}$. We can now write $\mathbf{p} = \mathbf{A^{-1}} * \mathbf{b}$. We obtain

$$\begin{bmatrix} p^{\mathrm{B}}_{0,1} \\ p^{\mathrm{B}}_{1,1} \\ \dots \\ p^{\mathrm{B}}_{n-1,1} \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & \dots & a_{n-1} \\ a_{n-1} & a_0 & \dots & a_{n-2} \\ \dots & \dots & \dots & \dots \\ a_1 & a_2 & \dots & a_0 \end{bmatrix} = \begin{bmatrix} \frac{2t}{n} - 2t\alpha d_{0,1} \\ \frac{2t}{n} - 2t\alpha d_{1,1} \\ \dots \\ \frac{2t}{n} - 2t\alpha d_{n-1,1} \end{bmatrix}$$

Thus, we can write

$$p^{\mathrm{B}*}_{i,1} = \left(\frac{2t}{n} * \sum_{j=0}^{n-1} a_j \right) - 2t \sum_{j=0}^{n-1} \alpha d_{i+j,1} a_j.$$

Given that $\sum_{j=0}^{n-1} a_j = \frac{1}{2}$, the previous equation can be rewritten as

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

The apex $*$ indicates equilibrium results in the equilibrium case. The equation above clearly shows that all firms' profits are decreasing in α and in all firms' partitions. Moreover, as $a_j > a_{j+1} \forall j \in 0, ..., \frac{n}{2}$, firm i's basic price is more influenced by partitions of firms closer to him, and the partition that most affects his strategy is his own.

Proof of Lemma 2. Suppose that firm i buys from DB_2. FOC of (5) with respect to $p^B_{i,2}$ is

$$p^{\mathrm{B}}_{i,2} = \frac{t}{2n} - \frac{t\beta\alpha d_{i,2}}{2} + \frac{p^{\mathrm{B}}_{i+1,1} + p^{\mathrm{B}}_{i-1,1}}{4}$$

We follow the same method as in the Proof of Lemma 1. The only difference is in the known term vector $\mathbf{b}$, where the i-th term will be $\frac{t}{2n} - \frac{t\beta\alpha d_{i,2}}{2}$. By inverting the matrix and solving for basic prices, we obtain

$$p^{\mathrm{B\ D}}_{i,2} = \frac{t}{n} - 2t \sum_{j=0}^{n-1} \alpha d_{i+j,1} a_j + 2t\alpha d_{i,1} a_0 - 2t\alpha\beta d_{i,2} a_0, \tag{A.2}$$

$$p^{\mathrm{B\ D}}_{i+1,1} = \frac{t}{n} - 2t \sum_{j=0}^{n-1} \alpha d_{i+1+j,1} a_j + 2t\alpha d_{i,1} a_1 - 2t\alpha\beta d_{i,2} a_1, \tag{A.3}$$

$$p^{\mathrm{B\ D}}_{i-1,1} = \frac{t}{n} - 2t \sum_{j=0}^{n-1} \alpha d_{i-1+j,1} a_j + 2t\alpha d_{i,1} a_1 - 2t\alpha\beta d_{i,2} a_1. \tag{A.4}$$

The apex D indicates equilibrium results in the subgame where firm i buys from DB_2. Since $a_0 > a_1$, All firms' prices when firm i buys from DB_2 are higher than when it buys from DB_1, and they are decreasing in β.

Proof of Proposition 1. We obtain DBs' profits as a function of the partitions they offer by replacing the prices computed in the two previous lemmas in equation (1) and (5). DB_1's profits are equal to

$$\max_{d_{0,1},d_{1,1},...,d_{n-1,1}} \pi_{DB_1} = \sum_{i=0}^{n-1} \pi_{i,1} - \pi_{i,2}.$$

Instead, DB_2 aims to maximize firms' profits when buying from him:

$$\max_{d_{0,2},d_{1,2},...,d_{n-1,2}} \sum_{i=0}^{n-1} \pi_{i,2}.$$

By computing each DBs' profits FOCs with respect to all partitions they offer, we find that their profits depend in the same way from each partition. Thus, in equilibrium, both DB_1 and DB_2 offer equally sized partitions to all firms, i.e., $d_{i,1} = d_1 \forall i \in \{0, ..., n-1\}$ and $d_{i,2} = d_2 \forall i \in \{0, ..., n-1\}$. By applying these properties to (A.1), (A.2), (A.3) and (A.4), we obtain

$$p_{i,1}^{\mathrm{B}*} = \frac{t}{n} - t\alpha d_1 \forall i \in \{0, ..., n-1\}; \tag{A.5}$$

$$p_{i,2}^{\mathrm{B\,D}} = \frac{t}{n} - t\alpha d_1 + 2t\alpha d_1 a_0 - 2t\alpha\beta d_2 a_0; \tag{A.6}$$

$$p_{i+1,1}^{\mathrm{B\,D}} = \frac{t}{n} - t\alpha d_1 + 2t\alpha d_1 a_1 - 2t\alpha\beta d_2 a_1; \tag{A.7}$$

$$p_{i-1,1}^{\mathrm{B\,D}} = \frac{t}{n} - t\alpha d_1 + 2t\alpha d_1 a_1 - 2t\alpha\beta d_2 a_1. \tag{A.8}$$

By replacing these prices in equations (1) and (5), we obtain

$$\pi_{i,1}^{*} = \frac{t}{n^2} - \frac{1}{2}\alpha t d_1^2 - F; \tag{A.9}$$

$$\pi_{i,2}^{D} = \frac{t}{2n^2}(2 + 2\alpha n(-1 + 2a_1)(d_1 - \beta d_2) - \alpha n^2(\beta d_2^2 + 4\alpha(-1 + 2a_0)(a_0 - a_1)(d_1 - \beta d_2)^2) - F. \tag{A.10}$$

By computing DBs' profits FOCs with respect to d_1 and d_2 and solving the equation system,

we obtain

$$d_1^* = d_2^* = d^* = \frac{1 - 2a_1}{n\alpha(1-\beta)(\frac{1}{\alpha(1-\beta)} + 4a_0 - 8a_0^2 - 4a_1 + 8a_0a_1)}. \tag{A.11}$$

FOCs of (A.11) with respect to α and β show that d^* is decreasing in the former and decreasing in the latter. Replacing (A.11) in π_{DB_1} and computing FOCs with respect to α and β highlight how DB_1's profits are increasing in α and decreasing β.

Proof of Proposition 2. The proof proceeds in two steps. First, we show that the number of entering firms is always lower than in the benchmark case. Second, we show how the number of entering firms is influenced by α and β.

Step 1.

The number of entering firms is given by equating to 0 firms' profits after paying for entry and data. In the benchmark model, firms' profits are equal to

$$\tilde{\pi}_i = \frac{t}{n^2} - F.$$

Comparing this profit function with (A.9), it is clear that firms' equilibrium profits (prior to paying for data) are already lower than the benchmark profits. As the price of data is positive, we conclude that the number of entering firms is always lower than in the benchmark.

Step 2.

Firms' profits after paying for data are equal to

$$\pi_{i,1}^* - (\pi_{i,1}^* - \pi_{i,2}^D) = \pi_{i,2}^D.$$

By replacing (A.11) in (A.10), we obtain

$$\pi_{i,2}^D = \frac{t}{n^2} + \frac{\alpha t(1-2a_1)^2(-2 + 4\alpha(-1+2a_0)(a_0-a_1)(\beta-1)^2 + \beta)}{2(n + 4\alpha n(-1+2a_0)(a_0-a_1)(-1+\beta))^2} - F. \tag{A.12}$$

FOCs of (A.12) with respect to α and β show that firms' profits after paying for data and entry are decreasing in the former and increasing in the latter. As higher profits imply a higher number of entering firms (since in equilibrium firms profits will be equal to 0), we can conclude that the number of entering firms is decreasing in α and increasing in β.

Proof of Proposition 3.

The proof proceeds in two steps: first, we show that, for any level of α, there exists a

threshold β^* such that, if $\beta > \beta^*$, then $CS^* > \tilde{CS}$. Second, we show that total welfare is always higher than in the benchmark case.

Step 1.

In equilibrium, all entering firms obtain same sized partitions and charge equal prices. Thus, indifferent consumers will be located in the middle points between firms. To compute total consumer surplus, we evaluate the consumer surplus of consumers located in $[0, \frac{1}{2n^*}]$, and multiply it by $2n^*$. We obtain

$$CS^* = 2n(\alpha \int_0^{\frac{d^*}{2}} u - tx - p_{0,1}^T(x)dx + (1-\alpha) \int_0^{\frac{d^*}{2}} u - tx - p_{0,1}^{B*}dx + \int_{\frac{d^*}{2}}^{\frac{1}{2n*}} u - tx - p_{0,1}^{B*}dx) = u - \frac{5t}{4n^*} + \frac{1}{2}\alpha n^* t d^{*2} \quad \text{(A.13)}$$

Data have two effects on CS. First, they directly affect it by lowering firms' prices (*competition effect*), which benefits consumers (third term in the right-hand side of (A.13)). Treating the number of entering firms as given, FOCs on the third term show that this effect is increasing in α, as more accurate data intensify competition, and in β, as a higher level of competition between DBs results in bigger partitions in equilibrium. Second, data indirectly affect CS by influencing the number of entering firms. A decrease in firm entry (*entry barrier effect*) increases firms' prices (see equation (A.5)), which in turn harms consumers. FOCs of (A.12) show that firms' profits after paying for entry and data, and thus the number of entering firms, are decreasing in α and increasing in β. We can conclude that the effect of β on CS is always positive, while the effect of α is ambiguous.

First, let us focus on the case where $\beta = 0$. When $\alpha = 0$, we have the standard Salop model, whereas when $\alpha = 1$ we have the model described in Abrardi et al. (2022). If $\alpha = 1$, CS is lower than in the benchmark, as shown in Abrardi et al. (2022). Since the effects of α on CS are monotonic, as priorly described, we can conclude that the effect of α alone on CS is negative, i.e. the *entry barrier effect* is stronger than the *competition effect*.

Second, suppose that $\alpha = \beta = 1$, i.e. information is perfect and the DB market exhibits perfect competition. By posing these conditions in (A.11) and (A.12) we find

$$d^* = \frac{1 - 2a_1}{n} \quad \text{(A.14)}$$

and

$$\pi_{i,2}^{D} = \frac{t}{n^2} - \frac{t(1-2a_1)^2}{2n^2} - F. \tag{A.15}$$

To find the number of entering firms, we must equate (A.15) to 0 and solve for n: However, as a_1 is exponential in n, the equation has no explicit solution. To estimate the number of entering firms, we must approximate $(1-2a_1)^2$. Our objective is to show that CS is higher than in the benchmark when $\alpha = \beta = 1$: thus, since $(1-2a_1)^2$ decreases firm i's profits and, in turn, firm entry, we search for a function that overestimates it. If $(1-2a_1)^2$ is overestimated, then firm entry and CS will be underestimated: if the approximated CS is still higher than the benchmark, we can conclude that the exact CS will also be higher than the benchmark. We approximate

$$(1-2a_1)^2 \approx -\frac{1}{n^2} + \frac{8}{3}(2-\sqrt{3}) + \frac{1}{20}. \tag{A.16}$$

By replacing (A.16) in (A.15) and solving for n, we obtain

$$n^* = \frac{\sqrt{-\frac{203t}{F} + \frac{160\sqrt{3}t}{F} + \frac{\sqrt{t(28800F+118009t-64960\sqrt{3}t)}}{F}}}{4\sqrt{15}} \tag{A.17}$$

By replacing (A.17) in (A.13) and comparing it with $\tilde{CS}$, we find that, when $\alpha = \beta = 1$, $CS^* > \tilde{CS}$. Since, when $\beta = 0$, $CS^* < \tilde{CS}$ for any α, and the effects of α and β on CS are monotonic, we can conclude that for any $\alpha \in [0,1]$ there exists a threshold value β^* such that if $\beta \geq \beta^*$, then $CS^* \geq \tilde{CS}$.

Step 2.

With regards to Total Welfare (TW), we recall that in equilibrium firms obtain 0 profits, i.e., $\pi_i^D(DB_2) = 0$. In equilibrium, DB_1's profits are equal to

$$\pi_{DB_1}^* = n^*(\pi_{i,1}^* - \pi_{i,2}^D) = n * (\frac{t}{n^{*2}} - \frac{1}{2}\alpha t d^{*2} - F). \tag{A.18}$$

We obtain TW by adding (A.18) with (A.13) and simplifying:

$$TW^* = u - \frac{t}{4n^*} - n^*F. \tag{A.19}$$

FOC of (A.19) with respect to n^* shows that TW is decreasing in n^* whenever $n^* \geq \frac{1}{2}\sqrt{\frac{t}{F}}$.

As argued in previous Propositions, the number of entering firms is directly proportional to firms' profits when buying from DB_2. FOC of (A.12) shows that firms' profits are minimized when $\alpha = 1, \beta = 0$, which corresponds to the scenario analyzed in Abrardi et al. (2022). They show that, when the DB sells data to all firms like in our model, the equilibrium number of entering firms is $\approx \frac{3}{4}\sqrt{\frac{t}{F}}$. As $\frac{3}{4}\sqrt{\frac{t}{F}} > \frac{1}{2}\sqrt{\frac{t}{F}}$, we can conclude that in our model TW is decreasing in n^*. Since $TW^* = T\tilde{W}$ when $\alpha = \beta = 0$, and since $n^* \leq \tilde{n} \quad \forall \quad \alpha, \beta \in [0,1]$ (as described in Proposition 2), $TW^* \geq T\tilde{W} \quad \forall \quad \alpha, \beta \in [0,1]$.

Proof of Proposition 4. To ease the exposition, the proof is organized in two steps. First, we compute firms' profits when buying both datasets $\pi_{i,12}$ and when not buying any datasets $\pi_{i,0}$. Then, we proceed to solve the DB's pricing game and compute market outcomes.

Step 1. Suppose that firm i buys both datasets, and that $d_{i,2} > d_{i,1}$, as in Figure 5.[7] We can rewrite firm i's profits as

$$\pi_{i,12} = \gamma \int_{\frac{i}{n}-\frac{d_{i,1}}{2}}^{\frac{i}{n}+\frac{d_{i,1}}{2}} p^{\mathrm{T}}_{i,12}(x)\,dx + (1-\gamma)d_{i,1}p^{\mathrm{B}}_{i,12} + \beta\alpha\left(\int_{\frac{i}{n}-\frac{d_{i,2}}{2}}^{\frac{i}{n}-\frac{d_{i,1}}{2}} p^{\mathrm{T}}_{i,12}(x)\,dx + \int_{\frac{i}{n}+\frac{d_{i,1}}{2}}^{\frac{i}{n}+\frac{d_{i,2}}{2}} p^{\mathrm{T}}_{i,12}(x)\,dx \right) + (1-\beta\alpha)(d_{i,2}-d_{i,1})p^{\mathrm{B}}_{i,12} + p^{\mathrm{B}}_{i,12}\left(\widehat{x}_{i,i+1} - \widehat{x}_{i-1,i} - d_{i,2}\right) - F, \quad \text{(A.20)}$$

where the first line represents profits over the segment where firm i has precision γ, the second line represents profits over the segments where firm i has precision $\beta\alpha$ and the third line represents the profits over unidentified consumers. To find equilibrium profits, we must solve the system of firms' reaction functions, as we did in the proof of Lemma 1. By applying the same method, we find

$$p^{\mathrm{B}*}_{i,12} = \frac{t}{n} - 2t\sum_{j=0}^{n-1} \gamma d_{i+j,1} + \beta\alpha(d_{i+j,2} - d_{i+j,1})a_j. \quad \text{(A.21)}$$

[7] The procedure is the same if $d_{i,2} < d_{i,1}$. However, in the proof of Proposition 5, we show that in equilibrium DB_2 offers larger partitions when datasets are *sub-additive*, and thus the equilibrium result would break the assumption that $d_{i,2} < d_{i,1}$.

Instead, suppose that firm i does not buy any dataset. Its profits then become

$$\pi_{i,0} = +p^{\mathrm{B}}_{i,0}\left(\widehat{x}_{i,i+1} - \widehat{x}_{i-1,i}\right) - F. \tag{A.22}$$

As in the proof of Lemma 2, we can follow the same method applied in the proof of Lemma 1 and simply modify the i-th term of the known term vector $\mathbf{b}$, which becomes $\frac{t}{2n}$. We thus obtain the following equilibrium prices:

$$p^{\mathrm{B\ D}}_{i,0} = \frac{t}{n} - 2t\sum_{j=0}^{n-1}\left(\gamma d_{i+j,1} + \beta\alpha(d_{i+j,2} - d_{i+j,1})a_j\right) - 2ta_0(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})) + 2t\frac{a_0}{n}, \tag{A.23}$$

$$p^{\mathrm{B\ D}}_{i+1,12} = \frac{t}{n} - 2t\sum_{j=0}^{n-1}\left(\gamma d_{i+j+1,1} + \beta\alpha(d_{i+j+1,2} - d_{i+j+1,1})a_j\right) - 2ta_1(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})) + 2t\frac{a_1}{n}, \tag{A.24}$$

$$p^{\mathrm{B\ D}}_{i-1,12} = \frac{t}{n} - 2t\sum_{j=0}^{n-1}\left(\gamma d_{i+j-1,1} + \beta\alpha(d_{i+j-1,2} - d_{i+j-1,1})a_j\right) - 2ta_1(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})) + 2t\frac{a_1}{n}, \tag{A.25}$$

which allow us to compute firm i's profits when not buying data.

Step 2.

As firms are symmetric, it is straightforward to demonstrate that both DB_1 and DB_2 will each offer same-sized partitions to each entering firm, i.e., $d_{i,1} = d_1$ and $d_{i,2} = d_2 \quad \forall \quad i \in \{0, ..., n-1\}$. Then, we can rewrite equilibrium firms' profits as

$$\pi^*_{i,12} = \frac{t}{n^2} - \frac{\alpha d_1^2}{2} - \frac{(\gamma - \alpha)d_2^2}{2} - F; \tag{A.26}$$

$$\pi^*_{i,0} = \frac{t}{n^2}(-1+n(1-2a_0)(\gamma d_1+\beta\alpha(d_2-d_1))(-1+2n(a_0-a_1)(\gamma d_1+\beta\alpha(d_2-d_1))-F. \tag{A.27}$$

Then, as described in Lemma 3, both DBs simultaneously maximize

$$ES_{12} = \pi^*_{i,12} - \pi^*_{i,0} \tag{A.28}$$

with respect to their own partition size, resulting in

$$d_1^{*sup} = d_2^{*sup} = d^{*sup} \frac{1 - 2a_1}{n(1 + 4\gamma a_0 - 8\gamma a_0^2 - 4\gamma a_1 + 8\gamma a_0 a_1)}. \tag{A.29}$$

As a continuum of Nash equilibrium exists, we can only compute total DB profits by replacing (A.29) in (A.28) and multiplying by n.

With regard to consumer surplus, in equilibrium, each firm obtains a partition of size d^{*sup} and accuracy γ. As in Proposition 3, we can thus write CS as

$$CS = u - \frac{5t}{4n} + \frac{1}{2}\gamma n t d^{*sup2}. \tag{A.30}$$

We obtain the equilibrium number of entering firms by posing $\pi_{i,0}^* = 0$ and solving for n. By using the same approximation approach as in Proposition 3, we obtain

$$n^{*sup} = \frac{3\sqrt{t}}{\sqrt{9F + 48\gamma F - 24\sqrt{3}\gamma F + 112\gamma^2 F - 64\sqrt{3}\gamma^2 F}}. \tag{A.31}$$

By replacing (A.31) in (A.30), we find that CS^{*sup} is monotonically decreasing in γ, and $CS^{*sup} = \tilde{CS}$ for $\gamma = 0$. Thus, we conclude that consumer surplus is always lower than in the benchmark model.

The same approach can be repeated for Total Welfare, which can be computed as

$$TW^{*sup} = n^{*sup} ES_{12} + CS^{*sup}. \tag{A.32}$$

We find that TW^{*sup} is monotonically increasing in γ, and $TW^{*sup} = \tilde{TW} for \gamma = 0$. Thus, we conclude that Total Welfare is always higher than in the benchmark model.

Proof of Proposition 5. The proof proceeds in two steps. First, we compute firms' profits when they only buy one dataset, as the computations for when they buy both datasets have already been made in Proposition 4. Second, we compute the equilibrium partitions' sizes, profits and welfare.

Step 1. Suppose that all firms except firm i buy both datasets, whereas firm i only buys $d_{i,1}$. As in the previous proofs, we can obtain firms' equilibrium prices by properly adjusting the i-th term of the known term vector $\mathbf{b}$. In this subgame, firm i only obtains a partition

of size $d_{i,1}$ and accuracy α, resulting in prices

$$p_{i,1}^{\text{B D}} = \frac{t}{n} - 2t\sum_{j=0}^{n-1}\Big(\gamma d_{i+j,1} + \beta\alpha(d_{i+j,2} - d_{i+j,1})a_j\Big) - 2ta_0(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})) + 2ta_0(\frac{1}{n} - \alpha d_{i,1}t), \quad \text{(A.33)}$$

$$p_{i+1,12}^{\text{B D}} = \frac{t}{n} - 2t\sum_{j=0}^{n-1}\Big(\gamma d_{i+j+1,1} + \beta\alpha(d_{i+j+1,2} - d_{i+j+1,1})a_j\Big) - 2ta_1(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})) + 2ta_1(\frac{1}{n} - \alpha d_{i,1}t), \quad \text{(A.34)}$$

$$p_{i-1,12}^{\text{B D}} = \frac{t}{n} - 2t\sum_{j=0}^{n-1}\Big(\gamma d_{i+j-1,1} + \beta\alpha(d_{i+j-1,2} - d_{i+j-1,1})a_j\Big) - 2ta_1(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})) + 2ta_1(\frac{1}{n} - \alpha d_{i,1}t), \quad \text{(A.35)}$$

which allow us to compute firm i's profits when it only buys from DB_1.

Similarly, we can obtain firms' equilibrium prices when firm i only buys from DB_2. We obtain

$$p_{i,2}^{\text{B D}} = \frac{t}{n} - 2t\sum_{j=0}^{n-1}\Big(\gamma d_{i+j,1} + \beta\alpha(d_{i+j,2} - d_{i+j,1})a_j\Big) - 2ta_0(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})) + 2ta_0(\frac{1}{n} - \beta\alpha d_{i,2}t), \quad \text{(A.36)}$$

$$p_{i+1,12}^{\text{B D}} = \frac{t}{n} - 2t\sum_{j=0}^{n-1}\Big(\gamma d_{i+j+1,1} + \beta\alpha(d_{i+j+1,2} - d_{i+j+1,1})a_j\Big) - 2ta_1(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})) + 2ta_1(\frac{1}{n} - \beta\alpha d_{i,2}t), \quad \text{(A.37)}$$

$$p_{i-1,12}^{\text{B D}} = \frac{t}{n} - 2t\sum_{j=0}^{n-1}\Big(\gamma d_{i+j-1,1} + \beta\alpha(d_{i+j-1,2} - d_{i+j-1,1})a_j\Big) - 2ta_1(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})) + 2ta_1(\frac{1}{n} - \beta\alpha d_{i,2}t). \quad \text{(A.38)}$$

Step 2. As firms are symmetric, it is straightforward to demonstrate that both DB_1 and DB_2 will each offer same-sized partitions to each entering firm, i.e., $d_{i,1} = d_1$ and $d_{i,2} = d_2 \quad \forall \quad i \in \{0, ..., n-1\}$. Then, we can rewrite equilibrium firms' profits as

$$\pi^*_{i,1} = \frac{t}{2n^2}\Big(\alpha d_1 n\big(2 + (-1 + 4(a_0 - a_1)(\alpha + \beta\alpha - \gamma))d_1 n + 4\alpha(a_1 - a_0)\beta d_2 n\big) - 2\big(-1 + (\gamma - \beta\alpha + 2\alpha a_0(1+\beta) - 2\gamma a_0)d_1 n + \alpha(1 - 2a_0)\beta d_2 n\big)\big(1 + 2(a_0 - a_1)(-\gamma d_1 + \alpha(d_1 + \beta d_1 - \beta d_2))n\big)\Big) - F, \quad \text{(A.39)}$$

$$\pi^*_{i,2} = \frac{t}{n^2} - \frac{t}{2}\Big(4(-1 + 2a_0)(a_0 - a_1)(\gamma - \beta\alpha)^2 d_1^2 + \beta\alpha d_2^2\Big) - \frac{t d_1(\beta\alpha - \gamma)(-1 + 2a_1)}{n} - F. \quad \text{(A.40)}$$

As described in Lemma 3, DBs set their prices as

$$w_1^{*sub} = \pi^*_{i,12} - \pi^*_{i,2} \qquad \text{and} \qquad w_2^{*sub} = \pi^*_{i,12} - \pi^*_{i,1}. \quad \text{(A.41)}$$

By maximizing DBs' profits with respect to their partitions' sizes, we obtain

$$d_1^{*sub} = \frac{1 - 2a_1}{n(1 - 4a_0\beta\alpha + 8a_0^2\beta\alpha + 4a_1\beta\alpha - 8a_0a_1\beta\alpha + 4a_0\gamma - 8a_0^2\gamma - 4a_1\gamma + 8a_0a_1\gamma)} \quad \text{(A.42)}$$

$$d_2^{*sub} = \frac{d_1^{*sub}}{\beta} \quad \text{(A.43)}$$

As $\beta \leq 1$, we can conclude that DB_2 always offers bigger partitions in equilibrium. Moreover, DB_1's profits are always higher than DB_2's for any $\beta < 1, \gamma \in [\alpha, \alpha + \beta\alpha]$.

With regard to welfare, by following the same approach as in Proposition 3, we find that consumer surplus is equal to

$$CS^{sub} = u - \frac{5t}{4n} + \frac{nt}{2}((\gamma - \beta\alpha)d_1^{*sub2} + \beta\alpha d_2^{*sub2}). \quad \text{(A.44)}$$

To find the number of entering firms, we bind firms' remaining profits after paying for data to zero, which are equal to:

$$\pi^*_{i,12} - (\pi^*_{i,12} - \pi^*_{i,2}) - (\pi^*_{i,12} - \pi^*_{i,1}) = \pi^*_{i,1} + \pi^*_{i,2} - \pi^*_{i,12}. \quad \text{(A.45)}$$

Following the same approach as in Proposition 3, we obtain

$$n^{*sub} = \Big((-14336\sqrt{3}+24832)\alpha^2\beta^2\gamma^2 t+(28672\sqrt{3}-49664)\alpha^3\beta^3\gamma t+(8640\sqrt{3}-14976)\alpha^2\beta^2\gamma t+$$
$$(24832-14336\sqrt{3})\alpha^4\beta^4 t+(4992-2880\sqrt{3})\alpha^3\beta^3 t+(4992-2880\sqrt{3})\alpha^3\beta^2 t+(576\sqrt{3}-1008)\alpha^2\beta^2 t+$$
$$(2016-1152\sqrt{3})\alpha^2\beta t+(576\sqrt{3}-1008)\alpha^2 t+(9984-5760\sqrt{3})\alpha\beta\gamma^2 t+(108\sqrt{3}-216)\alpha\beta t+$$
$$(108\sqrt{3}-216)\alpha t+(1008-576\sqrt{3})\gamma^2 t+(216-108\sqrt{3})\gamma t-81t\Big)^{\frac{1}{2}}/\Big(-24832\alpha^2\beta^2\gamma^2 F+$$
$$14336\sqrt{3}\alpha^2\beta^2\gamma^2 F-28672\sqrt{3}\alpha^3\beta^3\gamma F+49664\alpha^3\beta^3\gamma F-5760\sqrt{3}\alpha^2\beta^2\gamma F+$$
$$9984\alpha^2\beta^2\gamma F-24832\alpha^4\beta^4 F+14336\sqrt{3}\alpha^4\beta^4 F-1152\sqrt{3}\alpha^2\beta^2 F+2016\alpha^2\beta^2 F-$$
$$9984\alpha\beta\gamma^2 F+5760\sqrt{3}\alpha\beta\gamma^2 F-2016\alpha\beta\gamma F+1152\sqrt{3}\alpha\beta\gamma F-1008\gamma^2 F+$$
$$576\sqrt{3}\gamma^2 F-432\gamma F+216\sqrt{3}\gamma F-81F\Big)^{\frac{1}{2}} \quad \text{(A.46)}$$

By replacing n^{*sub} in (A.44) and computing FOCs with respect to β and γ, we find that CS^{*sub} is monotonically increasing in β and monotonically decreasing in γ. Moreover, by setting $\beta = 1$ and $\gamma = \alpha + \beta\alpha$, we find that $CS^{*sub} < \tilde{CS}$. Thus, we can conclude that for any value of β, there exists a threshold level $\bar{\gamma}$ such that, if $\gamma > \bar{\gamma}$, $CS^{*sub} < \tilde{CS}$.